

Masterclass

mc

Achim Klenke

Wahrscheinlichkeitstheorie

3. Auflage



Springer Spektrum

Wahrscheinlichkeitstheorie

Achim Klenke

Wahrscheinlichkeitstheorie

3., überarbeitete und ergänzte Auflage



Springer Spektrum

Prof. Dr. Achim Klenke
Institut für Mathematik
Universität Mainz
Mainz
Germany

ISBN 978-3-642-36017-6
DOI 10.1007/978-3-642-36018-3

ISBN 978-3-642-36018-3 (eBook)

Die Deutsche Nationalbibliothek verzeichnetet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Mathematics Subject Classification (2010): 60-01, 28-01, 60G05, 60J10, 60H05

Springer Spektrum
© Springer-Verlag Berlin Heidelberg 2006, 2008, 2013
Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier.

Springer Spektrum ist eine Marke von Springer DE. Springer DE ist Teil der Fachverlagsgruppe Springer Science+Business Media
www.springer-spektrum.de

Vorwort zur dritten Auflage

In der dritten Auflage konnten einige Fehler korrigiert werden. Außerdem wurde das Buch an einigen Stellen behutsam erweitert, ohne jedoch die Grundstruktur wesentlich zu verändern. Insbesondere sind weitere Übungsaufgaben und etliche Illustrationen hinzugekommen.

Ich möchte an dieser Stelle allen danken, die durch ihre Hinweise bei der Verbesserung der zweiten Auflage geholfen haben, insbesondere Michael Diether, Marlen Eckhoff, Christopher Grant, Matthias Hammer, Heiko Hoffmann, Martin Hutzenthaler, Martin Kolb, Manuel Mergens, Thal Nowik, Felix Schneider, Wolfgang Schwarz und Stephan Tolksdorf.

Mainz,
November 2012

Achim Klenke

Vorwort zur zweiten Auflage

Gegenüber der ersten Auflage sind Übungsaufgaben hinzugekommen und manche Beweise sind etwas effizienter geworden (insbesondere der von Bemerkung 22.6). Vor allem aber konnte ich durch die Rückmeldungen aufmerksamer Leser einige Fehler der ersten Auflage korrigieren.

Ich möchte an dieser Stelle allen danken, die bei der Verbesserung der ersten Auflage geholfen haben, speziell René Billing, Anne Eisenbürger, Katrin Hofmann-Credner, Patric Karl Glöde, Patrick Jahn, Arnulf Jentzen, L. Mayer, Felix Schneider, Henning Sulzbach, Sebastian Thiemann und Oliver Wesche sowie den Kollegen Ehrhard Behrends, Nina Gantert, Rudolf Grübel, Peter Mörters, Ralph Neininger und Martin Zerner.

Mainz,
Januar 2008

Achim Klenke

Vorwort zur ersten Auflage

Das vorliegende Buch basiert auf den vierstündigen Vorlesungen Stochastik I und Stochastik II, die ich in den vergangenen Jahren an der Universität zu Köln und an der Johannes Gutenberg-Universität in Mainz gehalten habe, und die an eine Vorlesung über elementare Stochastik anschließen. Eine gewisse Vertrautheit mit den Ideen der elementaren Stochastik wird zwar nicht formal vorausgesetzt, dem Leser jedoch empfohlen.

Ziel dieses Buches ist es, die zentralen Objekte und Konzepte der Wahrscheinlichkeitstheorie vorzustellen: Zufallsvariablen, Unabhängigkeit, Gesetze der großen Zahl und zentrale Grenzwertsätze, Martingale, Austauschbarkeit und unbegrenzte Teilbarkeit, Markovketten und -prozesse sowie den Zusammenhang mit der diskreten Potentialtheorie, Kopplung, Ergodentheorie, die Brown'sche Bewegung und das Itô-Integral, den Poisson'schen Punktprozess, Perkolation und die Theorie der großen Abweichungen sowie stochastische Differentialgleichungen.

Die Maß- und Integrationstheorie wird entwickelt, soweit sie für das Verständnis und die Formulierung der Wahrscheinlichkeitstheorie notwendig ist: Konstruktion von Maßen und Integralen, Satz von Radon-Nikodym und reguläre bedingte Verteilungen, Konvergenzsätze für Funktionen (Lebesgue) und Maße (Prohorov) und Konstruktion von Maßen in Produkträumen. Die einzelnen maßtheoretischen Kapitel kommen nicht als Block am Anfang des Buches, obwohl sie so geschrieben sind, dass das möglich wäre, nämlich unabhängig von den wahrscheinlichkeitstheoretischen Kapiteln, sondern abwechselnd mit wahrscheinlichkeitstheoretischen Kapiteln, die so gebaut sind, dass sie mit den gerade zur Verfügung stehenden Begriffen auskommen (beispielsweise kann man Perkolation studieren, ohne einen Integralbegriff an der Hand zu haben). Als einzige Ausnahme wird die systematische Konstruktion von unabhängigen Zufallsvariablen erst im 14ten Kapitel nachgeliefert. Ich verspreche mir von diesem Vorgehen eine Auflockerung des maßtheoretischen Stoffes, der von manchen als etwas trocken empfunden wird. Letztlich ist dieses genauso eine Geschmacksfrage wie diejenige, welches der beiden Themen als linke und welches als rechte Hand anzusehen ist.

Wer eine maßtheoretische Grundbildung hat, kann insbesondere das erste Kapitel beim ersten Lesen zunächst überspringen und braucht eventuell nur Einzelnes darin nachzuschlagen. Das Gleiche gilt für das vierte Kapitel (Integrationstheorie).

In den ersten acht Kapiteln wird das Fundament gelegt, das in allen weiteren Kapiteln benötigt wird. Danach können die sieben inhaltlichen Einheiten von Kapitel 9–12, 13, 14, 15–16, 17–19, 20 und 23 einigermaßen unabhängig voneinander gelesen werden. Das Kapitel zur Brown'schen Bewegung (21) greift auf die Kapitel 9–15 zurück. Danach sind unabhängig voneinander die Blöcke 22, 24 und 25–26 lesbar.

Ich danke all denjenigen, die das Manuskript gelesen und zahlreiche Verbesserungsvorschläge und Korrekturen angebracht haben: Den Mitarbeitern und Studenten Ro-

land Alkemper, Dirk Brüggemann, Anne Eisenbürger, Ortwin Lorenz, Mario Oeler, Marcus Schölpen, den Kollegen Wolfgang Bühler und Wolfgang König sowie besonders dem Münchener Kollegen Hans-Otto Georgii. Für weitere Hinweise auf Fehler unter math@aklenke.de wäre ich dankbar.

Außerdem möchte ich mich beim Springer-Verlag für die gute Zusammenarbeit bedanken.

Mainz,
November 2005

Achim Klenke

Inhaltsverzeichnis

1	Grundlagen der Maßtheorie	1
1.1	Mengensysteme	1
1.2	Mengenfunktionen	12
1.3	Fortsetzung von Maßen	18
1.4	Messbare Abbildungen	34
1.5	Zufallsvariablen	43
2	Unabhängigkeit	49
2.1	Unabhängigkeit von Ereignissen	49
2.2	Unabhängigkeit von Zufallsvariablen	56
2.3	Kolmogorov'sches 0-1 Gesetz	63
2.4	Beispiel: Perkolation	67
3	Erzeugendenfunktion	79
3.1	Definition und Beispiele	79
3.2	Poisson-Approximation	82
3.3	Verzweigungsprozesse	84
4	Das Integral	87
4.1	Konstruktion und einfache Eigenschaften	87
4.2	Monotone Konvergenz und Lemma von Fatou	95
4.3	Lebesgue-Integral versus Riemann-Integral	98
5	Momente und Gesetze der Großen Zahl	103
5.1	Momente	103
5.2	Schwaches Gesetz der Großen Zahl	110

5.3	Starkes Gesetz der Großen Zahl	113
5.4	Konvergenzrate im starken GGZ	122
5.5	Der Poissonprozess	125
6	Konvergenzsätze	133
6.1	Fast-überall- und stochastische Konvergenz	133
6.2	Gleichgradige Integrierbarkeit	138
6.3	Vertauschung von Integral und Ableitung	145
7	L^p-Räume und Satz von Radon-Nikodym	147
7.1	Definitionen	147
7.2	Ungleichungen und Satz von Fischer-Riesz	149
7.3	Hilberträume	155
7.4	Lebesgue'scher Zerlegungssatz	158
7.5	Ergänzung: Signierte Maße	162
7.6	Ergänzung: Dualräume	169
8	Bedingte Erwartungen	173
8.1	Elementare bedingte Wahrscheinlichkeiten	173
8.2	Bedingte Erwartungen	177
8.3	Reguläre Version der bedingten Verteilung	184
9	Martingale	193
9.1	Prozesse, Filtrationen, Stoppzeiten	193
9.2	Martingale	198
9.3	Diskretes stochastisches Integral	202
9.4	Diskreter Martingaldarstellungssatz und CRR Modell	204
10	Optional Sampling Sätze	209
10.1	Doob-Zerlegung und quadratische Variation	209
10.2	Optional Sampling und Optional Stopping	213
10.3	Gleichgradige Integrierbarkeit und Optional Sampling	217
11	Martingalkonvergenzsätze und Anwendungen	221

11.1 Die Doob'sche Ungleichung	221
11.2 Martingalkonvergenzsätze	223
11.3 Beispiel: Verzweigungsprozess	233
12 Rückwärtsmartingale und Austauschbarkeit	235
12.1 Austauschbare Familien von Zufallsvariablen	235
12.2 Rückwärtsmartingale	240
12.3 Satz von de Finetti	243
13 Konvergenz von Maßen	249
13.1 Wiederholung Topologie	249
13.2 Schwache und vage Konvergenz	256
13.3 Der Satz von Prohorov	265
13.4 Anwendung: Satz von de Finetti – anders angeschaut	275
14 W-Maße auf Produkträumen	279
14.1 Produkträume	280
14.2 Endliche Produkte und Übergangskerne	283
14.3 Satz von Ionescu-Tulcea und Projektive Familien	292
14.4 Markov'sche Halbgruppen	297
15 Charakteristische Funktion und Zentraler Grenzwertsatz	301
15.1 Trennende Funktionenklassen	301
15.2 Charakteristische Funktionen: Beispiele	308
15.3 Der Lévy'sche Stetigkeitssatz	315
15.4 Charakteristische Funktion und Momente	320
15.5 Der Zentrale Grenzwertsatz	326
15.6 Mehrdimensionaler Zentraler Grenzwertsatz	334
16 Unbegrenzt teilbare Verteilungen	337
16.1 Die Lévy-Khinchin Formel	337
16.2 Stabile Verteilungen	349
17 Markovketten	357

17.1	Begriffsbildung und Konstruktion	357
17.2	Diskrete Markovketten, Beispiele	364
17.3	Diskrete Markovprozesse in stetiger Zeit	368
17.4	Diskrete Markovketten, Rekurrenz und Transienz	373
17.5	Anwendung: Rekurrenz und Transienz von Irrfahrten	377
17.6	Invariante Verteilungen	384
17.7	Anwendung: Stochastische Ordnung und Kopplung	390
18	Konvergenz von Markovketten	397
18.1	Periodizität von Markovketten	397
18.2	Kopplung und Konvergenzsatz	401
18.3	Markovketten Monte Carlo Methode	406
18.4	Konvergenzgeschwindigkeit	413
19	Markovketten und elektrische Netzwerke	419
19.1	Harmonische Funktionen	419
19.2	Reversible Markovketten	423
19.3	Endliche Elektrische Netzwerke	424
19.4	Rekurrenz und Transienz	430
19.5	Netzwerkreduktion	436
19.6	Irrfahrt in zufälliger Umgebung	445
20	Ergodentheorie	449
20.1	Begriffsbildung	449
20.2	Ergodensätze	453
20.3	Beispiele	455
20.4	Anwendung: Rekurrenz von Irrfahrten	457
20.5	Mischung	460
20.6	Entropie	463
21	Die Brown'sche Bewegung	467
21.1	Stetige Modifikationen	467
21.2	Konstruktion und Pfadeigenschaften	474

21.3 Starke Markoveigenschaft	479
21.4 Ergänzung: Feller Prozesse	482
21.5 Konstruktion durch L^2 -Approximation	485
21.6 Der Raum $C([0, \infty))$	492
21.7 Konvergenz von W-Maßen auf $C([0, \infty))$	494
21.8 Satz von Donsker.....	497
21.9 Pfadweise Konvergenz von Verzweigungsprozessen*	501
21.10Quadratische Variation und lokale Martingale	507
22 Gesetz vom iterierten Logarithmus	519
22.1 Iterierter Logarithmus für die Brown'sche Bewegung.....	519
22.2 Skorohod'scher Einbettungssatz	522
22.3 Satz von Hartman-Wintner	527
23 Große Abweichungen	529
23.1 Satz von Cramér	530
23.2 Prinzip der großen Abweichungen	534
23.3 Satz von Sanov	539
23.4 Varadhan'sches Lemma und Freie Energie	543
24 Der Poisson'sche Punktprozess	551
24.1 Zufällige Maße	551
24.2 Eigenschaften des Poisson'schen Punktprozesses	555
24.3 Die Poisson-Dirichlet-Verteilung*	562
25 Das Itô-Integral	571
25.1 Das Itô-Integral bezüglich der Brown'schen Bewegung	571
25.2 Itô-Integral bezüglich Diffusionen.....	580
25.3 Die Itô-Formel	583
25.4 Dirichlet-Problem und Brown'sche Bewegung	591
25.5 Rekurrenz und Transienz der Brown'schen Bewegung	593
26 Stochastische Differentialgleichungen	597
26.1 Starke Lösungen	597

26.2 Schwache Lösungen und Martingalproblem	606
26.3 Eindeutigkeit schwacher Lösungen via Dualität	613
Literatur	621
Notation	631
Glossar englischer Ausdrücke	635
Namensregister	637
Sachregister	641

1 Grundlagen der Maßtheorie

In diesem Kapitel führen wir die Mengensysteme ein, die eine systematische Beobachtung von Ereignissen und zufälligen Beobachtungen in der Wahrscheinlichkeitstheorie erlauben. Ferner sollen Maße, insbesondere Wahrscheinlichkeitsmaße, auf solchen Mengensystemen konstruiert werden. Schließlich werden wir Zufallsvariablen als messbare Abbildungen definieren.

1.1 Mengensysteme

Im Folgenden ist stets $\Omega \neq \emptyset$ eine Menge und $\mathcal{A} \subset 2^\Omega$ (Potenzmenge von Ω) eine Familie von Teilmengen. Später wird die Menge Ω als Raum von Elementarereignissen interpretiert werden und \mathcal{A} als ein System von beobachtbaren Ereignissen. Wir wollen in diesem Abschnitt Mengensysteme, die abgeschlossen sind unter einfachen mengentheoretischen Verknüpfungen, mit Namen versehen und einfache Beziehungen zwischen solchen Systemen herstellen.

Definition 1.1. Das Mengensystem \mathcal{A} heißt

- \cap -stabil (sprich: schnittstabil) oder ein **π -System**, falls für je zwei Mengen $A, B \in \mathcal{A}$ gilt, dass auch $A \cap B \in \mathcal{A}$,
- σ - \cap -stabil (sigma-schnittstabil), falls für je abzählbar unendlich viele Mengen $A_1, A_2, \dots \in \mathcal{A}$ gilt, dass auch $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$,
- \cup -stabil (vereinigungsstabil), falls für je zwei Mengen $A, B \in \mathcal{A}$ gilt, dass auch $A \cup B \in \mathcal{A}$,
- σ - \cup -stabil (sigma-vereinigungsstabil), falls für je abzählbar unendlich viele Mengen $A_1, A_2, \dots \in \mathcal{A}$ gilt, dass auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$,
- \setminus -stabil (differenzmengenstabil), falls für je zwei Mengen $A, B \in \mathcal{A}$ gilt, dass auch $A \setminus B \in \mathcal{A}$,
- komplementstabil, falls mit jeder Menge $A \in \mathcal{A}$ auch $A^c := \Omega \setminus A \in \mathcal{A}$ gilt.

Definition 1.2 (σ -Algebra). Ein Mengensystem $\mathcal{A} \subset 2^\Omega$ heißt **σ -Algebra**, falls die folgenden drei Bedingungen erfüllt sind.

- (i) $\Omega \in \mathcal{A}$,
- (ii) \mathcal{A} ist komplementstabil,
- (iii) \mathcal{A} ist σ - \cup -stabil.

σ -Algebren sind die natürlichen Mengensysteme für zufällige Ereignisse, denn wie wir sehen werden, können wir diesen Ereignissen in konsistenter Weise Wahrscheinlichkeiten zuordnen.

Satz 1.3. Ist \mathcal{A} komplementstabil, so gelten die beiden folgenden Äquivalenzen.

$$\begin{aligned}\mathcal{A} \text{ ist } \cap\text{-stabil} &\iff \mathcal{A} \text{ ist } \cup\text{-stabil}, \\ \mathcal{A} \text{ ist } \sigma\text{-}\cap\text{-stabil} &\iff \mathcal{A} \text{ ist } \sigma\text{-}\cup\text{-stabil}.\end{aligned}$$

Beweis. Dies folgt direkt aus den de Morgan'schen Regeln (Erinnerung: $(\bigcup A_i)^c = \bigcap A_i^c$). Ist beispielsweise \mathcal{A} σ - \cap -stabil und sind $A_1, A_2, \dots \in \mathcal{A}$, so ist auch

$$\bigcup_{n=1}^{\infty} A_n = \left(\bigcap_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{A}.$$

Also ist \mathcal{A} auch σ - \cup -stabil. Die anderen Fälle folgen analog. \square

Satz 1.4. Ist \mathcal{A} \setminus -stabil, so gelten die folgenden Aussagen.

- (i) \mathcal{A} ist \cap -stabil.
- (ii) Falls \mathcal{A} σ - \cup -stabil ist, dann ist \mathcal{A} auch σ - \cap -stabil.
- (iii) Jede abzählbare (beziehungsweise endliche) Vereinigung von Mengen aus \mathcal{A} lässt sich als abzählbare (beziehungsweise endliche), disjunkte Vereinigung von Mengen in \mathcal{A} schreiben.

Beweis. (i) Seien $A, B \in \mathcal{A}$. Dann ist auch $A \cap B = A \setminus (A \setminus B) \in \mathcal{A}$.

(ii) Seien $A_1, A_2, \dots \in \mathcal{A}$. Dann ist

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=2}^{\infty} (A_1 \cap A_n) = \bigcap_{n=2}^{\infty} A_1 \setminus (A_1 \setminus A_n) = A_1 \setminus \bigcup_{n=2}^{\infty} (A_1 \setminus A_n) \in \mathcal{A}.$$

(iii) Seien $A_1, A_2, \dots \in \mathcal{A}$. Dann ist $\bigcup_{n=1}^{\infty} A_n$ als abzählbare, disjunkte Vereinigung in \mathcal{A} darstellbar durch

$$\bigcup_{n=1}^{\infty} A_n = A_1 \uplus (A_2 \setminus A_1) \uplus ((A_3 \setminus A_1) \setminus A_2) \uplus (((A_4 \setminus A_1) \setminus A_2) \setminus A_3) \uplus \dots \quad \square$$

Bemerkung 1.5. Manchmal bezeichnen wir, wie im obigen Beweis, die Vereinigung paarweise disjunkter Mengen mit dem Symbol \uplus . Dies soll lediglich der optischen Verdeutlichung dienen und ist keine neue Verknüpfung. \diamond

Definition 1.6. Ein Mengensystem $\mathcal{A} \subset 2^\Omega$ heißt **Algebra**, falls gilt:

- (i) $\Omega \in \mathcal{A}$,
- (ii) \mathcal{A} ist \setminus -stabil,
- (iii) \mathcal{A} ist \cup -stabil.

Offenbar ist in einer Algebra stets $\emptyset = \Omega \setminus \Omega$ enthalten. Diese Eigenschaft ist im Allgemeinen jedoch schwächer als (i) in Definition 1.6.

Satz 1.7. Ein Mengensystem $\mathcal{A} \subset 2^\Omega$ ist genau dann eine Algebra, wenn es folgende drei Eigenschaften hat:

- (i) $\Omega \in \mathcal{A}$,
- (ii) \mathcal{A} ist komplementstabil,
- (iii) \mathcal{A} ist \cap -stabil.

Beweis. Übung! \square

Definition 1.8. Ein Mengensystem $\mathcal{A} \subset 2^\Omega$ heißt **Ring**, falls gilt:

- (i) $\emptyset \in \mathcal{A}$,
- (ii) \mathcal{A} ist \setminus -stabil,
- (iii) \mathcal{A} ist \cup -stabil.

Ein Ring heißt σ -**Ring**, falls er σ - \cup -stabil ist.

Definition 1.9. Ein Mengensystem $\mathcal{A} \subset 2^\Omega$ heißt **Semiring** (oder **Halbring**), falls gilt:

- (i) $\emptyset \in \mathcal{A}$,
- (ii) für je zwei Mengen $A, B \in \mathcal{A}$ ist $B \setminus A$ endliche Vereinigung von paarweise disjunkten Mengen aus \mathcal{A} ,
- (iii) \mathcal{A} ist \cap -stabil.

Definition 1.10. Ein Mengensystem $\mathcal{A} \subset 2^\Omega$ heißt **Dynkin-System** (oder λ -System), falls gilt:

- (i) $\Omega \in \mathcal{A}$,
- (ii) für je zwei Mengen $A, B \in \mathcal{A}$ mit $A \subset B$ ist $B \setminus A \in \mathcal{A}$,
- (iii) für je abzählbar viele, paarweise disjunkte Mengen $A_1, A_2, \dots \in \mathcal{A}$ gilt
 $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Beispiele 1.11. (i) Ist Ω eine beliebige nichtleere Menge, so sind $\mathcal{A} = \{\emptyset, \Omega\}$ und $\mathcal{A} = 2^\Omega$ die trivialen Beispiele für Algebren, σ -Algebren und Dynkin-Systeme. Hingegen sind $\mathcal{A} = \{\emptyset\}$ und $\mathcal{A} = \{\Omega\}$ die trivialen Beispiele für Semiringe, Ringe und σ -Ringe.

- (ii) Sei $\Omega = \mathbb{R}$. Dann ist $\mathcal{A} = \{A \subset \mathbb{R} : A \text{ ist abzählbar}\}$ ein σ -Ring.
- (iii) $\mathcal{A} = \{(a, b] : a, b \in \mathbb{R}, a \leq b\}$ ist ein Semiring über $\Omega = \mathbb{R}$ (aber kein Ring).
- (iv) Die Menge endlicher Vereinigungen von beschränkten Intervallen ist ein Ring über $\Omega = \mathbb{R}$ (aber keine Algebra).
- (v) Die Menge endlicher Vereinigungen beliebiger (auch unbeschränkter) Intervalle ist eine Algebra über $\Omega = \mathbb{R}$ (aber keine σ -Algebra).

- (vi) Sei E eine endliche, nichtleere Menge und $\Omega := E^\mathbb{N}$ die Menge aller Folgen $\omega = (\omega_n)_{n \in \mathbb{N}}$ mit Werten in E . Für $\omega_1, \dots, \omega_n \in E$ sei

$$[\omega_1, \dots, \omega_n] := \{\omega' \in \Omega : \omega'_i = \omega_i \text{ für jedes } i = 1, \dots, n\}$$

die Menge aller Folgen, die mit den Werten $\omega_1, \dots, \omega_n$ beginnen. Sei $\mathcal{A}_0 = \{\emptyset\}$. Für $n \in \mathbb{N}$ setze

$$\mathcal{A}_n := \{[\omega_1, \dots, \omega_n] : \omega_1, \dots, \omega_n \in E\}. \quad (1.1)$$

Dann ist $\mathcal{A} := \bigcup_{n=0}^{\infty} \mathcal{A}_n$ ein Semiring, aber kein Ring (falls $\#E > 1$).

- (vii) Sei Ω eine beliebige nichtleere Menge. Dann ist

$$\mathcal{A} := \{A \subset \Omega : A \text{ oder } A^c \text{ ist endlich}\}$$

eine Algebra. Ist $\#\Omega = \infty$, so ist \mathcal{A} jedoch keine σ -Algebra.

- (viii) Sei Ω eine beliebige nichtleere Menge. Dann ist

$$\mathcal{A} := \{A \subset \Omega : A \text{ oder } A^c \text{ ist abzählbar}\}$$

eine σ -Algebra.

- (ix) Jede σ -Algebra ist auch ein Dynkin-System.

- (x) Sei $\Omega = \{1, 2, 3, 4\}$ und $\mathcal{A} = \{\emptyset, \{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \{1, 2, 3, 4\}\}$. Dann ist \mathcal{A} ein Dynkin-System, aber keine Algebra. \diamond

Satz 1.12 (Inklusionen zwischen Mengensystemen).

- (i) Jede σ -Algebra ist ein Dynkin-System, eine Algebra und ein σ -Ring.
- (ii) Jeder σ -Ring ist ein Ring, jeder Ring ein Semiring.
- (iii) Jede Algebra ist auch ein Ring. Eine Algebra auf einer endlichen Menge Ω ist auch eine σ -Algebra.

Beweis. (i) Das ist klar.

(ii) Sei \mathcal{A} ein Ring. Nach Satz 1.4 ist \mathcal{A} schmittstabil und damit ein Semiring.

(iii) Sei \mathcal{A} eine Algebra. Dann ist $\emptyset = \Omega \setminus \Omega \in \mathcal{A}$, also ist \mathcal{A} ein Ring. Ist zudem Ω endlich, so ist \mathcal{A} endlich und damit jede abzählbare Vereinigung in \mathcal{A} schon eine endliche Vereinigung. \square

Definition 1.13 (liminf und limsup). Es seien A_1, A_2, \dots Teilmengen von Ω . Dann heißen

$$\liminf_{n \rightarrow \infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m \quad \text{und} \quad \limsup_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

Limes inferior beziehungsweise *Limes superior* der Folge $(A_n)_{n \in \mathbb{N}}$.

Bemerkung 1.14. (i) Es gilt

$$\begin{aligned} \liminf_{n \rightarrow \infty} A_n &= \{\omega \in \Omega : \#\{n \in \mathbb{N} : \omega \notin A_n\} < \infty\}, \\ \limsup_{n \rightarrow \infty} A_n &= \{\omega \in \Omega : \#\{n \in \mathbb{N} : \omega \in A_n\} = \infty\}. \end{aligned}$$

Der Limes inferior ist also das Ereignis, dass *schließlich alle* der A_n eintreten, der Limes superior hingegen das Ereignis, dass unendlich viele der A_n eintreten. Insbesondere ist $A_* := \liminf_{n \rightarrow \infty} A_n \subset A^* := \limsup_{n \rightarrow \infty} A_n$.

(ii) Bezeichnen wir mit

$$\mathbb{1}_A(x) := \begin{cases} 1, & \text{falls } x \in A, \\ 0, & \text{falls } x \notin A, \end{cases} \quad (1.2)$$

die **Indikatorfunktion** auf der Menge A , so gilt

$$\mathbb{1}_{A_*} = \liminf_{n \rightarrow \infty} \mathbb{1}_{A_n}, \quad \mathbb{1}_{A^*} = \limsup_{n \rightarrow \infty} \mathbb{1}_{A_n}.$$

(iii) Ist $\mathcal{A} \subset 2^\Omega$ eine σ -Algebra und $A_n \in \mathcal{A}$ für jedes $n \in \mathbb{N}$, so ist $A_* \in \mathcal{A}$ und $A^* \in \mathcal{A}$. \diamond

Beweis. Übung! \square

Satz 1.15 (Schnitt von Mengensystemen). Ist I eine beliebige Indexmenge und \mathcal{A}_i eine σ -Algebra für jedes $i \in I$, so ist

$$\mathcal{A}_I := \{A \subset \Omega : A \in \mathcal{A}_i \text{ für jedes } i \in I\} = \bigcap_{i \in I} \mathcal{A}_i$$

eine σ -Algebra. Dies gilt analog für: Ringe, σ -Ringe, Algebren und Dynkin-Systeme; nicht aber für Semiringe.

Beweis. Wir führen den Beweis hier nur für σ -Algebren durch. Wir prüfen für \mathcal{A}_I die Punkte (i)–(iii) aus Definition 1.2.

- (i) Für jedes $i \in I$ ist $\Omega \in \mathcal{A}_i$. Also ist $\Omega \in \mathcal{A}_I$.
- (ii) Sei $A \in \mathcal{A}_I$. Dann ist $A \in \mathcal{A}_i$ für jedes $i \in I$. Also ist auch $A^c \in \mathcal{A}_i$ für jedes $i \in I$. Mithin ist $A^c \in \mathcal{A}_I$.
- (iii) Seien $A_1, A_2, \dots \in \mathcal{A}_I$. Dann ist $A_n \in \mathcal{A}_i$ für jedes $n \in \mathbb{N}$ und jedes $i \in I$. Also ist auch $A := \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_i$ für jedes $i \in I$ und damit $A \in \mathcal{A}_I$.

Gegenbeispiel für Semiringe: Seien $\Omega = \{1, 2, 3, 4\}$, $\mathcal{A}_1 = \{\emptyset, \Omega, \{1\}, \{2, 3\}, \{4\}\}$ und $\mathcal{A}_2 = \{\emptyset, \Omega, \{1\}, \{2\}, \{3, 4\}\}$. Dann sind \mathcal{A}_1 und \mathcal{A}_2 Semiringe, aber $\mathcal{A}_1 \cap \mathcal{A}_2 = \{\emptyset, \Omega, \{1\}\}$ ist keiner. \square

Satz 1.16 (Erzeugte σ -Algebra). Sei $\mathcal{E} \subset 2^\Omega$. Dann existiert eine kleinste σ -Algebra $\sigma(\mathcal{E})$ mit $\mathcal{E} \subset \sigma(\mathcal{E})$:

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathcal{A} \subset 2^\Omega \text{ ist } \sigma\text{-Algebra} \\ \mathcal{A} \supseteq \mathcal{E}}} \mathcal{A}.$$

$\sigma(\mathcal{E})$ heißt die von \mathcal{E} erzeugte σ -Algebra. \mathcal{E} heißt Erzeuger von $\sigma(\mathcal{E})$. Analog wird das von \mathcal{E} erzeugte Dynkin-System $\delta(\mathcal{E})$ definiert.

Beweis. $\mathcal{A} = 2^\Omega$ ist eine σ -Algebra mit $\mathcal{E} \subset \mathcal{A}$. Also ist der Schnitt nicht leer. Nach Satz 1.15 ist $\sigma(\mathcal{E})$ eine σ -Algebra, und dies ist offenbar die kleinste σ -Algebra, die \mathcal{E} enthält. Für Dynkin-Systeme geht der Beweis genauso. \square

Bemerkung 1.17. Es gelten die folgenden einfachen Aussagen.

- (i) $\mathcal{E} \subset \sigma(\mathcal{E})$.
- (ii) Gilt $\mathcal{E}_1 \subset \mathcal{E}_2$, so ist $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_2)$.
- (iii) \mathcal{A} ist genau dann σ -Algebra, wenn $\sigma(\mathcal{A}) = \mathcal{A}$.

Die analogen Aussagen gelten für Dynkin-Systeme. Ferner ist stets $\delta(\mathcal{E}) \subset \sigma(\mathcal{E})$. \diamond

Satz 1.18 (Schnittstabiles Dynkin-System). Ist $\mathcal{D} \subset 2^\Omega$ ein Dynkin-System, so gilt

$$\mathcal{D} \text{ ist } \cap\text{-stabil} \iff \mathcal{D} \text{ ist eine } \sigma\text{-Algebra.}$$

Beweis. „ \Leftarrow “ Dies ist klar.

„ \Rightarrow “ Wir prüfen die Eigenschaften (i)–(iii) aus Definition 1.2.

- (i) Offensichtlich ist $\Omega \in \mathcal{D}$.
- (ii) (Komplementstabilität) Sei $A \in \mathcal{D}$. Da $\Omega \in \mathcal{D}$ gilt, und nach Eigenschaft (ii) des Dynkin-Systems, ist $A^c = \Omega \setminus A \in \mathcal{D}$.
- (iii) ($\sigma\cup$ -Stabilität) Seien $A, B \in \mathcal{D}$. Nach Voraussetzung ist $A \cap B \in \mathcal{D}$, und es gilt trivialerweise $A \cap B \subset A$. Also ist $A \setminus B = A \setminus (A \cap B) \in \mathcal{D}$. Mithin ist $\mathcal{D} \setminus$ -stabil. Seien nun $A_1, A_2, \dots \in \mathcal{D}$. Nach Satz 1.4(iii) existieren paarweise disjunkte Mengen $B_1, B_2, \dots \in \mathcal{D}$ mit $\bigcup_{n=1}^{\infty} A_n = \biguplus_{n=1}^{\infty} B_n \in \mathcal{D}$. \square

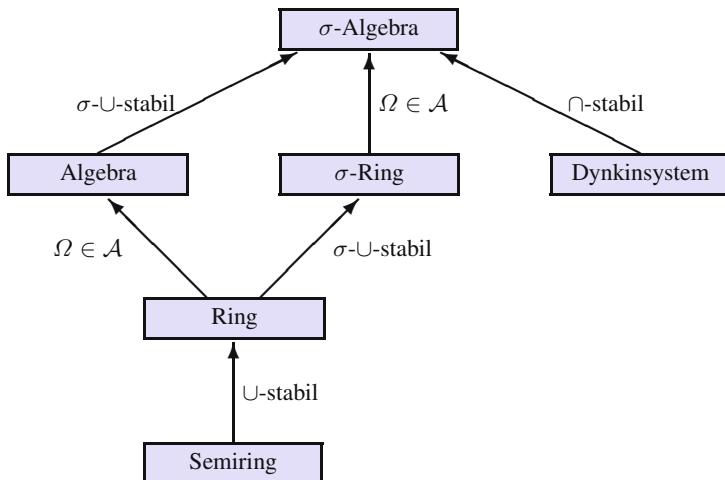


Abb. 1.1. Zusammenhang zwischen den Mengensystemen $\mathcal{A} \subset 2^\Omega$.

Satz 1.19 (Dynkin'scher $\pi-\lambda$ -Satz). Sei $\mathcal{E} \subset 2^\Omega$ ein \cap -stabiles Mengensystem. Dann gilt

$$\sigma(\mathcal{E}) = \delta(\mathcal{E}).$$

Beweis. „ \supseteq “ Dies ist klar nach Bemerkung 1.17.

„ \subseteq “ Zu zeigen ist: $\delta(\mathcal{E})$ ist eine σ -Algebra. Nach Satz 1.18 reicht es zu zeigen, dass $\delta(\mathcal{E}) \cap$ -stabil ist. Für $B \in \delta(\mathcal{E})$ sei

$$\mathcal{D}_B := \{A \in \delta(\mathcal{E}) : A \cap B \in \delta(\mathcal{E})\}.$$

Für die Schnittstabilität von $\delta(\mathcal{E})$ reicht es zu zeigen, dass

$$\delta(\mathcal{E}) \subset \mathcal{D}_B \quad \text{für jedes } B \in \delta(\mathcal{E}). \quad (1.3)$$

Wir zeigen, dass \mathcal{D}_E für jedes $E \in \delta(\mathcal{E})$ ein Dynkin-System ist, indem wir (i)–(iii) aus Definition 1.10 prüfen:

- (i) Offenbar ist $\Omega \cap E = E \in \delta(\mathcal{E})$, also ist $\Omega \in \mathcal{D}_E$.
- (ii) Für $A, B \in \mathcal{D}_E$ mit $A \subset B$ ist $(B \setminus A) \cap E = (B \cap E) \setminus (A \cap E) \in \delta(\mathcal{E})$.
- (iii) Seien $A_1, A_2, \dots \in \mathcal{D}_E$ paarweise disjunkt. Dann ist

$$\left(\bigcup_{n=1}^{\infty} A_n \right) \cap E = \biguplus_{n=1}^{\infty} (A_n \cap E) \in \delta(\mathcal{E}).$$

Nach Voraussetzung ist für $A, E \in \mathcal{E}$ auch $A \cap E \in \mathcal{E}$, also ist $\mathcal{E} \subset \mathcal{D}_E$, falls $E \in \mathcal{E}$ gilt. Nach Bemerkung 1.17(ii) ist daher auch $\delta(\mathcal{E}) \subset \mathcal{D}_E$ für $E \in \mathcal{E}$. Für $B \in \delta(\mathcal{E})$ und $E \in \mathcal{E}$ ist also $B \cap E \in \delta(\mathcal{E})$. Mithin gilt $E \in \mathcal{D}_B$ für jedes $B \in \delta(\mathcal{E})$, also $\mathcal{E} \subset \mathcal{D}_B$ für jedes $B \in \delta(\mathcal{E})$, und damit gilt (1.3). \square

Von besonderer Bedeutung sind σ -Algebren, die von Topologien erzeugt werden. Hier wiederum spielt natürlich der euklidische Raum \mathbb{R}^n die prominenteste Rolle, aber wir wollen auch den (unendlichdimensionalen) Raum $C([0, 1])$ der stetigen Funktionen $[0, 1] \rightarrow \mathbb{R}$ im Blick haben. Auf diesem Raum wird durch die Norm $\|f\|_\infty = \sup_{x \in [0, 1]} |f(x)|$ eine Topologie erzeugt. Zur Erinnerung bringen wir hier das Axiomensystem der Topologie.

Definition 1.20 (Topologie). Sei $\Omega \neq \emptyset$ eine beliebige Menge. Ein Mengensystem $\tau \subset 2^\Omega$ heißt **Topologie** auf Ω , falls folgende drei Eigenschaften gelten.

- (i) $\emptyset, \Omega \in \tau$.
- (ii) Sind $A, B \in \tau$, so ist auch $A \cap B \in \tau$.
- (iii) Ist $\mathcal{F} \subset \tau$ eine beliebige Familie, so ist auch $(\bigcup_{A \in \mathcal{F}} A) \in \tau$.

Das Paar (Ω, τ) heißt dann **topologischer Raum**. Die Mengen $A \in \tau$ heißen **offen**, die Mengen $A \subset \Omega$ mit $A^c \in \tau$ heißen **abgeschlossen**.

Anders als bei σ -Algebren sind bei Topologien nur endliche Schnitte, jedoch auch überabzählbare Vereinigungen erlaubt. Ist d eine Metrik auf Ω , und bezeichnet

$$B_r(x) = \{y \in \Omega : d(x, y) < r\}$$

die offene Kugel um $x \in \Omega$ mit Radius $r > 0$, so wird eine Topologie erzeugt durch

$$\tau = \left\{ \bigcup_{(x,r) \in F} B_r(x) : F \subset \Omega \times (0, \infty) \right\}.$$

Dies ist das gewöhnliche System offener Mengen, das man in den meisten Analysisbüchern findet.

Definition 1.21 (Borel'sche σ -Algebra). Sei (Ω, τ) ein topologischer Raum. Die von den offenen Mengen erzeugte σ -Algebra

$$\mathcal{B}(\Omega) := \mathcal{B}(\Omega, \tau) := \sigma(\tau)$$

heißt **Borel'sche σ -Algebra** auf Ω . Die Elemente $A \in \mathcal{B}(\Omega, \tau)$ heißen **Borel'sche Mengen** oder **Borel-messbare Mengen**.

Bemerkung 1.22. Wir sind meistens an $\mathcal{B}(\mathbb{R}^n)$ interessiert, wobei wir auf \mathbb{R}^n den euklidischen Abstand annehmen:

$$d(x, y) = \|x - y\|_2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

(i) Es gibt Teilmengen von \mathbb{R}^n , die keine Borel'schen Mengen sind. Diese sind kompliziert herzustellen, wie beispielsweise die **Vitali-Mengen**, die man in Analysisbüchern findet (siehe etwa [8]). Wir wollen hier auf diesen Aspekt nicht näher eingehen, sondern lediglich die - mathematisch unpräzise - Feststellung treffen, dass jede Menge, die man sich konstruktiv herstellen kann, auch Borel'sch ist.

(ii) Jede abgeschlossene Menge $C \subset \mathbb{R}^n$ ist in $\mathcal{B}(\mathbb{R}^n)$, denn es ist $C^c \in \tau$, also ist $C = (C^c)^c \in \sigma(\tau)$. Speziell ist $\{x\} \in \mathcal{B}(\mathbb{R}^n)$ für jedes $x \in \mathbb{R}^n$.

(iii) $\mathcal{B}(\mathbb{R}^n)$ ist keine Topologie. Sei nämlich $V \subset \mathbb{R}^n$, $V \notin \mathcal{B}(\mathbb{R}^n)$. Wäre $\mathcal{B}(\mathbb{R}^n)$ eine Topologie, so wären beliebige Vereinigungen Borel'scher Mengen wieder Borel'sch, also auch $V = \bigcup_{x \in V} \{x\} \in \mathcal{B}(\mathbb{R}^n)$. \diamond

Das Mengensystem der offenen Mengen, das die Borel'sche σ -Algebra erzeugt, ist in vielen Fällen unhandlich groß. Wir wollen daher andere Mengensysteme als Erzeuger von $\mathcal{B}(\mathbb{R}^n)$ identifizieren, mit denen wir in der Praxis besser arbeiten können. Hierzu wollen wir einerseits Mengen von einfacher Struktur, Quader etwa, betrachten, andererseits aber auch die Größe des Systems einschränken, indem wir abzählbare Mengensysteme betrachten. Wir führen folgende Notationen ein. Mit \mathbb{Q} bezeichnen wir die Menge der rationalen Zahlen, mit \mathbb{Q}^+ die Menge der strikt positiven rationalen Zahlen. Für $a, b \in \mathbb{R}^n$ schreiben wir

$$a < b, \quad \text{falls } a_i < b_i \quad \text{für jedes } i = 1, \dots, n. \quad (1.4)$$

Wir definieren für $a < b$ den offenen **Quader** als das kartesische Produkt

$$(a, b) := \bigtimes_{i=1}^n (a_i, b_i) := (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n) \quad (1.5)$$

und analog $[a, b]$, $(a, b]$ und $[a, b)$. Ferner schreiben wir $(-\infty, b) := \times_{i=1}^n (-\infty, b_i)$ und definieren analog $(-\infty, b]$ und so fort. Wir führen die folgenden Mengensysteme ein:

$$\begin{aligned}\mathcal{E}_1 &:= \{A \subset \mathbb{R}^n : A \text{ ist offen}\}, & \mathcal{E}_2 &:= \{A \subset \mathbb{R}^n : A \text{ ist abgeschlossen}\}, \\ \mathcal{E}_3 &:= \{A \subset \mathbb{R}^n : A \text{ ist kompakt}\}, & \mathcal{E}_4 &:= \{B_r(x) : x \in \mathbb{Q}^n, r \in \mathbb{Q}^+\}, \\ \mathcal{E}_5 &:= \{(a, b) : a, b \in \mathbb{Q}^n, a < b\}, & \mathcal{E}_6 &:= \{[a, b) : a, b \in \mathbb{Q}^n, a < b\}, \\ \mathcal{E}_7 &:= \{(a, b] : a, b \in \mathbb{Q}^n, a < b\}, & \mathcal{E}_8 &:= \{[a, b] : a, b \in \mathbb{Q}^n, a < b\}, \\ \mathcal{E}_9 &:= \{(-\infty, b) : b \in \mathbb{Q}^n\}, & \mathcal{E}_{10} &:= \{(-\infty, b] : b \in \mathbb{Q}^n\}, \\ \mathcal{E}_{11} &:= \{(a, \infty) : a \in \mathbb{Q}^n\}, & \mathcal{E}_{12} &:= \{[a, \infty) : a \in \mathbb{Q}^n\}.\end{aligned}$$

Satz 1.23. Die Borel'sche σ -Algebra $\mathcal{B}(\mathbb{R}^n)$ wird von jedem der Mengensysteme $\mathcal{E}_1, \dots, \mathcal{E}_{12}$ erzeugt: $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{E}_i)$ für jedes $i = 1, \dots, 12$.

Beweis. Wir zeigen nur exemplarisch ein paar der Identitäten.

(1) $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{E}_1)$ gilt per Definition.

(2) Sei $A \in \mathcal{E}_1$. Dann ist $A^c \in \mathcal{E}_2$, also $A = (A^c)^c \in \sigma(\mathcal{E}_2)$. Daher gilt $\mathcal{E}_1 \subset \sigma(\mathcal{E}_2)$ und dann (wegen Bemerkung 1.17) auch $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_2)$. Analog folgt aber $\sigma(\mathcal{E}_2) \subset \sigma(\mathcal{E}_1)$ und damit die Gleichheit.

(3) Jede kompakte Menge ist abgeschlossen. Also gilt $\sigma(\mathcal{E}_3) \subset \sigma(\mathcal{E}_2)$. Sei nun $A \in \mathcal{E}_2$. Dann sind die Mengen $A_K := A \cap [-K, K]^n$, $K \in \mathbb{N}$, kompakt, also ist die abzählbare Vereinigung $A = \bigcup_{K=1}^{\infty} A_K$ in $\sigma(\mathcal{E}_3)$. Es gilt also $\mathcal{E}_2 \subset \sigma(\mathcal{E}_3)$ und damit $\sigma(\mathcal{E}_2) = \sigma(\mathcal{E}_3)$.

(4) Offenbar ist $\mathcal{E}_4 \subset \mathcal{E}_1$, also $\sigma(\mathcal{E}_4) \subset \sigma(\mathcal{E}_1)$. Sei nun $A \subset \mathbb{R}^n$ offen. Für $x \in A$ sei $R(x) = \min(1, \sup\{r > 0 : B_r(x) \subset A\})$. Da A offen ist, folgt $R(x) > 0$. Sei $r(x) \in (R(x)/2, R(x)) \cap \mathbb{Q}$. Für jedes $y \in A$ und $x \in (B_{R(y)/3}(y)) \cap \mathbb{Q}^n$ ist nun $R(x) \geq R(y) - \|x - y\|_2 > \frac{2}{3}R(y)$, also $r(x) > \frac{1}{3}R(y)$, also $y \in B_{r(x)}(x)$. Also ist $A = \bigcup_{x \in A \cap \mathbb{Q}^n} B_{r(x)}(x)$ eine abzählbare Vereinigung von Mengen aus \mathcal{E}_4 und damit in $\sigma(\mathcal{E}_4)$. Es gilt also auch $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_4)$.

(5–12) Ähnliche Ausschöpfungsargumente wie in (4) funktionieren auch für die Quader. In (4) können statt der offenen Kugeln $B_r(x)$ offene Quader genommen werden. So folgt die Gleichheit mit $\sigma(\mathcal{E}_5)$. Man bemerke beispielsweise, dass

$$\bigtimes_{i=1}^n [a_i, b_i) = \bigcap_{k=1}^{\infty} \bigtimes_{i=1}^n \left(a_i - \frac{1}{k}, b_i\right) \in \sigma(\mathcal{E}_5).$$

Die anderen Inklusionen $\mathcal{E}_i \subset \sigma(\mathcal{E}_j)$ zeigt man analog. \square

Bemerkung 1.24. Jedes der Mengensysteme $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_5, \dots, \mathcal{E}_{12}$ (nicht aber \mathcal{E}_4) ist schnittstabil, mithin ist die Borel'sche σ -Algebra jeweils gleich dem erzeugten

Dynkin-System: $\mathcal{B}(\mathbb{R}^n) = \delta(\mathcal{E}_i)$ für $i = 1, 2, 3, 5, \dots, 12$. Die Mengensysteme $\mathcal{E}_4, \dots, \mathcal{E}_{12}$ sind zudem abzählbar. Dies ist eine Eigenschaft, die wir an späterer Stelle wieder benötigen werden. \diamond

Definition 1.25 (Spur eines Mengensystems). Es sei $\mathcal{A} \subset 2^\Omega$ ein beliebiges System von Teilmengen von Ω und $A \in 2^\Omega \setminus \{\emptyset\}$. Das Mengensystem

$$\mathcal{A}|_A := \{A \cap B : B \in \mathcal{A}\} \subset 2^A \quad (1.6)$$

heißt **Spur** von \mathcal{A} auf A , oder **Einschränkung** von \mathcal{A} auf A .

Satz 1.26. Ist \mathcal{A} eine σ -Algebra, oder eines der Mengensysteme aus den Definitionen 1.6 – 1.9 auf Ω , so ist $\mathcal{A}|_A$ ein Mengensystem vom selben Typ, allerdings auf A statt Ω . Für Dynkin-Systeme ist im Allgemeinen nicht richtig.

Beweis. Übung! \square

Übung 1.1.1. Sei \mathcal{A} ein Semiring. Man zeige: Jede abzählbare (beziehungsweise endliche) Vereinigung von Mengen aus \mathcal{A} lässt sich als abzählbare (beziehungsweise endliche), *disjunkte* Vereinigung von Mengen in \mathcal{A} schreiben. \clubsuit

Übung 1.1.2. Man zeige durch ein Gegenbeispiel, dass im Allgemeinen die Vereinigung $\mathcal{A} \cup \mathcal{A}'$ zweier σ -Algebren keine σ -Algebra ist. \clubsuit

Übung 1.1.3. Seien (Ω_1, d_1) und (Ω_2, d_2) metrische Räume, $f : \Omega_1 \rightarrow \Omega_2$ eine beliebige Abbildung und $U_f = \{x \in \Omega_1 : f \text{ ist unstetig in } x\}$ die Menge der Unstetigkeitsstellen. Man zeige: $U_f \in \mathcal{B}(\Omega_1)$.

Hinweis: Man zeige zunächst, dass für $\varepsilon > 0$ und $\delta > 0$ die Menge

$$U_f^{\delta, \varepsilon} := \{x \in \Omega_1 : \text{ es gibt } y, z \in B_\varepsilon(x) \text{ mit } d_2(f(y), f(z)) > \delta\}$$

(wobei $B_\varepsilon(x) = \{y \in \Omega_1 : d_1(x, y) < \varepsilon\}$) offen ist und konstruiere dann U_f aus solchen Mengen. \clubsuit

Übung 1.1.4. Sei Ω eine überabzählbare Menge und $\mathcal{A} = \sigma(\{\omega\} : \omega \in \Omega)$. Zeige:

$$\mathcal{A} = \{A \subset \Omega : A \text{ ist abzählbar oder } A^c \text{ ist abzählbar}\}. \quad \clubsuit$$

Übung 1.1.5. Sei \mathcal{A} ein Ring auf der Menge Ω . Man zeige: \mathcal{A} erfüllt die Axiome eines kommutativen Rings (im Sinne der Algebra) mit „ \cap “ als Multiplikation und „ \triangle “ als Addition. \clubsuit

1.2 Mengenfunktionen

Definition 1.27. Sei $\mathcal{A} \subset 2^\Omega$ und $\mu : \mathcal{A} \rightarrow [0, \infty]$ eine Mengenfunktion. μ heißt

- (i) **monoton**, falls für je zwei Mengen $A, B \in \mathcal{A}$ mit $A \subset B$ gilt, dass $\mu(A) \leq \mu(B)$,
- (ii) **additiv**, falls für je endlich viele paarweise disjunkte Mengen $A_1, \dots, A_n \in \mathcal{A}$ mit $\bigcup_{i=1}^n A_i \in \mathcal{A}$ gilt, dass $\mu\left(\biguplus_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$,
- (iii) **σ -additiv**, falls für je abzählbar viele paarweise disjunkte Mengen A_1, A_2, \dots aus \mathcal{A} mit $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$ gilt, dass $\mu\left(\biguplus_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mu(A_i)$,
- (iv) **subadditiv**, falls für je endlich viele Mengen $A, A_1, A_2, \dots, A_n \in \mathcal{A}$ mit $A \subset \bigcup_{i=1}^n A_i$ gilt, dass $\mu(A) \leq \sum_{i=1}^n \mu(A_i)$,
- (v) **σ -subadditiv**, falls für je abzählbar viele $A, A_1, A_2, \dots \in \mathcal{A}$ mit $A \subset \bigcup_{i=1}^\infty A_i$ gilt, dass $\mu(A) \leq \sum_{i=1}^\infty \mu(A_i)$.

Definition 1.28. Sei \mathcal{A} ein Semiring und $\mu : \mathcal{A} \rightarrow [0, \infty]$ eine Mengenfunktion mit $\mu(\emptyset) = 0$. μ heißt

- **Inhalt**, falls μ additiv ist,
- **Prämaß**, falls μ σ -additiv ist,
- **Maß**, falls μ ein Prämaß ist und \mathcal{A} eine σ -Algebra,
- **Wahrscheinlichkeitsmaß** (kurz **W-Maß**), falls μ ein Maß ist und $\mu(\Omega) = 1$.

Definition 1.29. Sei \mathcal{A} ein Semiring. Ein Inhalt μ auf \mathcal{A} heißt

- (i) **endlich**, falls $\mu(A) < \infty$ für jedes $A \in \mathcal{A}$,
- (ii) **σ -endlich**, falls es Mengen $\Omega_1, \Omega_2, \dots \in \mathcal{A}$ gibt mit $\Omega = \bigcup_{n=1}^\infty \Omega_n$ und $\mu(\Omega_n) < \infty$ für jedes $n \in \mathbb{N}$.

Beispiel 1.30 (Inhalte, Maße). (i) Sei $\omega \in \Omega$ und $\delta_\omega(A) = \mathbb{1}_A(\omega)$ (siehe (1.2)). Dann ist δ_ω ein Wahrscheinlichkeitsmaß auf jeder σ -Algebra $\mathcal{A} \subset 2^\Omega$ und heißt **Dirac-Maß** im Punkt ω , oder **Einheitsmasse**.

(ii) Sei Ω eine endliche, nichtleere Menge. Durch

$$\mu(A) := \frac{\#A}{\#\Omega} \quad \text{für } A \subset \Omega,$$

wird ein Wahrscheinlichkeitsmaß auf $\mathcal{A} = 2^\Omega$ definiert. μ heißt **Gleichverteilung** oder **uniforme Verteilung** auf Ω . Wir führen hierfür das Symbol $\mathcal{U}_\Omega := \mu$ ein. Das so definierte Tripel $(\Omega, \mathcal{A}, \mathcal{U}_\Omega)$ wird auch **Laplace-Raum** genannt.

(iii) Sei Ω abzählbar unendlich und

$$\mathcal{A} := \{A \subset \Omega : \#A < \infty \text{ oder } \#A^c < \infty\}.$$

Dann ist \mathcal{A} eine Algebra. Die durch

$$\mu(A) = \begin{cases} 0, & \text{falls } A \text{ endlich,} \\ \infty, & \text{falls } A^c \text{ endlich,} \end{cases}$$

auf \mathcal{A} definierte Mengenfunktion ist ein Inhalt, aber kein Prämaß, denn es gilt $\mu(\bigcup_{\omega \in \Omega} \{\omega\}) = \mu(\Omega) = \infty$, aber $\sum_{\omega \in \Omega} \mu(\{\omega\}) = 0$.

(iv) Sei $(\mu_n)_{n \in \mathbb{N}}$ eine Folge von Maßen (Prämaßen, Inhalten) und $(\alpha_n)_{n \in \mathbb{N}}$ eine Folge von nichtnegativen Zahlen. Dann ist auch $\mu := \sum_{n=1}^{\infty} \alpha_n \mu_n$ ein Maß (Prämaß, Inhalt).

(v) Sei Ω eine (höchstens) abzählbare, nichtleere Menge und $\mathcal{A} = 2^\Omega$. Ferner seien $(p_\omega)_{\omega \in \Omega}$ nichtnegative Zahlen. Dann wird durch $\mu(A) := \sum_{\omega \in A} p_\omega$ für jedes $A \subset \Omega$, ein σ -endliches Maß auf 2^Ω definiert. Wir nennen $p = (p_\omega)_{\omega \in \Omega}$ die **Gewichtsfunktion** von μ .

(vi) Ist in (v) speziell $\sum_{\omega \in \Omega} p_\omega = 1$, so ist μ ein Wahrscheinlichkeitsmaß. Wir interpretieren dann p_ω als Wahrscheinlichkeit des Elementarereignisses ω und nennen $p = (p_\omega)_{\omega \in \Omega}$ auch einen **Wahrscheinlichkeitsvektor**.

(vii) Ist in (v) speziell $p_\omega = 1$ für jedes $\omega \in \Omega$, so heißt μ das **Zählmaß** auf Ω . Ist Ω endlich, so ist auch μ endlich.

(viii) Sei \mathcal{A} der Ring endlicher Vereinigungen von Intervallen $(a, b] \subset \mathbb{R}$. Für $a_1 < b_1 < a_2 < b_2 < \dots < b_n$ und $A = \biguplus_{i=1}^n (a_i, b_i]$ setzen wir

$$\mu(A) = \sum_{i=1}^n (b_i - a_i).$$

μ ist ein σ -endlicher Inhalt auf \mathcal{A} (sogar ein Prämaß), denn es ist $\bigcup_{n=1}^{\infty} (-n, n] = \mathbb{R}$ und $\mu((-n, n]) = 2n < \infty$ für jedes $n \in \mathbb{N}$.

(ix) Sei $f : \mathbb{R} \rightarrow [0, \infty)$ stetig. Analog zu (viii) setze

$$\mu_f(A) = \sum_{i=1}^n \int_{a_i}^{b_i} f(x) dx.$$

μ_f ist ein σ -endlicher Inhalt auf \mathcal{A} (sogar ein Prämaß). Die Funktion f heißt **Dichte** und spielt hier eine ähnliche Rolle wie die Gewichtsfunktion p in (v). \diamond

Lemma 1.31 (Eigenschaften von Inhalten). Sei \mathcal{A} ein Semiring und μ ein Inhalt auf \mathcal{A} . Dann gelten die folgenden Aussagen.

- (i) Ist \mathcal{A} ein Ring, so ist $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$ für je zwei Mengen $A, B \in \mathcal{A}$.
- (ii) μ ist monoton. Ist \mathcal{A} ein Ring, so gilt genauer $\mu(B) = \mu(A) + \mu(B \setminus A)$ für je zwei Mengen $A, B \in \mathcal{A}$ mit $A \subset B$.
- (iii) μ ist subadditiv. Ist μ sogar σ -additiv, so ist μ auch σ -subadditiv.
- (iv) Ist \mathcal{A} ein Ring, so gilt für je abzählbar viele, paarweise disjunkte Mengen $A_1, A_2, \dots \in \mathcal{A}$ mit $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ stets $\sum_{n=1}^{\infty} \mu(A_n) \leq \mu\left(\bigcup_{n=1}^{\infty} A_n\right)$.

Beweis. (i) Es ist $A \cup B = A \uplus (B \setminus A)$ und $B = (A \cap B) \uplus (B \setminus A)$. Da μ additiv ist, folgt

$$\mu(A \cup B) = \mu(A) + \mu(B \setminus A) \quad \text{und} \quad \mu(B) = \mu(A \cap B) + \mu(B \setminus A).$$

Hieraus folgt sofort (i).

(ii) Sei $A \subset B$. Wegen $A \cap B = A$ folgt $\mu(B) = \mu(A \uplus (B \setminus A)) = \mu(A) + \mu(B \setminus A)$, falls $B \setminus A \in \mathcal{A}$ ist, insbesondere also, falls \mathcal{A} ein Ring ist. Ist nun \mathcal{A} nur ein Semiring, so ist $B \setminus A = \biguplus_{i=1}^n C_i$ für gewisses $n \in \mathbb{N}$ und paarweise disjunkte Mengen $C_1, \dots, C_n \in \mathcal{A}$. In diesem Fall ist $\mu(B) = \mu(A) + \sum_{i=1}^n \mu(C_i) \geq \mu(A)$, also ist μ monoton.

(iii) Seien $n \in \mathbb{N}$ und $A, A_1, \dots, A_n \in \mathcal{A}$ mit $A \subset \bigcup_{i=1}^n A_i$. Setze $B_1 = A_1$ und

$$B_k = A_k \setminus \bigcup_{i=1}^{k-1} A_i = \bigcap_{i=1}^{k-1} (A_k \setminus (A_k \cap A_i)) \quad \text{für } k = 2, \dots, n.$$

Per Definition des Semirings ist jedes $A_k \setminus (A_k \cap A_i)$ disjunkte Vereinigung endlich vieler Mengen in \mathcal{A} , also existiert ein $c_k \in \mathbb{N}$ und Mengen $C_{k,1}, \dots, C_{k,c_k} \in \mathcal{A}$ mit $\biguplus_{i=1}^{c_k} C_{k,i} = B_k \subset A_k$. Analog existieren $d_k \in \mathbb{N}$ und $D_{k,1}, \dots, D_{k,d_k} \in \mathcal{A}$ mit $A_k \setminus B_k = \biguplus_{i=1}^{d_k} D_{k,i}$. Da μ additiv ist, gilt

$$\mu(A_k) = \sum_{i=1}^{c_k} \mu(C_{k,i}) + \sum_{i=1}^{d_k} \mu(D_{k,i}) \geq \sum_{i=1}^{c_k} \mu(C_{k,i}).$$

Wiederum aufgrund von Additivität und Monotonie gilt

$$\begin{aligned} \mu(A) &= \mu\left(\biguplus_{k=1}^n \biguplus_{i=1}^{c_k} (C_{k,i} \cap A)\right) = \sum_{k=1}^n \sum_{i=1}^{c_k} \mu(C_{k,i} \cap A) \\ &\leq \sum_{k=1}^n \sum_{i=1}^{c_k} \mu(C_{k,i}) \leq \sum_{k=1}^n \mu(A_k). \end{aligned}$$

Also ist μ subadditiv. Die σ -Subadditivität folgt aus der σ -Additivität in analoger Weise.

(iv) Sei \mathcal{A} ein Ring und $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. Da μ additiv (und damit monoton) ist, gilt nach (ii)

$$\sum_{n=1}^m \mu(A_n) = \mu\left(\biguplus_{n=1}^m A_n\right) \leq \mu(A) \quad \text{für jedes } m \in \mathbb{N}.$$

Also ist $\sum_{n=1}^{\infty} \mu(A_n) \leq \mu(A)$. □

Bemerkung 1.32. In (iv) kann strikte Ungleichheit herrschen (siehe etwa Beispiel 1.30(iii)). Mit anderen Worten: Es gibt Inhalte, die keine Prämäße sind. ◇

Satz 1.33 (Einschluss- Ausschlussformel). Sei \mathcal{A} ein Ring und μ ein Inhalt. Dann gelten für $n \in \mathbb{N}$ und $A_1, \dots, A_n \in \mathcal{A}$ mit $\mu(A_1 \cup \dots \cup A_n) < \infty$ die Einschluss-Ausschlussformeln

$$\begin{aligned} \mu(A_1 \cup \dots \cup A_n) &= \sum_{k=1}^n (-1)^{k-1} \sum_{\{i_1, \dots, i_k\} \subset \{1, \dots, n\}} \mu(A_{i_1} \cap \dots \cap A_{i_k}), \\ \mu(A_1 \cap \dots \cap A_n) &= \sum_{k=1}^n (-1)^{k-1} \sum_{\{i_1, \dots, i_k\} \subset \{1, \dots, n\}} \mu(A_{i_1} \cup \dots \cup A_{i_k}), \end{aligned}$$

wobei sich die Summen über alle k -elementigen Teilmengen von $\{1, \dots, n\}$ erstrecken.

Beweis. Übung! Hinweis: Man verwende vollständige Induktion über n . □

Wir wollen die σ -Subadditivität durch eine Stetigkeitseigenschaft charakterisieren (Satz 1.36). Hierzu verabreden wir die folgende Sprechweise und Notation.

Definition 1.34. Sind A, A_1, A_2, \dots Mengen, so schreiben wir

- $A_n \uparrow A$, falls $A_1 \subset A_2 \subset \dots$ und $\bigcup_{n=1}^{\infty} A_n = A$,
- $A_n \downarrow A$, falls $A_1 \supset A_2 \supset A_3 \supset \dots$ und $\bigcap_{n=1}^{\infty} A_n = A$.

Wir sagen dann, dass $(A_n)_{n \in \mathbb{N}}$ gegen A aufsteigt beziehungsweise absteigt.

Definition 1.35 (Stetigkeit von Inhalten). Sei μ ein Inhalt auf dem Ring \mathcal{A} .

- (i) μ heißt **stetig von unten**, falls für jedes $A \in \mathcal{A}$ und jede Folge $(A_n)_{n \in \mathbb{N}}$ in \mathcal{A} mit $A_n \uparrow A$ gilt: $\mu(A_n) \xrightarrow{n \rightarrow \infty} \mu(A)$.
- (ii) μ heißt **stetig von oben**, falls für jedes $A \in \mathcal{A}$ und jede Folge $(A_n)_{n \in \mathbb{N}}$ in \mathcal{A} mit $A_n \downarrow A$ sowie $\mu(A_n) < \infty$ für jedes $n \in \mathbb{N}$ gilt: $\mu(A_n) \xrightarrow{n \rightarrow \infty} \mu(A)$.
- (iii) μ heißt **\emptyset -stetig**, falls (ii) für $A = \emptyset$ gilt.

Bei der Stetigkeit von oben wurde die Endlichkeitsbedingung eingeführt, weil sogar für das Zählmaß μ auf $(\mathbb{N}, 2^{\mathbb{N}})$ und $A_n := \{n, n+1, \dots\} \downarrow \emptyset$ sonst keine Gleichheit gelten kann.

Satz 1.36 (Stetigkeit und Prämäβ). Sei μ ein Inhalt auf einem Ring \mathcal{A} . Betrachte die folgenden fünf Eigenschaften.

- (i) μ ist σ -additiv (also ein Prämäβ).
- (ii) μ ist σ -subadditiv.
- (iii) μ ist stetig von unten.
- (iv) μ ist \emptyset -stetig.
- (v) μ ist stetig von oben.

Dann gelten die Implikationen (i) \iff (ii) \iff (iii) \implies (iv) \iff (v).

Ist μ endlich, so gilt auch (iv) \implies (iii).

Beweis. „(i) \implies (ii)“ Seien $A, A_1, A_2, \dots \in \mathcal{A}$ mit $A \subset \bigcup_{i=1}^{\infty} A_i$. Setze $B_1 = A_1$ und $B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i \in \mathcal{A}$ für $n = 2, 3, \dots$ Dann ist $A = \biguplus_{n=1}^{\infty} (A \cap B_n)$, also wegen der Monotonie von μ und der σ -Additivität von μ

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap B_n) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

Damit ist μ als σ -subadditiv erkannt.

„(ii) \implies (i)“ Dies folgt aus Lemma 1.31(iv).

„(i) \implies (iii)“ Sei μ ein Prämäβ und $A \in \mathcal{A}$ sowie $(A_n)_{n \in \mathbb{N}}$ eine Folge in \mathcal{A} mit $A_n \uparrow A$ sowie $A_0 = \emptyset$. Dann gilt

$$\mu(A) = \sum_{i=1}^{\infty} \mu(A_i \setminus A_{i-1}) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(A_i \setminus A_{i-1}) = \lim_{n \rightarrow \infty} \mu(A_n).$$

„(iii) \implies (i)“ Gelte nun (iii). Seien $B_1, B_2, \dots \in \mathcal{A}$ paarweise disjunkt, und gelte $B = \biguplus_{n=1}^{\infty} B_n \in \mathcal{A}$. Setze $A_n = \bigcup_{i=1}^n B_i$ für jedes $n \in \mathbb{N}$. Dann folgt aus (iii)

$$\mu(B) = \lim_{n \rightarrow \infty} \mu(A_n) = \sum_{i=1}^{\infty} \mu(B_i).$$

Also ist μ σ -additiv und damit ein Prämaß.

„(iv) \implies (v)“ Seien $A, A_1, A_2, \dots \in \mathcal{A}$ mit $A_n \downarrow A$ und $\mu(A_1) < \infty$. Setze $B_n = A_n \setminus A \in \mathcal{A}$ für jedes $n \in \mathbb{N}$. Dann gilt $B_n \downarrow \emptyset$. Es gilt also $\mu(A_n) - \mu(A) = \mu(B_n) \xrightarrow{n \rightarrow \infty} 0$.

„(v) \implies (iv)“ Dies ist trivial.

„(iii) \implies (iv)“ Seien $A_1, A_2, \dots \in \mathcal{A}$ mit $A_n \downarrow \emptyset$ und $\mu(A_1) < \infty$. Dann gilt $A_1 \setminus A_n \in \mathcal{A}$ für jedes $n \in \mathbb{N}$ und $A_1 \setminus A_n \uparrow A_1$, also

$$\mu(A_1) = \lim_{n \rightarrow \infty} \mu(A_1 \setminus A_n) = \mu(A_1) - \lim_{n \rightarrow \infty} \mu(A_n).$$

Wegen $\mu(A_1) < \infty$ ist $\lim_{n \rightarrow \infty} \mu(A_n) = 0$.

„(iv) \implies (iii)“ (für den Fall μ endlich) Es gelte nun $\mu(A) < \infty$ für jedes $A \in \mathcal{A}$, und μ sei \emptyset -stetig. Seien $A, A_1, A_2, \dots \in \mathcal{A}$ mit $A_n \uparrow A$. Dann gilt $A \setminus A_n \downarrow \emptyset$ und

$$\mu(A) - \mu(A_n) = \mu(A \setminus A_n) \xrightarrow{n \rightarrow \infty} 0.$$

Also gilt (iii). \square

Beispiel 1.37. (Vergleiche Beispiel 1.30(iii).) Sei Ω abzählbar unendlich und

$$\mathcal{A} = \{A \subset \Omega : \#A < \infty \text{ oder } \#A^c < \infty\},$$

$$\mu(A) = \begin{cases} 0, & \text{falls } A \text{ endlich,} \\ \infty, & \text{falls } A \text{ unendlich.} \end{cases}$$

Dann ist μ ein \emptyset -stetiger Inhalt, aber kein Prämaß. \diamond

Definition 1.38. (i) Ein Paar (Ω, \mathcal{A}) , bestehend aus einer nichtleeren Menge Ω und einer σ -Algebra $\mathcal{A} \subset 2^\Omega$, heißt **Messraum**. Die Mengen $A \in \mathcal{A}$ heißen **messbare Mengen**. Ist Ω höchstens abzählbar und $\mathcal{A} = 2^\Omega$, so heißt der Messraum $(\Omega, 2^\Omega)$ **diskret**.

(ii) Ein Tripel $(\Omega, \mathcal{A}, \mu)$ heißt **Maßraum**, wenn (Ω, \mathcal{A}) ein Messraum ist und μ ein Maß auf \mathcal{A} .

(iii) Ist zudem $\mu(\Omega) = 1$, so heißt $(\Omega, \mathcal{A}, \mu)$ ein **Wahrscheinlichkeitsraum**. In diesem Fall heißen die Mengen $A \in \mathcal{A}$ auch **Ereignisse**.

(iv) Den Raum aller endlichen Maße auf (Ω, \mathcal{A}) bezeichnen wir mit $\mathcal{M}_f(\Omega) := \mathcal{M}_f(\Omega, \mathcal{A})$, den der W-Maße mit $\mathcal{M}_1(\Omega) := \mathcal{M}_1(\Omega, \mathcal{A})$, schließlich den der σ -endlichen Maße mit $\mathcal{M}_\sigma(\Omega, \mathcal{A})$.

Übung 1.2.1. Sei $\mathcal{A} = \{(a, b] \cap \mathbb{Q} : a, b \in \mathbb{R}, a \leq b\}$. Definiere $\mu : \mathcal{A} \rightarrow [0, \infty)$ durch $\mu((a, b] \cap \mathbb{Q}) = b - a$. Man zeige, dass \mathcal{A} ein Semiring ist und μ ein unterhalb und oberhalb stetiger Inhalt auf \mathcal{A} , der jedoch nicht σ -additiv ist. ♣

1.3 Fortsetzung von Maßen

In diesem Abschnitt wollen wir Maße konstruieren, indem wir zunächst auf einem einfachen Mengensystem, nämlich einem Semiring, plausible Werte für einen Inhalt angeben und dann, nach Möglichkeit, diesen Inhalt zu einem Maß auf der erzeugten σ -Algebra fortsetzen. Bevor wir zu den konkreten Bedingungen kommen, unter denen das machbar ist, bringen wir zwei Beispiele.

Beispiel 1.39 (Lebesgue-Maß). Sei $n \in \mathbb{N}$ und

$$\mathcal{A} = \{(a, b] : a, b \in \mathbb{R}^n, a < b\}$$

der Semiring der halboffenen Quader $(a, b] \subset \mathbb{R}^n$ (vergleiche (1.5)). Das n -dimensionale Volumen des Quaders ist

$$\mu((a, b]) = \prod_{i=1}^n (b_i - a_i).$$

Können wir μ zu einem (eindeutig bestimmten) Maß auf der Borel'schen σ -Algebra $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{A})$ fortsetzen? Wir werden sehen, dass dies möglich ist. Das resultierende Maß heißt Lebesgue-Maß (manchmal auch Lebesgue-Borel-Maß) λ auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. ◇

Beispiel 1.40 (Produktmaß, Bernoulli-Maß). Wir wollen ein Wahrscheinlichkeitsmaß konstruieren für die unendliche, unabhängige Wiederholung eines Zufallsexperiments mit endlich vielen möglichen Ausgängen. Die Menge der Ausgänge sei E . Für $e \in E$ sei p_e die Wahrscheinlichkeit, dass e eintritt. Es gilt also $\sum_{e \in E} p_e = 1$. Die Ergebnisse dieser Experimente seien $\omega_1, \omega_2, \dots \in E$. Der Raum des gesamten Experiments ist daher $\Omega = E^{\mathbb{N}}$. Wie in Beispiel 1.11(vi) definieren wir

$$[\omega_1, \dots, \omega_n] := \{\omega' \in \Omega : \omega'_i = \omega_i \text{ für jedes } i = 1, \dots, n\} \quad (1.7)$$

als die Menge aller Folgen, die mit den Werten $\omega_1, \dots, \omega_n$ beginnen.

Sei $\mathcal{A}_0 = \{\emptyset\}$. Für $n \in \mathbb{N}$ definieren wir das Mengensystem der Zylindermengen, die nur von den ersten n Koordinaten abhängen,

$$\mathcal{A}_n := \{[\omega_1, \dots, \omega_n] : \omega_1, \dots, \omega_n \in E\}, \quad (1.8)$$

und setzen $\mathcal{A} := \bigcup_{n=0}^{\infty} \mathcal{A}_n$.

Wir interpretieren $[\omega_1, \dots, \omega_n]$ als das Ereignis, dass im ersten Experiment der Wert ω_1 herauskommt, im zweiten ω_2 und schließlich im n -ten Experiment der Wert ω_n . Die Ergebnisse der weiteren Experimente spielen für das Eintreten des Ereignisses keine Rolle. Für $\omega_1, \dots, \omega_n \in E$ soll die Wahrscheinlichkeit für $[\omega_1, \dots, \omega_n]$ das Produkt der einzelnen Wahrscheinlichkeiten sein (das verstehen wir intuitiv unter „Unabhängigkeit“)

$$\mu([\omega_1, \dots, \omega_n]) = \prod_{i=1}^n p_{\omega_i}.$$

Hierdurch wird ein Inhalt auf \mathcal{A} definiert, und unser Ziel ist es, μ in eindeutiger Weise zu einem Wahrscheinlichkeitsmaß auf $\sigma(\mathcal{A})$ fortzusetzen.

Bevor wir dies tun, treffen wir noch die folgenden Definition. Wir definieren eine (Ultra-)Metrik auf Ω durch

$$d(\omega, \omega') = \begin{cases} 2^{-\inf\{n \in \mathbb{N}: \omega_n \neq \omega'_n\}}, & \text{falls } \omega \neq \omega', \\ 0, & \text{falls } \omega = \omega'. \end{cases} \quad (1.9)$$

Dann ist (Ω, d) ein kompakter, metrischer Raum. Offenbar ist

$$[\omega_1, \dots, \omega_n] = B_{2^{-n}}(\omega) = \{\omega' \in \Omega : d(\omega, \omega') < 2^{-n}\}.$$

Das Komplement von $[\omega_1, \dots, \omega_n]$ ist die Vereinigung von $(\#E)^n - 1$ offenen Kugeln

$$[\omega_1, \dots, \omega_n]^c = \bigcup_{(\omega'_1, \dots, \omega'_n) \neq (\omega_1, \dots, \omega_n)} [\omega'_1, \dots, \omega'_n],$$

also offen. Damit ist $[\omega_1, \dots, \omega_n]$ abgeschlossen und kompakt, weil Ω kompakt ist. Ähnlich wie in Satz 1.23 kann man zeigen, dass $\sigma(\mathcal{A}) = \mathcal{B}(\Omega, d)$.

Übung: Man zeige die obigen Aussagen. \diamond

Das Hauptergebnis dieses Kapitels ist der Fortsetzungssatz für Maße, den wir hier in der Form von Carathéodory formulieren.

Satz 1.41 (Carathéodory). Sei $\mathcal{A} \subset 2^\Omega$ ein Ring und μ ein σ -endliches Prämaß auf \mathcal{A} . Dann kann μ auf genau eine Weise zu einem Maß $\tilde{\mu}$ auf $\sigma(\mathcal{A})$ fortgesetzt werden, und $\tilde{\mu}$ ist σ -endlich.

Den Beweis dieses Satzes müssen wir mit einigen Lemmata vorbereiten. Wir zeigen dann in Satz 1.53 eine etwas stärkere Aussage. Dort wird auch die griffige Formulierung „kann fortgesetzt werden“ präzisiert.

Lemma 1.42 (Eindeutigkeit durch schnittstabilen Erzeuger). Sei $(\Omega, \mathcal{A}, \mu)$ ein σ -endlicher Maßraum und $\mathcal{E} \subset \mathcal{A}$ ein schnitt stabiler Erzeuger von \mathcal{A} . Es gebe $\Omega_1, \Omega_2, \dots \in \mathcal{E}$ mit $\bigcup_{n=1}^{\infty} \Omega_n = \Omega$ und $\mu(\Omega_n) < \infty$ für jedes $n \in \mathbb{N}$. Dann ist μ durch die Werte $\mu(E)$, $E \in \mathcal{E}$, eindeutig festgelegt.

Ist μ ein W-Maß, so gilt die Folgerung auch ohne die Existenz der Folge $(\Omega_n)_{n \in \mathbb{N}}$.

Beweis. Sei ν ein weiteres σ -endliches Maß auf (Ω, \mathcal{A}) mit der Eigenschaft

$$\mu(E) = \nu(E) \quad \text{für jedes } E \in \mathcal{E}.$$

Sei $E \in \mathcal{E}$ mit $\mu(E) < \infty$. Betrachte das Mengensystem

$$\mathcal{D}_E = \{A \in \mathcal{A} : \mu(A \cap E) = \nu(A \cap E)\}.$$

Um zu zeigen, dass \mathcal{D}_E ein Dynkin-System ist, prüfen wir die Eigenschaften aus Definition 1.10:

- (i) Offensichtlich ist $\Omega \in \mathcal{D}_E$.
- (ii) Seien $A, B \in \mathcal{D}_E$ mit $A \supset B$. Dann ist

$$\begin{aligned} \mu((A \setminus B) \cap E) &= \mu(A \cap E) - \mu(B \cap E) \\ &= \nu(A \cap E) - \nu(B \cap E) = \nu((A \setminus B) \cap E). \end{aligned}$$

Also ist $A \setminus B \in \mathcal{D}_E$.

- (iii) Seien $A_1, A_2, \dots \in \mathcal{D}_E$ paarweise disjunkt sowie $A = \bigcup_{n=1}^{\infty} A_n$. Dann ist

$$\mu(A \cap E) = \sum_{n=1}^{\infty} \mu(A_n \cap E) = \sum_{n=1}^{\infty} \nu(A_n \cap E) = \nu(A \cap E),$$

also $A \in \mathcal{D}_E$.

Offenbar ist $\mathcal{E} \subset \mathcal{D}_E$, also $\delta(\mathcal{E}) \subset \mathcal{D}_E$. Da \mathcal{E} schnittstabil ist, ist nach Satz 1.19

$$\mathcal{A} \supset \mathcal{D}_E \supset \delta(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{A}.$$

Also ist $\mathcal{D}_E = \mathcal{A}$.

Für jedes $A \in \mathcal{A}$ und $E \in \mathcal{E}$ mit $\mu(E) < \infty$ gilt also $\mu(A \cap E) = \nu(A \cap E)$. Seien nun $\Omega_1, \Omega_2, \dots \in \mathcal{E}$ mit $\bigcup_{n=1}^{\infty} \Omega_n = \Omega$ und $\mu(\Omega_n) < \infty$ für jedes $n \in \mathbb{N}$. Setze $E_n := \bigcup_{i=1}^n \Omega_i$, $n \in \mathbb{N}$, und $E_0 = \emptyset$. Dann ist $E_n = \biguplus_{i=1}^n (E_{i-1}^c \cap \Omega_i)$. Für jedes $A \in \mathcal{A}$ und $n \in \mathbb{N}$ bekommen wir also

$$\mu(A \cap E_n) = \sum_{i=1}^n \mu((A \cap E_{i-1}^c) \cap \Omega_i) = \sum_{i=1}^n \nu((A \cap E_{i-1}^c) \cap \Omega_i) = \nu(A \cap E_n).$$

Da $E_n \uparrow \Omega$ gilt und μ und ν von unten stetig sind, gilt für $A \in \mathcal{A}$

$$\mu(A) = \lim_{n \rightarrow \infty} \mu(A \cap E_n) = \lim_{n \rightarrow \infty} \nu(A \cap E_n) = \nu(A).$$

Der Zusatz ist trivial, denn $\tilde{\mathcal{E}} := \mathcal{E} \cup \{\Omega\}$ ist ebenfalls ein schnittstabiler Erzeuger von \mathcal{A} , und der Wert $\mu(\Omega) = 1$ ist bekannt. Es kann also die konstante Folge $E_n = \Omega$, $n \in \mathbb{N}$, gewählt werden. Man beachte jedoch, dass es nicht reicht zu fordern, dass μ endlich ist, weil dann im Allgemeinen die Gesamtmasse $\mu(\Omega)$ nicht eindeutig festgelegt ist (siehe Beispiel 1.45(ii)). \square

Beispiel 1.43. Sei $\Omega = \mathbb{Z}$ und $\mathcal{E} = \{E_n : n \in \mathbb{Z}\}$, wobei $E_n = (-\infty, n] \cap \mathbb{Z}$. \mathcal{E} ist schnittstabil und $\sigma(\mathcal{E}) = 2^\Omega$. Also ist ein endliches Maß μ auf $(\Omega, 2^\Omega)$ eindeutig festgelegt durch die Werte $\mu(E_n), n \in \mathbb{Z}$.

Ein σ -endliches Maß auf \mathbb{Z} ist jedoch durch die Werte auf \mathcal{E} noch nicht eindeutig bestimmt: Sei μ das Zählmaß auf \mathbb{Z} und $\nu = 2\mu$. Dann ist $\mu(E) = \infty = \nu(E)$ für jedes $E \in \mathcal{E}$. Um μ und ν zu unterscheiden, brauchen wir also einen Erzeuger, der Mengen endlichen Maßes (für μ) enthält. Tun es die Mengen $\tilde{F}_n = [-n, n] \cap \mathbb{Z}, n \in \mathbb{N}$? In der Tat ist für jedes σ -endliche Maß μ jetzt $\mu(\tilde{F}_n) < \infty$ für jedes $n \in \mathbb{N}$. Allerdings erzeugen die \tilde{F}_n nicht 2^Ω (sondern welche σ -Algebra?). Wir können aber die Definition so modifizieren: $F_n = [-n/2, (n+1)/2] \cap \mathbb{Z}$. Dann ist $\sigma(\{F_n, n \in \mathbb{N}\}) = 2^\Omega$, also $\mathcal{E} = \{F_n, n \in \mathbb{N}\}$ ein schnittstabiler Erzeuger von 2^Ω und $\mu(F_n) < \infty$ für jedes $n \in \mathbb{N}$. Wegen $F_n \uparrow \Omega$ sind die Bedingungen des Satzes erfüllt. \diamond

Beispiel 1.44 (Verteilungsfunktion). Ein W-Maß μ auf dem Raum $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ ist durch Angabe der Werte $\mu((-\infty, b])$ auf den Mengen $(-\infty, b] = \times_{i=1}^n (-\infty, b_i], b \in \mathbb{R}^n$, eindeutig festgelegt, da diese Mengen einen schnittstabilen Erzeuger bilden (Satz 1.23). Speziell ist ein W-Maß μ auf \mathbb{R} durch Angabe der **Verteilungsfunktion** $F : \mathbb{R} \rightarrow [0, 1], x \mapsto \mu((-\infty, x])$ eindeutig bestimmt. \diamond

Beispiel 1.45. (i) Sei $\Omega = \{1, 2, 3, 4\}$ und $\mathcal{E} = \{\{1, 2\}, \{2, 3\}\}$. Offenbar gilt $\sigma(\mathcal{E}) = 2^\Omega$, jedoch ist \mathcal{E} nicht schnittstabil. Tatsächlich ist hier ein W-Maß μ durch Angabe der Werte $\mu(\{1, 2\}) = \mu(\{2, 3\}) = \frac{1}{2}$ nicht eindeutig festgelegt. Es gibt beispielsweise die Möglichkeiten $\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_3$ oder $\mu' = \frac{1}{2}\delta_2 + \frac{1}{2}\delta_4$.

(ii) Sei $\Omega = \{1, 2\}$ und $\mathcal{E} = \{\{1\}\}$. Dann ist \mathcal{E} ein schnittstabiler Erzeuger von 2^Ω , und ein W-Maß μ ist durch Angabe von $\mu(\{1\})$ eindeutig festgelegt. Allerdings gilt dies nicht für endliche Maße im Allgemeinen, denn $\mu = 0$ und $\nu = \delta_2$ sind zwei endliche Maße, die auf \mathcal{E} übereinstimmen. \diamond

Definition 1.46 (Äußeres Maß). Eine Mengenfunktion $\mu^* : 2^\Omega \rightarrow [0, \infty]$ heißt **äußeres Maß**, falls gilt:

- (i) $\mu^*(\emptyset) = 0$,
- (ii) μ^* ist monoton,
- (iii) μ^* ist σ -subadditiv.

Lemma 1.47. Sei $\mathcal{A} \subset 2^\Omega$ ein beliebiges Mengensystem mit $\emptyset \in \mathcal{A}$ und μ eine nichtnegative Mengenfunktion auf \mathcal{A} mit $\mu(\emptyset) = 0$. Für $A \subset \Omega$ sei

$$\mathcal{U}(A) = \left\{ \mathcal{F} \subset \mathcal{A} : \mathcal{F} \text{ ist höchstens abzählbar und } A \subset \bigcup_{F \in \mathcal{F}} F \right\}$$

die Menge der abzählbaren Überdeckungen \mathcal{F} von A mit Mengen F aus \mathcal{A} . Setze

$$\mu^*(A) := \inf \left\{ \sum_{F \in \mathcal{F}} \mu(F) : \mathcal{F} \in \mathcal{U}(A) \right\},$$

wobei $\inf \emptyset = \infty$. Dann ist μ^* ein äußeres Maß. Ist μ zudem σ -subadditiv, so gilt $\mu^*(A) = \mu(A)$ für jedes $A \in \mathcal{A}$.

Beweis. Wir weisen die Eigenschaften (i)–(iii) des äußeren Maßes nach.

- (i) Wegen $\emptyset \in \mathcal{A}$ ist $\{\emptyset\} \in \mathcal{U}(\emptyset)$, also ist $\mu^*(\emptyset) = 0$.
- (ii) Ist $A \subset B$, so ist $\mathcal{U}(A) \supset \mathcal{U}(B)$, also ist $\mu^*(A) \leq \mu^*(B)$.
- (iii) Sei $A_n \subset \Omega$ für jedes $n \in \mathbb{N}$ und $A \subset \bigcup_{n=1}^{\infty} A_n$. Wir müssen zeigen, dass $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$. Ohne Einschränkung sei $\mu^*(A_n) < \infty$ und damit $\mathcal{U}(A_n) \neq \emptyset$ für jedes $n \in \mathbb{N}$. Wähle $\varepsilon > 0$ und zu jedem $n \in \mathbb{N}$ eine Überdeckung $\mathcal{F}_n \in \mathcal{U}(A_n)$ mit

$$\sum_{F \in \mathcal{F}_n} \mu(F) \leq \mu^*(A_n) + \varepsilon 2^{-n}.$$

Dann ist $\mathcal{F} := \bigcup_{n=1}^{\infty} \mathcal{F}_n \in \mathcal{U}(A)$ und

$$\mu^*(A) \leq \sum_{F \in \mathcal{F}} \mu(F) \leq \sum_{n=1}^{\infty} \sum_{F \in \mathcal{F}_n} \mu(F) \leq \sum_{n=1}^{\infty} \mu^*(A_n) + \varepsilon.$$

Sei $A \in \mathcal{A}$. Wegen $\{A\} \in \mathcal{U}(A)$ ist $\mu^*(A) \leq \mu(A)$. Ist μ σ -subadditiv, so gilt für jedes $\mathcal{F} \in \mathcal{U}(A)$, dass $\sum_{F \in \mathcal{F}} \mu(F) \geq \mu(A)$ ist, also auch $\mu^*(A) \geq \mu(A)$. \square

Definition 1.48 (μ^* -messbare Mengen). Sei μ^* ein äußeres Maß. Eine Menge $A \in 2^{\Omega}$ heißt μ^* -messbar, falls

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) = \mu^*(E) \quad \text{für jedes } E \in 2^{\Omega}. \quad (1.10)$$

Wir schreiben $\mathcal{M}(\mu^*) = \{A \in 2^{\Omega} : A \text{ ist } \mu^*\text{-messbar}\}$.

Lemma 1.49. Es ist $A \in \mathcal{M}(\mu^*)$ genau dann, wenn

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) \leq \mu^*(E) \quad \text{für jedes } E \in 2^{\Omega}.$$

Beweis. Da μ^* subadditiv ist, gilt stets die andere Ungleichung. \square

Lemma 1.50. $\mathcal{M}(\mu^*)$ ist eine Algebra.

Beweis. Wir prüfen die Eigenschaften (i)–(iii) der Algebra aus Satz 1.7.

- (i) $\Omega \in \mathcal{M}(\mu^*)$ ist klar.

(ii) (Komplementstabilität) Per Definition ist $A \in \mathcal{M}(\mu^*) \iff A^c \in \mathcal{M}(\mu^*)$.

(iii) (Schnittstabilität) Seien $A, B \in \mathcal{M}(\mu^*)$ und $E \in 2^\Omega$. Dann ist

$$\begin{aligned} & \mu^*((A \cap B) \cap E) + \mu^*((A \cap B)^c \cap E) \\ &= \mu^*(A \cap B \cap E) + \mu^*((A^c \cap B \cap E) \cup (A^c \cap B^c \cap E) \cup (A \cap B^c \cap E)) \\ &\leq \mu^*(A \cap B \cap E) + \mu^*(A^c \cap B \cap E) \\ &\quad + \mu^*(A^c \cap B^c \cap E) + \mu^*(A \cap B^c \cap E) \\ &= \mu^*(B \cap E) + \mu^*(B^c \cap E) \\ &= \mu^*(E). \end{aligned}$$

Dabei haben wir in der vorletzten Gleichung $A \in \mathcal{M}(\mu^*)$ benutzt und in der letzten $B \in \mathcal{M}(\mu^*)$. \square

Lemma 1.51. Ein äußeres Maß μ^* ist σ -additiv auf $\mathcal{M}(\mu^*)$.

Beweis. Seien $A, B \in \mathcal{M}(\mu^*)$ mit $A \cap B = \emptyset$. Dann ist

$$\mu^*(A \cup B) = \mu^*(A \cap (A \cup B)) + \mu^*(A^c \cap (A \cup B)) = \mu^*(A) + \mu^*(B).$$

Induktiv folgt die (endliche) Additivität. Da μ^* per Definition σ -subadditiv ist, folgt nach Satz 1.36, dass μ^* auch σ -additiv ist. \square

Lemma 1.52. Ist μ^* ein äußeres Maß, so ist $\mathcal{M}(\mu^*)$ eine σ -Algebra. Speziell ist μ^* ein Maß auf $\mathcal{M}(\mu^*)$.

Beweis. Nach Lemma 1.50 ist $\mathcal{M}(\mu^*)$ eine Algebra, also insbesondere schnittstabil. Nach Satz 1.18 reicht es zu zeigen, dass $\mathcal{M}(\mu^*)$ ein Dynkin-System ist.

Seien also $A_1, A_2, \dots \in \mathcal{M}(\mu^*)$ paarweise disjunkt und $A := \biguplus_{n=1}^{\infty} A_n$. Zu zeigen ist $A \in \mathcal{M}(\mu^*)$, also

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) \leq \mu^*(E) \quad \text{für jedes } E \in 2^\Omega. \quad (1.11)$$

Setze $B_n = \bigcup_{i=1}^n A_i$ für jedes $n \in \mathbb{N}$. Es gilt für jedes $n \in \mathbb{N}$

$$\begin{aligned} \mu^*(E \cap B_{n+1}) &= \mu^*((E \cap B_{n+1}) \cap B_n) + \mu^*((E \cap B_{n+1}) \cap B_n^c) \\ &= \mu^*(E \cap B_n) + \mu^*(E \cap A_{n+1}), \end{aligned}$$

und induktiv $\mu^*(E \cap B_n) = \sum_{i=1}^n \mu^*(E \cap A_i)$. Wegen der Monotonie von μ^* folgt

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B_n) + \mu^*(E \cap A^c)$$

$$= \sum_{i=1}^n \mu^*(E \cap A_i) + \mu^*(E \cap A^c).$$

Indem wir $n \rightarrow \infty$ gehen lassen, folgt mit der σ -Subadditivität von μ^*

$$\mu^*(E) \geq \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \cap A^c) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Also gilt (1.11), und der Beweis ist komplett. \square

Wir zeigen nun einen Satz, der mit schwächeren Voraussetzungen auskommt als der Satz von Carathéodory (Satz 1.41) und diesen impliziert.

Satz 1.53 (Fortsetzungssatz für Maße). Sei \mathcal{A} ein Semiring und $\mu : \mathcal{A} \rightarrow [0, \infty]$ eine additive, σ -subadditive, σ -endliche Mengenfunktion mit $\mu(\emptyset) = 0$.

Dann existiert ein eindeutig bestimmtes, σ -endliches Maß $\tilde{\mu} : \sigma(\mathcal{A}) \rightarrow [0, \infty]$ mit $\tilde{\mu}(A) = \mu(A)$ für jedes $A \in \mathcal{A}$.

Beweis. Da \mathcal{A} schnittstabil ist, folgt die Eindeutigkeit aus Lemma 1.42.

Um die Existenz zu zeigen, definieren wir wie in Lemma 1.47

$$\mu^*(A) := \inf \left\{ \sum_{F \in \mathcal{F}} \mu(F) : \mathcal{F} \in \mathcal{U}(A) \right\} \quad \text{für jedes } A \in 2^{\mathcal{Q}}.$$

Nach Lemma 1.47 ist μ^* ein äußeres Maß und $\mu^*(A) = \mu(A)$ für jedes $A \in \mathcal{A}$. Wir müssen zeigen, dass $\mathcal{M}(\mu^*) \supset \sigma(\mathcal{A})$ gilt. Da $\mathcal{M}(\mu^*)$ eine σ -Algebra ist (Lemma 1.52), reicht es, $\mathcal{A} \subset \mathcal{M}(\mu^*)$ zu zeigen.

Seien also $A \in \mathcal{A}$ und $E \in 2^{\mathcal{Q}}$ mit $\mu^*(E) < \infty$. Sei $\varepsilon > 0$. Dann gibt es $E_1, E_2, \dots \in \mathcal{A}$ mit

$$E \subset \bigcup_{n=1}^{\infty} E_n \quad \text{und} \quad \sum_{n=1}^{\infty} \mu(E_n) \leq \mu^*(E) + \varepsilon.$$

Setze $B_n := E_n \cap A \in \mathcal{A}$. Da \mathcal{A} ein Semiring ist, gibt es zu jedem $n \in \mathbb{N}$ ein $m_n \in \mathbb{N}$ sowie $C_n^1, \dots, C_n^{m_n} \in \mathcal{A}$ mit $E_n \setminus A = E_n \setminus B_n = \biguplus_{k=1}^{m_n} C_n^k$. Also ist

$$E \cap A \subset \bigcup_{n=1}^{\infty} B_n, \quad E \cap A^c \subset \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{m_n} C_n^k \quad \text{und} \quad E_n = B_n \uplus \biguplus_{k=1}^{m_n} C_n^k.$$

Nach der Definition des äußeren Maßes und weil μ (endlich) additiv ist, gilt

$$\begin{aligned}
\mu^*(E \cap A) + \mu^*(E \cap A^c) &\leq \sum_{n=1}^{\infty} \mu(B_n) + \sum_{n=1}^{\infty} \sum_{k=1}^{m_n} \mu(C_n^k) \\
&= \sum_{n=1}^{\infty} \left(\mu(B_n) + \sum_{k=1}^{m_n} \mu(C_n^k) \right) \\
&= \sum_{n=1}^{\infty} \mu(E_n) \\
&\leq \mu^*(E) + \varepsilon.
\end{aligned}$$

Daher ist $\mu^*(E \cap A) + \mu^*(E \cap A^c) \leq \mu^*(E)$ und damit $A \in \mathcal{M}(\mu^*)$, also ist $\mathcal{A} \subset \mathcal{M}(\mu^*)$. Setze nun $\tilde{\mu} : \sigma(\mathcal{A}) \rightarrow [0, \infty]$, $A \mapsto \mu^*(A)$. Nach Lemma 1.51 ist $\tilde{\mu}$ ein Maß und $\tilde{\mu}$ ist σ -endlich, weil μ σ -endlich ist. \square

Beispiel 1.54 (Lebesgue-Maß, Fortsetzung von Beispiel 1.39). Wir wollen das auf den Quadern $\mathcal{A} = \{(a, b] : a, b \in \mathbb{R}^n, a < b\}$ eingeführte Volumen $\mu((a, b]) = \prod_{i=1}^n (b_i - a_i)$ zu einem Maß auf der Borel'schen σ -Algebra $\mathcal{B}(\mathbb{R}^n)$ fortsetzen. Um die Voraussetzungen von Satz 1.53 zu prüfen, müssen wir nur noch zeigen, dass μ σ -subadditiv ist. Seien also $(a, b], (a(1), b(1)], (a(2), b(2)], \dots \in \mathcal{A}$ mit

$$(a, b] \subset \bigcup_{k=1}^{\infty} (a(k), b(k)].$$

Wir müssen zeigen, dass

$$\mu((a, b]) \leq \sum_{k=1}^{\infty} \mu((a(k), b(k)]). \quad (1.12)$$

Hierzu benutzen wir ein Kompaktheitsargument, um (1.12) auf die endliche Additivität zurück zu führen. Sei also $\varepsilon > 0$, und sei für jedes $k \in \mathbb{N}$ ein $b_{\varepsilon}(k) > b(k)$ so gewählt, dass

$$\mu((a(k), b_{\varepsilon}(k)]) \leq \mu((a(k), b(k)]) + \varepsilon 2^{-k-1}.$$

Ferner sei $a_{\varepsilon} \in (a, b)$ so gewählt, dass $\mu((a_{\varepsilon}, b]) \geq \mu((a, b]) - \frac{\varepsilon}{2}$. Nun ist $[a_{\varepsilon}, b]$ kompakt und

$$\bigcup_{k=1}^{\infty} (a(k), b_{\varepsilon}(k)) \supset \bigcup_{k=1}^{\infty} (a(k), b(k)] \supset (a, b] \supset [a_{\varepsilon}, b].$$

Also existiert ein K_0 mit $\bigcup_{k=1}^{K_0} (a(k), b_{\varepsilon}(k)) \supset (a_{\varepsilon}, b]$. Da μ (endlich) subadditiv ist (Lemma 1.31(iii)), folgt

$$\begin{aligned}\mu((a, b]) &\leq \frac{\varepsilon}{2} + \mu((a_\varepsilon, b]) \leq \frac{\varepsilon}{2} + \sum_{k=1}^{K_0} \mu((a(k), b_\varepsilon(k)]) \\ &\leq \frac{\varepsilon}{2} + \sum_{k=1}^{K_0} (\varepsilon 2^{-k-1} + \mu((a(k), b(k)])) \leq \varepsilon + \sum_{k=1}^{\infty} \mu((a(k), b(k))).\end{aligned}$$

Da $\varepsilon > 0$ beliebig war, folgt (1.12) und damit die σ -Subadditivität von μ . \diamond

Zusammen mit Satz 1.53 haben wir den folgenden Satz gezeigt.

Satz 1.55 (Lebesgue-Maß). Es existiert ein eindeutig bestimmtes Maß λ^n auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ mit der Eigenschaft

$$\lambda^n((a, b]) = \prod_{i=1}^n (b_i - a_i) \quad \text{für alle } a, b \in \mathbb{R}^n \text{ mit } a < b.$$

λ^n heißt **Lebesgue-Maß** auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, oder **Lebesgue-Borel-Maß**.

Beispiel 1.56 (Lebesgue-Stieltjes-Maß). Sei $\Omega = \mathbb{R}$ und $\mathcal{A} = \{(a, b] : a, b \in \mathbb{R}, a \leq b\}$. \mathcal{A} ist ein Semiring und $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$, wo $\mathcal{B}(\mathbb{R})$ die Borel'sche σ -Algebra auf \mathbb{R} ist. Ferner sei $F : \mathbb{R} \rightarrow \mathbb{R}$ monoton wachsend und rechtsseitig stetig. Wir definieren eine Mengenfunktion

$$\tilde{\mu}_F : \mathcal{A} \rightarrow [0, \infty), \quad (a, b] \mapsto F(b) - F(a).$$

Offensichtlich ist $\tilde{\mu}_F(\emptyset) = 0$, und $\tilde{\mu}_F$ ist additiv.

Seien $(a, b], (a(1), b(1)], (a(2), b(2)], \dots \in \mathcal{A}$ mit $(a, b] \subset \bigcup_{n=1}^{\infty} (a(n), b(n)]$. Sei $\varepsilon > 0$, und sei $a_\varepsilon \in (a, b)$ so gewählt, dass $F(a_\varepsilon) - F(a) < \varepsilon/2$. Dies geht, weil F als rechtsstetig angenommen wurde. Ferner sei für jedes $k \in \mathbb{N}$ ein $b_\varepsilon(k) > b(k)$ so gewählt, dass $F(b_\varepsilon(k)) - F(b(k)) < \varepsilon 2^{-k-1}$. Wie in Beispiel 1.54 kann man jetzt zeigen, dass $\tilde{\mu}_F((a, b]) \leq \varepsilon + \sum_{k=1}^{\infty} \tilde{\mu}_F((a(k), b(k))]$. Es folgt, dass $\tilde{\mu}_F$ σ -subadditiv ist. Nach Satz 1.53 können wir $\tilde{\mu}_F$ auf eindeutige Weise zu einem σ -endlichen Maß μ_F auf $\mathcal{B}(\mathbb{R})$ fortsetzen. \diamond

Definition 1.57 (Lebesgue-Stieltjes-Maß). Das Maß μ_F auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit

$$\mu_F((a, b]) = F(b) - F(a) \quad \text{für alle } a, b \in \mathbb{R} \text{ mit } a < b$$

heißt **Lebesgue-Stieltjes-Maß** zur Funktion F .

Beispiel 1.58. Wichtige Spezialfälle für das Lebesgue-Stieltjes-Maß sind:

- (i) Ist $F(x) = x$, so ist $\mu_F = \lambda^1$ das Lebesgue-Maß auf \mathbb{R} .
- (ii) Sei $f : \mathbb{R} \rightarrow [0, \infty)$ stetig und $F(x) = \int_0^x f(t) dt$ für $x \in \mathbb{R}$. Dann ist μ_F die Fortsetzung des in Beispiel 1.30(ix) definierten Prämaßes mit **Dichte** f .

(iii) Sind $x_1, x_2, \dots \in \mathbb{R}$ und $\alpha_n \geq 0$ für $n \in \mathbb{N}$ mit $\sum_{n=1}^{\infty} \alpha_n < \infty$, so gehört zu $F = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{[x_n, \infty)}$ das endliche Maß $\mu_F = \sum_{n=1}^{\infty} \alpha_n \delta_{x_n}$.

(iv) Sind $x_1, x_2, \dots \in \mathbb{R}$, so ist $\mu = \sum_{n=1}^{\infty} \delta_{x_n}$ ein σ -endliches Maß. μ ist genau dann ein Lebesgue-Stieltjes-Maß, wenn die Folge $(x_n)_{n \in \mathbb{N}}$ keinen Häufungspunkt hat. Hat nämlich $(x_n)_{n \in \mathbb{N}}$ keinen Häufungspunkt, so ist nach dem Satz von Bolzano-Weierstraß $\#\{n \in \mathbb{N} : x_n \in [-K, K]\} < \infty$ für jedes $K > 0$. Setzen wir $F(x) = \#\{n \in \mathbb{N} : x_n \in [0, x]\}$ für $x \geq 0$ und $F(x) = -\#\{n \in \mathbb{N} : x_n \in [x, 0]\}$, so ist $\mu = \mu_F$. Ist nun andererseits μ ein Lebesgue-Stieltjes-Maß, also $\mu = \mu_F$ für ein F , dann ist $\#\{n \in \mathbb{N} : x_n \in (-K, K)\} = F(K) - F(-K) < \infty$ für jedes $K > 0$, also hat $(x_n)_{n \in \mathbb{N}}$ keinen Häufungspunkt.

(v) Gilt $\lim_{x \rightarrow \infty} (F(x) - F(-x)) = 1$, so ist μ_F ein W-Maß. \diamond

Den Fall, wo μ_F ein W-Maß ist, wollen wir noch weiter untersuchen.

Definition 1.59 (Verteilungsfunktion). Eine rechtsseitig stetige, monoton wachsende Funktion $F : \mathbb{R} \rightarrow [0, 1]$ mit $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$ und $F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1$ heißt Verteilungsfunktion. Gilt statt $F(\infty) = 1$ lediglich $F(\infty) \leq 1$, so heißt F uneigentliche Verteilungsfunktion.

Ist μ ein (Sub-)W-Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so heißt $F_\mu : x \mapsto \mu((-\infty, x])$ die Verteilungsfunktion von μ .

Offenbar ist F_μ rechtsseitig stetig und $F(-\infty) = 0$, weil μ stetig von oben und endlich ist (Satz 1.36). Auf Grund der Stetigkeit von unten ist $F(\infty) = \mu(\mathbb{R})$, also ist F_μ tatsächlich eine (uneigentliche) Verteilungsfunktion, wenn μ ein (Sub-)W-Maß ist.

Die Argumentation aus Beispiel 1.56 liefert nun den folgenden Satz.

Satz 1.60. Die Abbildung $\mu \mapsto F_\mu$ ist eine Bijektion von der Menge der W-Mäße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ auf die Menge der Verteilungsfunktionen, beziehungsweise von der Menge der Sub-W-Mäße auf die der uneigentlichen Verteilungsfunktionen.

Wir sehen also, dass jedes endliche Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ein Lebesgue-Stieltjes-Maß für eine gewisse Funktion F ist. Für σ -endliche Maße ist dies im Allgemeinen falsch, wie wir in Beispiel 1.58(iv) gesehen haben.

Wir kommen nun zu einem Satz, der Satz 1.55 mit dem Lebesgue-Stieltjes-Maß kombiniert. Später werden wir sehen, dass dieser Satz in größerer Allgemeinheit gültig ist. Speziell kann man auf die Bedingung verzichten, dass die einzelnen Faktoren vom Lebesgue-Stieltjes-Typ sind.

Satz 1.61 (Endliche Produkte von Maßen). Sei $n \in \mathbb{N}$, und seien μ_1, \dots, μ_n endliche Maße oder, allgemeiner, Lebesgue-Stieltjes-Maße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann existiert ein eindeutig bestimmtes, σ -endliches Maß μ auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ mit

$$\mu((a, b]) = \prod_{i=1}^n \mu_i((a_i, b_i]) \quad \text{für alle } a, b \in \mathbb{R}^n \text{ mit } a < b.$$

Wir nennen $\mu =: \bigotimes_{i=1}^n \mu_i$ das **Produktmaß** zu den Maßen μ_1, \dots, μ_n .

Beweis. Dies geht völlig analog zum Beweis von Satz 1.55. Man muss sich vergegenwissen, dass die Intervalle $(a, b_\varepsilon]$ und so weiter, so gewählt werden können, dass $\mu((a, b_\varepsilon]) < \mu((a, b]) + \varepsilon$. Hierzu wird die Rechtsstetigkeit der zu den μ_i gehörigen wachsenden Funktion F_i verwendet. Wir überlassen die Details zur Übung. \square

Bemerkung 1.62. Wir werden später in Satz 14.14 sehen, dass die Aussage auch für beliebige σ -endliche Maße μ_1, \dots, μ_n auf beliebigen (auch unterschiedlichen) Messräumen gilt. Wir können auch unendliche (sogar überabzählbare) Produkte betrachten, wenn wir voraussetzen, dass alle Faktoren Wahrscheinlichkeitsräume sind (Satz 14.36). \diamond

Beispiel 1.63 (Unendliches Produktmaß, Fortsetzung von Beispiel 1.40). Sei E eine endliche Menge und $\Omega = E^{\mathbb{N}}$ der Raum der Folgen mit Werten in E . Ferner sei $(p_e)_{e \in E}$ ein Wahrscheinlichkeitsvektor. Der auf $\mathcal{A} = \{[\omega_1, \dots, \omega_n] : \omega_1, \dots, \omega_n \in E, n \in \mathbb{N}\}$ definierte Inhalt

$$\mu([\omega_1, \dots, \omega_n]) = \prod_{i=1}^n p_{\omega_i}$$

soll nun zu einem Maß auf $\sigma(\mathcal{A})$ fortgesetzt werden. Um die Voraussetzungen von Satz 1.53 zu prüfen, müssen wir zeigen, dass μ σ -subadditiv ist. Wie im vorangehenden Beispiel geht dies mit Hilfe eines Kompaktheitsarguments.

Seien also $A, A_1, A_2, \dots \in \mathcal{A}$ und $A \subset \bigcup_{n=1}^{\infty} A_n$. Es reicht zu zeigen, dass es ein $N \in \mathbb{N}$ gibt mit der Eigenschaft

$$A \subset \bigcup_{n=1}^N A_n. \tag{1.13}$$

Dann ist nämlich aufgrund der endlichen Subadditivität von μ (Lemma 1.31(iii)) schon $\mu(A) \leq \sum_{n=1}^N \mu(A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$, also ist μ σ -subadditiv.

Wir geben nun zwei Beweise für (1.13) an.

1. Beweis. Wie in Beispiel 1.40 angemerkt, ist Ω mit der von der Metrik d in (1.9) erzeugten Produkttopologie kompakt, und jedes $A \in \mathcal{A}$ ist abgeschlossen und damit auch kompakt. Da jedes der A_n zugleich offen ist, gibt es eine endliche Teilüberdeckung von A , mithin gilt (1.13).

2. Beweis. Wir zeigen nun auf *elementare* Weise die Gültigkeit von (1.13). Das Vorgehen imitiert den Beweis dafür, dass Ω kompakt ist. Wir setzen $B_n := A \setminus \bigcup_{i=1}^n A_i$, nehmen an, dass $B_n \neq \emptyset$ für jedes $n \in \mathbb{N}$ und führen dies zum Widerspruch. Nach dem Dirichlet'schen Schubfachprinzip (E ist endlich) können wir ein $\omega_1 \in E$ auswählen, sodass $[\omega_1] \cap B_n \neq \emptyset$ für unendlich viele $n \in \mathbb{N}$. Wegen $B_1 \supset B_2 \supset \dots$ folgt

$$[\omega_1] \cap B_n \neq \emptyset \quad \text{für jedes } n \in \mathbb{N}.$$

Wähle nun sukzessive $\omega_2, \omega_3, \dots \in E$ so aus, dass

$$[\omega_1, \dots, \omega_k] \cap B_n \neq \emptyset \quad \text{für alle } k, n \in \mathbb{N}.$$

B_n ist disjunkte Vereinigung von gewissen Mengen $C_{n,1}, \dots, C_{n,m_n} \in \mathcal{A}$. Daher existiert zu jedem $n \in \mathbb{N}$ ein $i_n \in \{1, \dots, m_n\}$ mit $[\omega_1, \dots, \omega_k] \cap C_{n,i_n} \neq \emptyset$ für unendlich viele $k \in \mathbb{N}$. Wegen $[\omega_1] \supset [\omega_1, \omega_2] \supset \dots$ folgt

$$[\omega_1, \dots, \omega_k] \cap C_{n,i_n} \neq \emptyset \quad \text{für alle } k, n \in \mathbb{N}.$$

Für festes $n \in \mathbb{N}$ und großes k ist $[\omega_1, \dots, \omega_k] \subset C_{n,i_n}$, also ist $\omega = (\omega_1, \omega_2, \dots) \in C_{n,i_n} \subset B_n$. Es folgt im Widerspruch zur Annahme, dass $\bigcap_{n=1}^{\infty} B_n \neq \emptyset$. \diamond

Zusammen mit Satz 1.53 haben wir den folgenden Satz gezeigt.

Satz 1.64 (Produktmaß, Bernoulli-Maß). Sei E eine endliche, nichtleere Menge und $\Omega = E^{\mathbb{N}}$ sowie $(p_e)_{e \in E}$ ein Wahrscheinlichkeitsvektor. Dann gibt es ein eindeutig bestimmtes W-Maß μ auf $\sigma(\mathcal{A}) = \mathcal{B}(\Omega)$ mit

$$\mu([\omega_1, \dots, \omega_n]) = \prod_{i=1}^n p_{\omega_i} \quad \text{für alle } \omega_1, \dots, \omega_n \in E \text{ und } n \in \mathbb{N}.$$

Wir nennen μ das **Produktmaß** oder **Bernoulli-Maß** auf Ω mit Gewichten $(p_e)_{e \in E}$.

Wir schreiben auch $(\sum_{e \in E} p_e \delta_e)^{\otimes \mathbb{N}} := \mu$.

Ferner nennen wir $(2^E)^{\otimes \mathbb{N}} := \sigma(\mathcal{A})$ die **Produkt- σ -Algebra** auf Ω .

Auf Produktmaße gehen wir systematisch noch einmal in Kapitel 14 ein.

Der Fortsetzungssatz liefert uns einen abstrakten Existenz- und Eindeutigkeitssatz für Maße, die wir zuvor nur auf einem Semiring \mathcal{A} definiert hatten. Der folgende Satz zeigt, wie gut wir das Maß von $\sigma(\mathcal{A})$ -messbaren Mengen durch endliche, beziehungsweise abzählbare Operationen mit Mengen aus \mathcal{A} annähern können.

Wir schreiben

$$A \triangle B := (A \setminus B) \cup (B \setminus A), \quad \text{für } A, B \subset \Omega, \quad (1.14)$$

für die **symmetrische Differenz** zweier Mengen A und B .

Satz 1.65 (Approximationssatz für Maße). Sei $\mathcal{A} \subset 2^\Omega$ ein Semiring und μ ein Maß auf $\sigma(\mathcal{A})$, das σ -endlich auf \mathcal{A} ist.

- (i) Zu $A \in \sigma(\mathcal{A})$ und $\varepsilon > 0$ gibt es paarweise disjunkte Mengen $A_1, A_2, \dots \in \mathcal{A}$ mit $A \subset \bigcup_{n=1}^{\infty} A_n$ und $\mu\left(\bigcup_{n=1}^{\infty} A_n \setminus A\right) < \varepsilon$.
- (ii) Zu $A \in \sigma(\mathcal{A})$ mit $\mu(A) < \infty$ und $\varepsilon > 0$ gibt es $n \in \mathbb{N}$ und paarweise disjunkte Mengen $A_1, \dots, A_n \in \mathcal{A}$ mit $\mu\left(A \triangle \bigcup_{k=1}^n A_k\right) < \varepsilon$.
- (iii) Zu jedem $A \in \mathcal{M}(\mu^*)$ gibt es $A_-, A_+ \in \sigma(\mathcal{A})$ mit $A_- \subset A \subset A_+$ und $\mu(A_+ \setminus A_-) = 0$.

Bemerkung 1.66. Nach (iii) gelten (i) und (ii) auch für $A \in \mathcal{M}(\mu^*)$ (mit μ^* statt μ). Ist \mathcal{A} eine Algebra, so gilt in (ii) für jedes $A \in \sigma(\mathcal{A})$ sogar $\inf_{B \in \mathcal{A}} \mu(A \triangle B) = 0$. \diamond

Beweis. (ii) Da μ auf $\sigma(\mathcal{A})$ mit dem äußeren Maß μ^* übereinstimmt und $\mu(A)$ endlich ist, gibt es nach Definition von μ^* (siehe Lemma 1.47) eine Überdeckung $B_1, B_2, \dots \in \mathcal{A}$ von A mit

$$\mu(A) \geq \sum_{i=1}^{\infty} \mu(B_i) - \varepsilon/2.$$

Sei $n \in \mathbb{N}$ mit $\sum_{i=n+1}^{\infty} \mu(B_i) < \frac{\varepsilon}{2}$ (dies existiert, weil $\mu(A) < \infty$). Für je drei Mengen C, D, E gilt

$$C \triangle D = (D \setminus C) \cup (C \setminus D) \subset (D \setminus C) \cup (C \setminus (D \cup E)) \cup E \subset (C \triangle (D \cup E)) \cup E.$$

Mit $C = A$, $D = \bigcup_{i=1}^n B_i$ und $E = \bigcup_{i=n+1}^{\infty} B_i$ erhalten wir

$$\begin{aligned} \mu\left(A \triangle \bigcup_{i=1}^n B_i\right) &\leq \mu\left(A \triangle \bigcup_{i=1}^{\infty} B_i\right) + \mu\left(\bigcup_{i=n+1}^{\infty} B_i\right) \\ &\leq \mu\left(\bigcup_{i=1}^{\infty} B_i\right) - \mu(A) + \frac{\varepsilon}{2} \leq \varepsilon. \end{aligned}$$

Schreibe nun

$$\bigcup_{i=1}^n B_i = B_1 \uplus \biguplus_{i=2}^n \bigcap_{j=1}^{i-1} (B_i \setminus B_j) =: \biguplus_{i=1}^k A_i$$

für ein gewisses $k \in \mathbb{N}$ und gewisse $A_1, \dots, A_k \in \mathcal{A}$ (Semiring-Eigenschaft).

(i) Sei $A \in \sigma(\mathcal{A})$ und $E_n \uparrow \Omega$, $E_n \in \sigma(\mathcal{A})$ mit $\mu(E_n) < \infty$ für jedes $n \in \mathbb{N}$. Wähle zu $n \in \mathbb{N}$ eine Überdeckung $(B_{n,m})_{m \in \mathbb{N}}$ von $A \cap E_n$ mit

$$\mu(A \cap E_n) \geq \sum_{m=1}^{\infty} \mu(B_{n,m}) - 2^{-n} \varepsilon.$$

(Dies ist möglich nach Definition des äußeren Maßes μ^* , das auf \mathcal{A} mit μ übereinstimmt.) Schreibe $\bigcup_{m,n=1}^{\infty} B_{n,m} = \biguplus_{n=1}^{\infty} A_n$ für gewisse $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ (Übung 1.1.1). Dann ist

$$\begin{aligned} \mu\left(\biguplus_{n=1}^{\infty} A_n \setminus A\right) &= \mu\left(\bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} B_{n,m} \setminus A\right) \\ &\leq \mu\left(\bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} (B_{n,m} \setminus (A \cap E_n))\right) \\ &\leq \sum_{n=1}^{\infty} \left(\left(\sum_{m=1}^{\infty} \mu(B_{n,m}) \right) - \mu(A \cap E_n) \right) \leq \varepsilon. \end{aligned}$$

(iii) Sei $A \in \mathcal{M}(\mu^*)$ und $(E_n)_{n \in \mathbb{N}}$ wie oben. Wähle zu $m, n \in \mathbb{N}$ ein $A_{n,m} \in \sigma(\mathcal{A})$ mit $A_{n,m} \supset A \cap E_n$ und $\mu^*(A_{n,m}) \leq \mu^*(A \cap E_n) + \frac{2^{-n}}{m}$.

Setze $A_m := \bigcup_{n=1}^{\infty} A_{n,m} \in \sigma(\mathcal{A})$. Dann ist $A_m \supset A$ und $\mu^*(A_m \setminus A) \leq \frac{1}{m}$. Setze $A_+ := \bigcap_{m=1}^{\infty} A_m$. Dann ist $\sigma(\mathcal{A}) \ni A_+ \supset A$ und $\mu^*(A_+ \setminus A) = 0$. Wähle analog $(A_-)^c \in \sigma(\mathcal{A})$ mit $(A_-)^c \supset A^c$ und $\mu^*((A_-)^c \setminus A^c) = 0$. Dann ist $A_+ \supset A \supset A_-$ und $\mu(A_+ \setminus A_-) = \mu^*(A_+ \setminus A_-) = \mu^*(A_+ \setminus A) + \mu^*(A \setminus A_-) = 0$. \square

Bemerkung 1.67 (Regularität von Maßen). (Vergleiche auch Satz 13.6 auf Seite 252.) Sei λ^n das Lebesgue-Maß auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Sei \mathcal{A} der Semiring der Quadrate der Form $(a, b] \subset \mathbb{R}^n$. Nach Satz 1.23 ist $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{A})$. Nach dem Approximationssatz gibt es zu $A \in \mathcal{B}(\mathbb{R}^n)$ und $\varepsilon > 0$ abzählbar viele $A_1, A_2, \dots \in \mathcal{A}$ mit $A \subset \bigcup_{i=1}^{\infty} A_i$ und

$$\lambda^n\left(\bigcup_{i=1}^{\infty} A_i \setminus A\right) < \varepsilon/2.$$

Zu jedem A_i existiert ein *offener* Quader $B_i \supset A_i$ mit $\lambda^n(B_i \setminus A_i) < \varepsilon 2^{-i-1}$ (Stetigkeit von oben von λ^n). Daher ist $U = \bigcup_{i=1}^{\infty} B_i$ eine offene Menge $U \supset A$ mit

$$\lambda^n(U \setminus A) < \varepsilon.$$

Diese Eigenschaft von λ^n heißt **Regularität von außen**.

Ist $\lambda^n(A)$ endlich, so gibt es zu $\varepsilon > 0$ eine kompakte Menge $K \subset A$ mit

$$\lambda^n(A \setminus K) < \varepsilon.$$

Diese Eigenschaft von λ^n heißt **Regularität von innen**. In der Tat: Sei $N > 0$ mit $\lambda^n(A) - \lambda^n(A \cap [-N, N]^n) < \varepsilon/2$. Wähle eine offene Menge $U \supset (A \cap [-N, N]^n)^c$ mit $\lambda^n(U \setminus (A \cap [-N, N]^n)^c) < \varepsilon/2$ und setze $K := [-N, N]^n \setminus U \subset A$. \diamond

Definition 1.68 (Nullmenge). Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.

(i) Eine Menge $A \in \mathcal{A}$ heißt **μ -Nullmenge**, oder kurz **Nullmenge**, falls $\mu(A) = 0$. Mit \mathcal{N}_μ bezeichnen wir das System aller Teilmengen von μ -Nullmengen.

(ii) Sei $E(\omega)$ eine Eigenschaft, die dem Punkt $\omega \in \Omega$ zukommen kann. Wir sagen, dass E **μ -fast überall** (f.ü.) gilt oder für **fast alle** (f.a.) ω , falls es eine Nullmenge N gibt, sodass $E(\omega)$ für jedes $\omega \in \Omega \setminus N$ gilt. Ist $A \in \mathcal{A}$, so sagen wir, dass E fast überall auf A gilt, falls es eine Nullmenge N gibt, sodass $E(\omega)$ für jedes $\omega \in A \setminus N$ gilt.

Ist $\mu = P$ ein W-Maß, so sagen wir dann auch, dass E **P-fast sicher** (f.s.) gilt, beziehungsweise fast sicher auf A .

(iii) Sind $A, B \in \mathcal{A}$, so schreiben wir $A = B \pmod{\mu}$, falls $\mu(A \Delta B) = 0$.

Definition 1.69. Ein Maßraum $(\Omega, \mathcal{A}, \mu)$ heißt **vollständig**, falls $\mathcal{N}_\mu \subset \mathcal{A}$.

Bemerkung 1.70 (Vervollständigung eines Maßraums). Sei $(\Omega, \mathcal{A}, \mu)$ ein σ -endlicher Maßraum. Es gibt genau eine kleinste σ -Algebra $\mathcal{A}^* \supset \mathcal{A}$ und eine Fortsetzung μ^* von μ auf \mathcal{A}^* , sodass $(\Omega, \mathcal{A}^*, \mu^*)$ vollständig ist. $(\Omega, \mathcal{A}^*, \mu^*)$ heißt die **Vervollständigung** von $(\Omega, \mathcal{A}, \mu)$. In der Notation des Beweises von Satz 1.53 ist

$$\left(\Omega, \mathcal{M}(\mu^*), \mu^* \Big|_{\mathcal{M}(\mu^*)} \right)$$

diese Vervollständigung.

Ferner ist $\mathcal{M}(\mu^*) = \sigma(\mathcal{A} \cup \mathcal{N}_\mu) = \{A \cup N : A \in \mathcal{A}, N \in \mathcal{N}_\mu\}$ und $\mu^*(A \cup N) = \mu(A)$ für jedes $A \in \mathcal{A}$ und $N \in \mathcal{N}_\mu$.

Da wir diese Aussagen im Folgenden nicht benötigen werden, verzichten wir auf den Beweis und verweisen auf die gängigen Maßtheoriebücher, etwa [49].

Beispiel 1.71. Ist λ das Lebesgue-Maß (genauer: das Lebesgue-Borel-Maß) auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, so lässt sich λ eindeutig fortsetzen zu einem Maß λ^* auf

$$\mathcal{B}^*(\mathbb{R}^n) = \sigma(\mathcal{B}(\mathbb{R}^n) \cup \mathcal{N}),$$

wo \mathcal{N} die Menge der Teilmengen der Lebesgue-Borel'schen Nullmengen bezeichnet. $\mathcal{B}^*(\mathbb{R}^n)$ heißt σ -Algebra der Lebesgue-messbaren Mengen. Zur Unterscheidung wird manchmal λ das **Lebesgue-Borel-Maß** genannt und λ^* das **Lebesgue-Maß**. Wir werden diese Unterscheidung im Folgenden aber nicht benötigen. \diamond

Beispiel 1.72. Sei $\mu = \delta_\omega$ auf einem Messraum (Ω, \mathcal{A}) . Ist $\{\omega\} \in \mathcal{A}$, so ist die Vervollständigung $\mathcal{A}^* = 2^\Omega$, $\mu^* = \delta_\omega$. Im Extremfall der trivialen σ -Algebra $\mathcal{A} = \{\emptyset, \Omega\}$ hingegen ist $\mathcal{N}_\mu = \{\emptyset\}$, also die Vervollständigung $\mathcal{A}^* = \{\emptyset, \Omega\}$, $\mu^* = \delta_\omega$. Man beachte, dass man auf dieser trivialen σ -Algebra die Dirac-Maße zu verschiedenen Punkten aus Ω nicht unterscheiden kann. \diamond

Definition 1.73. Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $\Omega' \in \mathcal{A}$. Dann wird durch

$$\mu|_{\Omega'}(A) := \mu(A) \quad \text{für } A \in \mathcal{A} \text{ mit } A \subset \Omega'$$

ein Maß auf der Spur- σ -Algebra $\mathcal{A}|_{\Omega'}$ definiert. Dieses Maß nennen wir die **Einschränkung** von μ auf Ω' .

Beispiel 1.74. Die Einschränkung des Lebesgue-Borel-Maßes λ von $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ auf $[0, 1]$ ist ein W-Maß auf $([0, 1], \mathcal{B}(\mathbb{R})|_{[0,1]})$. Allgemeiner nennen wir für messbares $A \in \mathcal{B}(\mathbb{R})$ die Einschränkung $\lambda|_A$ das Lebesgue-Maß auf A . Oftmals wird als Symbol wieder λ verwendet, weil wir nicht zu viele kleinliche Unterscheidungen treffen wollen.

Wir sehen später (Korollar 1.84), dass $\mathcal{B}(\mathbb{R})|_A = \mathcal{B}(A)$, wobei $\mathcal{B}(A)$ die Borel'sche σ -Algebra auf A ist, die von den in A (relativ) offenen Mengen erzeugt wird. \diamond

Beispiel 1.75 (Gleichverteilung). Ist $A \in \mathcal{B}(\mathbb{R}^n)$ mit n -dimensionalem Lebesgue-Maß $\lambda^n(A) \in (0, \infty)$, so wird durch

$$\mu(B) := \frac{\lambda^n(B)}{\lambda^n(A)} \quad \text{für } B \in \mathcal{B}(\mathbb{R}^n), B \subset A,$$

ein W-Maß auf $\mathcal{B}(\mathbb{R}^n)|_A$ definiert. Wir nennen μ die **uniforme Verteilung** oder **Gleichverteilung** auf A und schreiben $\mathcal{U}_A := \mu$. \diamond

Übung 1.3.1. Man zeige die folgende Verallgemeinerung von Beispiel 1.58(iv): Ein Maß $\sum_{n=1}^{\infty} \alpha_n \delta_{x_n}$ ist genau dann ein Lebesgue-Stieltjes Maß zu einer geeigneten Funktion F , wenn $\sum_{n: |x_n| \leq K} \alpha_n < \infty$ für jedes $K > 0$ gilt. \clubsuit

Übung 1.3.2. Sei Ω eine überabzählbare Menge und $\omega_0 \in \Omega$ ein beliebiges Element. Sei $\mathcal{A} = \sigma(\{\omega\} : \omega \in \Omega \setminus \{\omega_0\})$.

(i) Charakterisiere \mathcal{A} ähnlich wie in Übung 1.1.4 (Seite 11).

(ii) Zeige, dass $(\Omega, \mathcal{A}, \delta_{\omega_0})$ vollständig ist. ♣

Übung 1.3.3. Sei $(\mu_n)_{n \in \mathbb{N}}$ eine Folge von endlichen Maßen auf dem Messraum (Ω, \mathcal{A}) . Für jedes $A \in \mathcal{A}$ existiere der Grenzwert $\mu(A) := \lim_{n \rightarrow \infty} \mu_n(A)$.

Man zeige: μ ist ein Maß auf (Ω, \mathcal{A}) .

Hinweis: Zu zeigen ist insbesondere die \emptyset -Stetigkeit von μ . ♣

1.4 Messbare Abbildungen

Eine Zwangshandlung in der Mathematik ist es, Homomorphismen zwischen Objekten anzugeben, also strukturerhaltende Abbildungen. Für topologische Räume sind dies die stetigen Abbildungen, für Messräume die messbaren Abbildungen.

Seien im Folgenden stets (Ω, \mathcal{A}) und (Ω', \mathcal{A}') Messräume.

Definition 1.76 (Messbare Abbildungen).

(i) Eine Abbildung $X : \Omega \rightarrow \Omega'$ heißt $\mathcal{A} - \mathcal{A}'$ -messbar (oder kurz: messbar), falls $X^{-1}(\mathcal{A}') := \{X^{-1}(A') : A' \in \mathcal{A}'\} \subset \mathcal{A}$ ist, falls also

$$X^{-1}(A') \in \mathcal{A} \quad \text{für jedes } A' \in \mathcal{A}'.$$

Ist X messbar, so schreiben wir auch $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$.

(ii) Ist $\Omega' = \mathbb{R}$ und $\mathcal{A}' = \mathcal{B}(\mathbb{R})$ die Borel'sche σ -Algebra auf \mathbb{R} , so heißt $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ kurz eine reelle \mathcal{A} -messbare Abbildung.

Beispiel 1.77. (i) Die Identität $\text{id} : \Omega \rightarrow \Omega$ ist $\mathcal{A} - \mathcal{A}$ -messbar.

(ii) Sei $\mathcal{A} = 2^\Omega$ oder $\mathcal{A}' = \{\emptyset, \Omega'\}$. Jedes $X : \Omega \rightarrow \Omega'$ ist dann $\mathcal{A} - \mathcal{A}'$ -messbar.

(iii) Sei $A \subset \Omega$. Die Indikatorfunktion $\mathbb{1}_A : \Omega \rightarrow \{0, 1\}$ ist genau dann $\mathcal{A} - 2^{\{0,1\}}$ -messbar, wenn $A \in \mathcal{A}$. ◇

Satz 1.78 (Erzeugte σ -Algebra). Sei (Ω', \mathcal{A}') ein Messraum und Ω eine nichtleere Menge sowie $X : \Omega \rightarrow \Omega'$ eine Abbildung. Das Urbild

$$X^{-1}(\mathcal{A}') := \{X^{-1}(A') : A' \in \mathcal{A}'\} \tag{1.15}$$

ist die kleinste σ -Algebra, bezüglich der X messbar ist. Wir nennen $\sigma(X) := X^{-1}(\mathcal{A}')$ die von X erzeugte σ -Algebra auf Ω .

Beweis. Übung! □

Wir wollen nun σ -Algebren betrachten, die von mehreren Abbildungen erzeugt werden.

Definition 1.79 (Erzeugte σ -Algebra). Sei Ω eine nichtleere Menge. Sei I eine beliebige Indexmenge, und für jedes $i \in I$ sei $(\Omega_i, \mathcal{A}_i)$ ein Messraum sowie $X_i : \Omega \rightarrow \Omega_i$ eine beliebige Abbildung. Dann heißt

$$\sigma(X_i, i \in I) := \sigma \left(\bigcup_{i \in I} \sigma(X_i) \right) = \sigma \left(\bigcup_{i \in I} X_i^{-1}(\mathcal{A}_i) \right)$$

die von $(X_i, i \in I)$ erzeugte σ -Algebra auf Ω . Dies ist die kleinste σ -Algebra, bezüglich der jedes X_i messbar ist.

Wie bei stetigen oder linearen Abbildungen gibt es eine Verknüpfungseigenschaft.

Satz 1.80 (Verknüpfung von Abbildungen). Sind (Ω, \mathcal{A}) , (Ω', \mathcal{A}') und $(\Omega'', \mathcal{A}'')$ Messräume sowie $X : \Omega \rightarrow \Omega'$ messbar und $X' : \Omega' \rightarrow \Omega''$ messbar, so ist die Abbildung $Y := X' \circ X : \Omega \rightarrow \Omega''$, $\omega \mapsto X'(X(\omega))$ messbar bezüglich $\mathcal{A} - \mathcal{A}''$.

Beweis. Es ist $Y^{-1}(\mathcal{A}'') = X^{-1}((X')^{-1}(\mathcal{A}'')) \subset X^{-1}(\mathcal{A}') \subset \mathcal{A}$. □

Praktisch kann man die Messbarkeit einer Abbildung X kaum prüfen, indem man sämtliche Urbilder $X^{-1}(A')$, $A' \in \mathcal{A}'$ auf Messbarkeit hin untersucht. Dafür sind die meisten σ -Algebren \mathcal{A}' einfach zu groß. Glücklicherweise reicht hier die Betrachtung eines Erzeugers von \mathcal{A}' aus:

Satz 1.81 (Messbarkeit auf einem Erzeuger). Für jedes System $\mathcal{E}' \subset \mathcal{A}'$ von \mathcal{A}' -messbaren Mengen gilt $\sigma(X^{-1}(\mathcal{E}')) = X^{-1}(\sigma(\mathcal{E}'))$ und damit

$$X \text{ ist } \mathcal{A} - \sigma(\mathcal{E}')\text{-messbar} \iff X^{-1}(E') \in \mathcal{A} \quad \text{für jedes } E' \in \mathcal{E}'.$$

Ist speziell $\sigma(\mathcal{E}') = \mathcal{A}'$, dann gilt

$$X \text{ ist } \mathcal{A} - \mathcal{A}'\text{-messbar} \iff X^{-1}(\mathcal{E}') \subset \mathcal{A}.$$

Beweis. Offenbar ist $X^{-1}(\mathcal{E}') \subset X^{-1}(\sigma(\mathcal{E}')) = \sigma(X^{-1}(\mathcal{E}'))$. Also ist auch

$$\sigma(X^{-1}(\mathcal{E}')) \subset X^{-1}(\sigma(\mathcal{E}')).$$

Für die andere Inklusion betrachten wir das Mengensystem

$$\mathcal{A}'_0 := \{A' \in \sigma(\mathcal{E}') : X^{-1}(A') \in \sigma(X^{-1}(\mathcal{E}'))\}$$

und zeigen zunächst, dass \mathcal{A}'_0 eine σ -Algebra ist, indem wir die Punkte (i)–(iii) aus Definition 1.2 prüfen:

- (i) Offensichtlich ist $\Omega' \in \mathcal{A}'_0$.
- (ii) (Komplementstabilität) Ist $A' \in \mathcal{A}'_0$, so ist

$$X^{-1}((A')^c) = (X^{-1}(A'))^c \in \sigma(X^{-1}(\mathcal{E})),$$

also $(A')^c \in \mathcal{A}'_0$.

- (iii) (σ - \cup -Stabilität) Seien $A'_1, A'_2, \dots \in \mathcal{A}'_0$. Dann ist

$$X^{-1}\left(\bigcup_{n=1}^{\infty} A'_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A'_n) \in \sigma(X^{-1}(\mathcal{E})),$$

also ist $\bigcup_{n=1}^{\infty} A'_n \in \mathcal{A}'_0$.

Wegen $\mathcal{E}' \subset \mathcal{A}'_0$ ist $\mathcal{A}'_0 = \sigma(\mathcal{E}')$, also $X^{-1}(A') \in \sigma(X^{-1}(\mathcal{E}'))$ für jedes $A' \in \sigma(\mathcal{E}')$ und damit $X^{-1}(\sigma(\mathcal{E}')) \subset \sigma(X^{-1}(\mathcal{E}'))$. \square

Korollar 1.82 (Messbarkeit von verknüpften Abbildungen). Sei I eine nichtleere Indexmenge sowie (Ω, \mathcal{A}) , (Ω', \mathcal{A}') und $(\Omega_i, \mathcal{A}_i)$ Messräume, $i \in I$. Sei ferner $(X_i : i \in I)$ eine Familie messbarer Abbildungen $X_i : \Omega' \rightarrow \Omega_i$ mit der Eigenschaft $\mathcal{A}' = \sigma(X_i : i \in I)$. Dann gilt: Eine Abbildung $Y : \Omega \rightarrow \Omega'$ ist genau dann \mathcal{A} -messbar, wenn $X_i \circ Y$ messbar ist bezüglich \mathcal{A} - \mathcal{A}_i für jedes $i \in I$.

Beweis. Ist Y messbar, so ist nach Satz 1.80 jedes $X_i \circ Y$ messbar. Sei nun jede der zusammengesetzten Abbildungen $X_i \circ Y$ messbar bezüglich \mathcal{A} - \mathcal{A}_i . Die Menge $\mathcal{E}' := \{X_i^{-1}(A'') : A'' \in \mathcal{A}_i, i \in I\}$ ist nach Voraussetzung ein Erzeuger von \mathcal{A}' , und es gilt $Y^{-1}(A') \in \mathcal{A}$ für jedes $A' \in \mathcal{E}'$ wegen der Messbarkeit aller $X_i \circ Y$. Nach Satz 1.81 ist also Y messbar. \square

Wir erinnern an den Begriff der Spur eines Mengensystems aus Definition 1.25.

Korollar 1.83 (Spur der erzeugten σ -Algebra). Ist $\mathcal{E} \subset 2^\Omega$ und $A \subset \Omega$ nichtleer, so gilt $\sigma(\mathcal{E}|_A) = \sigma(\mathcal{E})|_A$.

Beweis. Sei $X : A \hookrightarrow \Omega$, $\omega \mapsto \omega$ die Inklusionsabbildung. Dann ist $X^{-1}(B) = A \cap B$ für jedes $B \subset \Omega$. Nach Satz 1.81 ist

$$\begin{aligned} \sigma(\mathcal{E}|_A) &= \sigma(\{E \cap A : E \in \mathcal{E}\}) \\ &= \sigma(\{X^{-1}(E) : E \in \mathcal{E}\}) = \sigma(X^{-1}(\mathcal{E})) \\ &= X^{-1}(\sigma(\mathcal{E})) = \{A \cap B : B \in \sigma(\mathcal{E})\} = \sigma(\mathcal{E})|_A. \end{aligned} \quad \square$$

Zur Erinnerung: Für eine Teilmenge $A \subset \Omega$ eines topologischen Raums (Ω, τ) ist $\tau|_A$ die Topologie der in A relativ offenen Mengen. Mit $\mathcal{B}(\Omega, \tau) = \sigma(\tau)$ bezeichnen wir die Borel'sche σ -Algebra auf (Ω, τ) .

Korollar 1.84 (Spur der Borel'schen σ -Algebra). Sei (Ω, τ) ein topologischer Raum und $A \subset \Omega$ eine beliebige nichtleere Teilmenge von Ω . Dann gilt

$$\mathcal{B}(\Omega, \tau)|_A = \mathcal{B}(A, \tau|_A).$$

Beispiel 1.85. (i) Ist Ω' abzählbar, so ist $X : \Omega \rightarrow \Omega'$ genau dann $\mathcal{A} - 2^{\Omega'}$ -messbar, wenn $X^{-1}(\{\omega'\}) \in \mathcal{A}$ für jedes $\omega' \in \Omega'$. Für überabzählbare Ω' ist dies im Allgemeinen falsch. (Man betrachte etwa $\Omega = \Omega' = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $X(\omega) = \omega$ für jedes $\omega \in \Omega$. Offenbar ist $X^{-1}(\{\omega\}) = \{\omega\} \in \mathcal{B}(\mathbb{R})$. Ist andererseits $A \subset \mathbb{R}$ nicht in $\mathcal{B}(\mathbb{R})$, so ist $A \in 2^{\mathbb{R}}$, jedoch $X^{-1}(A) \notin \mathcal{B}(\mathbb{R})$.)

(ii) Für $x \in \mathbb{R}$ verabreden wir folgende Schreibweisen für das Ab- und Aufrunden

$$\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \leq x\} \quad \text{und} \quad \lceil x \rceil := \min\{k \in \mathbb{Z} : k \geq x\}. \quad (1.16)$$

Die Abbildungen $\mathbb{R} \rightarrow \mathbb{Z}$, $x \mapsto \lfloor x \rfloor$ und $x \mapsto \lceil x \rceil$ sind messbar bezüglich $\mathcal{B}(\mathbb{R}) - 2^{\mathbb{Z}}$, denn für jedes $k \in \mathbb{Z}$ sind die Urbilder $\{x \in \mathbb{R} : \lfloor x \rfloor = k\} = [k, k+1)$ und $\{x \in \mathbb{R} : \lceil x \rceil = k\} = (k-1, k]$ in $\mathcal{B}(\mathbb{R})$. Nach dem Verknüpfungssatz (Satz 1.80) sind dann für jede messbare Abbildung $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ auch die Abbildungen $\lfloor f \rfloor$ und $\lceil f \rceil$ messbar bezüglich $\mathcal{A} - 2^{\mathbb{Z}}$.

(iii) Eine Abbildung $X : \Omega \rightarrow \mathbb{R}^d$ ist genau dann $\mathcal{A} - \mathcal{B}(\mathbb{R}^d)$ -messbar, wenn

$$X^{-1}((-\infty, a]) \in \mathcal{A} \quad \text{für jedes } a \in \mathbb{R}^d,$$

denn $\sigma((-\infty, a], a \in \mathbb{R}^d) = \mathcal{B}(\mathbb{R}^d)$ nach Satz 1.23. Analog gilt dies auch für die anderen Mengensysteme $\mathcal{E}_1, \dots, \mathcal{E}_{12}$ aus Satz 1.23. \diamond

Beispiel 1.86. Sei $d(x, y) = \|x - y\|_2$ der gewöhnliche euklidische Abstand auf \mathbb{R}^n und $\mathcal{B}(\mathbb{R}^n, d) = \mathcal{B}(\mathbb{R}^n)$ die Borel'sche σ -Algebra zu der von d erzeugten Topologie. Für jede Teilmenge A von \mathbb{R}^n ist dann $\mathcal{B}(A, d) = \mathcal{B}(\mathbb{R}^n, d)|_A$. \diamond

Wir wollen die reellen Zahlen um die Punkte $-\infty$ und $+\infty$ erweitern und definieren

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}.$$

Topologisch wollen wir $\overline{\mathbb{R}}$ als die so genannte Zweipunktkompaktifizierung ansehen, indem wir $\overline{\mathbb{R}}$ als topologisch isomorph zu $[-1, 1]$ betrachten, beispielsweise vermöge der Abbildung

$$\varphi : [-1, 1] \rightarrow \overline{\mathbb{R}}, \quad x \mapsto \begin{cases} \tan(\pi x/2), & \text{falls } x \in (-1, 1), \\ -\infty, & \text{falls } x = -1, \\ \infty, & \text{falls } x = +1. \end{cases}$$

In der Tat wird durch $\bar{d}(x, y) = |\varphi^{-1}(x) - \varphi^{-1}(y)|$ für $x, y \in \overline{\mathbb{R}}$ eine Metrik auf $\overline{\mathbb{R}}$ definiert, sodass φ und φ^{-1} stetig sind (also ist φ ein topologischer Isomorphismus). Mit $\bar{\tau}$ bezeichnen wir die induzierte Topologie auf $\overline{\mathbb{R}}$, mit τ die gewöhnliche Topologie auf \mathbb{R} .

Korollar 1.87. Es gilt $\bar{\tau}|_{\mathbb{R}} = \tau$, und daher gilt $\mathcal{B}(\overline{\mathbb{R}})|_{\mathbb{R}} = \mathcal{B}(\mathbb{R})$.

Ist speziell $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ messbar, so ist X in kanonischer Weise auch eine $\overline{\mathbb{R}}$ -wertige messbare Abbildung.

Mit $\overline{\mathbb{R}}$ haben wir also eine echte Erweiterung der reellen Zahlen geschaffen, und die Inklusion $\mathbb{R} \hookrightarrow \overline{\mathbb{R}}$ ist messbar.

Satz 1.88 (Messbarkeit stetiger Abbildungen). Sind (Ω, τ) und (Ω', τ') topologische Räume und $f : \Omega \rightarrow \Omega'$ stetig, dann ist f auch $\mathcal{B}(\Omega) - \mathcal{B}(\Omega')$ -messbar.

Beweis. Wegen $\mathcal{B}(\Omega') = \sigma(\tau')$ reicht es nach Satz 1.81 zu zeigen, dass $f^{-1}(A') \in \sigma(\tau)$ für jedes $A' \in \tau'$. Da f stetig ist, gilt aber sogar $f^{-1}(A') \in \tau$ für jedes $A' \in \tau'$. \square

Für $x, y \in \overline{\mathbb{R}}$ verabreden wir folgende Notationen

$$\begin{aligned} x \vee y &= \max(x, y) && \text{(Maximum),} \\ x \wedge y &= \min(x, y) && \text{(Minimum),} \\ x^+ &= \max(x, 0) && \text{(Positivteil),} \\ x^- &= \max(-x, 0) && \text{(Negativteil),} \\ |x| &= \max(x, -x) = x^- + x^+ && \text{(Absolutbetrag),} \\ \text{sign}(x) &= \mathbb{1}_{\{x>0\}} - \mathbb{1}_{\{x<0\}} && \text{(Vorzeichenfunktion).} \end{aligned}$$

Analog bezeichnen wir für reelle messbare Abbildungen beispielsweise $X^+ = \max(X, 0)$. Die Abbildungen $x \mapsto x^+$, $x \mapsto x^-$ und $x \mapsto |x|$ sind stetig (und damit nach dem vorangehenden Satz messbar), die Abbildung $x \mapsto \text{sign}(x)$ ist offenbar auch messbar. Wir erhalten also (zusammen mit Korollar 1.82):

Korollar 1.89. Ist X eine reelle oder $\overline{\mathbb{R}}$ -wertige messbare Abbildung, so sind auch die Abbildungen $X^-, X^+, |X|$ und $\text{sign}(X)$ messbar.

Satz 1.90 (Koordinatenabbildungen sind messbar). Sei (Ω, \mathcal{A}) ein Messraum und $f_1, \dots, f_n : \Omega \rightarrow \mathbb{R}$ Abbildungen sowie $f := (f_1, \dots, f_n) : \Omega \rightarrow \mathbb{R}^n$. Dann gilt

$$f \text{ ist } \mathcal{A} - \mathcal{B}(\mathbb{R}^n)\text{-messbar} \iff \text{jedes } f_i \text{ ist } \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-messbar.}$$

Die Aussage gilt analog für $f_i : \Omega \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$.

Beweis. Für $b \in \mathbb{R}^n$ ist $f^{-1}((-\infty, b)) = \bigcap_{i=1}^n f_i^{-1}((-\infty, b_i))$. Ist jedes f_i messbar, so ist also $f^{-1}((-\infty, b)) \in \mathcal{A}$. Die Quader $(-\infty, b)$, $b \in \mathbb{R}^n$, erzeugen aber $\mathcal{B}(\mathbb{R}^n)$, und daher ist dann f messbar. Sei nun f messbar. Für $i = 1, \dots, n$ sei $\pi_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $x \mapsto x_i$ die i -te Projektion. Offenbar ist π_i stetig also $\mathcal{B}(\mathbb{R}^n) - \mathcal{B}(\mathbb{R})$ -messbar. Nach Satz 1.80 ist auch $f_i = \pi_i \circ f$ messbar. \square

Für den folgenden Satz vereinbaren wir die Konvention $\frac{x}{0} := 0$ für jedes $x \in \mathbb{R}$.

Satz 1.91. Sei (Ω, \mathcal{A}) ein Messraum und $f, g : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ sowie $h : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ messbar. Dann sind auch die Abbildungen $f + g$, $f - g$, $f \cdot h$ und f/h messbar.

Beweis. Die Abbildung $\pi : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$, $(x, \alpha) \mapsto \alpha \cdot x$ ist stetig, also messbar. Nach Satz 1.90 ist $(f, h) : \Omega \rightarrow \mathbb{R}^n \times \mathbb{R}$ messbar, also auch die zusammengesetzte Abbildung $f \cdot h = \pi \circ (f, h)$. Analog folgt die Messbarkeit von $f + g$ und $f - g$.

Um die Messbarkeit von f/h zu zeigen, definieren wir $H : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto 1/x$. Nach unserer Konvention ist $H(0) = 0$. Dann ist $f/h = f \cdot H \circ h$. Es reicht also zu zeigen, dass H messbar ist. Offenbar ist $H|_{\mathbb{R} \setminus \{0\}}$ stetig. Für offenes $U \subset \mathbb{R}$ ist auch $U \setminus \{0\}$ offen und damit $H^{-1}(U \setminus \{0\}) \in \mathcal{B}(\mathbb{R})$. Ferner ist $H^{-1}(\{0\}) = \{0\}$. Also ist schließlich $H^{-1}(U) = H^{-1}(U \setminus \{0\}) \cup (U \cap \{0\}) \in \mathcal{B}(\mathbb{R})$. \square

Satz 1.92. Sind X_1, X_2, \dots messbare Abbildungen $(\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, dann sind auch die folgenden Abbildungen messbar:

$$\inf_{n \in \mathbb{N}} X_n, \quad \sup_{n \in \mathbb{N}} X_n, \quad \liminf_{n \rightarrow \infty} X_n, \quad \limsup_{n \rightarrow \infty} X_n.$$

Beweis. Für jedes $a \in \overline{\mathbb{R}}$ gilt

$$\left(\inf_{n \in \mathbb{N}} X_n \right)^{-1} ([-\infty, a)) = \bigcup_{n=1}^{\infty} X_n^{-1}([-\infty, a)) \in \mathcal{A}.$$

Nach Satz 1.81 folgt hieraus die Messbarkeit von $\inf_{n \in \mathbb{N}} X_n$. Analog geht der Beweis für $\sup_{n \in \mathbb{N}} X_n$.

Für $n \in \mathbb{N}$ setzen wir $Y_n := \inf_{m \geq n} X_m$. Dann ist Y_n messbar, und damit auch $\liminf_{n \rightarrow \infty} X_n := \sup Y_n$. Analog folgt der Beweis für den Limes superior. \square

Ein wichtiges Beispiel für messbare Abbildungen $(\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sind Elementarfunktionen.

Definition 1.93 (Elementarfunktion). Sei (Ω, \mathcal{A}) ein Messraum. Eine Abbildung $f : \Omega \rightarrow \mathbb{R}$ heißt **Elementarfunktion**, wenn es ein $n \in \mathbb{N}$ und paarweise disjunkte, messbare Mengen $A_1, \dots, A_n \in \mathcal{A}$ sowie Zahlen $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ gibt mit

$$f = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}.$$

Bemerkung 1.94. Eine messbare Abbildung, die nur endlich viele Werte annimmt, ist eine Elementarfunktion. (Übung!) \diamond

Definition 1.95. Sind f, f_1, f_2, \dots Abbildungen $\Omega \rightarrow \overline{\mathbb{R}}$ mit

$$f_1(\omega) \leq f_2(\omega) \leq \dots \text{ und } \lim_{n \rightarrow \infty} f_n(\omega) = f(\omega) \text{ für jedes } \omega \in \Omega,$$

so schreiben wir $f_n \uparrow f$ und sagen, dass $(f_n)_{n \in \mathbb{N}}$ punktweise monoton aufsteigend gegen f konvergiert. Analog schreiben wir $f_n \downarrow f$, falls $(-f_n) \uparrow (-f)$.

Satz 1.96. Sei (Ω, \mathcal{A}) ein Messraum und $f : \Omega \rightarrow [0, \infty]$ messbar. Dann gelten die folgenden Aussagen.

(i) Es gibt eine Folge nichtnegativer Elementarfunktionen $(f_n)_{n \in \mathbb{N}}$ mit $f_n \uparrow f$.

(ii) Es gibt $A_1, A_2, \dots \in \mathcal{A}$ und $\alpha_1, \alpha_2, \dots \geq 0$ mit $f = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{A_n}$.

Beweis. (i) Für $n \in \mathbb{N}_0$ definiere $f_n = (2^{-n} \lfloor 2^n f \rfloor) \wedge n$. Dann ist f_n messbar (nach Satz 1.92 und Beispiel 1.85(ii)) und nimmt höchstens $n2^n + 1$ Werte an, ist also eine Elementarfunktion. Offenbar gilt $f_n \uparrow f$.

(ii) Seien f_n wie oben, $B_{n,i} := \{\omega : f_n(\omega) - f_{n-1}(\omega) = i 2^{-n}\}$ und $\beta_{n,i} = i 2^{-n}$ für $n \in \mathbb{N}$ und $i = 1, \dots, 2^n$. Dann ist $f_n - f_{n-1} = \sum_{i=1}^{2^n} \beta_{n,i} \mathbb{1}_{B_{n,i}}$. Nach Umnummerierung $(n, i) \mapsto m$ erhalten wir $(\alpha_m)_{m \in \mathbb{N}}$ und $(A_m)_{m \in \mathbb{N}}$, sodass

$$f = f_0 + \sum_{n=1}^{\infty} (f_n - f_{n-1}) = \sum_{m=1}^{\infty} \alpha_m \mathbb{1}_{A_m}. \quad \square$$

Als Korollar zu dieser Strukturaussage für messbare $[0, \infty]$ -wertige Abbildungen zeigen wir das Faktorisierungslemma.

Korollar 1.97 (Faktorisierungslemma). Seien (Ω', \mathcal{A}') ein Messraum und Ω eine nichtleere Menge. Sei $f : \Omega \rightarrow \Omega'$ eine Abbildung. Eine Abbildung $g : \Omega \rightarrow \overline{\mathbb{R}}$ ist genau dann messbar bezüglich $\sigma(f) - \mathcal{B}(\overline{\mathbb{R}})$, wenn es eine messbare Abbildung $\varphi : (\Omega', \mathcal{A}') \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ gibt mit $g = \varphi \circ f$.

Beweis. „ \Leftarrow “ Ist φ messbar und $g = \varphi \circ f$, so ist g messbar nach Satz 1.80.

„ \Rightarrow “ Sei nun g messbar bezüglich $\sigma(f) - \mathcal{B}(\overline{\mathbb{R}})$. Wir betrachten zunächst den Fall, wo g nichtnegativ ist. Dann existieren messbare Mengen $A_1, A_2, \dots \in \sigma(f)$ sowie Zahlen $\alpha_1, \alpha_2, \dots \in [0, \infty)$ mit $g = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{A_n}$. Nach der Definition

von $\sigma(f)$ gibt es für jedes $n \in \mathbb{N}$ eine Menge $B_n \in \mathcal{A}'$ mit $f^{-1}(B_n) = A_n$, also mit $\mathbb{1}_{A_n} = \mathbb{1}_{B_n} \circ f$. Wir definieren nun $\varphi : \Omega' \rightarrow \overline{\mathbb{R}}$ durch

$$\varphi = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{B_n}.$$

Offenbar ist φ messbar bezüglich $\mathcal{A}' - \mathcal{B}(\overline{\mathbb{R}})$, und es gilt $g = \varphi \circ f$.

Sei nun der allgemeine Fall betrachtet, wo g auch negative Werte annehmen kann. Dann existieren messbare Abbildungen φ^- und φ^+ mit $g^- = \varphi^- \circ f$ und $g^+ = \varphi^+ \circ f$. Daher leistet $\varphi := \varphi^+ - \varphi^-$ das Gewünschte. \square

Mit einer messbaren Abbildung wird auch ein Maß von einem Raum auf einen anderen transportiert.

Definition 1.98 (Bildmaß). Seien (Ω, \mathcal{A}) und (Ω', \mathcal{A}') Messräume und μ ein Maß auf (Ω, \mathcal{A}) . Ferner sei $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ messbar. Das durch

$$\mu \circ X^{-1} : \mathcal{A}' \rightarrow [0, \infty], \quad A' \mapsto \mu(X^{-1}(A'))$$

definierte Maß auf (Ω', \mathcal{A}') heißt **Bildmaß** von μ unter X .

Beispiel 1.99. Sei μ ein Maß auf \mathbb{Z}^2 und $X : \mathbb{Z}^2 \rightarrow \mathbb{Z}$, $(x, y) \mapsto x + y$. Dann ist

$$\mu \circ X^{-1}(\{x\}) = \sum_{y \in \mathbb{Z}} \mu(\{(x-y, y)\}). \quad \diamond$$

Beispiel 1.100. Ist $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine bijektive lineare Abbildung und λ das Lebesgue-Maß auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, so ist $\lambda \circ L^{-1} = |\det(L)|^{-1} \lambda$. Dies ist klar, weil für $a, b \in \mathbb{R}^n$ mit $a < b$ das Spat (oder Parallelepiped) $L^{-1}((a, b])$ das Volumen $|\det(L^{-1})| \prod_{i=1}^n (b_i - a_i)$ hat. \diamond

Als Verallgemeinerung des letzten Beispiels geben wir hier ohne Beweis den Transformationssatz für Maße mit stetigen Dichten unter differenzierbaren Abbildungen an. Den Beweis findet man in Lehrbüchern zur Analysis II unter dem Stichwort „Transformationssatz“ oder „Substitutionsregel“ (siehe etwa [8] oder [49]).

Satz 1.101 (Dichtetransformationsformel im \mathbb{R}^n). Es sei μ ein Maß auf \mathbb{R}^n mit stetiger (oder stückweise stetiger) Dichte $f : \mathbb{R}^n \rightarrow [0, \infty)$, das heißt

$$\mu((-\infty, x]) = \int_{-\infty}^{x_1} dt_1 \cdots \int_{-\infty}^{x_n} dt_n f(t_1, \dots, t_n) \quad \text{für jedes } x \in \mathbb{R}^n.$$

Sei $A \subset \mathbb{R}^n$ eine offene (oder abgeschlossene) Menge mit $\mu(\mathbb{R}^n \setminus A) = 0$. Ferner sei $B \subset \mathbb{R}^n$ offen oder abgeschlossen sowie $\varphi : A \rightarrow B$ bijektiv und stetig differenzierbar mit Ableitung φ' . Dann hat das Bildmaß $\mu \circ \varphi^{-1}$ die Dichte

$$f_\varphi(x) = \begin{cases} \frac{f(\varphi^{-1}(x))}{|\det(\varphi'(\varphi^{-1}(x)))|}, & \text{falls } x \in B, \\ 0, & \text{falls } x \in \mathbb{R}^n \setminus B. \end{cases}$$

Übung 1.4.1. Sei $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto |x|$. Zeige: Eine Borel-messbare Abbildung $g : \mathbb{R} \rightarrow \mathbb{R}$ ist genau dann messbar bezüglich $\sigma(f) = f^{-1}(\mathcal{B}(\mathbb{R}))$, wenn g gerade ist. ♣

Übung 1.4.2. Man zeige: Ist $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $f : \Omega \rightarrow \mathbb{R}$ messbar und $g = f$ μ -fast überall, so braucht g nicht messbar zu sein. ♣

Übung 1.4.3. Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ differenzierbar mit Ableitung f' . Zeige: f' ist $\mathcal{B}(\mathbb{R})$ -messbar. ♣

Übung 1.4.4. (Vergleiche Beispiele 1.40 und 1.63.) Sei $\Omega = \{0, 1\}^{\mathbb{N}}$ und $\mathcal{A} = (2^{\{0,1\}})^{\otimes \mathbb{N}}$ die σ -Algebra, die von den Zylindermengen $\{[\omega_1, \dots, \omega_n] : n \in \mathbb{N}, \omega_1, \dots, \omega_n \in \{0, 1\}\}$ erzeugt wird. Ferner sei $\mu = (\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)^{\otimes \mathbb{N}}$ das Bernoulli-Maß auf Ω mit gleichen Gewichten auf 0 und 1. Für $n \in \mathbb{N}$ sei $X_n : \Omega \rightarrow \{0, 1\}$, $\omega \mapsto \omega_n$ die n -te Koordinatenabbildung, und sei

$$U(\omega) = \sum_{n=1}^{\infty} X_n(\omega) 2^{-n} \quad \text{für } \omega \in \Omega.$$

- (i) Zeige: $\mathcal{A} = \sigma(X_n : n \in \mathbb{N})$.
- (ii) Zeige: U ist \mathcal{A} - $\mathcal{B}([0, 1])$ messbar.
- (iii) Bestimme das Bildmaß $\mu \circ U^{-1}$ auf $([0, 1], \mathcal{B}([0, 1]))$.
- (iv) Man gebe ein $\Omega_0 \in \mathcal{A}$ an, sodass $\tilde{U} := U|_{\Omega_0}$ bijektiv ist.
- (v) Man zeige, dass \tilde{U}^{-1} messbar ist bezüglich $\mathcal{B}([0, 1])$ - $\mathcal{A}|_{\Omega_0}$.
- (vi) Welche Interpretation hat die Abbildung $X_n \circ \tilde{U}^{-1}$? ♣

Übung 1.4.5 (Satz von Lusin). Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ Borel-messbar. Man zeige: Für jedes $\varepsilon > 0$ existiert eine abgeschlossene Menge $C \subset \mathbb{R}$ mit $\lambda(\mathbb{R} \setminus C) < \varepsilon$, sodass die Einschränkung $f|_C$ von f auf C stetig ist. (Merke: Dies heißt natürlich nicht, dass f in jedem Punkte $x \in C$ stetig wäre.)

Anleitung: Man zeige die Aussage zunächst mit Hilfe der inneren Regularität des Lebesgue-Maßes λ (Bemerkung 1.67) für Indikatorfunktionen messbarer Mengen und approximiere mit solchen die Abbildung f auf einer geeigneten Menge C gleichmäßig. ♣

1.5 Zufallsvariablen

In diesem Abschnitt werden wir messbare Abbildungen als *Zufallsvariablen* auffassen, die zufällige Beobachtungen beschreiben. Wir definieren den Begriff der Verteilung von Zufallsvariablen.

Im Folgenden sei stets $(\Omega, \mathcal{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum. Die Mengen $A \in \mathcal{A}$ heißen **Ereignisse**. $\mathbf{P}[A]$ wird als die Wahrscheinlichkeit interpretiert, dass A eintritt. Oft ist allerdings nicht der Wahrscheinlichkeitsraum selbst betrachtbar, sondern nur gewisse Beobachtungsgrößen. Wir wollen also Wahrscheinlichkeiten dafür definieren, dass Zufallsgrößen bestimmte Werte annehmen und einen Kalkül für, zum Beispiel, Summen von Zufallsgrößen entwickeln.

Definition 1.102 (Zufallsvariablen). Sei (Ω', \mathcal{A}') ein Messraum und $X : \Omega \rightarrow \Omega'$ messbar.

- (i) X heißt **Zufallsvariable** mit Werten in (Ω', \mathcal{A}') . Ist $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so nennen wir X eine reelle Zufallsvariable oder schlicht Zufallsvariable.
- (ii) Ist $A' \in \mathcal{A}'$, so schreiben wir $\{X \in A'\} := X^{-1}(A')$ und $\mathbf{P}[X \in A'] := \mathbf{P}[X^{-1}(A')]$. Speziell schreiben wir $\{X \geq 0\} := X^{-1}([0, \infty))$ und analog $\{X \leq b\}$ und so weiter.

Definition 1.103 (Verteilungen). Sei X eine Zufallsvariable.

- (i) Das W-Maß $\mathbf{P}_X := \mathbf{P} \circ X^{-1}$ heißt **Verteilung** von X .
- (ii) Ist X eine reelle Zufallsvariable, so heißt die Abbildung $F_X : x \mapsto \mathbf{P}[X \leq x]$ die **Verteilungsfunktion von X** (eigentlich von \mathbf{P}_X). Ist $\mu = \mathbf{P}_X$, so schreiben wir auch $X \sim \mu$ und sagen, dass X nach μ verteilt ist.
- (iii) Eine Familie $(X_i)_{i \in I}$ heißt **identisch verteilt**, falls $\mathbf{P}_{X_i} = \mathbf{P}_{X_j}$ für alle $i, j \in I$. Wir schreiben $X \stackrel{\mathcal{D}}{=} Y$, falls $\mathbf{P}_X = \mathbf{P}_Y$ (\mathcal{D} für distribution).

Satz 1.104. Zu jeder Verteilungsfunktion F existiert eine reelle Zufallsvariable X mit $F_X = F$.

Beweis. Wir müssen explizit einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$ und eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ angeben mit $F_X = F$.

Die einfachste Möglichkeit ist, $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ zu wählen, $X : \mathbb{R} \rightarrow \mathbb{R}$ die identische Abbildung und \mathbf{P} das Lebesgue-Stieltjes Maß mit Verteilungsfunktion F (siehe Beispiel 1.56).

Eine andere Möglichkeit, die zudem etwas lehrreicher ist, beruht darauf, zunächst unabhängig vom konkreten F eine Art Standard-Wahrscheinlichkeitsraum zu definieren, auf dem eine uniform auf $(0, 1)$ verteilte Zufallsvariable definiert ist, die

dann vermöge der Umkehrabbildung F^{-1} zu einer Zufallsvariablen X mit Verteilungsfunktion F transformiert wird: Wir wählen $\Omega := (0, 1)$, $\mathcal{A} := \mathcal{B}(\mathbb{R})|_{\Omega}$ und \mathbf{P} das Lebesgue-Maß auf (Ω, \mathcal{A}) (siehe Beispiel 1.74). Definiere die (linksstetige) Inverse von F

$$F^{-1}(t) := \inf\{x \in \mathbb{R} : F(x) \geq t\} \quad \text{für } t \in (0, 1).$$

Dann ist

$$F^{-1}(t) \leq x \iff t \leq F(x).$$

Speziell ist $\{t : F^{-1}(t) \leq x\} = (0, F(x)] \cap (0, 1)$, also ist $F^{-1} : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ messbar und

$$\mathbf{P}[\{t : F^{-1}(t) \leq x\}] = F(x).$$

Mithin ist $X := F^{-1}$ die gewünschte Zufallsvariable. \square

Beispiel 1.105. Wir geben zu verschiedenen Wahrscheinlichkeitsverteilungen auf \mathbb{R} reelle Zufallsvariablen X mit ebendieser Verteilung an. (Der konkrete Ort in diesem Buch dient lediglich als Vorwand, um ein paar der wichtigsten Verteilungen einzuführen, auf die wir bei späteren Gelegenheiten immer wieder zurückkommen.)

(i) Ist $p \in [0, 1]$ und $\mathbf{P}[X = 1] = p$, $\mathbf{P}[X = 0] = 1 - p$, so heißt $\mathbf{P}_X =: \text{Ber}_p$ die **Bernoulli-Verteilung** mit Parameter p . Formal ist

$$\text{Ber}_p = (1 - p)\delta_0 + p\delta_1,$$

und die Verteilungsfunktion ist

$$F_X(x) = \begin{cases} 0, & \text{falls } x < 0, \\ 1 - p, & \text{falls } x \in [0, 1), \\ 1, & \text{falls } x \geq 1. \end{cases}$$

(ii) Ist $p \in [0, 1]$ und $n \in \mathbb{N}$ sowie $X : \Omega \rightarrow \{0, \dots, n\}$ mit

$$\mathbf{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k},$$

so heißt $\mathbf{P}_X =: b_{n,p}$ die **Binomialverteilung** mit Parametern n und p . Formal ist

$$b_{n,p} = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta_k.$$

(iii) Ist $p \in (0, 1]$ und $X : \Omega \rightarrow \mathbb{N}_0$ mit

$$\mathbf{P}[X = n] = p(1 - p)^n \quad \text{für jedes } n \in \mathbb{N}_0,$$

so heißt $\gamma_p := b_{1,p}^- := \mathbf{P}_X$ die **geometrische Verteilung**¹ mit Parameter p . Formal können wir schreiben:

$$\gamma_p = \sum_{n=0}^{\infty} p(1-p)^n \delta_n.$$

Die Verteilungsfunktion ist $F(x) = 1 - (1-p)^{\lfloor x+1 \rfloor \vee 0}$ für $x \in \mathbb{R}$.

Wir können $X + 1$ als die Wartezeit auf den ersten Erfolg bei „unabhängigen“ Zufallsexperimenten auffassen, die jeweils mit Wahrscheinlichkeit p zum Erfolg führen. In der Tat: Sei $\Omega = \{0, 1\}^{\mathbb{N}}$ und \mathbf{P} das Produktmaß $((1-p)\delta_0 + p\delta_1)^{\otimes \mathbb{N}}$ (Satz 1.64) sowie $\mathcal{A} = \sigma([\omega_1, \dots, \omega_n] : \omega_1, \dots, \omega_n \in \{0, 1\}, n \in \mathbb{N})$. Wir setzen

$$X(\omega) := \inf\{n \in \mathbb{N} : \omega_n = 1\} - 1,$$

mit der Konvention $\inf \emptyset = \infty$. Offenbar ist jede der Abbildungen

$$X_n : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto \begin{cases} n-1, & \text{falls } \omega_n = 1, \\ \infty, & \text{falls } \omega_n = 0, \end{cases}$$

$\mathcal{A} - \mathcal{B}(\overline{\mathbb{R}})$ -messbar und $X = \inf_{n \in \mathbb{N}} X_n$. Also ist X auch $\mathcal{A} - \mathcal{B}(\overline{\mathbb{R}})$ -messbar, also eine Zufallsvariable. Sei $\omega^0 := (0, 0, \dots) \in \Omega$. Dann ist $\mathbf{P}[X \geq n] = \mathbf{P}[[\omega_1^0, \dots, \omega_n^0]] = (1-p)^n$. Also ist

$$\mathbf{P}[X = n] = \mathbf{P}[X \geq n] - \mathbf{P}[X \geq n+1] = (1-p)^n - (1-p)^{n+1} = p(1-p)^n.$$

(iv) Seien $r > 0$ (nicht notwendigerweise ganzzahlig) und $p \in (0, 1]$. Mit

$$b_{r,p}^- := \sum_{k=0}^{\infty} \binom{-r}{k} (-1)^k p^r (1-p)^k \delta_k \tag{1.17}$$

bezeichnen wir die **negative Binomialverteilung** oder **Pascal-Verteilung** mit Parametern r und p . (Hierbei ist $\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!}$ für $x \in \mathbb{R}$ und $k \in \mathbb{N}$ der verallgemeinerte Binomialkoeffizient.) Für $r \in \mathbb{N}$ ist $b_{r,p}^-$, ähnlich wie im vorangehenden Beispiel, die Verteilung der Wartezeit auf den r -ten Erfolg bei unabhängigen Versuchen. Wir werden hierauf in Beispiel 3.4(iv) zurückkommen.

(v) Ist $\lambda \in [0, \infty)$ und $X : \Omega \rightarrow \mathbb{N}_0$ mit

$$\mathbf{P}[X = n] = e^{-\lambda} \frac{\lambda^n}{n!} \quad \text{für jedes } n \in \mathbb{N}_0,$$

so heißt $\mathbf{P}_X =: \text{Poi}_\lambda$ die **Poisson-Verteilung** mit Parameter λ .

(vi) Die **hypergeometrische Verteilung** mit Parametern $B, W, n \in \mathbb{N}$

¹ Obacht: Manche Autoren nennen die um Eins verschobene Verteilung auf \mathbb{N} die geometrische Verteilung.

$$\text{Hyp}_{B,W;n}(\{b\}) = \frac{\binom{B}{b} \binom{W}{n-b}}{\binom{B+W}{n}} \quad \text{für } b \in \{0, \dots, n\}, \quad (1.18)$$

gibt die Wahrscheinlichkeit dafür an, aus einer Urne mit B schwarzen und W weißen Kugeln bei n -maligen Ziehen ohne Zurücklegen genau b schwarze Kugeln zu ziehen. Mit ein bisschen Kombinatorik lässt sich dies leicht auf die Situation mit k Farben und B_i Kugeln der Farbe $i = 1, \dots, k$ verallgemeinern. Die Wahrscheinlichkeit, dass unter n gezogenen Kugeln exakt b_i von jeder Farbe $i = 1, \dots, k$ sind, ist gegeben durch die **verallgemeinerte hypergeometrische Verteilung**

$$\text{Hyp}_{B_1, \dots, B_k; n}(\{(b_1, \dots, b_k)\}) = \frac{\binom{B_1}{b_1} \cdots \binom{B_k}{b_k}}{\binom{B_1 + \dots + B_k}{n}}. \quad (1.19)$$

(vii) Seien $\mu \in \mathbb{R}$, $\sigma^2 > 0$ und X reell mit

$$\mathbf{P}[X \leq x] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt \quad \text{für } x \in \mathbb{R}.$$

Dann heißt $\mathbf{P}_X =: \mathcal{N}_{\mu, \sigma^2}$ Gauß'sche **Normalverteilung** mit Parametern μ und σ^2 .

(viii) Ist $X \geq 0$ reell und $\theta > 0$ sowie

$$\mathbf{P}[X \leq x] = \mathbf{P}[X \in [0, x]] = \int_0^x \theta e^{-\theta t} dt \quad \text{für } x \geq 0,$$

so heißt \mathbf{P}_X **Exponentialverteilung** mit Parameter θ (kurz: \exp_θ).

(ix) Ist $X \mathbb{R}^d$ -wertig, $\mu \in \mathbb{R}^d$, Σ eine positiv definite $d \times d$ Matrix und

$$\mathbf{P}[X \leq x] = \det(2\pi \Sigma)^{-1/2} \int_{(-\infty, x]} \exp\left(-\frac{1}{2} \langle t - \mu, \Sigma^{-1}(t - \mu) \rangle\right) \lambda^d(dt)$$

für $x \in \mathbb{R}^d$ (wobei $\langle \cdot, \cdot \rangle$ das Skalarprodukt im \mathbb{R}^d bezeichnet), so heißt $\mathbf{P}_X =: \mathcal{N}_{\mu, \Sigma}$ die d -dimensionale Normalverteilung mit Parametern μ und Σ . \diamond

Definition 1.106. Hat die Verteilungsfunktion $F : \mathbb{R}^n \rightarrow [0, 1]$ die Gestalt

$$F(x) = \int_{-\infty}^{x_1} dt_1 \cdots \int_{-\infty}^{x_n} dt_n f(t_1, \dots, t_n) \quad \text{für } x = (x_1, \dots, x_n) \in \mathbb{R}^n,$$

für eine integrierbare Funktion $f : \mathbb{R}^n \rightarrow [0, \infty)$, so heißt f die **Dichte** der Verteilung.

Beispiel 1.107. (i) Für $\theta, r > 0$ heißt die Verteilung $\Gamma_{\theta,r}$ auf $[0, \infty)$ mit Dichte

$$x \mapsto \frac{\theta^r}{\Gamma(r)} x^{r-1} e^{-\theta x}$$

(wo Γ die Gamma-Funktion bezeichnet) **Gamma-Verteilung** mit Größenparameter θ und Formparameter r .

(ii) Für $r, s > 0$ heißt die Verteilung $\beta_{r,s}$ auf $[0, 1]$ mit Dichte

$$x \mapsto \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1} (1-x)^{s-1}$$

Beta-Verteilung mit Parametern r und s .

(iii) Für $a > 0$ heißt die Verteilung Cau_a auf \mathbb{R} mit Dichte

$$x \mapsto \frac{1}{a\pi} \frac{1}{1 + (x/a)^2}$$

Cauchy-Verteilung mit Parameter a . ◇

Übung 1.5.1. Man leite (1.17) nach der Interpretation als Wartezeit kombinatorisch her unter Benutzung der Identität $\binom{-n}{k}(-1)^k = \binom{n+k-1}{k}$. ♣

Übung 1.5.2. Man gebe ein Beispiel an für zwei normalverteilte X und Y , sodass (X, Y) nicht (zweidimensional) normalverteilt ist. ♣

Übung 1.5.3. Man zeige mit Hilfe von Satz 1.101 (Transformationsformel für Dichten):

- (i) Ist $X \sim \mathcal{N}_{\mu, \sigma^2}$ und sind $a \in \mathbb{R} \setminus \{0\}$ und $b \in \mathbb{R}$, so ist $(aX + b) \sim \mathcal{N}_{a\mu + b, a^2\sigma^2}$.
- (ii) Ist $X \sim \exp_\theta$ und $a > 0$, so ist $aX \sim \exp_{\theta/a}$. ♣

Übung 1.5.4. Man zeige, dass $F : \mathbb{R}^2 \rightarrow [0, 1]$ genau dann die Verteilungsfunktion eines (eindeutig bestimmten) Wahrscheinlichkeitsmaßes μ auf $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ ist, falls

- (i) F monoton wachsend und rechtsstetig ist,
- (ii) $F(-x) \rightarrow 0$ und $F(x) \rightarrow 1$ für $x \rightarrow \infty$ gilt, und
- (iii) $F((y_1, y_2)) - F((y_1, x_2)) - F((x_1, y_2)) + F((x_1, x_2)) \geq 0$ für alle $x_1 \leq y_1$ und $x_2 \leq y_2$ gilt. ♣

Übung 1.5.5. (i) Seien F und G zwei Verteilungsfunktionen auf \mathbb{R} . Verwende Übung 1.5.4 um zu zeigen, dass $(x, y) \mapsto F(x) \wedge G(y)$ eine Verteilungsfunktion auf \mathbb{R}^2 ist.

(ii) Man gebe ein Beispiel an für zwei Verteilungsfunktionen F und G auf \mathbb{R}^2 , so dass $(x, y) \mapsto F(x) \wedge G(y)$ *keine* Verteilungsfunktion auf \mathbb{R}^4 ist.

Hinweis: Man leite mit Hilfe der Einschluss- Ausschlussformel (Satz 1.33) ein ähnliches Kriterium her wie in Übung 1.5.4(iii). 

2 Unabhängigkeit

Die Maßtheorie aus dem vorigen Kapitel ist eine lineare Theorie, die keine Abhängigkeitsstrukturen zwischen Ereignissen oder Zufallsvariablen kennt. Wir betreten das Gebiet der Wahrscheinlichkeitstheorie genau an dieser Stelle mit der Definition der Unabhängigkeit von Ereignissen und schließlich von Zufallsvariablen. Die Unabhängigkeit ist ein zentraler Begriff der Wahrscheinlichkeitstheorie, die Quantifizierung von Abhängigkeiten eines ihrer wichtigen Anliegen.

Fortan ist stets $(\Omega, \mathcal{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, und die Mengen $A \in \mathcal{A}$ sind die Ereignisse. Sobald wir die Phase hinter uns gelassen haben, in der wir W-Räume explizit konstruieren, wird der konkrete W-Raum in den Hintergrund treten, und die beobachtbaren Größen, also Zufallsvariablen, werden an Bedeutung gewinnen. Das fett gedruckte \mathbf{P} steht dann für das universelle Objekt des W-Maßes, und Wahrscheinlichkeiten $\mathbf{P}[\cdot]$ bezüglich \mathbf{P} werden stets mit eckigen Klammern geschrieben.

2.1 Unabhängigkeit von Ereignissen

Wir wollen zwei Ereignisse A und B als (stochastisch) unabhängig betrachten, wenn das Eintreten von A nicht die Wahrscheinlichkeit beeinflusst, dass zudem B eintritt. Etwas formaler können wir diesen intuitiven Begriff fassen, indem wir A und B als unabhängig betrachten, wenn

$$\mathbf{P}[A \cap B] = \mathbf{P}[A] \cdot \mathbf{P}[B]. \quad (2.1)$$

Beispiel 2.1 (Zweifacher Würfelwurf). Wir betrachten das Zufallsexperiment des zweifachen Würfelwurfs. Es ist also $\Omega = \{1, \dots, 6\}^2$ mit der σ -Algebra $\mathcal{A} = 2^\Omega$ und der Gleichverteilung $\mathbf{P} = \mathcal{U}_\Omega$ (siehe Beispiel 1.30(ii)).

(i) Zwei Ereignisse A und B sollten unabhängig sein, wenn A nur vom Ergebnis des ersten Wurfes abhängt, B hingegen nur vom Ergebnis des zweiten Wurfes. Formal beschreiben wir diese Situation, indem wir annehmen, dass es Mengen $\tilde{A}, \tilde{B} \subset \{1, \dots, 6\}$ gibt mit

$$A = \tilde{A} \times \{1, \dots, 6\} \quad \text{und} \quad B = \{1, \dots, 6\} \times \tilde{B}.$$

Wir prüfen jetzt, dass A und B tatsächlich (2.1) erfüllen. Dazu berechnen wir $\mathbf{P}[A] = \frac{\#A}{36} = \frac{\#\tilde{A}}{6}$ und $\mathbf{P}[B] = \frac{\#B}{36} = \frac{\#\tilde{B}}{6}$. Ferner ist

$$\mathbf{P}[A \cap B] = \frac{\#(\tilde{A} \times \tilde{B})}{36} = \frac{\#\tilde{A}}{6} \cdot \frac{\#\tilde{B}}{6} = \mathbf{P}[A] \cdot \mathbf{P}[B].$$

(ii) Stochastische Unabhängigkeit kann auch in weniger augenfälligen Situationen auftreten. Sei hierzu A das Ereignis, dass die Augensumme ungerade ist, $A = \{(\omega_1, \omega_2) \in \Omega : \omega_1 + \omega_2 \in \{3, 5, 7, 9, 11\}\}$, und B das Ereignis, dass der erste Wurf höchstens eine Drei bringt, $B = \{(\omega_1, \omega_2) \in \Omega : \omega_1 \in \{1, 2, 3\}\}$. Obwohl beide Ereignisse anscheinend etwas miteinander zu tun haben, sind sie stochastisch unabhängig, denn es gilt, wie man leicht prüft, $\mathbf{P}[A] = \mathbf{P}[B] = \frac{1}{2}$ und $\mathbf{P}[A \cap B] = \frac{1}{4}$. \diamond

Wann sind nun *drei* Ereignisse A_1, A_2, A_3 unabhängig? Hierzu muss natürlich jedes der Paare (A_1, A_2) , (A_1, A_3) und (A_2, A_3) unabhängig sein. Jedoch wollen wir auch sicherstellen, dass beispielsweise das Eintreten von A_1 und A_2 nicht die Wahrscheinlichkeit für das zusätzliche Eintreten von A_3 beeinflusst. Wir müssen also mehr als nur Paare betrachten.

Formal nennen wir daher drei Ereignisse A_1, A_2 und A_3 (stochastisch) unabhängig, falls

$$\mathbf{P}[A_i \cap A_j] = \mathbf{P}[A_i] \cdot \mathbf{P}[A_j] \quad \text{für alle } i, j \in \{1, 2, 3\}, i \neq j, \quad (2.2)$$

und

$$\mathbf{P}[A_1 \cap A_2 \cap A_3] = \mathbf{P}[A_1] \cdot \mathbf{P}[A_2] \cdot \mathbf{P}[A_3]. \quad (2.3)$$

Man beachte, dass (2.3) nicht aus (2.2) folgt (und (2.2) nicht aus (2.3)).

Beispiel 2.2 (Dreifacher Würfelwurf). Wir betrachten den dreifachen Wurf eines Würfels. Sei also $\Omega = \{1, \dots, 6\}^3$ ausgestattet mit der diskreten σ -Algebra $\mathcal{A} = 2^\Omega$ und der Gleichverteilung $\mathbf{P} = \mathcal{U}_\Omega$ (siehe Beispiel 1.30(ii)).

(i) Hängt für $i = 1, 2, 3$ das Ereignis A_i nur vom i -ten Wurf ab, so sind die Ereignisse A_1, A_2 und A_3 unabhängig. In der Tat können wir sie wie im vorangehenden Beispiel für gewisse Mengen $\tilde{A}_1, \tilde{A}_2, \tilde{A}_3 \subset \{1, \dots, 6\}$ schreiben als

$$\begin{aligned} A_1 &= \tilde{A}_1 \times \{1, \dots, 6\}^2, \\ A_2 &= \{1, \dots, 6\} \times \tilde{A}_2 \times \{1, \dots, 6\}, \\ A_3 &= \{1, \dots, 6\}^2 \times \tilde{A}_3. \end{aligned}$$

Die Gültigkeit von (2.2) folgt wie in Beispiel 2.1(i). Um (2.3) zu zeigen, berechnen wir

$$\mathbf{P}[A_1 \cap A_2 \cap A_3] = \frac{\#(\tilde{A}_1 \times \tilde{A}_2 \times \tilde{A}_3)}{216} = \prod_{i=1}^3 \frac{\#\tilde{A}_i}{6} = \prod_{i=1}^3 \mathbf{P}[A_i].$$

(ii) Wir betrachten nun die folgenden drei Ereignisse

$$\begin{aligned} A_1 &:= \{\omega \in \Omega : \omega_1 = \omega_2\}, \\ A_2 &:= \{\omega \in \Omega : \omega_2 = \omega_3\}, \\ A_3 &:= \{\omega \in \Omega : \omega_1 = \omega_3\}. \end{aligned}$$

Dann ist $\#A_1 = \#A_2 = \#A_3 = 36$, also $\mathbf{P}[A_1] = \mathbf{P}[A_2] = \mathbf{P}[A_3] = \frac{1}{6}$. Ferner ist $\#(A_i \cap A_j) = 6$, falls $i \neq j$, also $\mathbf{P}[A_i \cap A_j] = \frac{1}{36}$. Daher gilt (2.2). Jedoch ist $\#(A_1 \cap A_2 \cap A_3) = 6$, also $\mathbf{P}[A_1 \cap A_2 \cap A_3] = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$, mithin ist (2.3) verletzt, und die Ereignisse A_1, A_2, A_3 sind nicht unabhängig. \diamond

Um für größere Familien von Ereignissen Unabhängigkeit zu definieren, müssen wir die Gültigkeit von Produktformeln wie (2.2) und (2.3) nunmehr nicht nur für Paare und Tripel fordern, sondern für alle endlichen Teilmengen. Wir treffen daher die folgende Definition.

Definition 2.3 (Unabhängigkeit von Ereignissen). Sei I eine beliebige Indexmenge, und sei $(A_i)_{i \in I}$ eine beliebige Familie von Ereignissen. Die Familie $(A_i)_{i \in I}$ heißt **unabhängig**, falls für jede endliche Teilmenge $J \subset I$ gilt, dass

$$\mathbf{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbf{P}[A_j].$$

Das wichtigste Beispiel für eine unendlich große, unabhängige Familie von Ereignissen wird durch die unendliche (unabhängige) Wiederholung eines Zufallsexperiments gegeben.

Beispiel 2.4. Sei E eine endliche Menge (die möglichen Ausgänge des einzelnen Experiments) und $(p_e)_{e \in E}$ ein Wahrscheinlichkeitsvektor auf E . Sei (wie in Satz 1.64) der Wahrscheinlichkeitsraum $\Omega = E^{\mathbb{N}}$ ausgestattet mit der σ -Algebra $\mathcal{A} = \sigma(\{[\omega_1, \dots, \omega_n] : \omega_1, \dots, \omega_n \in E, n \in \mathbb{N}\})$ und $\mathbf{P} = (\sum_{e \in E} p_e \delta_e)^{\otimes \mathbb{N}}$ das Produktmaß (oder Bernoulli-Maß) auf (Ω, \mathcal{A}) , also $\mathbf{P}[[\omega_1, \dots, \omega_n]] = \prod_{i=1}^n p_{\omega_i}$. Sei $\tilde{A}_i \subset E$ für jedes $i \in \mathbb{N}$, und A_i das Ereignis, dass \tilde{A}_i im i -ten Durchgang des Experiments auftritt, also

$$A_i = \{\omega \in \Omega : \omega_i \in \tilde{A}_i\} = \biguplus_{(\omega_1, \dots, \omega_i) \in E^{i-1} \times \tilde{A}_i} [\omega_1, \dots, \omega_i].$$

Nach unserer Intuition sollte die Familie $(A_i)_{i \in \mathbb{N}}$ unabhängig sein, wenn die Definition der Unabhängigkeit sinnvoll sein soll. Wir weisen jetzt nach, dass dies in der Tat richtig ist. Sei $J \subset \mathbb{N}$ endlich und $n := \max J$. Wir setzen formal $B_j = A_j$ und $\tilde{B}_j = \tilde{A}_j$ für $j \in J$ und $B_j = \Omega$ und $\tilde{B}_j = E$ für $j \in \{1, \dots, n\} \setminus J$. Dann ist

$$\begin{aligned}\mathbf{P}\left[\bigcap_{j \in J} A_j\right] &= \mathbf{P}\left[\bigcap_{j \in J} B_j\right] = \mathbf{P}\left[\bigcap_{j=1}^n B_j\right] \\ &= \sum_{e_1 \in \tilde{B}_1} \cdots \sum_{e_n \in \tilde{B}_n} \prod_{j=1}^n p_{e_j} = \prod_{j=1}^n \left(\sum_{e \in \tilde{B}_j} p_e \right) = \prod_{j \in J} \left(\sum_{e \in \tilde{A}_j} p_e \right).\end{aligned}$$

Dies gilt speziell natürlich für $\#J = 1$, also ist $\mathbf{P}[A_i] = \sum_{e \in \tilde{A}_i} p_e$ für jedes $i \in \mathbb{N}$. Es folgt

$$\mathbf{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbf{P}[A_j]. \quad (2.4)$$

Da dies für alle endlichen $J \subset \mathbb{N}$ gilt, ist die Familie $(A_i)_{i \in \mathbb{N}}$ unabhängig. \diamond

Sind A und B unabhängig, so sind auch A^c und B unabhängig, denn $\mathbf{P}[A^c \cap B] = \mathbf{P}[B] - \mathbf{P}[A \cap B] = \mathbf{P}[B] - \mathbf{P}[A]\mathbf{P}[B] = (1 - \mathbf{P}[A])\mathbf{P}[B] = \mathbf{P}[A^c]\mathbf{P}[B]$. Wir wollen diese Beobachtung etwas verallgemeinern und als Satz festhalten.

Satz 2.5. Sei I eine beliebige Indexmenge, und sei $(A_i)_{i \in I}$ eine Familie von Ereignissen. Setze $B_i^0 = A_i$ und $B_i^1 = A_i^c$ für $i \in I$. Dann sind folgende drei Aussagen äquivalent.

- (i) Die Familie $(A_i)_{i \in I}$ ist unabhängig.
- (ii) Es gibt ein $\alpha \in \{0, 1\}^I$, sodass die Familie $(B_i^{\alpha_i})_{i \in I}$ unabhängig ist.
- (iii) Für jedes $\alpha \in \{0, 1\}^I$ ist die Familie $(B_i^{\alpha_i})_{i \in I}$ unabhängig.

Beweis. Übung! \square

Beispiel 2.6 (Euler'sche Primzahlformel). Die **Riemann'sche Zetafunktion** ist definiert durch die Dirichlet-Reihe

$$\zeta(s) := \sum_{n=1}^{\infty} n^{-s} \quad \text{für } s \in (1, \infty).$$

Die Euler'sche Primzahlformel ist die Produktdarstellung

$$\zeta(s) = \prod_{p \in \mathcal{P}} (1 - p^{-s})^{-1}, \quad (2.5)$$

wobei $\mathcal{P} := \{p \in \mathbb{N} : p \text{ ist Primzahl}\}$ ist.

Wir beweisen die Produktdarstellung probabilistisch. Sei $\Omega = \mathbb{N}$ und (für festes s) \mathbf{P} definiert durch

$$\mathbf{P}[\{n\}] = \zeta(s)^{-1} n^{-s} \quad \text{für } n \in \mathbb{N}.$$

Sei $p\mathbb{N} = \{pn : n \in \mathbb{N}\}$ und $\mathcal{P}_n = \{p \in \mathcal{P} : p \leq n\}$. Wir fassen $p\mathbb{N} \subset \Omega$ als Ereignis auf und bemerken, dass $(p\mathbb{N}, p \in \mathcal{P})$ unabhängig ist. In der Tat: Für $k \in \mathbb{N}$ und unterschiedliche $p_1, \dots, p_k \in \mathcal{P}$ ist $\bigcap_{i=1}^k (p_i\mathbb{N}) = (p_1 \cdots p_k)\mathbb{N}$, also

$$\begin{aligned}
\mathbf{P}\left[\bigcap_{i=1}^k (p_i \mathbb{N})\right] &= \sum_{n=1}^{\infty} \mathbf{P}[\{p_1 \cdots p_k n\}] \\
&= \zeta(s)^{-1} (p_1 \cdots p_k)^{-s} \sum_{n=1}^{\infty} n^{-s} \\
&= (p_1 \cdots p_k)^{-s} = \prod_{i=1}^k \mathbf{P}[p_i \mathbb{N}].
\end{aligned}$$

Nach Satz 2.5 ist nun auch $((p\mathbb{N})^c, p \in \mathcal{P})$ unabhängig. Deshalb gilt

$$\begin{aligned}
\zeta(s)^{-1} &= \mathbf{P}[\{1\}] = \mathbf{P}\left[\bigcap_{p \in \mathcal{P}} (p\mathbb{N})^c\right] \\
&= \lim_{n \rightarrow \infty} \mathbf{P}\left[\bigcap_{p \in \mathcal{P}_n} (p\mathbb{N})^c\right] \\
&= \lim_{n \rightarrow \infty} \prod_{p \in \mathcal{P}_n} (1 - \mathbf{P}[p\mathbb{N}]) = \prod_{p \in \mathcal{P}} (1 - p^{-s}).
\end{aligned}$$

Damit ist (2.5) gezeigt. \diamond

Wenn wir einen Würfel unendlich oft werfen, wie groß ist die Wahrscheinlichkeit, dass unendlich oft (also: immer wieder mal) eine Sechs geworfen wird? Diese Wahrscheinlichkeit sollte Eins sein, denn sonst gäbe es einen letzten Zeitpunkt, zu dem eine Sechs fällt und danach nicht wieder. Dies wäre zumindest nicht sehr plausibel.

Man erinnere sich daran, wie wir mit Hilfe des Limes superior (Definition 1.13) formalisiert hatten, dass unendlich viele Ereignisse aus einer Familie von Ereignissen eintreten. Der folgende Satz bestätigt nun unsere oben geäußerte Vermutung und gibt zudem Auskunft darüber, unter welchen Bedingungen wir *nicht* erwarten können, dass unendlich viele der Ereignisse eintreten.

Satz 2.7 (Lemma von Borel-Cantelli). Seien A_1, A_2, \dots Ereignisse, und sei $A^* = \limsup_{n \rightarrow \infty} A_n$.

(i) Ist $\sum_{n=1}^{\infty} \mathbf{P}[A_n] < \infty$, so ist $\mathbf{P}[A^*] = 0$. (Hier kann \mathbf{P} ein beliebiges Maß auf (Ω, \mathcal{A}) sein.)

(ii) Ist $(A_n)_{n \in \mathbb{N}}$ unabhängig und $\sum_{n=1}^{\infty} \mathbf{P}[A_n] = \infty$, so ist $\mathbf{P}[A^*] = 1$.

Beweis. (i) Da \mathbf{P} stetig von oben und σ -subadditiv ist, ist nach Voraussetzung

$$\mathbf{P}[A^*] = \lim_{n \rightarrow \infty} \mathbf{P}\left[\bigcup_{m=n}^{\infty} A_m\right] \leq \lim_{n \rightarrow \infty} \sum_{m=n}^{\infty} \mathbf{P}[A_m] = 0.$$

(ii) Nach den de Morgan'schen Regeln und der Stetigkeit von \mathbf{P} von unten gilt daher

$$\mathbf{P}[(A^*)^c] = \mathbf{P}\left[\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n^c\right] = \lim_{m \rightarrow \infty} \mathbf{P}\left[\bigcap_{n=m}^{\infty} A_n^c\right].$$

Da \mathbf{P} auch stetig von oben ist, gilt aber für jedes $m \in \mathbb{N}$ (wegen $\log(1-x) \leq -x$ für $x \in [0, 1]$)

$$\begin{aligned} \mathbf{P}\left[\bigcap_{n=m}^{\infty} A_n^c\right] &= \lim_{N \rightarrow \infty} \mathbf{P}\left[\bigcap_{n=m}^N A_n^c\right] = \prod_{n=m}^{\infty} (1 - \mathbf{P}[A_n]) \\ &= \exp\left(\sum_{n=m}^{\infty} \log(1 - \mathbf{P}[A_n])\right) \leq \exp\left(-\sum_{n=m}^{\infty} \mathbf{P}[A_n]\right) = 0. \quad \square \end{aligned}$$

Beispiel 2.8. Wir betrachten den unendlich oft wiederholten Würfelwurf und fragen nach der Wahrscheinlichkeit, dass unendlich oft die Sechs auftritt. Es ist also $\Omega = \{1, \dots, 6\}^{\mathbb{N}}$, $\mathcal{A} = (2^{\{1, \dots, 6\}})^{\otimes \mathbb{N}}$ die Produkt- σ -Algebra und $\mathbf{P} = \left(\sum_{e \in \{1, \dots, 6\}} \frac{1}{6} \delta_e\right)^{\otimes \mathbb{N}}$ das Bernoulli-Maß (vergleiche Satz 1.64). Ferner sei $A_n = \{\omega \in \Omega : \omega_n = 6\}$ das Ereignis, dass beim n -ten Wurf eine Sechs auftritt. Dann ist $A^* = \limsup_{n \rightarrow \infty} A_n$ das Ereignis, dass unendlich oft eine Sechs auftritt (siehe Bemerkung 1.14). Ferner ist $(A_n)_{n \in \mathbb{N}}$ eine unabhängige Familie mit $\sum_{n=1}^{\infty} \mathbf{P}[A_n] = \sum_{n=1}^{\infty} \frac{1}{6} = \infty$ und deshalb nach dem Lemma von Borel-Cantelli $\mathbf{P}[A^*] = 1$. \diamond

Beispiel 2.9. Wir werfen einen Würfel nur einmal und definieren A_n für jedes $n \in \mathbb{N}$ als das Ereignis, dass bei diesem (einen) Wurf eine Sechs geworfen wurde. Man bemerke, dass $A_1 = A_2 = A_3 = \dots$. Dann ist $\sum_{n \in \mathbb{N}} \mathbf{P}[A_n] = \infty$, jedoch $\mathbf{P}[A^*] = \mathbf{P}[A_1] = \frac{1}{6}$. Dies zeigt, dass in Teil (ii) des Lemmas von Borel-Cantelli nicht ohne weiteres auf die Unabhängigkeit verzichtet werden kann. \diamond

Beispiel 2.10. Sei $\Lambda \in (0, \infty)$ und $0 \leq \lambda_n \leq \Lambda$ für $n \in \mathbb{N}$. Ferner seien $X_n, n \in \mathbb{N}$, Poisson-verteilte Zufallsvariablen mit Parametern λ_n . Dann gilt

$$\mathbf{P}[X_n \geq n \text{ für unendlich viele } n] = 0.$$

Es ist nämlich

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbf{P}[X_n \geq n] &= \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \mathbf{P}[X_n = m] = \sum_{m=1}^{\infty} \sum_{n=1}^m \mathbf{P}[X_n = m] \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^m e^{-\lambda_n} \frac{\lambda_n^m}{m!} \leq \sum_{m=1}^{\infty} m \frac{\Lambda^m}{m!} = \Lambda e^{\Lambda} < \infty. \quad \diamond \end{aligned}$$

Da im obigen Satz, zumindest im Falle unabhängiger Ereignisse, nur die Wahrscheinlichkeiten $\mathbf{P}[A^*] = 0$ und $\mathbf{P}[A^*] = 1$ auftreten können, zählt das Lemma von Borel-Cantelli zu den so genannten 0-1 Gesetzen. Wir werden später weitere 0-1 Gesetze kennen lernen (siehe beispielsweise Satz 2.37).

Wir wollen jetzt den Begriff der Unabhängigkeit von Familien von Ereignissen auf Familien von Ereignissystemen ausdehnen.

Definition 2.11 (Unabhängigkeit von Mengensystemen). Sei I eine beliebige Indexmenge und $\mathcal{E}_i \subset \mathcal{A}$ für jedes $i \in I$. Die Familie $(\mathcal{E}_i)_{i \in I}$ heißt **unabhängig**, falls für jede endliche Teilmenge $J \subset I$ und für jede Wahl von $E_j \in \mathcal{E}_j$, $j \in J$, gilt, dass

$$\mathbf{P}\left[\bigcap_{j \in J} E_j\right] = \prod_{j \in J} \mathbf{P}[E_j]. \quad (2.6)$$

Beispiel 2.12. Sei $(\Omega, \mathcal{A}, \mathbf{P})$ wie in Beispiel 2.4 der Produktraum der unendlichen Wiederholung des Experiments mit Ausgängen in der endlichen Menge E mit Wahrscheinlichkeitsvektor $p = (p_e)_{e \in E}$. Setze für $i \in \mathbb{N}$

$$\mathcal{E}_i = \{\{\omega \in \Omega : \omega_i \in A\} : A \subset E\}.$$

Dann ist für jede Wahl von $A_i \in \mathcal{E}_i$, $i \in \mathbb{N}$, die Familie $(A_i)_{i \in \mathbb{N}}$ unabhängig, also ist $(\mathcal{E}_i)_{i \in \mathbb{N}}$ unabhängig. \diamond

Satz 2.13. (i) Sei I endlich, und für jedes $i \in I$ sei $\mathcal{E}_i \subset \mathcal{A}$ mit $\Omega \in \mathcal{E}_i$. Dann gilt

$$(\mathcal{E}_i)_{i \in I} \text{ ist unabhängig} \iff (2.6) \text{ gilt für } J = I.$$

(ii) $(\mathcal{E}_i)_{i \in I}$ ist unabh. $\iff ((\mathcal{E}_j)_{j \in J}$ ist unabh. für alle endlichen $J \subset I$).

(iii) Ist $(\mathcal{E}_i \cup \{\emptyset\}) \cap$ -stabil, dann gilt

$$(\mathcal{E}_i)_{i \in I} \text{ ist unabhängig} \iff (\sigma(\mathcal{E}_i))_{i \in I} \text{ ist unabhängig.}$$

(iv) Sei K eine beliebige Menge und $(I_k)_{k \in K}$ paarweise disjunkte Teilmengen von I . Ist $(\mathcal{E}_i)_{i \in I}$ unabhängig, dann ist auch $(\bigcup_{i \in I_k} \mathcal{E}_i)_{k \in K}$ unabhängig.

Beweis. (i), \implies Dies ist trivial.

(i), \impliedby Für $J \subset I$ und $j \in I \setminus J$ wähle $E_j = \Omega$.

(ii) Dies ist trivial.

(iii), \impliedby Dies ist trivial.

(iii) „ \implies “ Sei $J \subset I$ endlich. Wir zeigen: Für je zwei endliche Mengen J und J' mit $J \subset J' \subset I$ gilt

$$\mathbf{P}\left[\bigcap_{i \in J'} E_i\right] = \prod_{i \in J'} \mathbf{P}[E_i] \text{ für jede Wahl } \begin{cases} E_i \in \sigma(\mathcal{E}_i), & \text{falls } i \in J, \\ E_i \in \mathcal{E}_i, & \text{falls } i \in J' \setminus J. \end{cases} \quad (2.7)$$

Mit $J' = J$ ist dies genau die zu zeigende Aussage.

Wir führen den Beweis von (2.7) durch vollständige Induktion nach $\#J$. Für $\#J = 0$ gilt (2.7) nach Voraussetzung des Satzes.

Es gelte nun (2.7) für jedes J mit $\#J = n$ und jedes endliche $J' \supset J$. Sei solch ein J gewählt und $j \in I \setminus J$. Sei $J' \supset \tilde{J} := J \cup \{j\}$. Wir zeigen nun die Gültigkeit von (2.7) mit \tilde{J} statt mit J . Wegen $\#\tilde{J} = n + 1$ ist damit der Induktionsschritt gezeigt.

Sei $E_i \in \sigma(\mathcal{E}_i)$ für jedes $i \in J$ und $E_i \in \mathcal{E}_i$ für jedes $i \in J' \setminus (J \cup \{j\})$. Wir definieren Maße μ und ν auf (Ω, \mathcal{A}) durch

$$\mu : E_j \mapsto \mathbf{P}\left[\bigcap_{i \in J'} E_i\right] \quad \text{und} \quad \nu : E_j \mapsto \prod_{i \in J'} \mathbf{P}[E_i].$$

Nach Induktionsvoraussetzung (2.7) gilt $\mu(E_j) = \nu(E_j)$ für jedes $E_j \in \mathcal{E}_j \cup \{\emptyset, \Omega\}$. Da $\mathcal{E}_j \cup \{\emptyset\}$ schnittstabil ist, gilt nach Lemma 1.42 auch $\mu(E_j) = \nu(E_j)$ für jedes $E_j \in \sigma(\mathcal{E}_j)$, das heißt, es gilt (2.7) mit $J \cup \{j\}$ statt J .

(iv) Dies ist trivial, weil (2.6) nur für $J \subset I$ mit

$$\#(J \cap I_k) \leq 1 \quad \text{für jedes } k \in K,$$

nachgewiesen werden muss. □

2.2 Unabhängigkeit von Zufallsvariablen

Nachdem wir Unabhängigkeit von Ereignissen behandelt haben, wollen wir auch Unabhängigkeit von Zufallsvariablen betrachten. Auch hier läuft die Definition auf eine Produktformel hinaus. Formal können wir jedoch die Unabhängigkeit der von Zufallsvariablen erzeugten σ -Algebren als Definition heranziehen. Wir können dann Verteilungen von Summen unabhängiger Zufallsvariablen mittels Faltung ausrechnen. Da wir an dieser Stelle noch keinen allgemeinen Integralbegriff zur Verfügung haben, bringen wir die Faltung zunächst nur für Zufallsvariablen mit ganzzahligen Werten.

Sei I eine beliebige Indexmenge, und für jedes $i \in I$ sei $(\Omega_i, \mathcal{A}_i)$ ein Messraum sowie $X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega_i, \mathcal{A}_i)$ eine Zufallsvariable mit erzeugter σ -Algebra $\sigma(X_i) = X_i^{-1}(\mathcal{A}_i)$.

Definition 2.14 (Unabhängigkeit von Zufallsvariablen). Die Familie $(X_i)_{i \in I}$ von Zufallsvariablen heißt **unabhängig**, falls die Familie $(\sigma(X_i))_{i \in I}$ von σ -Algebren unabhängig ist.

Wir schreiben, dass $(X_i)_{i \in I}$ „u.i.v.“ ist, für „unabhängig und identisch verteilt“ (englisch: „i.i.d.“ für „independent and identically distributed“), falls $(X_i)_{i \in I}$ unabhängig ist und $\mathbf{P}_{X_i} = \mathbf{P}_{X_j}$ für alle $i, j \in I$ gilt.

Bemerkung 2.15. (i) Offenbar ist die Familie $(X_i)_{i \in I}$ genau dann unabhängig, wenn für jede endliche Menge $J \subset I$ und für jede Wahl $A_j \in \mathcal{A}_j$, $j \in J$ gilt, dass

$$\mathbf{P}\left[\bigcap_{j \in J} \{X_j \in A_j\}\right] = \prod_{j \in J} \mathbf{P}[X_j \in A_j].$$

Im folgenden Satz werden wir sehen, dass es reicht, diese Bedingung für A_j aus einem \cap -stabilen Erzeuger von \mathcal{A}_j nachzuprüfen.

(ii) Ist $(\tilde{\mathcal{A}}_i)_{i \in I}$ eine unabhängige Familie von σ -Algebren und ist jedes X_i messbar bezüglich $\tilde{\mathcal{A}}_i - \mathcal{A}_i$, so ist $(X_i)_{i \in I}$ unabhängig. Dies ist klar, weil $\sigma(X_i) \subset \tilde{\mathcal{A}}_i$, also die Bedingung an die Unabhängigkeit von $(X_i)_{i \in I}$ schwächer ist als die Bedingung an die Unabhängigkeit von $(\tilde{\mathcal{A}}_i)_{i \in I}$.

(iii) Für jedes $i \in I$ sei $(\Omega'_i, \mathcal{A}'_i)$ ein weiterer Messraum sowie $f_i : (\Omega_i, \mathcal{A}_i) \rightarrow (\Omega'_i, \mathcal{A}'_i)$ eine messbare Abbildung. Ist $(X_i)_{i \in I}$ unabhängig, so ist $(f_i \circ X_i)_{i \in I}$ unabhängig. Diese Aussage ist ein Spezialfall von (ii), weil $f_i \circ X_i$ messbar ist bezüglich $\sigma(X_i) - \mathcal{A}'_i$ (siehe Satz 1.80). \diamond

Satz 2.16 (Unabhängigkeit von Erzeugern). Für jedes $i \in I$ sei $\mathcal{E}_i \subset \mathcal{A}_i$ ein schnittstabiler Erzeuger von \mathcal{A}_i . Ist $(X_i^{-1}(\mathcal{E}_i))_{i \in I}$ unabhängig, so ist $(X_i)_{i \in I}$ unabhängig.

Beweis. Nach Satz 1.81 ist $X_i^{-1}(\mathcal{E}_i)$ ein schnittstabiler Erzeuger der σ -Algebra $X_i^{-1}(\mathcal{A}_i) = \sigma(X_i)$. Mit Satz 2.13 folgt die Aussage. \square

Beispiel 2.17. Sei E eine höchstens abzählbare Menge, und seien $(X_i)_{i \in I}$ Zufallsvariablen mit Werten in $(E, 2^E)$. In diesem Falle ist $(X_i)_{i \in I}$ genau dann unabhängig, wenn für jede endliche Teilmenge $J \subset I$ und jede Wahl von $x_j \in E$, $j \in J$, gilt, dass

$$\mathbf{P}[X_j = x_j \text{ für jedes } j \in J] = \prod_{j \in J} \mathbf{P}[X_j = x_j].$$

Dies ist klar, weil $\{\{x\} : x \in E\} \cup \{\emptyset\}$ ein schnittstabiler Erzeuger von 2^E ist, also $\{X_i^{-1}(\{x_i\}) : x_i \in E\} \cup \{\emptyset\}$ ein schnittstabiler Erzeuger von $\sigma(X_i)$ ist (Satz 1.81). \diamond

Beispiel 2.18. Sei E eine endliche Menge und $p = (p_e)_{e \in E}$ ein Wahrscheinlichkeitsvektor. Wir wollen das zu E und p gehörige Zufallsexperiment unendlich oft unabhängig wiederholen (siehe Beispiel 1.40 und Satz 1.64). Sei $\Omega = E^{\mathbb{N}}$ der unendliche Produktraum und \mathcal{A} die von den endlichen Zylindermengen (siehe (1.8)) erzeugte σ -Algebra sowie $\mathbf{P} = (\sum_{e \in E} p_e \delta_e)^{\otimes \mathbb{N}}$ das Bernoulli-Maß. Ferner sei für jedes $n \in \mathbb{N}$

$$X_n : \Omega \rightarrow E, \quad (\omega_m)_{m \in \mathbb{N}} \mapsto \omega_n,$$

die Projektion auf die n -te Koordinate. Mit anderen Worten: Zu jedem Elementarereignis $\omega \in \Omega$ liefert $X_n(\omega)$ das Ergebnis des n -ten Experiments. Dann gilt nach (2.4) (in Beispiel 2.4) für $n \in \mathbb{N}$ und $x \in E^n$

$$\begin{aligned} \mathbf{P}[X_j = x_j \text{ für jedes } j = 1, \dots, n] &= \mathbf{P}[[x_1, \dots, x_n]] = \mathbf{P}\left[\bigcap_{j=1}^n X_j^{-1}(\{x_j\})\right] \\ &= \prod_{j=1}^n \mathbf{P}[X_j^{-1}(\{x_j\})] = \prod_{j=1}^n \mathbf{P}[X_j = x_j], \end{aligned}$$

sowie $\mathbf{P}[X_j = x_j] = p_{x_j}$. Nach Satz 2.13(i) sind also (X_1, \dots, X_n) unabhängig und nach Satz 2.13(ii) auch $(X_n)_{n \in \mathbb{N}}$. \diamond

Speziell haben wir den folgenden Satz gezeigt.

Satz 2.19. Sei E eine endliche Menge und $(p_e)_{e \in E}$ ein Wahrscheinlichkeitsvektor auf E . Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$ und eine unabhängige Familie $(X_n)_{n \in \mathbb{N}}$ von E -wertigen Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbf{P})$ mit $\mathbf{P}[X_n = e] = p_e$ für jedes $e \in E$.

Wir werden später sehen, dass wir auf die Endlichkeit von E verzichten können und auch unterschiedliche Verteilungen zulassen können. Für den Moment gibt uns dieser Satz aber genügend Beispiele für abzählbare Familien von unabhängigen Zufallsvariablen an die Hand.

Wir wollen nun einfache Kriterien zur Prüfung der Unabhängigkeit von Zufallsvariablen herleiten, die sich mit Hilfe von Verteilungsfunktionen beziehungsweise Dichten ausdrücken lassen.

Definition 2.20. Für jedes $i \in I$ sei X_i eine reelle Zufallsvariable. Für jede endliche Teilmenge $J \subset I$ sei

$$F_J := F_{(X_j)_{j \in J}} : \mathbb{R}^J \rightarrow [0, 1],$$

$$x \mapsto \mathbf{P}[X_j \leq x_j \text{ für jedes } j \in J] = \mathbf{P}\left[\bigcap_{j \in J} X_j^{-1}((-\infty, x_j])\right].$$

Dann heißt F_J die **gemeinsame Verteilungsfunktion** von $(X_j)_{j \in J}$. Das W-Maß $\mathbf{P}_{(X_j)_{j \in J}}$ auf \mathbb{R}^J heißt **gemeinsame Verteilung** von $(X_j)_{j \in J}$.

Satz 2.21. Eine Familie $(X_i)_{i \in I}$ reeller Zufallsvariablen ist genau dann unabhängig, wenn für jedes endliche $J \subset I$ und jedes $x = (x_j)_{j \in J} \in \mathbb{R}^J$ gilt, dass

$$F_J(x) = \prod_{j \in J} F_{\{j\}}(x_j). \quad (2.8)$$

Beweis. Das Mengensystem $\{(-\infty, b], b \in \mathbb{R}\}$ ist ein schnittstabilier Erzeuger der Borel'schen σ -Algebra $\mathcal{B}(\mathbb{R})$ (siehe Satz 1.23). Die Gleichung (2.8) besagt nun aber, dass für jede Wahl von reellen Zahlen $(x_i)_{i \in I}$ die Ereignisse $(X_i^{-1}((-\infty, x_i]))_{i \in I}$ unabhängig sind. Nach Satz 2.16 folgt daher die Aussage dieses Satzes. \square

Korollar 2.22. Zusätzlich zur Situation von Satz 2.21 nehmen wir an, dass jedes F_J eine stetige **Dichte** $f_J = f_{(X_j)_{j \in J}}$ hat (die so genannte **gemeinsame Dichte** der $(X_j)_{j \in J}$), das heißt, es gibt eine stetige Abbildung $f_J : \mathbb{R}^J \rightarrow [0, \infty)$ mit

$$F_J(x) = \int_{-\infty}^{x_{j_1}} dt_1 \cdots \int_{-\infty}^{x_{j_n}} dt_n f_J(t_1, \dots, t_n) \quad \text{für jedes } x \in \mathbb{R}^J,$$

(wobei $J = \{j_1, \dots, j_n\}$). Dann ist die Familie $(X_i)_{i \in I}$ genau dann unabhängig, wenn für jedes endliche $J \subset I$ gilt

$$f_J(x) = \prod_{j \in J} f_j(x_j) \quad \text{für jedes } x \in \mathbb{R}^J. \quad (2.9)$$

Korollar 2.23. Seien $n \in \mathbb{N}$ und μ_1, \dots, μ_n W-Maße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann existiert ein W-Raum $(\Omega, \mathcal{A}, \mathbf{P})$ und eine unabhängige Familie von Zufallsvariablen $(X_i)_{i=1, \dots, n}$ auf $(\Omega, \mathcal{A}, \mathbf{P})$ mit $\mathbf{P}_{X_i} = \mu_i$ für jedes $i = 1, \dots, n$.

Beweis. Sei $\Omega = \mathbb{R}^n$ und $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$ sowie $\mathbf{P} = \bigotimes_{i=1}^n \mu_i$ das Produktmaß der μ_i (siehe Satz 1.61). Ferner sei $X_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $(x_1, \dots, x_n) \mapsto x_i$ die Projektion auf die i -te Koordinate für jedes $i = 1, \dots, n$. Dann ist für jedes $i = 1, \dots, n$

$$\begin{aligned} F_{\{i\}}(x) &= \mathbf{P}[X_i \leq x] = \mathbf{P}[\mathbb{R}^{i-1} \times (-\infty, x] \times \mathbb{R}^{n-i}] \\ &= \mu_i((-\infty, x]) \cdot \prod_{j \neq i} \mu_j(\mathbb{R}) = \mu_i((-\infty, x]). \end{aligned}$$

Also gilt tatsächlich $\mathbf{P}_{X_i} = \mu_i$. Ferner ist für $x_1, \dots, x_n \in \mathbb{R}$

$$F_{\{1, \dots, n\}}((x_1, \dots, x_n)) = \mathbf{P}\left[\bigtimes_{i=1}^n (-\infty, x_i]\right] = \prod_{i=1}^n \mu_i((-\infty, x_i]) = \prod_{i=1}^n F_{\{i\}}(x_i).$$

Nach Satz 2.21 (und Satz 2.13(i)) folgt die Unabhängigkeit von $(X_i)_{i=1, \dots, n}$. \square

Beispiel 2.24. Seien X_1, \dots, X_n unabhängige, exponentialverteilte Zufallsvariablen mit Parametern $\theta_1, \dots, \theta_n \in (0, \infty)$. Dann ist $F_{\{i\}}(x) = \int_0^x \theta_i \exp(-\theta_i t) dt = 1 - \exp(-\theta_i x)$ für $x \geq 0$ und daher

$$F_{\{1, \dots, n\}}((x_1, \dots, x_n)) = \prod_{i=1}^n (1 - e^{-\theta_i x_i}).$$

Betrachte nun die Zufallsvariable $Y = \max(X_1, \dots, X_n)$. Dann ist

$$\begin{aligned} F_Y(x) &= \mathbf{P}[X_i \leq x \text{ für jedes } i = 1, \dots, n] \\ &= F_{\{1, \dots, n\}}((x, \dots, x)) = \prod_{i=1}^n (1 - e^{-\theta_i x}). \end{aligned}$$

Für die Zufallsvariable $Z := \min(X_1, \dots, X_n)$ hat die Verteilungsfunktion eine geschlossene Form

$$\begin{aligned} F_Z(x) &= 1 - \mathbf{P}[Z > x] \\ &= 1 - \mathbf{P}[X_i > x \text{ für jedes } i = 1, \dots, n] \\ &= 1 - \prod_{i=1}^n e^{-\theta_i x} = 1 - \exp(-(\theta_1 + \dots + \theta_n) x). \end{aligned}$$

Mit anderen Worten: Z ist exponentialverteilt mit Parameter $\theta_1 + \dots + \theta_n$. ◇

Beispiel 2.25. Seien $\mu_i \in \mathbb{R}$ und $\sigma_i^2 > 0$, $i \in I$, sowie $(X_i)_{i \in I}$ reell mit gemeinsamen Dichtefunktionen (für endliches $J \subset I$)

$$f_J(x) = \prod_{j \in J} (2\pi\sigma_j^2)^{-\frac{1}{2}} \exp\left(-\sum_{j \in J} \frac{(x_j - \mu_j)^2}{2\sigma_j^2}\right) \quad \text{für } x \in \mathbb{R}^J.$$

Dann sind die $(X_i)_{i \in I}$ unabhängig, und X_i ist normalverteilt mit Parametern (μ_i, σ_i^2) .

Für jedes endliche $I = \{i_1, \dots, i_n\}$ (mit paarweise unterschiedlichen i_1, \dots, i_n) ist der Vektor $Y = (X_{i_1}, \dots, X_{i_n})$ n -dimensional normalverteilt mit $\mu = \mu^I := (\mu_{i_1}, \dots, \mu_{i_n})$ und $\Sigma = \Sigma^I$ die Diagonalmatrix mit Einträgen $\sigma_{i_1}^2, \dots, \sigma_{i_n}^2$ (vergleiche Beispiel 1.105(ix)). ◇

Satz 2.26. Sei K eine beliebige Menge und I_k , $k \in K$, beliebige paarweise disjunkte Indexmengen sowie $I = \bigcup_{k \in K} I_k$.

Ist die Familie $(X_i)_{i \in I}$ unabhängig, dann sind auch die σ -Algebren $(\sigma(X_j, j \in I_k))_{k \in K}$ unabhängig.

Beweis. Sei für $k \in K$

$$\mathcal{Z}_k = \left\{ \bigcap_{j \in I_k} A_j : A_j \in \sigma(X_j), \#\{j \in I_k : A_j \neq \Omega\} < \infty \right\}$$

der Semiring der endlichdimensionalen rechteckigen Zylinder. Offenbar ist \mathcal{Z}_k schnittstabil und $\sigma(\mathcal{Z}_k) = \sigma(X_j, j \in I_k)$. Also reicht es nach Satz 2.13(iii) zu zeigen, dass $(\mathcal{Z}_k)_{k \in K}$ unabhängig ist. Nach Satz 2.13(ii) können wir sogar annehmen, dass K endlich ist.

Für $k \in K$ seien nun $B_k \in \mathcal{Z}_k$ und $J_k \subset I_k$ endlich mit $B_k = \bigcap_{j \in J_k} A_j$ für gewisse $A_j \in \sigma(X_j)$. Setze $J = \bigcup_{k \in K} J_k$. Dann ist

$$\mathbf{P}\left[\bigcap_{k \in K} B_k\right] = \mathbf{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbf{P}[A_j] = \prod_{k \in K} \prod_{j \in J_k} \mathbf{P}[A_j] = \prod_{k \in K} \mathbf{P}[B_k]. \quad \square$$

Beispiel 2.27. Sind $(X_n)_{n \in \mathbb{N}}$ unabhängige, reelle Zufallsvariablen, dann sind auch $(Y_n)_{n \in \mathbb{N}} = (X_{2n} - X_{2n-1})_{n \in \mathbb{N}}$ unabhängig. In der Tat ist für jedes $n \in \mathbb{N}$ die Zufallsvariable Y_n schon messbar bezüglich $\sigma(X_{2n}, X_{2n-1})$ nach Satz 1.91, und $(\sigma(X_{2n}, X_{2n-1}))_{n \in \mathbb{N}}$ ist unabhängig nach Satz 2.26. \diamond

Beispiel 2.28. Seien $(X_{m,n})_{(m,n) \in \mathbb{N}^2}$ unabhängige Bernoulli-Zufallsvariablen mit Parameter $p \in (0, 1)$. Sei

$$Y_m := \inf \{n \in \mathbb{N} : X_{m,n} = 1\} - 1$$

die Wartezeit auf den ersten „Erfolg“ in der m -ten Zeile der Matrix $(X_{m,n})_{m,n}$. Dann sind $(Y_m)_{m \in \mathbb{N}}$ unabhängige, geometrisch verteilte Zufallsvariablen mit Parameter p (siehe Beispiel 1.105(iii)). Denn:

$$\{Y_m \leq k\} = \bigcup_{l=1}^{k+1} \{X_{m,l} = 1\} \in \sigma(X_{m,l}, l = 1, \dots, k+1) \subset \sigma(X_{m,l}, l \in \mathbb{N}).$$

Also ist Y_m messbar bezüglich $\sigma(X_{m,l}, l \in \mathbb{N})$, und damit ist $(Y_m)_{m \in \mathbb{N}}$ unabhängig. Ferner ist

$$\mathbf{P}[Y_m > k] = \mathbf{P}[X_{m,l} = 0, l = 1, \dots, k+1] = \prod_{l=1}^{k+1} \mathbf{P}[X_{m,l} = 0] = (1-p)^{k+1}.$$

Es folgt $\mathbf{P}[Y_m = k] = \mathbf{P}[Y_m > k-1] - \mathbf{P}[Y_m > k] = p(1-p)^k$. \diamond

Definition 2.29 (Faltung). Seien μ und ν W-Maße auf $(\mathbb{Z}, 2^\mathbb{Z})$. Wir definieren die Faltung $\mu * \nu$ als das W-Maß auf $(\mathbb{Z}, 2^\mathbb{Z})$ mit

$$(\mu * \nu)(\{n\}) = \sum_{m=-\infty}^{\infty} \mu(\{m\}) \nu(\{n-m\}).$$

Wir definieren die n -te Faltungspotenz rekursiv durch $\mu^{*1} = \mu$ und

$$\mu^{*(n+1)} = \mu^{*n} * \mu.$$

Bemerkung 2.30. Es gilt $\mu * \nu = \nu * \mu$. ◊

Satz 2.31. Sind X und Y unabhängige \mathbb{Z} -wertige Zufallsvariablen, so gilt $\mathbf{P}_{X+Y} = \mathbf{P}_X * \mathbf{P}_Y$.

Beweis. Für jedes $n \in \mathbb{Z}$ gilt

$$\begin{aligned} \mathbf{P}_{X+Y}(\{n\}) &= \mathbf{P}[X + Y = n] \\ &= \mathbf{P}\left[\biguplus_{m \in \mathbb{Z}} (\{X = m\} \cap \{Y = n - m\})\right] \\ &= \sum_{m \in \mathbb{Z}} \mathbf{P}[\{X = m\} \cap \{Y = n - m\}] \\ &= \sum_{m \in \mathbb{Z}} \mathbf{P}_X[\{m\}] \mathbf{P}_Y[\{n - m\}] = (\mathbf{P}_X * \mathbf{P}_Y)[\{n\}]. \quad \square \end{aligned}$$

Auf Grund dieses Satzes liegt es nahe, die Faltung von zwei Wahrscheinlichkeitsmaßen auf \mathbb{R}^n (oder allgemeiner: auf abelschen Gruppen) als die Verteilung der Summe zweier unabhängiger Zufallsvariablen mit den entsprechenden Verteilungen zu definieren. Wir werden später eine andere Definition kennen lernen, die natürlich zu dieser äquivalent ist, jedoch auf den Integralbegriff zurückgreift, der hier noch nicht verfügbar ist (siehe Definition 14.17).

Definition 2.32 (Faltung von Maßen). Seien μ und ν W-Maße auf \mathbb{R}^n , und seien X und Y unabhängige Zufallsvariablen mit $\mathbf{P}_X = \mu$ und $\mathbf{P}_Y = \nu$. Dann definieren wir die **Faltung** von μ und ν durch $\mu * \nu = \mathbf{P}_{X+Y}$.

Iterativ definieren wir die Faltungspotenzen μ^{*k} für $k \in \mathbb{N}$ sowie $\mu^{*0} = \delta_0$.

Beispiel 2.33. Seien X und Y unabhängig und Poisson-verteilt mit Parametern $\mu, \lambda \geq 0$. Dann gilt

$$\begin{aligned} \mathbf{P}[X + Y = n] &= e^{-\mu} e^{-\lambda} \sum_{m=0}^n \frac{\mu^m}{m!} \frac{\lambda^{n-m}}{(n-m)!} \\ &= e^{-(\mu+\lambda)} \frac{1}{n!} \sum_{m=0}^n \binom{n}{m} \mu^m \lambda^{n-m} = e^{-(\mu+\lambda)} \frac{(\mu + \lambda)^n}{n!}. \end{aligned}$$

Also ist $\text{Poi}_\mu * \text{Poi}_\lambda = \text{Poi}_{\mu+\lambda}$. ◊

Übung 2.2.1. Seien X und Y unabhängige Zufallsvariablen mit $X \sim \exp_\theta$ und $Y \sim \exp_\rho$ für gewisse $\theta, \rho > 0$. Man zeige:

$$\mathbf{P}[X < Y] = \frac{\theta}{\theta + \rho}. \quad \clubsuit$$

Übung 2.2.2 (Box-Muller Methode). Seien U und V unabhängige, uniform auf $[0, 1]$ verteilte Zufallsvariablen. Setze

$$X := \sqrt{-2 \log(U)} \cos(2\pi V) \quad \text{und} \quad Y := \sqrt{-2 \log(U)} \sin(2\pi V).$$

Man zeige: X und Y sind unabhängig und $\mathcal{N}_{0,1}$ -verteilt.

Hinweis: Man berechne zunächst die Verteilung von $\sqrt{-2 \log(U)}$ und benutze die Transformationsformel für Dichten (Satz 1.101) sowie Polarkoordinatentransformation. \clubsuit

Übung 2.2.3 (Multinomialverteilung). Sei $m \in \mathbb{N}$ und sei $p = (p_1, \dots, p_m)$ ein Wahrscheinlichkeitsvektor auf $\{1, \dots, m\}$. Seien X_1, \dots, X_n unabhängige Zufallsvariablen mit Werten in $1, \dots, m$ und Verteilung p . Wir definieren eine \mathbb{N}_0^m -wertige Zufallsvariable $Y = (Y_1, \dots, Y_m)$ durch

$$Y_i := \#\{k = 1, \dots, n : X_k = i\} \quad \text{für } i = 1, \dots, m.$$

Man zeige:

$$\mathbf{P}[Y = k] = \text{Mul}_{n,p}(\{k\}) := \binom{n}{k} p^k \quad (2.10)$$

für $k = (k_1, \dots, k_m) \in \mathbb{N}_0^m$ mit $k_1 + \dots + k_m = n$. Dabei ist

$$\binom{n}{k} = \binom{n}{k_1, \dots, k_m} = \frac{n!}{k_1! \cdots k_m!}$$

der **Multinomialkoeffizient** und $p^k = p_1^{k_1} \cdots p_m^{k_m}$. Die Verteilung $\text{Mul}_{n,p}$ auf \mathbb{N}_0^m heißt Multinomialverteilung mit Parametern n und p . \clubsuit

2.3 Kolmogorov'sches 0-1 Gesetz

Mit dem Lemma von Borel-Cantelli haben wir bereits ein 0-1 Gesetz für unabhängige Ereignisse kennen gelernt. Wir kommen jetzt zu einem weiteren 0-1 Gesetz für unabhängige Ereignisse, beziehungsweise σ -Algebren. Um dies zu formulieren, müssen wir zunächst den Begriff der terminalen σ -Algebra einführen.

Definition 2.34 (Terminale σ -Algebra). Sei I eine abzählbar unendliche Indexmenge und $(\mathcal{A}_i)_{i \in I}$ eine Familie von σ -Algebren. Dann heißt

$$\mathcal{T}((\mathcal{A}_i)_{i \in I}) := \bigcap_{\substack{J \subset I \\ \#J < \infty}} \sigma \left(\bigcup_{j \in I \setminus J} \mathcal{A}_j \right)$$

die **terminale σ -Algebra** (oder das tail σ -field) von $(\mathcal{A}_i)_{i \in I}$. Sind $(A_i)_{i \in I}$ Ereignisse, so setzen wir

$$\mathcal{T}((A_i)_{i \in I}) := \mathcal{T}((\{\emptyset, A_i, A_i^c, \Omega\})_{i \in I}).$$

Ist $(X_i)_{i \in I}$ eine Familie von Zufallsvariablen, so setzen wir $\mathcal{T}((X_i)_{i \in I}) := \mathcal{T}((\sigma(X_i))_{i \in I})$.

In der terminalen σ -Algebra sind alle diejenigen Ereignisse A enthalten, deren Eintreten von jeder festen endlichen Teilstammfamilie der X_i nicht abhängt. Anders gesagt: für jede endliche Teilstammfamilie der X_i können wir die Werte dieser X_i willkürlich ändern, ohne dass sich dadurch am Eintreten von A etwas änderte.

Satz 2.35. Seien J_1, J_2, \dots endliche Mengen mit $J_n \uparrow I$. Dann ist $\mathcal{T}((\mathcal{A}_i)_{i \in I}) = \bigcap_{n=1}^{\infty} \sigma(\bigcup_{m \in I \setminus J_n} \mathcal{A}_m)$.

Ist speziell $I = \mathbb{N}$, so ist $\mathcal{T}((\mathcal{A}_n)_{n \in \mathbb{N}}) = \bigcap_{n=1}^{\infty} \sigma(\bigcup_{m=n}^{\infty} \mathcal{A}_m)$.

Von dem Fall $I = \mathbb{N}$, wo I als Menge von Zeitpunkten interpretiert wird, röhrt der Begriff *terminale σ -Algebra* her. Wie aus dem Satz hervorgeht, liegen in \mathcal{T} nur die Ereignisse, die nicht von den ersten endlich vielen Zeitpunkten abhängen, über deren Eintreten man also erst *terminal* eine Aussage machen kann.

Beweis. „ \subset “ Dies ist klar.

„ \supset “ Sei $J_n \uparrow I$ mit endlichen Mengen $J_n \subset I$, $n \in \mathbb{N}$, und sei $J \subset I$ endlich. Dann existiert ein $N \in \mathbb{N}$ mit $J \subset J_N$, und es ist

$$\begin{aligned} \bigcap_{n=1}^{\infty} \sigma \left(\bigcup_{m \in I \setminus J_n} \mathcal{A}_m \right) &\subset \bigcap_{n=1}^N \sigma \left(\bigcup_{m \in I \setminus J_n} \mathcal{A}_m \right) \\ &= \sigma \left(\bigcup_{m \in I \setminus J_N} \mathcal{A}_m \right) \subset \sigma \left(\bigcup_{m \in I \setminus J} \mathcal{A}_m \right). \end{aligned}$$

Die linke Seite hängt nicht von J ab, also können wir den Schnitt über alle endlichen J bilden und erhalten

$$\bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{m \in I \setminus J_n} \mathcal{A}_m\right) \subset \mathcal{T}((\mathcal{A}_i)_{i \in I}). \quad \square$$

Es ist vielleicht nicht so ohne weiteres klar, dass es überhaupt noch interessante Ereignisse in der terminalen σ -Algebra gibt. Vielleicht ist nicht einmal *a priori* klar, dass nicht etwa $\mathcal{T} = \{\emptyset, \Omega\}$ gilt. Daher geben wir jetzt erst einmal einfache Beispiele für terminale Ereignisse beziehungsweise terminal messbare Zufallsvariablen an. In Abschnitt 2.4 werden wir ein weiteres Beispiel kennen lernen.

Beispiel 2.36. (i) Seien A_1, A_2, \dots Ereignisse. Dann sind die Ereignisse $A_* := \liminf_{n \rightarrow \infty} A_n$ und $A^* := \limsup_{n \rightarrow \infty} A_n$ in $\mathcal{T}((A_n)_{n \in \mathbb{N}})$. Setzen wir nämlich $B_n := \bigcap_{m=n}^{\infty} A_m$ für $n \in \mathbb{N}$, dann gilt $B_n \uparrow A_*$ und $B_n \in \sigma((A_m)_{m \geq N})$ für jedes $n \geq N$. Also ist $A_* \in \sigma((A_m)_{m \geq N})$ für jedes $N \in \mathbb{N}$ und damit $A_* \in \mathcal{T}((A_n)_{n \in \mathbb{N}})$. Für A^* geht dies analog.

(ii) Ist $(X_n)_{n \in \mathbb{N}}$ eine Familie $\overline{\mathbb{R}}$ -wertiger Zufallsvariablen, so sind auch die Abbildungen $X_* := \liminf_{n \rightarrow \infty} X_n$ und $X^* := \limsup_{n \rightarrow \infty} X_n$ messbar bezüglich $\mathcal{T}((X_n)_{n \in \mathbb{N}})$. In der Tat: Setzen wir $Y_n := \sup_{m \geq n} X_m$, so ist für jedes $N \in \mathbb{N}$ die Zufallsvariable $X^* = \inf_{n \geq 1} Y_n = \inf_{n \geq N} Y_n$ messbar bezüglich $\mathcal{T}_N := \sigma(X_n, n \geq N)$, also auch bezüglich $\mathcal{T}((X_n)_{n \in \mathbb{N}}) = \bigcap_{n=1}^{\infty} \mathcal{T}_n$.

Für X_* geht dies analog.

(iii) Seien $(X_n)_{n \in \mathbb{N}}$ reellwertige Zufallsvariablen. Dann sind die **Cesàro-Limiten**

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i \quad \text{und} \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i$$

messbar bezüglich $\mathcal{T}((X_n)_{n \in \mathbb{N}})$. Um dies zu zeigen, wählen wir ein $N \in \mathbb{N}$ und beachten, dass

$$X_* := \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=N}^n X_i$$

$\sigma((X_n)_{n \geq N})$ -messbar ist. Da dies für jedes N gilt, ist X_* auch $\mathcal{T}((X_n)_{n \in \mathbb{N}})$ -messbar. Für den Limes superior geht das analog. \diamond

Satz 2.37 (Kolmogorov'sches 0-1 Gesetz). Sei I eine abzählbar unendliche Indexmenge, und sei $(\mathcal{A}_i)_{i \in I}$ eine unabhängige Familie von σ -Algebren. Dann ist die terminale σ -Algebra \mathbf{P} -trivial, das heißt, es gilt

$$\mathbf{P}[A] \in \{0, 1\} \quad \text{für jedes } A \in \mathcal{T}((\mathcal{A}_i)_{i \in I}).$$

Beweis. Es reicht, den Fall $I = \mathbb{N}$ zu betrachten. Für $n \in \mathbb{N}$ sei

$$\mathcal{F}_n := \left\{ \bigcap_{k=1}^n A_k : A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n \right\}.$$

Dann ist $\mathcal{F} := \bigcup_{n=1}^{\infty} \mathcal{F}_n$ ein Semiring und $\sigma(\mathcal{F}) = \sigma(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n)$. In der Tat ist für jedes $n \in \mathbb{N}$ und $A_n \in \mathcal{A}_n$ auch $A_n \in \mathcal{F}$, also $\sigma(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n) \subset \sigma(\mathcal{F})$. Andererseits ist $\mathcal{F}_m \subset \sigma(\bigcup_{n=1}^m \mathcal{A}_n) \subset \sigma(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n)$ für jedes $m \in \mathbb{N}$, also $\mathcal{F} \subset \sigma(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n)$.

Sei $A \in \mathcal{T}((\mathcal{A}_n)_{n \in \mathbb{N}})$, und sei $\varepsilon > 0$. Nach dem Approximationssatz für Maße (Satz 1.65) existiert ein $N \in \mathbb{N}$ und paarweise disjunkte Mengen $F_1, \dots, F_N \in \mathcal{F}$ mit $\mathbf{P}[A \Delta (F_1 \cup \dots \cup F_N)] < \varepsilon$. Nach Definition von \mathcal{F} existiert ein $n \in \mathbb{N}$ mit $F_1, \dots, F_N \in \mathcal{F}_n$. Also gilt $F := F_1 \cup \dots \cup F_N \in \sigma(\mathcal{A}_1 \cup \dots \cup \mathcal{A}_n)$. Offenbar ist $A \in \sigma(\bigcup_{m=n+1}^{\infty} \mathcal{A}_m)$, also ist A unabhängig von F . Daher ist

$$\varepsilon > \mathbf{P}[A \setminus F] = \mathbf{P}[A \cap (\Omega \setminus F)] = \mathbf{P}[A](1 - \mathbf{P}[F]) \geq \mathbf{P}[A](1 - \mathbf{P}[A] - \varepsilon).$$

Da $\varepsilon > 0$ beliebig war, folgt $0 = \mathbf{P}[A](1 - \mathbf{P}[A])$. \square

Korollar 2.38. Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge unabhängiger Ereignisse. Dann gilt

$$\mathbf{P}\left[\limsup_{n \rightarrow \infty} A_n\right] \in \{0, 1\} \quad \text{und} \quad \mathbf{P}\left[\liminf_{n \rightarrow \infty} A_n\right] \in \{0, 1\}.$$

Beweis. Dies ist im Grunde eine Schlussfolgerung aus dem Lemma von Borel-Cantelli. Allerdings folgt es auch direkt aus dem Kolmogorov'schen 0-1 Gesetz, da Limes superior und Limes inferior in der terminalen σ -Algebra liegen. \square

Korollar 2.39. Sei $(X_n)_{n \in \mathbb{N}}$ eine unabhängige Familie von $\overline{\mathbb{R}}$ -wertigen Zufallsvariablen. Dann sind $X_* := \liminf_{n \rightarrow \infty} X_n$ und $X^* := \limsup_{n \rightarrow \infty} X_n$ fast sicher konstant, das heißt, es gibt $x_*, x^* \in \overline{\mathbb{R}}$ mit $\mathbf{P}[X_* = x_*] = 1$ und $\mathbf{P}[X^* = x^*] = 1$.

Falls alle X_i sogar reellwertig sind, so sind auch die Cesàro-Limiten

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i \quad \text{und} \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i$$

fast sicher konstant.

Beweis. Sei $X_* := \liminf_{n \rightarrow \infty} X_n$. Für jedes $x \in \overline{\mathbb{R}}$ ist $\{X_* \leq x\} \in \mathcal{T}((X_n)_{n \in \mathbb{N}})$, also $\mathbf{P}[X_* \leq x] \in \{0, 1\}$. Setze

$$x_* := \inf\{x \in \mathbb{R} : \mathbf{P}[X_* \leq x] = 1\} \in \overline{\mathbb{R}}.$$

Ist $x_* = \infty$, so ist offenbar

$$\mathbf{P}[X_* < \infty] = \lim_{n \rightarrow \infty} \mathbf{P}[X_* \leq n] = 0.$$

Ist $x_* \in \mathbb{R}$, so ist

$$\mathbf{P}[X_* \leq x_*] = \lim_{n \rightarrow \infty} \mathbf{P}\left[X_* \leq x_* + \frac{1}{n}\right] = 1$$

und

$$\mathbf{P}[X_* < x_*] = \lim_{n \rightarrow \infty} \mathbf{P}\left[X_* \leq x_* - \frac{1}{n}\right] = 0.$$

Ist $x_* = -\infty$, so ist

$$\mathbf{P}[X_* > -\infty] = \lim_{n \rightarrow \infty} \mathbf{P}[X_* > -n] = 0.$$

Für den Limes superior sowie für die Cesàro-Limiten geht dies analog. \square

Übung 2.3.1. Man zeige: Ist $(X_n)_{n \in \mathbb{N}}$ eine unabhängige Familie von Zufallsvariablen mit $\mathbf{P}[X_n = -1] = \mathbf{P}[X_n = +1] = \frac{1}{2}$, und ist $S_n = X_1 + \dots + X_n$ für jedes $n \in \mathbb{N}$, so ist $\limsup_{n \rightarrow \infty} S_n = \infty$ fast sicher. ♣

2.4 Beispiel: Perkolation

Wir betrachten das d -dimensionale Gitter \mathbb{Z}^d , wobei jeder Punkt durch je eine Kante mit seinen $2d$ nächsten Nachbarpunkten verbunden ist. Sind $x, y \in \mathbb{Z}^d$ nächste Nachbarn, das heißt $\|x - y\|_2 = 1$, so schreiben wir $k = \langle x, y \rangle = \langle y, x \rangle$ für die Kante, die x und y verbindet. Formal ist die Kantenmenge eine Teilmenge der zweielementigen Teilmengen von \mathbb{Z}^d :

$$K = \{\{x, y\} : x, y \in \mathbb{Z}^d \text{ mit } \|x - y\|_2 = 1\}.$$

Etwas allgemeiner ist ein (ungerichteter) **Graph** G ein Paar $G = (V, K)$, wobei V eine Menge ist (die Menge der Knoten oder Punkte des Graphen) und $K \subset \{\{x, y\} : x, y \in V, x \neq y\}$ eine Teilmenge aller zweielementigen Teilmengen von V (die Menge der Kanten).

Da wir unter einer Kante intuitiv eine Verbindung zwischen zwei Punkten x und y verstehen (und nicht das ungeordnete Paar $\{x, y\}$), verwenden wir ein anderes Symbol als die Mengenklammern und schreiben $\langle x, y \rangle$ statt $\{x, y\}$.

Dieses Gitter ist für uns der Ausgangspunkt für ein stochastisches Modell eines porösen Mediums. Wir stellen uns die Kanten als Röhren vor, entlang derer Wasser fließen kann. Nun soll das Medium allerdings nicht völlig homogen wasser-durchlässig sein, sondern eine amorphe Struktur besitzen, etwa wie Bimsstein. Zu diesem Zweck wollen wir zufällig einen gewissen Anteil $1 - p$ (wobei $p \in [0, 1]$ ein Parameter ist) der Kanten zerstören, sodass das Wasser nur durch die verbliebenen Kanten fließen kann. Die Frage, die sich stellt, ist, bei welchen Werten von p die intakten Röhren unendlich große verbundene Systeme bilden und bei welchen Werten alle verbundenen Systeme nur endliche Größe haben.

Wir kommen jetzt zur formalen Beschreibung des Modells. Wir wählen einen Parameter $p \in [0, 1]$ und eine unabhängige Familie identisch verteilter Zufallsvariablen $(X_k^p)_{k \in K}$ mit $X_k^p \sim \text{Ber}_p$, also $\mathbf{P}[X_k^p = 1] = 1 - \mathbf{P}[X_k^p = 0] = p$ für jedes $k \in K$. Dann definieren wir

$$K^p := \{k \in K : X_k^p = 1\} \quad (2.11)$$

als die Menge der intakten (oder *offenen*) Kanten. Entsprechend nennen wir die Kanten $K \setminus K^p$ defekt (oder *geschlossen*). Auf diese Weise haben wir einen (zufälligen) Teilgraphen (\mathbb{Z}^d, K^p) von (\mathbb{Z}^d, K) hergestellt. Wir nennen (\mathbb{Z}^d, K^p) auch ein Perkolationsmodell (genauer: ein Modell für **Kantenperkolation**, im Gegensatz zu **Punktperkolation**, wo die einzelnen Punkte geschlossen oder offen sind). Ein (offener) Pfad (der Länge n) in diesem Teilgraphen ist eine Folge $\pi = (x_0, x_1, \dots, x_n)$ von Punkten in \mathbb{Z}^d mit $\langle x_{i-1}, x_i \rangle \in K^p$ für jedes $i = 1, \dots, n$. Wir sagen, dass zwei Punkte $x, y \in \mathbb{Z}^d$ durch einen offenen Pfad verbunden werden können, wenn es ein $n \in \mathbb{N}$ und einen offenen Pfad (x_0, x_1, \dots, x_n) mit $x_0 = x$ und $x_n = y$ gibt. In diesem Fall schreiben wir $x \longleftrightarrow_p y$. Offenbar ist „ \longleftrightarrow_p “ eine Äquivalenzrelation, jedoch eine zufällige, weil sie von den Werten der Zufallsvariablen $(X_k^p)_{k \in K}$ abhängt. Für $x \in \mathbb{Z}^d$ nennen wir

$$C^p(x) := \{y \in \mathbb{Z}^d : x \longleftrightarrow_p y\} \quad (2.12)$$

den (zufälligen) offenen Cluster von x , also die Zusammenhangskomponente von x in dem Graphen (\mathbb{Z}^d, K^p) .

Lemma 2.40. *Für je zwei Punkte $x, y \in \mathbb{Z}^d$ ist $\mathbb{1}_{\{x \longleftrightarrow_p y\}}$ eine Zufallsvariable. Insbesondere ist $\#C^p(x)$ eine Zufallsvariable für jedes $x \in \mathbb{Z}^d$.*

Beweis. Ohne Einschränkung können wir annehmen, dass $x = 0$ ist. Wir setzen $f_{y,n} = 1$, falls es einen offenen Pfad von 0 nach y der Länge höchstens n gibt, und $f_{y,n} = 0$ sonst. Offenbar ist $f_{y,n} \uparrow \mathbb{1}_{\{0 \longleftrightarrow_p y\}}$ für $n \rightarrow \infty$, also reicht es, die Messbarkeit von $f_{y,n}$ zu zeigen. Sei $B_n := \{-n, -n+1, \dots, n-1, n\}^d$ und

$$K_n := \{k \in K : k \cap B_n \neq \emptyset\}.$$

Dann ist $Y_n := (X_k^p : k \in K_n) : \Omega \rightarrow \{0, 1\}^{K_n}$ messbar (bezüglich $2^{(\{0,1\}^{K_n})}$) nach Satz 1.90. Nun ist aber $f_{y,n}$ eine Funktion von Y_n , sagen wir $f_{y,n} = g_{y,n} \circ Y_n$ für gewisses $g_{y,n} : \{0, 1\}^{K_n} \rightarrow \{0, 1\}$. Nach dem Verknüpfungssatz (Satz 1.80) ist daher $f_{y,n}$ messbar.

Schließlich ist auch $\#C^p(x) = \sum_{y \in \mathbb{Z}^d} \mathbb{1}_{\{x \longleftrightarrow_p y\}}$ messbar. \square

Definition 2.41. *Wir sagen, dass Perkolation eintritt, falls es (wenigstens) einen unendlich großen, offenen Cluster gibt und nennen*

$$\begin{aligned} \psi(p) &:= \mathbf{P}[es \text{ gibt einen unendlich großen, offenen Cluster}] \\ &= \mathbf{P}\left[\bigcup_{x \in \mathbb{Z}^d} \{\#C^p(x) = \infty\}\right] \end{aligned}$$

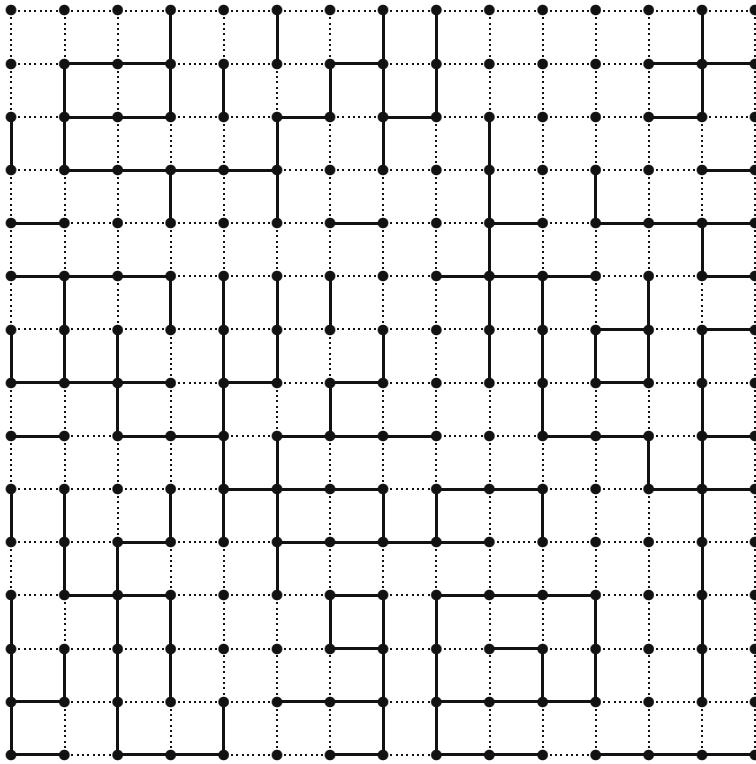


Abb. 2.1. Perkolation auf einem 15×15 Gitter, $p = 0.42$.

die Perkolationswahrscheinlichkeit. Wir definieren weiterhin die Funktion

$$\theta(p) := \mathbf{P}[\#C^p(0) = \infty]$$

als die Wahrscheinlichkeit, dass der Ursprung in einem unendlich großen, offenen Cluster liegt.

Auf Grund der Translationsinvarianz des Gitters ist

$$\theta(p) = \mathbf{P}[\#C^p(y) = \infty] \quad \text{für jedes } y \in \mathbb{Z}^d. \quad (2.13)$$

Die Grundfrage lautet: Wie groß sind $\theta(p)$ und $\psi(p)$ in Abhängigkeit von p ?

Wir machen die folgende, intuitiv leicht einsehbare Beobachtung.

Satz 2.42. Die Abbildung $[0, 1] \rightarrow [0, 1]$, $p \mapsto \theta(p)$ ist monoton wachsend.

Beweis. Obwohl die Aussage offensichtlich erscheint, wollen wir einen formalen Beweis geben, weil er ein wichtiges Beweisprinzip, das der Kopplung, verwendet.

Seien $p, p' \in [0, 1]$ mit $p < p'$. Es sei $(Y_k)_{k \in K}$ eine unabhängige Familie von Zufallsvariablen mit $\mathbf{P}[Y_k \leq q] = q$ für jedes $k \in K$ und $q \in \{p, p', 1\}$. Wir könnten an dieser Stelle zum Beispiel annehmen, dass jedes $Y_k \sim \mathcal{U}_{[0,1]}$ uniform auf $[0, 1]$ verteilt ist. Da der Existenzbeweis für unabhängige Familien mit solcher Verteilung noch aussteht, und da wir die reichhaltigere Struktur der uniformen Verteilung hier nicht benötigen, begnügen wir uns damit, dass jedes Y_k nur Werte in $\{p, p', 1\}$ annimmt, dass also gilt

$$\mathbf{P}[Y_k = q] = \begin{cases} p, & \text{falls } q = p, \\ p' - p, & \text{falls } q = p', \\ 1 - p', & \text{falls } q = 1. \end{cases}$$

Eine solche Familie $(Y_k)_{k \in K}$ existiert nach Satz 2.19. Wir setzen dann für $q \in \{p, p'\}$ und $k \in K$

$$X_k^q := \begin{cases} 1, & \text{falls } Y_k \leq q, \\ 0, & \text{sonst.} \end{cases}$$

Offenbar ist für $q \in \{p, p'\}$ die Familie $(X_k^q)_{k \in K}$ unabhängig (Bemerkung 2.15(iii)) und $X_k^q \sim \text{Ber}_q$. Ferner ist $X_k^p \leq X_k^{p'}$ für jedes $k \in K$. Wir nennen dieses Verfahren, zwei Familien von Zufallsvariablen auf einem Wahrscheinlichkeitsraum herzustellen eine *Kopplung*.

Offenbar gilt nun $C^p(x) \subset C^{p'}(x)$ für jedes $x \in \mathbb{Z}^d$, also auch $\theta(p) \leq \theta(p')$. \square

Mit Hilfe des Kolmogorov'schen 0-1 Gesetzes können wir folgenden Satz zeigen.

Satz 2.43. Für jedes $p \in [0, 1]$ gilt $\psi(p) = \begin{cases} 0, & \text{falls } \theta(p) = 0, \\ 1, & \text{falls } \theta(p) > 0. \end{cases}$

Beweis. Ist $\theta(p) = 0$, so ist nach (2.13)

$$\psi(p) \leq \sum_{y \in \mathbb{Z}^d} \mathbf{P}[\#C^p(y) = \infty] = \sum_{y \in \mathbb{Z}^d} \theta(p) = 0.$$

Sei nun $A = \bigcup_{y \in \mathbb{Z}^d} \{\#C^p(y) = \infty\}$. Offenbar ändert es nichts am Eintreten von A , wenn endlich viele Kanten ihren Zustand verändern. Das heißt $A \in \sigma((X_k^p)_{k \in K \setminus F})$ für jedes endliche $F \subset K$. Nach Satz 2.35 ist A also in der terminalen σ -Algebra $\mathcal{T}((X_k^p)_{k \in K})$. Nach dem Kolmogorov'schen 0-1 Gesetz (Satz 2.37) gilt also $\psi(p) = \mathbf{P}[A] \in \{0, 1\}$. Ist nun $\theta(p) > 0$, so folgt wegen $\psi(p) \geq \theta(p)$ schon $\psi(p) = 1$. \square

Aufgrund der Monotonie können wir nun die folgende Definition treffen.

Definition 2.44. Der kritische Wert p_c für das Auftreten von Perkolation wird definiert als

$$\begin{aligned} p_c &= \inf\{p \in [0, 1] : \theta(p) > 0\} = \sup\{p \in [0, 1] : \theta(p) = 0\} \\ &= \inf\{p \in [0, 1] : \psi(p) = 1\} = \sup\{p \in [0, 1] : \psi(p) = 0\}. \end{aligned}$$

Satz 2.45. Für $d = 1$ ist $p_c = 1$. Für $d \geq 2$ ist $p_c(d) \in [\frac{1}{2d-1}, \frac{2}{3}]$.

Beweis. Sei zunächst $d = 1$ und $p < 1$. Sei $A^- := \{X_{(n,n+1)}^p = 0 \text{ für ein } n < 0\}$ und $A^+ := \{X_{(n,n+1)}^p = 0 \text{ für ein } n > 0\}$ sowie $A = A^- \cap A^+$. Nach dem Lemma von Borel-Cantelli gilt $\mathbf{P}[A^-] = \mathbf{P}[A^+] = 1$. Also ist $\theta(p) = \mathbf{P}[A^c] = 0$.

Wir betrachten nun den Fall $d \geq 2$.

Untere Schranke. Wir zeigen zunächst $p_c \geq \frac{1}{2d-1}$. Offenbar gilt für jedes $n \in \mathbb{N}$

$$\mathbf{P}[\#C^p(0) = \infty] \leq \mathbf{P}[\text{es gibt ein } x \in C^p(0) \text{ mit } \|x\|_\infty = n].$$

Wir schätzen nun die Wahrscheinlichkeit dafür ab, dass ein Punkt $x \in C^p(0)$ im Abstand n vom Ursprung existiert. Jeder solche Punkt wird durch einen selbstüberschneidungsfreien Pfad π , der im Ursprung beginnt und eine Länge $m \geq n$ hat, an den Ursprung angebunden. Sei $\Pi_{0,m}$ die Menge solcher Pfade. Offenbar ist $\#\Pi_{0,m} \leq 2d \cdot (2d-1)^{m-1}$, denn im ersten Schritt gibt es $2d$ Möglichkeiten für den Pfad, in jedem weiteren Schritt höchstens $2d-1$, da der Schritt zurück zu einer Selbstüberschneidung führt. Weiter ist für jedes $\pi \in \Pi_{0,m}$ die Wahrscheinlichkeit, dass π nur offene Kanten benutzt

$$\mathbf{P}[\pi \text{ ist offen}] = p^m.$$

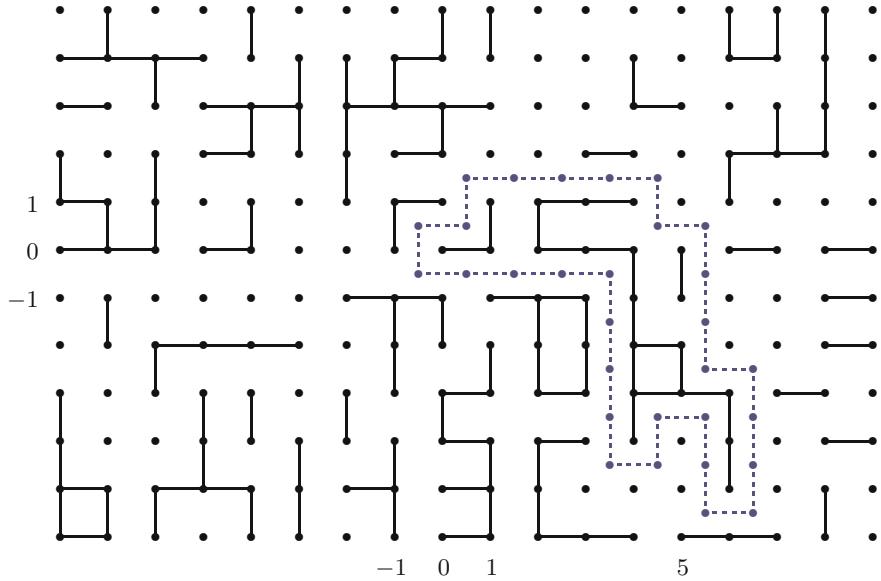
Also ist für $p < \frac{1}{2d-1}$

$$\begin{aligned} \theta(p) &\leq \sum_{m=n}^{\infty} \sum_{\pi \in \Pi_{0,m}} \mathbf{P}[\pi \text{ ist offen}] \\ &\leq \frac{2d}{2d-1} \sum_{m=n}^{\infty} ((2d-1)p)^m \\ &= \frac{2d}{(2d-1)(1-(2d-1)p)} ((2d-1)p)^n \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Mithin gilt $p_c \geq \frac{1}{2d-1}$.

Obere Schranke. Wir können \mathbb{Z}^d als Teilmenge $\mathbb{Z}^d \times \{0\} \subset \mathbb{Z}^{d+1}$ auffassen. Tritt daher für ein p Perkolation in \mathbb{Z}^d auf, so auch in \mathbb{Z}^{d+1} . Daher gilt für die entsprechenden kritischen Werte $p_c(d+1) \leq p_c(d)$.

Es reicht also, den Fall $d = 2$ zu betrachten. Hier zeigen wir $p_c \leq \frac{2}{3}$. Wir geben ein Konturargument an, das von Peierls für ein Magnetismusmodell (das Ising Modell, siehe Beispiel 18.16 und speziell (18.9)) entwickelt wurde (siehe [136]).

Abb. 2.2. Kontur des Clusters C_5 .

Für $N \in \mathbb{N}$ schreiben wir (vergleiche (2.12) mit $x = (i, 0)$)

$$C_N := \bigcup_{i=0}^N C^p((i, 0))$$

für die Menge der Punkte, die eine offene Verbindung in die Menge $\{0, \dots, N\} \times \{0\}$ haben. Dann ist wegen der Subadditivität der Wahrscheinlichkeit (und wegen $\mathbf{P}[\#C^p((i, 0)) = \infty] = \theta(p)$ für jedes $i \in \mathbb{Z}$)

$$\theta(p) = \frac{1}{N+1} \sum_{i=0}^N \mathbf{P}[\#C^p((i, 0)) = \infty] \geq \frac{1}{N+1} \mathbf{P}[\#C_N = \infty].$$

Wir betrachten nun Konturen im dualen Graphen $(\tilde{\mathbb{Z}}^2, \tilde{K})$, die C_N umschließen, falls $\#C_N < \infty$. Der duale Graph ist dabei definiert durch

$$\begin{aligned} \tilde{\mathbb{Z}}^2 &= \left(\frac{1}{2}, \frac{1}{2}\right) + \mathbb{Z}^2, \\ \tilde{K} &= \left\{ \{x, y\} : x, y \in \tilde{\mathbb{Z}}^2, \|x - y\|_2 = 1 \right\}. \end{aligned}$$

Eine Kante \tilde{k} im dualen Graphen $(\tilde{\mathbb{Z}}^2, \tilde{K})$ kreuzt also genau eine Kante k in (\mathbb{Z}^2, K) . Wir nennen \tilde{k} offen, falls k offen ist, und sonst geschlossen. Ein Kreis

γ ist ein selbstüberschneidungsfreier Pfad in $(\tilde{\mathbb{Z}}^2, \tilde{K})$, bei dem Anfangs- und Endpunkt übereinstimmen. Eine Kontur der Menge C_N ist ein minimaler Kreis, der C_N umschließt. Minimal heißt dabei, dass die umschlossene Fläche minimal ist (siehe Abb. 2.2).

Für $n \geq 2N$ sei

$$\Gamma_n = \left\{ \gamma : \gamma \text{ ist ein Kreis der Länge } n \text{ und umschließt } \{0, \dots, N\} \times \{0\} \right\}.$$

Wir wollen eine obere Abschätzung für $\#\Gamma_n$ angeben. Dafür wählen wir für $\gamma \in \Gamma_n$ einen Punkt aus γ willkürlich als Startpunkt aus, nämlich den oberen Punkt $(m + \frac{1}{2}, \frac{1}{2})$ des rechten x -Achsendurchgangs von γ (in Abb. 2.2 ist dies der Punkt $(5 + \frac{1}{2}, \frac{1}{2})$). Offenbar ist $m \geq N$ und $m \leq n$ weil der Ursprung von γ umschlossen wird. Ausgehend von $(m + \frac{1}{2}, \frac{1}{2})$ gibt es für jede weitere Kante von γ jeweils höchstens drei Möglichkeiten. Also ist

$$\#\Gamma_n \leq n \cdot 3^n.$$

Der Kreis γ heißt geschlossen, wenn er nur (in \tilde{K}) geschlossene Kanten benutzt. Eine Kontur von C_N muss automatisch geschlossen sein und eine Länge größer als $2N$ haben. Daher gilt für $p > \frac{2}{3}$

$$\begin{aligned} \mathbf{P}[\#C_N < \infty] &= \sum_{n=2N}^{\infty} \mathbf{P}[\text{es gibt einen geschlossenen Kreis } \gamma \in \Gamma_n] \\ &\leq \sum_{n=2N}^{\infty} n \cdot (3(1-p))^n \xrightarrow{N \rightarrow \infty} 0. \end{aligned}$$

Es folgt $p_c \leq \frac{2}{3}$. □

Im Allgemeinen ist der Wert von p_c nicht bekannt und extrem schwer zu bestimmen. Im Fall der Kantenperkolation in \mathbb{Z}^2 ist allerdings ein genaues Ergebnis bekannt, da man hier das starke Hilfsmittel der Selbstdualität des Graphen (\mathbb{Z}^2, K) zur Verfügung hat. (Ist $G = (V, K)$ ein planarer Graph, also einer, den man mit überschneidungsfreien Kanten in den \mathbb{R}^2 einbetten kann, so hat der duale Graph als Punktmenge die Menge der Flächen von G und als Kante zwischen zwei solchen Punkten, diejenige Kante aus K , die die beiden Flächenstücke trennt. Offenbar ist das zweidimensionale Gitter als Graph isomorph zu seinem dualen Graphen. Man beachte, dass man die Kontur in Abb. 2.2 als geschlossenen Pfad im dualen Graphen auffassen kann.) Wir zitieren hier den Satz von Kesten [102].

Satz 2.46 (Kesten (1980)). *Für Kantenperkolation in \mathbb{Z}^2 ist die kritische Wahrscheinlichkeit $p_c = \frac{1}{2}$, und es gilt $\theta(p_c) = 0$.*

Beweis. Siehe etwa das Buch von Grimmett [69, Seite 287ff]. □

Es wird vermutet, dass $\theta(p_c) = 0$ in jeder Dimension $d \geq 2$ gilt. Rigoros bewiesen ist dies allerdings nur für $d = 2$ und $d \geq 19$ (siehe [73]).

Eindeutigkeit des unendlichen Clusters*

Es sei p so gewählt, dass $\theta(p) > 0$ ist. Wir haben gesehen, dass es mit Wahrscheinlichkeit 1 *mindestens* einen unendlich großen, offenen Cluster gibt. Wir wollen nun zeigen, dass es *genau* einen gibt.

Sei also $N \in \{0, 1, \dots, \infty\}$ die (zufällige) Anzahl von unendlich großen Clustern.

Satz 2.47 (Eindeutigkeit des unendlichen großen Clusters). *Für jedes $p \in [0, 1]$ gilt $\mathbf{P}_p[N \leq 1] = 1$.*

Beweis. Diese Aussage wurde erstmals von Aizenman, Kesten und Newman gezeigt [2, 3]. Wir folgen der einfacheren Beweisidee von Burton und Keane [28], wie sie etwa in [69, Abschnitt 8.2] beschrieben wird.

In den Fällen $p = 1$ und $\theta(p) = 0$ (speziell also im Fall $p = 0$) ist die Aussage trivial. Seien nun also $p \in (0, 1)$ und $\theta(p) > 0$.

1. Schritt. Wir zeigen zunächst:

$$\mathbf{P}_p[N = m] = 1 \quad \text{für ein } m = 0, 1, \dots, \infty. \quad (2.14)$$

Wir benötigen ein 0-1 Gesetz, ähnlich dem Kolmogorov'schen. Allerdings ist N nicht messbar bezüglich der terminalen σ -Algebra, wir müssen also etwas subtiler vorgehen. Sei $e_1 = (1, 0, \dots, 0)$ der erste Einheitsvektor in \mathbb{Z}^d . Auf der Kantenmenge K definieren wir die Translation $\tau : K \rightarrow K$ durch $\tau(\langle x, y \rangle) = \langle x + e_1, y + e_1 \rangle$.

Sei

$$K_0 := \{ \langle (x_1, \dots, x_d), (y_1, \dots, y_d) \rangle \in K : x_1 = 0, y_1 \geq 0 \}$$

die Menge aller Kanten in \mathbb{Z}^d , die zwei Punkte in $\{0\} \times \mathbb{Z}^{d-1}$ verbinden oder einen Punkt aus $\{0\} \times \mathbb{Z}^{d-1}$ mit einem aus $\{1\} \times \mathbb{Z}^{d-1}$ verbinden. Offenbar sind die Mengen $(\tau^n(K_0), n \in \mathbb{Z})$ disjunkt und $K = \bigcup_{n \in \mathbb{Z}} \tau^n(K_0)$. Daher sind die Zufallsvariablen $Y_n := (X_{\tau^n(k)}^p)_{k \in K_0}, n \in \mathbb{Z}$, unabhängig und identisch verteilt (mit Werten in $\{0, 1\}^{K_0}$). Setze $Y = (Y_n)_{n \in \mathbb{Z}}$ und $\tau(Y) = (Y_{n+1})_{n \in \mathbb{Z}}$. Sei $A_m \in \{0, 1\}^K$ definiert durch

$$\{Y \in A_m\} = \{N = m\}.$$

Offenbar ändert sich der Wert von N nicht, wenn wir alle Kanten gleichzeitig verschieben. Es gilt also $\{Y \in A_m\} = \{\tau(Y) \in A_m\}$. Ein Ereignis mit dieser Eigenschaft nennen wir invariant. Mit einem Argument ähnlich dem für das Kolmogorov'sche 0-1 Gesetz kann man zeigen, dass invariante Ereignisse (die durch u.i.v. Zufallsvariablen definiert werden) nur die Wahrscheinlichkeiten 0 oder 1 haben können (für einen formalen Beweis siehe Beispiel 20.26).

2. Schritt. Wir zeigen:

$$\mathbf{P}_p[N = m] = 0 \quad \text{für jedes } m \in \mathbb{N} \setminus \{1\}. \quad (2.15)$$

Sei also $m = 2, 3, \dots$. Wir nehmen an, dass $\mathbf{P}[N = m] = 1$ gilt und führen dies zum Widerspruch.

Für $L \in \mathbb{N}$ setzen wir $B_L := \{-L, \dots, L\}^d$ und bezeichnen mit $K_L = \{k = \langle x, y \rangle \in K : x, y \in B_L\}$ die Menge der Kanten, deren beide Endpunkte in B_L liegen. Für $i = 0, 1$ sei $D_L^i := \{X_k^p = i \text{ für alle } k \in K_L\}$. Sei N_L^1 die Anzahl der unendlichen Cluster, wenn wir (unabhängig vom Wert von X_k^p) jede Kante k in K_L als offen betrachten. Analog definieren wir N_L^0 , wobei wir hier die Kanten in K_L als geschlossen betrachten. Wegen $\mathbf{P}_p[D_L^i] > 0$, und wegen $N = m$ fast sicher, gilt $N_L^i = m$ fast sicher für $i = 0, 1$.

Sei

$$A_L^2 := \bigcup_{x^1, x^2 \in B_L \setminus B_{L-1}} \{C^p(x^1) \cap C^p(x^2) = \emptyset\} \cap \{\#C^p(x^1) = \#C^p(x^2) = \infty\}$$

das Ereignis, dass es zwei Punkte auf dem Rand von B_L gibt, die in unterschiedlichen, unendlich großen, offenen Clustern sitzen. Offenbar gilt $A_L^2 \uparrow \{N \geq 2\}$ für $L \rightarrow \infty$.

Sei $A_{L,0}^2$ ähnlich wie A_L^2 definiert, jedoch wollen wir alle Kanten $k \in K_L$ als geschlossen betrachten, egal ob $X_k^p = 1$ oder $X_k^p = 0$ ist. Tritt A_L^2 ein, so gibt es zwei Punkte x^1, x^2 auf dem Rand von B_L und zu jedem $i = 1, 2$ einen unendlich langen selbstüberschneidungsfreien, offenen Pfad π_{x^i} , der in x^i startet und x^{3-i} vermeidet. Es gilt also $A_L^2 \subset A_{L,0}^2$. Wähle nun L so groß, dass $\mathbf{P}[A_{L,0}^2] > 0$ ist.

Tritt $A_{L,0}^2$ ein und werden alle Kanten in B_L geöffnet, so werden mindestens zwei der unendlich großen, offenen Cluster durch Kanten in B_L verbunden, die Gesamtzahl der unendlich großen, offenen Cluster also um mindestens Eins verringert. Es folgt $\mathbf{P}_p[N_L^1 \leq N_L^0 - 1] \geq \mathbf{P}_p[A_{L,0}^2] > 0$, was einen Widerspruch bedeutet.

3. Schritt. Da wir im zweiten Schritt bereits gezeigt haben, dass N fast sicher keinen *endlichen* Wert größer als 1 annimmt, brauchen wir nun nur noch zu zeigen, dass N fast sicher nicht den Wert ∞ annimmt. Wir zeigen hier, dass in der Tat gilt:

$$\mathbf{P}_p[N \geq 3] = 0. \quad (2.16)$$

Dieses ist der schwierigste Teil. Wir nehmen an, dass $\mathbf{P}_p[N \geq 3] > 0$ gilt und führen dies zum Widerspruch.

Wir nennen einen Punkt $x \in \mathbb{Z}^d$ einen **Trifurkationspunkt**, falls x in einem unendlich großen, offenen Cluster $C^p(x)$ liegt, genau drei offene Kanten zu x führen und die Wegnahme dieser drei Kanten $C^p(x)$ in drei unendlich große, *disjunkte* Cluster zerteilt. Mit T bezeichnen wir die Menge der Trifurkationspunkte und schreiben $T_L := T \cap B_L$. Sei $r := \mathbf{P}_p[0 \in T]$. Aufgrund der Translationsinvarianz gilt $(\#B_L)^{-1} \mathbf{E}_p[\#T_L] = r$ für jedes L . Sei

$$A_L^3 := \bigcup_{x^1, x^2, x^3 \in B_L \setminus B_{L-1}} \left(\bigcap_{i \neq j} \{C^p(x^i) \cap C^p(x^j) = \emptyset\} \right) \cap \left(\bigcap_{i=1}^3 \{\#C^p(x^i) = \infty\} \right)$$

das Ereignis, dass es drei Punkte auf dem Rand von B_L gibt, die in unterschiedlichen, unendlich großen, offenen Clustern sitzen. Offenbar gilt $A_L^3 \uparrow \{N \geq 3\}$ für $L \rightarrow \infty$.

Analog zu $A_{L,0}^2$ definieren wir $A_{L,0}^3$ als das Ereignis, dass es drei Punkte auf dem Rand von B_L gibt, die in unterschiedlichen unendlich großen, offenen Clustern sitzen, wenn wir alle Kanten in K_L als geschlossen ansehen. Wie oben ist $A_L^3 \subset A_{L,0}^3$.

Für drei unterschiedliche Punkte $x^1, x^2, x^3 \in B_L \setminus B_{L-1}$ sei F_{x^1, x^2, x^3} das Ereignis, dass es zu jedem $i = 1, 2, 3$ einen unendlich langen selbstüberschneidungsfreien, offenen Pfad π_{x^i} gibt, der in x^i startet, nur Kanten aus $K^p \setminus K_L$ benutzt und die anderen $x^j, j \neq i$, vermeidet. Dann gilt

$$A_{L,0}^3 \subset \bigcup_{\substack{x^1, x^2, x^3 \in B_L \setminus B_{L-1} \\ \text{paarweise unterschiedlich}}} F_{x^1, x^2, x^3}.$$

Sei L so groß, dass $\mathbf{P}_p[A_{L,0}^3] \geq \mathbf{P}_p[N \geq 3]/2 > 0$ gilt. Wähle drei unterschiedliche Punkte $x^1, x^2, x^3 \in B_L \setminus B_{L-1}$ mit $\mathbf{P}_p[F_{x^1, x^2, x^3}] > 0$.

Tritt F_{x^1, x^2, x^3} ein, so können wir einen Punkt $y \in B_L$ finden, von dem aus drei disjunkte (nicht notwendigerweise offene) Pfade π_1, π_2 und π_3 zu den Punkten x^1, x^2 und x^3 führen. Sei G_{y, x^1, x^2, x^3} das Ereignis, dass in K_L genau diejenigen Kanten offen sind, die zu diesen Pfaden gehören, und alle anderen geschlossen. Die Ereignisse F_{x^1, x^2, x^3} und G_{y, x^1, x^2, x^3} sind unabhängig, und y ist ein Trifurkationspunkt, falls beide eintreten. Daher ist

$$r = \mathbf{P}_p[y \in T] \geq \mathbf{P}_p[F_{x^1, x^2, x^3}] \cdot (p \wedge (1-p))^{\#K_L} > 0.$$

Wir zeigen nun, dass $r = 0$ sein muss, was die Annahme $\mathbf{P}_p[N \geq 3] > 0$ ad absurdum führt. Auf der Menge \tilde{K}_L aller Kanten, die mindestens einen Endpunkt in B_L liegen haben, führen wir eine Äquivalenzrelation R ein, indem wir zwei Kanten als äquivalent betrachten, wenn es einen offenen Pfad gibt, der keinen Trifurkationspunkt trifft, und der von jeder der beiden Kanten mindestens einen Endpunkt trifft. Mit $U_L := \tilde{K}_L/R$ bezeichnen wir die Menge der Äquivalenzklassen. (Man beachte, dass die drei an einen Trifurkationspunkt angrenzenden Kanten stets in drei unterschiedlichen Äquivalenzklassen liegen.) Wir machen $H_L := U_L \cup T_L$ zu einem Graphen, indem wir zwei Punkte $x \in T_L$ und $u \in U_L$ als benachbart ansehen, wenn es eine Kante $k \in u$ gibt, deren einer Endpunkt x ist. Jeder Punkt $x \in T_L$ ist mit genau drei Punkten aus U_L benachbart. Die Punkte aus U_L können isoliert sein (keine Nachbarn haben), oder an beliebig viele Punkte aus T_L angrenzen, nicht jedoch an Punkte aus U_L .

Eine Schleife ist ein selbstüberschneidungsfreier, endlicher Pfad, der zu seinem Startpunkt zurückkehrt. Der Graph H_L ist schleifenfrei. In der Tat: gäbe es in H_L einen selbstüberschneidungsfreien Pfad (h_0, h_1, \dots, h_n) , der in $h_0 = h_n = x \in T_L$ beginnt und endet, so müssten $h_1, h_{n-1} \in U_L$ liegen und unterschiedlich sein. Nach

Definition des Trifurkationspunktes x müssen h_1 und h_{n-1} in unterschiedlichen offenen Clustern liegen, wenn x entfernt wird. Andererseits gibt es eine Verbindung der Kanten in h_1 mit denen in h_{n-1} entlang des Weges h_1, \dots, h_{n-1} , der x nicht trifft. Da dies ein Widerspruch ist, kann es einen solchen Pfad nicht geben. Aus einer Schleife (g_0, \dots, g_m) , die in $g_0 = g_m \in U_L$ beginnt und endet, kann eine Schleife $(g_1, g_2, \dots, g_m, g_1)$ konstruiert werden, die in $g_1 \in T_L$ beginnt und endet. Da wir eine solche Schleife schon ausgeschlossen haben, ist der Graph H_L als schleifenfrei erkannt.

Wir schreiben $\deg_{H_L}(h)$ für den *Grad* von $h \in H_L$, also für die Anzahl der Nachbarn von h in H_L . Ein Punkt h mit $\deg_{H_L}(h) = 1$ heißt ein *Blatt* von H_L . Offenbar können nur die Punkte aus U_L Blätter sein. Sei nun $Z \subset H_L$ eine Zusammenhangskomponente, die mindestens einen Punkt $x \in T_L$ enthält. Da Z ein Baum ist (zusammenhängend und schleifenfrei), gilt

$$\#Z - 1 = \frac{1}{2} \sum_{h \in Z} \deg_{H_L}(h).$$

Diese Formel lässt sich leicht so umstellen, dass man die Anzahl der Blätter bestimmen kann:

$$\begin{aligned} \#\{u \in Z : \deg_{H_L}(u) = 1\} &= 2 + \sum_{h \in Z} (\deg_{H_L}(h) - 2)^+ \\ &\geq 2 + \#\{h \in Z : \deg_{H_L}(h) \geq 3\} \\ &= 2 + \#(Z \cap T_L). \end{aligned}$$

Indem wir über die Zusammenhangskomponenten Z von H_L mit mindestens einem Punkt in T_L summieren, erhalten wir

$$\#\{u \in H_L : \deg_{H_L}(u) = 1\} \geq \#T_L.$$

Nun enthält jedes Blatt $u \in H_L$ eine Kante, die an ein $x \in T_L$ angrenzt. Also liegen die Kanten von u in einem unendlich großen, offenen Cluster von K^p , und es gibt wenigstens eine Kante $k \in u$ mit mindestens einem Endpunkt in $B_L \setminus B_{L-1}$. Für unterschiedliche Blätter sind dies stets unterschiedliche Punkte, da die Blätter zu *disjunkten* Clustern gehören. Speziell erhalten wir die Abschätzung

$$\#T_L \leq \#(B_L \setminus B_{L-1}).$$

Es folgt

$$\frac{\#T_L}{\#B_L} \leq \frac{\#(B_L \setminus B_{L-1})}{\#B_L} \leq \frac{d}{L} \xrightarrow{L \rightarrow \infty} 0.$$

Wegen $r = (\#B_L)^{-1} \mathbf{E}_p[\#T_L] \leq d/L$ folgt $r = 0$. (Man beachte, dass wir hier im Vorgriff auf Kapitel 5 den Erwartungswert $\mathbf{E}_p[\#T_L]$ benutzt haben.) \square

3 Erzeugendenfunktion

Ein wichtiges Prinzip in der Mathematik ist es, eine Klasse von Objekten, die man betrachten möchte, in eine andere Klasse von Objekten, mit denen man besser rechnen kann, hinein abzubilden. Diese Abbildung kann eineindeutig sein, etwa bei der Zuordnung von Matrizen zu linearen Abbildungen, oder auch nur manche Eigenschaften eindeutig abbilden, etwa bei Determinanten.

Zu der zweiten Kategorie gehören in der Wahrscheinlichkeitstheorie die Kenngrößen wie Median, Erwartungswert und Varianz von Zufallsvariablen. Zur ersten Kategorie hingegen charakteristische Funktionen, Laplace-Transformierte und Erzeugendenfunktionen, die enge Verwandte sind und ihre Nützlichkeit daraus ziehen, dass Addition von unabhängigen Zufallsvariablen in Multiplikation übergeht. Bevor wir in späteren Kapiteln insbesondere die charakteristischen Funktionen ausgiebig behandeln, wollen wir wichtige Grundideen in der einfacheren Situation der Erzeugendenfunktionen, deren Anwendung auf \mathbb{N}_0 -wertige Zufallsvariablen beschränkt ist, kennen lernen.

3.1 Definition und Beispiele

Definition 3.1 (Erzeugendenfunktion). Sei X eine \mathbb{N}_0 -wertige Zufallsvariable. Die Abbildung $\psi_{\mathbf{P}_X} = \psi_X$, die erklärt wird durch

$$\psi_X : [0, 1] \rightarrow [0, 1], \quad z \mapsto \sum_{n=0}^{\infty} \mathbf{P}[X = n] z^n, \quad (3.1)$$

heißt **Erzeugendenfunktion** von \mathbf{P}_X (oder etwas lax: von X).

Satz 3.2. (i) ψ_X ist stetig und in $(0, 1)$ unendlich oft stetig differenzierbar. Es gilt für $n \in \mathbb{N}$ und die n -te Ableitung $\psi_X^{(n)}$

$$\lim_{z \uparrow 1} \psi_X^{(n)}(z) = \sum_{k=n}^{\infty} \mathbf{P}[X = k] \cdot k(k-1) \cdots (k-n+1), \quad (3.2)$$

wobei beide Seiten $= +\infty$ sein können.

- (ii) Die Verteilung \mathbf{P}_X von X ist durch ψ_X eindeutig charakterisiert.
- (iii) Für jedes $r \in (0, 1)$ ist ψ_X durch die Angabe abzählbar vieler Werte $\psi_X(x_i)$, $x_i \in [0, r]$, $i \in \mathbb{N}$, eindeutig festgelegt. Konvergiert die Reihe in (3.1) auch für ein $z > 1$, so ist dies auch für jedes $r \in (0, z)$ richtig, und es gilt

$$\lim_{x \uparrow 1} \psi_X^{(n)}(x) = \psi_X^{(n)}(1) < \infty \quad \text{für } n \in \mathbb{N}.$$

ψ_X ist dann durch Angabe von $\psi_X^{(n)}(1)$, $n \in \mathbb{N}$, eindeutig charakterisiert.

Beweis. Das folgt aus der elementaren Theorie der Potenzreihen. Für den ersten Teil von (iii), siehe z.B. [158, Satz 8.5] \square

Satz 3.3 (Multiplikativität der Erzeugendenfunktion). Sind X_1, \dots, X_n unabhängig und \mathbb{N}_0 -wertig, so ist

$$\psi_{X_1+\dots+X_n} = \prod_{i=1}^n \psi_{X_i}.$$

Beweis. Für $z \in [0, 1)$ können wir $\psi_{X_1}(z) \psi_{X_2}(z)$ als Cauchy-Produkt schreiben

$$\begin{aligned} \psi_{X_1}(z) \psi_{X_2}(z) &= \left(\sum_{n=0}^{\infty} \mathbf{P}[X_1 = n] z^n \right) \left(\sum_{n=0}^{\infty} \mathbf{P}[X_2 = n] z^n \right) \\ &= \sum_{n=0}^{\infty} z^n \left(\sum_{m=0}^n \mathbf{P}[X_1 = m] \mathbf{P}[X_2 = n-m] \right) \\ &= \sum_{n=0}^{\infty} z^n \sum_{m=0}^n \mathbf{P}[X_1 = m, X_2 = n-m] \\ &= \sum_{n=0}^{\infty} \mathbf{P}[X_1 + X_2 = n] z^n = \psi_{X_1+X_2}(z). \end{aligned}$$

Induktiv folgt die Aussage für jedes $n \geq 2$. \square

Beispiel 3.4. (i) Sei X $b_{n,p}$ -verteilt für gewisse $n \in \mathbb{N}$ und $p \in [0, 1]$. Dann ist

$$\psi_X(z) = \sum_{m=0}^n \binom{n}{m} p^m (1-p)^{n-m} z^m = (pz + (1-p))^n. \quad (3.3)$$

(ii) Sind X, Y unabhängig und $b_{m,p}$ beziehungsweise $b_{n,p}$ -verteilt, so ist nach Satz 3.3

$$\psi_{X+Y}(z) = (pz + (1-p))^m (pz + (1-p))^n = (pz + (1-p))^{m+n}.$$

Also ist nach Satz 3.2(ii) $X + Y$ $b_{m+n,p}$ -verteilt und damit (nach Satz 2.31)

$$b_{m,p} * b_{n,p} = b_{m+n,p}.$$

(iii) Seien X und Y unabhängig und Poisson-verteilt mit Parametern $\lambda \geq 0$ und $\mu \geq 0$, also $\mathbf{P}[X = n] = e^{-\lambda} \lambda^n / n!$ für $n \in \mathbb{N}_0$. Dann ist

$$\psi_{\text{Poi}_\lambda}(z) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{(\lambda z)^n}{n!} = e^{\lambda(z-1)}. \quad (3.4)$$

Also hat $X + Y$ die Erzeugendenfunktion

$$\psi_{\text{Poi}_\lambda}(z) \cdot \psi_{\text{Poi}_\mu}(z) = e^{\lambda(z-1)} e^{\mu(z-1)} = \psi_{\text{Poi}_{\lambda+\mu}}(z),$$

und daher ist $X + Y \sim \text{Poi}_{\lambda+\mu}$. Es folgt

$$\text{Poi}_\lambda * \text{Poi}_\mu = \text{Poi}_{\lambda+\mu}. \quad (3.5)$$

(iv) Seien $X_1, \dots, X_n \sim \gamma_p$ unabhängig und geometrisch verteilt mit Parameter $p \in (0, 1)$. Wir setzen $Y = X_1 + \dots + X_n$. Es ist für $z \in [0, 1]$

$$\psi_{X_1}(z) = \sum_{k=0}^{\infty} p(1-p)^k z^k = \frac{p}{1 - (1-p)z}. \quad (3.6)$$

Nach der verallgemeinerten binomischen Formel (siehe Lemma 3.5 mit $\alpha = -n$), Satz 3.3 und (3.6) ist

$$\begin{aligned} \psi_Y(z) &= \psi_{X_1}(z)^n = \frac{p^n}{(1 - (1-p)z)^n} \\ &= \sum_{k=0}^{\infty} p^n \binom{-n}{k} (-1)^k (1-p)^k z^k \\ &= \sum_{k=0}^{\infty} b_{n,p}^-(\{k\}) z^k, \end{aligned}$$

wobei für beliebiges $r \in (0, \infty)$ und $p \in (0, 1]$

$$b_{r,p}^- = \sum_{k=0}^{\infty} \binom{-r}{k} (-1)^k p^r (1-p)^k \delta_k \quad (3.7)$$

die negative Binomialverteilung mit Parametern r und p ist. Nach dem Eindeutigkeitssatz für Erzeugendenfunktionen ist damit $Y \sim b_{n,p}^-$, also (siehe Definition 2.29 für die n -te Faltungspotenz) $b_{n,p}^- = \gamma_p^{*n}$. \diamond

Lemma 3.5 (Verallgemeinerter binomischer Lehrsatz). Für $\alpha \in \mathbb{R}$ und $k \in \mathbb{N}_0$ definieren wir den Binomialkoeffizienten

$$\binom{\alpha}{k} := \frac{\alpha \cdot (\alpha - 1) \cdots (\alpha - k + 1)}{k!}. \quad (3.8)$$

Es gilt die erweiterte binomische Formel:

$$(1 + x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \quad \text{für jedes } x \in \mathbb{C} \text{ mit } |x| < 1. \quad (3.9)$$

Speziell gilt

$$\frac{1}{\sqrt{1-x}} = \sum_{n=0}^{\infty} \binom{2n}{n} 4^{-n} x^n \quad \text{für jedes } x \in \mathbb{C} \text{ mit } |x| < 1. \quad (3.10)$$

Beweis. Die Abbildung $f : x \mapsto (1 + x)^\alpha$ ist holomorph bis auf eventuell eine Singularität bei $x = -1$, ist also um 0 in eine Potenzreihe entwickelbar mit Radius mindestens 1:

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k \quad \text{für } |x| < 1.$$

Für $k \in \mathbb{N}_0$ ist die k -te Ableitung $f^{(k)}(0) = \alpha(\alpha - 1) \cdots (\alpha - k + 1)$, also folgt (3.9).

Der Zusatz folgt, weil für $\alpha = -1/2$ gilt, dass $\binom{-1/2}{n} = \binom{2n}{n} (-4)^{-n}$. □

Übung 3.1.1. Man zeige $b_{r,p}^- * b_{s,p}^- = b_{r+s,p}^-$ für $r, s \in (0, \infty)$ und $p \in (0, 1]$. ♣

Übung 3.1.2. Man gebe ein Beispiel an für zwei Erzeugendenfunktionen, die an abzählbar vielen Stellen $x_i \in (0, 1)$, $i \in \mathbb{N}$, übereinstimmen, aber nicht gleich sind. (Es kann also in Satz 3.2(iii) nicht auf die Bedingung $\psi(z) < \infty$ für ein $z > 1$ verzichtet werden.) ♣

3.2 Poisson-Approximation

Die Konvergenz von Wahrscheinlichkeitsmaßen untersuchen wir in systematischer Weise in Kapitel 13. Hier betrachten wir die Konvergenz von Wahrscheinlichkeitsmaßen auf \mathbb{N}_0 und übersetzen sie im folgenden Lemma in die Konvergenz der zugehörigen Erzeugendenfunktionen. Als Anwendung zeigen wir, unter welchen Bedingungen die Summe unabhängiger Bernoulli-Zufallsvariablen (auch mit unterschiedlichen Erfolgsparametern) gegen die Poissonverteilung konvergiert.

Lemma 3.6. Seien μ und $(\mu_n)_{n \in \mathbb{N}}$ W-Maße auf $(\mathbb{N}_0, 2^{\mathbb{N}_0})$ mit Erzeugendenfunktionen ψ und ψ_n , $n \in \mathbb{N}$. Dann sind äquivalent

- (i) $\mu_n(\{k\}) \xrightarrow{n \rightarrow \infty} \mu(\{k\})$ für jedes $k \in \mathbb{N}_0$,
- (ii) $\mu_n(A) \xrightarrow{n \rightarrow \infty} \mu(A)$ für jedes $A \subset \mathbb{N}_0$,
- (iii) $\psi_n(z) \xrightarrow{n \rightarrow \infty} \psi(z)$ für jedes $z \in [0, 1]$,
- (iv) $\psi_n(z) \xrightarrow{n \rightarrow \infty} \psi(z)$ für jedes $z \in [0, \eta)$ für ein $\eta \in (0, 1)$.

Gilt eine der vier Bedingungen, so schreiben wir $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ und sagen $(\mu_n)_{n \in \mathbb{N}}$ konvergiere schwach gegen μ .

Beweis. (i) \implies (ii) Sei $\varepsilon > 0$ und $N \in \mathbb{N}$ so gewählt, dass $\mu(\{N+1, N+2, \dots\}) < \frac{\varepsilon}{4}$. Für hinreichend großes $n_0 \in \mathbb{N}$ ist ferner

$$\sum_{k=0}^N |\mu_n(\{k\}) - \mu(\{k\})| < \frac{\varepsilon}{4} \quad \text{für jedes } n \geq n_0.$$

Speziell ist für $n \geq n_0$ auch $\mu_n(\{N+1, N+2, \dots\}) < \frac{\varepsilon}{2}$. Also ist für $n \geq n_0$

$$\begin{aligned} |\mu_n(A) - \mu(A)| &\leq \mu_n(\{N+1, N+2, \dots\}) + \mu(\{N+1, N+2, \dots\}) \\ &\quad + \sum_{k \in A \cap \{0, \dots, N\}} |\mu_n(\{k\}) - \mu(\{k\})| \\ &< \varepsilon. \end{aligned}$$

(ii) \implies (i) Dies ist trivial.

(i) \iff (iii) \iff (iv) Dies folgt aus elementarer Theorie der Potenzreihen. \square

Seien $(p_{n,k})_{n,k \in \mathbb{N}}$ Zahlen mit $p_{n,k} \in [0, 1]$, sodass der Grenzwert

$$\lambda := \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} p_{n,k} \in (0, \infty) \tag{3.11}$$

existiert und so, dass $\lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} p_{n,k}^2 = 0$ ist (beispielsweise $p_{n,k} = \lambda/n$ für $k \leq n$ und $p_{n,k} = 0$ für $k > n$). Für jedes $n \in \mathbb{N}$ sei $(X_{n,k})_{k \in \mathbb{N}}$ eine unabhängige Familie von Zufallsvariablen mit $X_{n,k} \sim \text{Ber}_{p_{n,k}}$. Setze

$$S^n := \sum_{l=1}^{\infty} X_{n,l} \quad \text{und} \quad S_k^n := \sum_{l=1}^k X_{n,l} \quad \text{für } k \in \mathbb{N}.$$

Satz 3.7 (Poisson-Approximation). Unter den obigen Annahmen konvergieren die Verteilungen $(\mathbf{P}_{S^n})_{n \in \mathbb{N}}$ schwach gegen die Poisson-Verteilung Poi_λ .

Beweis. Die Poisson-Verteilung hat die Erzeugendenfunktion $\psi(z) = e^{\lambda(z-1)}$ (siehe (3.4)). Andererseits sind $S^n - S_k^n$ und S_k^n unabhängig für jedes $k \in \mathbb{N}$, also $\psi_{S^n} = \psi_{S_k^n} \cdot \psi_{S^n - S_k^n}$. Nun ist für jedes $z \in [0, 1]$

$$1 \geq \frac{\psi_{S^n}(z)}{\psi_{S_k^n}(z)} = \psi_{S^n - S_k^n}(z) \geq 1 - \mathbf{P}[S^n - S_k^n \geq 1] \geq 1 - \sum_{l=k+1}^{\infty} p_{n,l} \xrightarrow{k \rightarrow \infty} 1,$$

also

$$\begin{aligned} \psi_{S^n}(z) &= \lim_{k \rightarrow \infty} \psi_{S_k^n}(z) = \prod_{l=1}^{\infty} (p_{n,l} z + (1 - p_{n,l})) \\ &= \exp \left(\sum_{l=1}^{\infty} \log (1 + p_{n,l}(z - 1)) \right). \end{aligned}$$

Für $|x| < \frac{1}{2}$ ist $|\log(1+x) - x| \leq x^2$. Nach Voraussetzung gilt $\max_{l \in \mathbb{N}} p_{n,l} \rightarrow 0$ für $n \rightarrow \infty$, also ist für hinreichend großes n

$$\begin{aligned} &\left| \left(\sum_{l=1}^{\infty} \log (1 + p_{n,l}(z - 1)) \right) - \left((z - 1) \sum_{l=1}^{\infty} p_{n,l} \right) \right| \\ &\leq \sum_{l=1}^{\infty} p_{n,l}^2 \leq \left(\sum_{l=1}^{\infty} p_{n,l} \right) \max_{l \in \mathbb{N}} p_{n,l} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Zusammen mit (3.11) folgt

$$\lim_{n \rightarrow \infty} \psi_{S^n}(z) = \lim_{n \rightarrow \infty} \exp \left((z - 1) \sum_{l=1}^{\infty} p_{n,l} \right) = e^{\lambda(z-1)}. \quad \square$$

3.3 Verzweigungsprozesse

Seien T, X_1, X_2, \dots unabhängige, \mathbb{N}_0 -wertige Zufallsvariablen. Wie sieht die Verteilung von $S := \sum_{n=1}^T X_n$ aus? Zunächst bemerken wir, dass S messbar ist, denn

$$\{S = k\} = \bigcup_{n=0}^{\infty} \{T = n\} \cap \{X_1 + \dots + X_n = k\}.$$

Satz 3.8. Sind die X_1, X_2, \dots zusätzlich identisch verteilt, so ist die Erzeugendenfunktion von S gegeben durch $\psi_S(z) = \psi_T(\psi_{X_1}(z))$.

Beweis. Für jedes $z \in [0, 1]$ gilt

$$\begin{aligned}\psi_S(z) &= \sum_{k=0}^{\infty} \mathbf{P}[S = k] z^k \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \mathbf{P}[T = n] \mathbf{P}[X_1 + \dots + X_n = k] z^k \\ &= \sum_{n=0}^{\infty} \mathbf{P}[T = n] \psi_{X_1}(z)^n = \psi_T(\psi_{X_1}(z)).\end{aligned}\quad \square$$

Wir nehmen jetzt an, dass $p_0, p_1, p_2, \dots \in [0, 1]$ sind mit $\sum_{k=0}^{\infty} p_k = 1$. Sei $(X_{n,i})_{n,i \in \mathbb{N}_0}$ eine unabhängige Familie von Zufallsvariablen mit $\mathbf{P}[X_{n,i} = k] = p_k$ für alle $i, k, n \in \mathbb{N}_0$.

Setze $Z_0 = 1$ und

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n-1,i} \quad \text{für } n \in \mathbb{N}.$$

Wir geben die folgende Interpretation an: Z_n ist die Anzahl von Individuen in der n -ten Generation einer sich zufällig entwickelnden Population. Das i -te Individuum aus der n -ten Generation hat $X_{n,i}$ Nachkommen (in der $(n+1)$ -ten Generation).

Definition 3.9. $(Z_n)_{n \in \mathbb{N}_0}$ heißt **Galton-Watson-Prozess** oder **Verzweigungsprozess** mit Nachkommenverteilung $(p_k)_{k \in \mathbb{N}_0}$.

Ein wichtiges Hilfsmittel bei der Untersuchung von Verzweigungsprozessen sind Erzeugendenfunktionen. Sei also

$$\psi(z) = \sum_{k=0}^{\infty} p_k z^k$$

die Erzeugendenfunktion der Nachkommenverteilung und ψ' deren Ableitung. Wir definieren die n -te Iterierte von ψ durch

$$\psi_1 := \psi, \quad \psi_n := \psi \circ \psi_{n-1} \quad \text{für } n = 2, 3, \dots$$

Sei schließlich ψ_{Z_n} die Erzeugendenfunktion von Z_n .

Lemma 3.10. Es gilt $\psi_n = \psi_{Z_n}$ für jedes $n \in \mathbb{N}$.

Beweis. Für $n = 1$ ist dies per Definition richtig. Für $n \in \mathbb{N}$ folgt mit Satz 3.8 induktiv $\psi_{Z_{n+1}} = \psi \circ \psi_{Z_n} = \psi \circ \psi_n = \psi_{n+1}$. \square

Offenbar ist die Wahrscheinlichkeit $q_n := \mathbf{P}[Z_n = 0]$, dass Z zur Zeit n schon ausgestorben ist, wachsend in n . Wir bezeichnen mit

$$q := \lim_{n \rightarrow \infty} \mathbf{P}[Z_n = 0]$$

die Aussterbewahrscheinlichkeit.

Unter welchen Bedingungen ist $q = 0$, $q = 1$, oder $q \in (0, 1)$? Offenbar ist $q \geq p_0$. Ist andererseits $p_0 = 0$, so ist Z_n wachsend in n , also $q = 0$.

Satz 3.11 (Aussterbewahrscheinlichkeit des Galton-Watson-Prozesses).

- (i) Es gilt $F := \{r \in [0, 1] : \psi(r) = r\} = \{q, 1\}$.
- (ii) $q < 1 \iff \lim_{z \uparrow 1} \psi'(z) > 1 \iff \sum_{k=1}^{\infty} k p_k > 1$.

Beweis. (i) Es gilt $\psi(1) = 1$, also $1 \in F$. Offenbar gilt

$$q_n = \psi_n(0) = \psi(q_{n-1}).$$

Wir wissen, dass $q_n \uparrow q$ gilt. Da ψ stetig ist, gilt

$$\psi(q) = \psi\left(\lim_{n \rightarrow \infty} q_n\right) = \lim_{n \rightarrow \infty} \psi(q_n) = \lim_{n \rightarrow \infty} q_{n+1} = q.$$

Also ist $q \in F$. Ist $r \in F$ ein beliebiger Fixpunkt von ψ , so gilt $r \geq 0 = q_0$. Da ψ monoton wachsend ist, folgt $r = \psi(r) \geq \psi(q_0) = q_1$ und induktiv $r \geq q_n$ für alle $n \in \mathbb{N}_0$, also $r \geq q$. Mithin ist $q = \min F$.

(ii) Für die erste Äquivalenz unterscheiden wir zwei Fälle.

1. Fall: $\lim_{z \uparrow 1} \psi'(z) \leq 1$. Da ψ strikt konvex ist, gilt in diesem Fall $\psi(z) > z$ für alle $z \in [0, 1)$, also $F = \{1\}$ und damit $q = 1$.

2. Fall: $\lim_{z \uparrow 1} \psi'(z) > 1$. Da ψ strikt konvex ist und $\psi(0) \geq 0$, gibt es genau ein $r \in [0, 1)$ mit $\psi(r) = r$, also ist $F = \{r, 1\}$ und damit $q = \min F = r$.

Die zweite Äquivalenz in (ii) folgt aus (3.2). □

Als weiterführende Literatur sei etwa auf [5] verwiesen.

4 Das Integral

Nach dem Begriff des Maßraums und der messbaren Abbildung ist das Integral messbarer reeller Abbildungen bezüglich allgemeiner Maße, nicht nur des Lebesgue-Maßes, wie es in den meisten Analysis-Vorlesungen behandelt wird, ein Eckstein der systematischen Wahrscheinlichkeitstheorie, der es uns beispielsweise erlaubt, Erwartungswerte und höhere Momente zu definieren.

In diesem Kapitel definieren wir das Integral durch Approximation mit Elementarfunktionen und leiten einfache Eigenschaften her wie das Lemma von Fatou. Die anderen Konvergenzsätze für Integrale folgen in den Kapiteln 6 und 7.

4.1 Konstruktion und einfache Eigenschaften

Sei im Folgenden stets $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Wir bezeichnen mit \mathbb{E} den Vektorraum der Elementarfunktionen (siehe Definition 1.93) auf (Ω, \mathcal{A}) und mit $\mathbb{E}^+ := \{f \in \mathbb{E} : f \geq 0\}$ den Kegel (woher der Name?) der nichtnegativen Elementarfunktionen. Gilt

$$f = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i} \quad (4.1)$$

für gewisses $m \in \mathbb{N}$ und für $\alpha_1, \dots, \alpha_m \in (0, \infty)$ sowie paarweise disjunkte Mengen $A_1, \dots, A_m \in \mathcal{A}$, so sagen wir, dass (4.1) eine Normaldarstellung der Elementarfunktion f ist.

Lemma 4.1. Sind $f = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i}$ und $f = \sum_{j=1}^n \beta_j \mathbb{1}_{B_j}$ zwei Normaldarstellungen von $f \in \mathbb{E}^+$, so gilt

$$\sum_{i=1}^m \alpha_i \mu(A_i) = \sum_{j=1}^n \beta_j \mu(B_j).$$

Beweis. Ist $\mu(A_i \cap B_j) > 0$ für gewisse i und j , so ist $A_i \cap B_j \neq \emptyset$, und für jedes $\omega \in A_i \cap B_j$ ist $f(\omega) = \alpha_i = \beta_j$. Außerdem ist offenbar $A_i \subset \bigcup_{j=1}^n B_j$, falls $\alpha_i \neq 0$ und $B_j \subset \bigcup_{i=1}^m A_i$, falls $\beta_j \neq 0$. Es folgt

$$\begin{aligned} \sum_{i=1}^m \alpha_i \mu(A_i) &= \sum_{i=1}^m \sum_{j=1}^n \alpha_i \mu(A_i \cap B_j) \\ &= \sum_{i=1}^m \sum_{j=1}^n \beta_j \mu(A_i \cap B_j) = \sum_{j=1}^n \beta_j \mu(B_j). \end{aligned} \quad \square$$

Dieses Lemma erlaubt uns, die folgende Definition zu treffen (weil der definierte Wert $I(f)$ von der gewählten Normaldarstellung nicht abhängt).

Definition 4.2. Wir definieren eine Abbildung $I : \mathbb{E}^+ \rightarrow [0, \infty]$ durch

$$I(f) = \sum_{i=1}^m \alpha_i \mu(A_i),$$

falls f die Normaldarstellung $f = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i}$ hat.

Lemma 4.3. Die Abbildung I ist positiv linear und monoton: Seien $f, g \in \mathbb{E}^+$ und $\alpha \geq 0$. Dann gelten die folgenden Aussagen.

- (i) $I(\alpha f) = \alpha I(f)$.
- (ii) $I(f + g) = I(f) + I(g)$.
- (iii) Ist $f \leq g$, so ist $I(f) \leq I(g)$.

Beweis. Übung. \square

Definition 4.4 (Integral). Ist $f : \Omega \rightarrow [0, \infty]$ messbar, so definieren wir das **Integral** von f bezüglich μ durch

$$\int f d\mu := \sup \{I(g) : g \in \mathbb{E}^+, g \leq f\}.$$

Bemerkung 4.5. Nach Lemma 4.3(iii) ist $I(f) = \int f d\mu$ für jedes $f \in \mathbb{E}^+$. Also ist das Integral eine Fortsetzung der Abbildung I von \mathbb{E}^+ auf die Menge der nicht-negativen messbaren Funktionen. \diamond

Sind $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ Abbildungen, so schreiben wir $f \leq g$, falls $f(\omega) \leq g(\omega)$ für jedes $\omega \in \Omega$ gilt. Analog verwenden wir die Schreibweise $f \geq 0$ und so fort. Hingegen schreiben wir „ $f \leq g$ fast überall“, falls die schwächere Bedingung gilt, dass eine μ -Nullmenge N existiert mit $f(\omega) \leq g(\omega)$ für jedes $\omega \in N^c$.

Lemma 4.6. Seien f, g, f_1, f_2, \dots messbare Abbildungen $\Omega \rightarrow [0, \infty]$. Dann gilt

- (i) (Monotonie) Ist $f \leq g$, dann ist $\int f d\mu \leq \int g d\mu$.
- (ii) (Monotone Konvergenz) Gilt $f_n \uparrow f$, dann konvergieren auch die Integrale $\int f_n d\mu \uparrow \int f d\mu$.
- (iii) (Linearität) Sind $\alpha, \beta \in [0, \infty]$, so gilt

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu,$$

wobei wir die Konvention $\infty \cdot 0 := 0$ benutzen.

Beweis. (i) Dies folgt direkt aus der Definition des Integrals.

(ii) Nach (i) gilt

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \sup_{n \in \mathbb{N}} \int f_n d\mu \leq \int f d\mu.$$

Wir müssen also nur noch $\int f d\mu \leq \sup_{n \in \mathbb{N}} \int f_n d\mu$ zeigen.

Sei $g \in \mathbb{E}^+$ mit $g \leq f$. Es reicht zu zeigen, dass

$$\sup_{n \in \mathbb{N}} \int f_n d\mu \geq \int g d\mu. \quad (4.2)$$

Die Elementarfunktion g habe die Normaldarstellung $g = \sum_{i=1}^N \alpha_i \mathbb{1}_{A_i}$, wobei $\alpha_1, \dots, \alpha_N \in (0, \infty)$ sind und $A_1, \dots, A_N \in \mathcal{A}$ paarweise disjunkt sind. Für jedes $\varepsilon > 0$ und $n \in \mathbb{N}$ definieren wir die Menge

$$B_n^\varepsilon = \{f_n \geq (1 - \varepsilon)g\}.$$

Wegen $f_n \uparrow f \geq g$ gilt $B_n^\varepsilon \uparrow \Omega$ für jedes $\varepsilon > 0$. Also gilt nach (i) für $\varepsilon > 0$

$$\begin{aligned} \int f_n d\mu &\geq \int ((1 - \varepsilon)g \mathbb{1}_{B_n^\varepsilon}) d\mu \\ &= \sum_{i=1}^N (1 - \varepsilon) \alpha_i \mu(A_i \cap B_n^\varepsilon) \\ &\xrightarrow{n \rightarrow \infty} \sum_{i=1}^N (1 - \varepsilon) \alpha_i \mu(A_i) = (1 - \varepsilon) \int g d\mu. \end{aligned}$$

Da $\varepsilon > 0$ beliebig war, folgt (4.2) und damit die Aussage (ii).

(iii) Nach Satz 1.96 ist jede nichtnegative messbare Abbildung monotoner Limes von Elementarfunktionen. Es gibt also Folgen $(f_n)_{n \in \mathbb{N}}$ und $(g_n)_{n \in \mathbb{N}}$ in \mathbb{E}^+ mit

$f_n \uparrow f$ und $g_n \uparrow g$. Es gilt dann aber auch $(\alpha f_n + \beta g_n) \uparrow \alpha f + \beta g$. Nach (ii) und Lemma 4.3 gilt daher

$$\begin{aligned} \int (\alpha f + \beta g) d\mu &= \lim_{n \rightarrow \infty} \int (\alpha f_n + \beta g_n) d\mu \\ &= \alpha \lim_{n \rightarrow \infty} \int f_n d\mu + \beta \lim_{n \rightarrow \infty} \int g_n d\mu = \alpha \int f d\mu + \beta \int g d\mu. \quad \square \end{aligned}$$

Für messbares $f : \Omega \rightarrow \overline{\mathbb{R}}$ ist $f^+ \leq |f|$ und $f^- \leq |f|$, also gilt auch $\int f^\pm d\mu \leq \int |f| d\mu$. Ist speziell $\int |f| d\mu < \infty$, so ist auch $\int f^- d\mu < \infty$ und $\int f^+ d\mu < \infty$. Daher können wir die folgende Definition treffen, die abschließend das Integral für messbare Funktionen erklärt.

Definition 4.7 (Integral für messbare Funktionen). Eine messbare Funktion $f : \Omega \rightarrow \overline{\mathbb{R}}$ heißt **μ -integrierbar**, falls $\int |f| d\mu < \infty$. Wir schreiben

$$\mathcal{L}^1(\mu) := \mathcal{L}^1(\Omega, \mathcal{A}, \mu) := \left\{ f : \Omega \rightarrow \overline{\mathbb{R}} : f \text{ ist messbar und } \int |f| d\mu < \infty \right\}.$$

Für $f \in \mathcal{L}^1(\mu)$ definieren wir das Integral von f bezüglich μ durch

$$\int f(\omega) \mu(d\omega) := \int f d\mu := \int f^+ d\mu - \int f^- d\mu. \quad (4.3)$$

Ist lediglich $\int f^- d\mu < \infty$ oder $\int f^+ d\mu < \infty$, so definieren wir ebenfalls $\int f d\mu$ durch (4.3), wobei wir dann die Werte $+\infty$ beziehungsweise $-\infty$ zulassen.

Ist $A \in \mathcal{A}$, so schreiben wir $\int_A f d\mu := \int (f \mathbb{1}_A) d\mu$.

Satz 4.8. Sei $f : \Omega \rightarrow [0, \infty]$ messbar.

(i) Es ist $f = 0$ fast überall genau dann, wenn $\int f d\mu = 0$ gilt.

(ii) Ist $\int f d\mu < \infty$, so gilt $f < \infty$ fast überall.

Beweis. (i) „ \Rightarrow “ Es gelte $f = 0$ fast überall. Sei $N = \{\omega : f(\omega) > 0\}$. Dann ist $f \leq \infty \cdot \mathbb{1}_N$ und $n \mathbb{1}_N \uparrow \infty \cdot \mathbb{1}_N$. Es folgt nach Lemma 4.6(i) und (ii)

$$0 \leq \int f d\mu \leq \int (\infty \cdot \mathbb{1}_N) d\mu = \lim_{n \rightarrow \infty} \int n \mathbb{1}_N d\mu = 0.$$

„ \Leftarrow “ Sei $N_n = \{f \geq \frac{1}{n}\}$, $n \in \mathbb{N}$. Dann ist $N_n \uparrow N$ und

$$0 = \int f d\mu \geq \int \frac{1}{n} \mathbb{1}_{N_n} d\mu = \frac{\mu(N_n)}{n}.$$

Also gilt $\mu(N_n) = 0$ für jedes $n \in \mathbb{N}$ und damit $\mu(N) = 0$.

(ii) Sei $A = \{\omega : f(\omega) = \infty\}$. Für $n \in \mathbb{N}$ ist $\frac{1}{n}f \mathbb{1}_{\{f \geq n\}} \geq \mathbb{1}_{\{f \geq n\}}$, also nach Lemma 4.6(i)

$$\mu(A) = \int \mathbb{1}_A d\mu \leq \int \mathbb{1}_{\{f \geq n\}} d\mu \leq \frac{1}{n} \int f \mathbb{1}_{\{f \geq n\}} d\mu \leq \frac{1}{n} \int f d\mu \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Satz 4.9 (Eigenschaften des Integrals). Seien $f, g \in \mathcal{L}^1(\mu)$.

(i) (Monotonie) Ist $f \leq g$ fast überall, so ist $\int f d\mu \leq \int g d\mu$.

Ist speziell $f = g$ fast überall, so ist $\int f d\mu = \int g d\mu$.

(ii) (Dreiecksungleichung) Es gilt stets $|\int f d\mu| \leq \int |f| d\mu$.

(iii) (Linearität) Sind $\alpha, \beta \in \mathbb{R}$, dann ist $\alpha f + \beta g \in \mathcal{L}^1(\mu)$ und

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

Diese Gleichung gilt auch, wenn höchstens eines der Integrale $\int f d\mu$ und $\int g d\mu$ einen der Werte $\pm\infty$ annimmt.

Beweis. (i) Es gilt $f^+ \leq g^+$ und $f^- \geq g^-$ f.ü., also ist nach Lemma 4.6(i)

$$\int f^+ d\mu \leq \int g^+ d\mu \quad \text{und} \quad \int f^- d\mu \geq \int g^- d\mu.$$

Es folgt

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu \leq \int g^+ d\mu - \int g^- d\mu = \int g d\mu.$$

(ii) Wegen $f^+ + f^- = |f|$ ist nach Lemma 4.6(iii)

$$\begin{aligned} \left| \int f d\mu \right| &= \left| \int f^+ d\mu - \int f^- d\mu \right| \leq \int f^+ d\mu + \int f^- d\mu \\ &= \int (f^+ + f^-) d\mu = \int |f| d\mu. \end{aligned}$$

(iii) Wegen $|\alpha f + \beta g| \leq |\alpha| \cdot |f| + |\beta| \cdot |g|$ ist nach Lemma 4.6(i) und (iii) auch $\alpha f + \beta g \in \mathcal{L}^1(\mu)$. Um die Linearität zu zeigen, reicht es die drei folgenden Eigenschaften zu prüfen.

(a) $\int(f + g) d\mu = \int f d\mu + \int g d\mu$.

(b) Für $\alpha \geq 0$ ist $\int \alpha f d\mu = \alpha \int f d\mu$.

(c) $\int(-f) d\mu = -\int f d\mu$.

Zu (a): Es ist $(f + g)^+ - (f + g)^- = f + g = f^+ - f^- + g^+ - g^-$, also ist $(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+$. Nach Lemma 4.6(iii) gilt

$$\int (f + g)^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int (f + g)^- d\mu + \int f^+ d\mu + \int g^+ d\mu,$$

also ist

$$\begin{aligned} \int (f + g) d\mu &= \int (f + g)^+ d\mu - \int (f + g)^- d\mu \\ &= \int f^+ d\mu - \int f^- d\mu + \int g^+ d\mu - \int g^- d\mu \\ &= \int f d\mu + \int g d\mu. \end{aligned}$$

Zu (b): Für $\alpha \geq 0$ ist

$$\int \alpha f d\mu = \int \alpha f^+ d\mu - \int \alpha f^- d\mu = \alpha \int f^+ d\mu - \alpha \int f^- d\mu = \alpha \int f d\mu.$$

Zu (c): Es ist

$$\begin{aligned} \int (-f) d\mu &= \int (-f)^+ d\mu - \int (-f)^- d\mu \\ &= \int f^- d\mu - \int f^+ d\mu = - \int f d\mu. \end{aligned}$$

Der Zusatz ist simpel und verbleibt zur Übung. \square

Satz 4.10 (Bildmaß). Seien (Ω, \mathcal{A}) und (Ω', \mathcal{A}') Messräume, μ ein Maß auf (Ω, \mathcal{A}) und $X : \Omega \rightarrow \Omega'$ messbar. Sei $\mu' = \mu \circ X^{-1}$ das Bildmaß von μ unter X und $f : \Omega' \rightarrow \mathbb{R}$ integrierbar bezüglich μ' . Dann ist $f \circ X \in \mathcal{L}^1(\mu)$ und

$$\int (f \circ X) d\mu = \int f d(\mu \circ X^{-1}).$$

Ist speziell X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbf{P})$, so ist

$$\int f(x) \mathbf{P}[X \in dx] := \int f(x) \mathbf{P}_X[dx] = \int f d\mathbf{P}_X = \int f(X(\omega)) \mathbf{P}[d\omega].$$

Beweis. Übung! \square

Beispiel 4.11 (Diskreter Maßraum). Sei (Ω, \mathcal{A}) ein diskreter Messraum und $\mu = \sum_{\omega \in \Omega} \alpha_\omega \delta_\omega$ für gewisse Zahlen $\alpha_\omega \geq 0$, $\omega \in \Omega$. Eine Abbildung $f : \Omega \rightarrow \mathbb{R}$ ist genau dann integrierbar, wenn $\sum_{\omega \in \Omega} |f(\omega)| \alpha_\omega < \infty$ ist. In diesem Fall gilt

$$\int f d\mu = \sum_{\omega \in \Omega} f(\omega) \alpha_\omega. \quad \diamond$$

Definition 4.12 (Lebesgue-Integral). Sei λ das Lebesgue-Maß auf \mathbb{R}^n und $f : \mathbb{R}^n \rightarrow \mathbb{R}$ messbar bezüglich $\mathcal{B}^*(\mathbb{R}^n) - \mathcal{B}(\mathbb{R})$ (wobei $\mathcal{B}^*(\mathbb{R}^n)$ die Lebesgue'sche σ -Algebra ist, siehe Beispiel 1.71) und λ -integrierbar. Dann nennen wir

$$\int f d\lambda$$

das **Lebesgue-Integral** von f . Ist $A \in \mathcal{B}(\mathbb{R}^n)$ und $f : \mathbb{R}^n \rightarrow \mathbb{R}$ messbar (oder $f : A \rightarrow \mathbb{R}$ messbar bezüglich $\mathcal{B}^*(\mathbb{R}^n)|_A - \mathcal{B}(\mathbb{R})$ und damit $f \mathbf{1}_A$ messbar bezüglich $\mathcal{B}^*(\mathbb{R}^n) - \mathcal{B}(\mathbb{R})$), so schreiben wir

$$\int_A f d\lambda := \int f \mathbf{1}_A d\lambda.$$

Definition 4.13. Sei μ ein Maß auf (Ω, \mathcal{A}) und $f : \Omega \rightarrow [0, \infty)$ messbar. Wir sagen, dass das durch

$$\nu(A) := \int (\mathbf{1}_A f) d\mu \quad \text{für } A \in \mathcal{A}$$

definierte Maß $f\mu := \nu$ die **Dichte** f bezüglich μ hat.

Bemerkung 4.14. Wir müssen noch zeigen, dass ν ein Maß ist und prüfen hierzu die Bedingung von Satz 1.36 nach. Offenbar ist $\nu(\emptyset) = 0$. Endliche Additivität folgt aus der Additivität des Integrals (Lemma 4.6(iii)) und Stetigkeit von unten aus dem Satz von der monotonen Konvergenz (Satz 4.20). \diamond

Satz 4.15. Es ist $g \in \mathcal{L}^1(f\mu)$ genau dann, wenn $(gf) \in \mathcal{L}^1(\mu)$. In diesem Fall gilt

$$\int g d(f\mu) = \int (gf) d\mu.$$

Beweis. Die Aussage gilt zunächst für Indikatorfunktionen und wird dann mit den üblichen Argumenten auf Elementarfunktionen, nichtnegative Funktionen sowie schließlich auf messbare Funktionen fortgesetzt. \square

Definition 4.16. Für messbares $f : \Omega \rightarrow \overline{\mathbb{R}}$ definieren wir

$$\|f\|_p := \left(\int |f|^p d\mu \right)^{1/p}, \quad \text{falls } p \in [1, \infty),$$

und

$$\|f\|_\infty := \inf \{K \geq 0 : \mu(\{|f| > K\}) = 0\}.$$

Ferner definieren wir für jedes $p \in [1, \infty]$ den Vektorraum

$$\mathcal{L}^p(\mu) := \left\{ f : \Omega \rightarrow \overline{\mathbb{R}} \text{ ist messbar und } \|f\|_p < \infty \right\}.$$

Satz 4.17. Die Abbildung $\|\cdot\|_1$ ist eine Pseudonorm auf $\mathcal{L}^1(\mu)$, das heißt, es gilt für $f, g \in \mathcal{L}^1(\mu)$ und $\alpha \in \mathbb{R}$

$$\begin{aligned}\|\alpha f\|_1 &= |\alpha| \cdot \|f\|_1 \\ \|f + g\|_1 &\leq \|f\|_1 + \|g\|_1 \\ \|f\|_1 &\geq 0 \text{ für alle } f \text{ und } \|f\|_1 = 0, \text{ falls } f = 0 \text{ f.ü.}\end{aligned}\tag{4.4}$$

Beweis. Die erste und dritte Aussage folgen aus Satz 4.9(iii) und Satz 4.8(i). Die zweite folgt aus Satz 4.9(i), denn es ist $|f + g| \leq |f| + |g|$, also

$$\|f + g\|_1 = \int |f + g| d\mu \leq \int |f| d\mu + \int |g| d\mu = \|f\|_1 + \|g\|_1. \quad \square$$

Bemerkung 4.18. Tatsächlich ist $\|\cdot\|_p$ für jedes $p \in [1, \infty]$ eine Pseudonorm auf $\mathcal{L}^p(\mu)$. Linearität und Positivität sind klar, und die Dreiecksungleichung ist die Minkowski'sche Ungleichung, die wir in Satz 7.17 zeigen werden. ◇

Satz 4.19. Seien $\mu(\Omega) < \infty$ und $1 \leq p' \leq p \leq \infty$. Dann ist $\mathcal{L}^p(\mu) \subset \mathcal{L}^{p'}(\mu)$, und die kanonische Inklusion $i : \mathcal{L}^p(\mu) \hookrightarrow \mathcal{L}^{p'}(\mu)$, $f \mapsto f$ ist stetig.

Beweis. Sei $f \in \mathcal{L}^\infty(\mu)$ und $p' \in [1, \infty)$. Dann ist $|f|^{p'} \leq \|f\|_\infty^{p'}$ fast überall, also

$$\int |f|^{p'} d\mu \leq \int \|f\|_\infty^{p'} d\mu = \|f\|_\infty^{p'} \cdot \mu(\Omega) < \infty.$$

Für $f, g \in \mathcal{L}^\infty(\mu)$ ist also $\|f - g\|_{p'} \leq \mu(\Omega)^{1/p'} \|f - g\|_\infty$ und damit ist i stetig.

Seien nun $p, p' \in [1, \infty)$ mit $p' < p$ sowie $f \in \mathcal{L}^p(\mu)$. Dann ist $|f|^{p'} \leq 1 + |f|^p$, also gilt

$$\int |f|^{p'} d\mu \leq \mu(\Omega) + \int |f|^p d\mu < \infty.$$

Seien schließlich $f, g \in \mathcal{L}^p(\mu)$. Für jedes $c > 0$ ist

$$|f - g|^{p'} = |f - g|^{p'} \mathbb{1}_{\{|f - g| \leq c\}} + |f - g|^{p'} \mathbb{1}_{\{|f - g| > c\}} \leq c^{p'} + c^{p'-p} |f - g|^p.$$

Speziell erhalten wir mit $c = \|f - g\|_p$

$$\|f - g\|_{p'} \leq \left(c^{p'} \mu(\Omega) + c^{p'-p} \|f - g\|_p^p \right)^{1/p'} = (1 + \mu(\Omega))^{1/p'} \|f - g\|_p.$$

Also ist i auch in diesem Falle stetig. ◻

Übung 4.1.1 (Folgenräume). Wir nehmen jetzt nicht mehr an, dass $\mu(\Omega) < \infty$ ist. Man zeige: Gibt es ein $a > 0$, sodass für jedes $A \in \mathcal{A}$ entweder $\mu(A) = 0$ oder $\mu(A) \geq a$ gilt, so gilt die zu Satz 4.19 umgekehrte Inklusion

$$\mathcal{L}^{p'}(\mu) \subset \mathcal{L}^p(\mu), \quad \text{falls } 1 \leq p' \leq p \leq \infty. \tag{4.5}$$



Übung 4.1.2. Sei $1 \leq p' < p \leq \infty$ und μ σ -endlich, aber nicht endlich. Man zeige, dass $\mathcal{L}^p(\mu) \setminus \mathcal{L}^{p'}(\mu) \neq \emptyset$. ♣

4.2 Monotone Konvergenz und Lemma von Fatou

Unter welchen Bedingungen ist die Vertauschbarkeit von Limes und Integral zulässig? Wir geben hier zwei einfache Kriterien für die Vertauschbarkeit an, die uns wichtige Anwendungen wie Gesetze der großen Zahl (Kapitel 5) ermöglichen. Allgemeinere Kriterien werden in Kapitel 6 hergeleitet.

Satz 4.20 (Monotone Konvergenz, Satz von Beppo Levi). Seien $f_1, f_2, \dots \in \mathcal{L}^1(\mu)$ und $f : \Omega \rightarrow \overline{\mathbb{R}}$ messbar. Es gelte $f_n \uparrow f$ f.ü. für $n \rightarrow \infty$. Dann gilt

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu,$$

wobei beide Seiten den Wert $+\infty$ annehmen können.

Beweis. Sei $N \subset \Omega$ eine Nullmenge, sodass $f_n(\omega) \uparrow f(\omega)$ für alle $\omega \in N^c$. Die Funktionen $f'_n := (f_n - f_1) \mathbb{1}_{N^c}$ und $f' := (f - f_1) \mathbb{1}_{N^c}$ sind nichtnegativ und erfüllen $f'_n \uparrow f'$. Nach Lemma 4.6(ii) gilt $\int f'_n d\mu \xrightarrow{n \rightarrow \infty} \int f' d\mu$. Wegen $f_n = f'_n + f_1$ f.ü. und $f = f' + f_1$ f.ü. folgt nach Satz 4.9(iii)

$$\int f_n d\mu = \int f_1 d\mu + \int f'_n d\mu \xrightarrow{n \rightarrow \infty} \int f_1 d\mu + \int f' d\mu = \int f d\mu. \quad \square$$

Satz 4.21 (Lemma von Fatou). Seien $f \in \mathcal{L}^1(\mu)$ und f_1, f_2, \dots messbar mit $f_n \geq f$ f.ü. für jedes $n \in \mathbb{N}$. Dann gilt

$$\int \left(\liminf_{n \rightarrow \infty} f_n \right) d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu.$$

Beweis. Indem wir $(f_n - f)_{n \in \mathbb{N}}$ betrachten, können wir annehmen, dass $f_n \geq 0$ f.ü. für jedes $n \in \mathbb{N}$. Setze

$$g_n := \inf_{m \geq n} f_m.$$

Dann gilt $g_n \uparrow \liminf_{m \rightarrow \infty} f_m$, also nach dem Satz von der monotonen Konvergenz (Lemma 4.6(ii)) sowie der Monotonie $g_n \leq f_n$ (also $\int g_n d\mu \leq \int f_n d\mu$)

$$\int \liminf_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int g_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu. \quad \square$$

Beispiel 4.22 (Petersburger Spiel). Wir wollen durch ein Beispiel zeigen, dass auf die Voraussetzung der Existenz einer integrierbaren Minorante im Lemma von Fatou nicht verzichtet werden kann. Wir betrachten ein Glücksspiel in einem Casino, bei dem in jeder Runde ein vom Spieler gewählter Einsatz entweder verdoppelt zurückgezahlt wird oder verloren geht. Dies ist etwa beim Roulette der Fall, wo der Spieler zum Beispiel auf „Rot“ setzen kann. Kommt eine rote Zahl, so gewinnt der Spieler seinen Einsatz verdoppelt zurück, ansonsten verliert er ihn. Es gibt 37 Felder, von denen 18 rot sind, 18 schwarz und eines grün (die Null). Die Gewinnchance sollte also $p = \frac{18}{37} < \frac{1}{2}$ betragen. Dieses Glücksspiel werde unendlich oft unabhängig hintereinander ausgeführt. Wir können es also auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$ realisieren, wobei $\Omega = \{-1, 1\}^{\mathbb{N}}$, $\mathcal{A} = (2^{\{-1, 1\}})^{\otimes \mathbb{N}}$ die von den Zylin dern $[\omega_1, \dots, \omega_n]$ erzeugte σ -Algebra ist und $\mathbf{P} = ((1-p)\delta_{-1} + p\delta_1)^{\otimes \mathbb{N}}$ das Produktmaß. Wir bezeichnen mit $D_n : \Omega \rightarrow \{-1, 1\}$, $\omega \mapsto \omega_n$ das Ergebnis der n -ten Runde für jedes $n \in \mathbb{N}$. Macht der Spieler in der i -ten Runde den (zufälligen) Einsatz H_i , so beträgt die Summe der Gewinne nach der n -ten Runde

$$S_n = \sum_{i=1}^n H_i D_i.$$

Wir nehmen nun an, dass der Spieler die folgende Strategie verfolgt: In der ersten Runde ist der Einsatz $H_1 = 1$. Gewinnt er, so setzt er in den folgenden Spielen gar nicht mehr, also ist $H_n = 0$ für jedes $n \geq 2$, falls $D_1 = 1$. Verliert er hingegen, so setzt er in der zweiten Runde den doppelten Einsatz, also ist $H_2 = 2$, falls $D_1 = -1$. Gibt die zweite Runde einen Gewinn, so setzt er ab der dritten Runde gar nicht mehr, andernfalls verdoppelt er wiederum seinen Einsatz in der dritten Runde und so weiter. Wir erhalten also als Strategie

$$H_n = \begin{cases} 0, & \text{falls es ein } i \in \{1, \dots, n-1\} \text{ gibt mit } D_i = 1, \\ 2^{n-1}, & \text{sonst.} \end{cases}$$

Man beachte, dass H_n nur von D_1, \dots, D_{n-1} abhängt, also messbar ist bezüglich $\sigma(D_1, \dots, D_{n-1})$. Dies ist offenbar ein wichtige Forderung an jede Spielstrategie, da man die Entscheidung über den Einsatz aufgrund der vorhandenen Kenntnis zum jeweiligen Zeitpunkt treffen muss und nicht in die Zukunft blicken kann.

Die Wahrscheinlichkeit, dass bis zum Zeitpunkt n kein Spiel gewonnen wurde ist $(1-p)^n$, also ist $\mathbf{P}[S_n = 1 - 2^n] = (1-p)^n$ und $\mathbf{P}[S_n = 1] = 1 - (1-p)^n$. Man erwartet also im Mittel einen Gewinn von

$$\int S_n d\mathbf{P} = (1-p)^n(1 - 2^n) + (1 - (1-p)^n) = 1 - (2(1-p))^n \leq 0,$$

da $p \leq \frac{1}{2}$ ist (in den profitablen Spielbanken). Wir setzen nun

$$S = \begin{cases} -\infty, & \text{falls } -1 = D_1 = D_2 = \dots, \\ 1, & \text{sonst.} \end{cases}$$

Dann gilt $S_n \xrightarrow{n \rightarrow \infty} S$ f.s., jedoch ist $\lim_{n \rightarrow \infty} \int S_n d\mathbf{P} < \int S d\mathbf{P} = 1$, weil $S = 1$ fast sicher gilt. Nach dem Lemma von Fatou ist dies nur möglich, wenn es keine integrierbare Minorante zur Folge $(S_n)_{n \in \mathbb{N}}$ gibt. Setzen wir $\tilde{S} := \inf\{S_n : n \in \mathbb{N}\}$, so gilt in der Tat

$$\mathbf{P}[\tilde{S} = 1 - 2^{n-1}] = \mathbf{P}[D_1 = \dots = D_{n-1} = -1 \text{ und } D_n = 1] = p(1-p)^{n-1},$$

also $\int \tilde{S} d\mathbf{P} = \sum_{n=1}^{\infty} (1 - 2^{n-1}) p(1-p)^{n-1} = -\infty$, weil $p \leq \frac{1}{2}$. \diamond

Übung 4.2.1. Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f \in \mathcal{L}^1(\mu)$. Man zeige: Zu jedem $\varepsilon > 0$ gibt es ein $A \in \mathcal{A}$ mit $\mu(A) < \infty$ und $|\int_A f d\mu - \int f d\mu| < \varepsilon$. \clubsuit

Übung 4.2.2. Seien $f_1, f_2, \dots \in \mathcal{L}^1(\mu)$ nichtnegativ und so, dass $\lim_{n \rightarrow \infty} \int f_n d\mu$ existiert. Es gebe ein messbares f mit $f_n \xrightarrow{n \rightarrow \infty} f$ μ -fast überall. Man zeige: Es gilt $f \in \mathcal{L}^1(\mu)$ und

$$\lim_{n \rightarrow \infty} \int |f_n - f| d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu - \int f d\mu. \quad \clubsuit$$

Übung 4.2.3. Sei $f \in \mathcal{L}^1([0, \infty), \lambda)$ eine Lebesgue-integrierbare Funktion auf $[0, \infty)$. Man zeige, dass für λ -fast alle $t \in [0, \infty)$ die Reihe $\sum_{n=1}^{\infty} f(nt)$ absolut konvergiert. \clubsuit

Übung 4.2.4. Sei λ das Lebesgue-Maß auf \mathbb{R} und A eine Borel'sche Menge mit $\lambda(A) < \infty$. Man zeige: Für jedes $\varepsilon > 0$ gibt es eine kompakte Menge $C \subset A$, eine abgeschlossene Menge $D \subset \mathbb{R} \setminus A$ und eine stetige Abbildung $\varphi : \mathbb{R} \rightarrow [0, 1]$ mit $\mathbb{1}_C \leq \varphi \leq \mathbb{1}_{\mathbb{R} \setminus D}$, sodass $\|\mathbb{1}_A - \varphi\|_1 < \varepsilon$.

Hinweis: Man verwende die Regularität des Lebesgue-Maßes (Bemerkung 1.67). \clubsuit

Übung 4.2.5. Sei λ das Lebesgue-Maß auf \mathbb{R} , $p \in [1, \infty)$ und $f \in \mathcal{L}^p(\lambda)$. Man zeige: Für jedes $\varepsilon > 0$ existiert eine stetige Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$ mit $\|f - h\|_p < \varepsilon$.

Hinweis: Verwende Übung 4.2.4 um die Aussage zunächst für Indikatorfunktionen messbarer Mengen zu zeigen, dann für einfache Funktionen und schließlich für allgemeines $f \in \mathcal{L}^p(\lambda)$. \clubsuit

Übung 4.2.6. Sei λ das Lebesgue-Maß auf \mathbb{R} , $p \in [1, \infty)$ und $f \in \mathcal{L}^p(\lambda)$. Eine Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$ heißt **Treppenfunktion**, wenn es Zahlen $n \in \mathbb{N}$, $t_0 < t_1 < \dots < t_n$ und $\alpha_1, \dots, \alpha_n$ gibt mit $h = \sum_{k=1}^n \alpha_k \mathbb{1}_{(t_{k-1}, t_k]}$.

Man zeige: Für jedes $\varepsilon > 0$ existiert eine Treppenfunktion h mit $\|f - h\|_p < \varepsilon$.

Hinweis: Man zeige die Aussage zunächst für messbare Indikatorfunktionen mit Hilfe des Approximationssatzes für Maße (Satz 1.65), angewandt auf den Semiring der linksoffenen beschränkten Intervalle. Danach verwende man die Approximationssargumente wie in Übung 4.2.5. \clubsuit

4.3 Lebesgue-Integral versus Riemann-Integral

Wir zeigen, dass für Riemann-integrierbare Funktionen das Lebesgue-Integral und das Riemann-Integral übereinstimmen.

Sei $I = [a, b] \subset \mathbb{R}$ ein Intervall und λ das Lebesgue-Maß auf I . Ferner betrachten wir Folgen $t = (t^n)_{n \in \mathbb{N}}$ von Partitionen $t^n = (t_i^n)_{i=0, \dots, n}$ von I , also $a = t_0^n < t_1^n < \dots < t_n^n = b$, die immer feiner werden, das heißt, mit $\max\{t_i^n - t_{i-1}^n : i = 1, \dots, n\} \xrightarrow{n \rightarrow \infty} 0$. Wir nehmen an, dass für jedes $n \in \mathbb{N}$ die Partition t^{n+1} eine Verfeinerung von t^n ist, dass also gilt $\{t_0^n, \dots, t_n^n\} \subset \{t_0^{n+1}, \dots, t_{n+1}^{n+1}\}$.

Für jede Funktion $f : I \rightarrow \mathbb{R}$ und jedes $n \in \mathbb{N}$ definieren wir die n -te Untersumme beziehungsweise Obersumme durch

$$U_n^t(f) := \sum_{i=1}^n (t_i^n - t_{i-1}^n) \inf f([t_{i-1}^n, t_i^n]),$$

$$O_n^t(f) := \sum_{i=1}^n (t_i^n - t_{i-1}^n) \sup f([t_{i-1}^n, t_i^n]).$$

Bekanntlich heißt eine Funktion $f : I \rightarrow \mathbb{R}$ Riemann-integrierbar, wenn es ein t gibt, sodass die Limiten von Obersummen und Untersummen übereinstimmen und endlich sind. Der Grenzwert hängt dann nicht vom gewählten t ab, und man definiert

$$\int_a^b f(x) dx := \lim_{n \rightarrow \infty} U_n^t(f) = \lim_{n \rightarrow \infty} O_n^t(f) \quad (4.6)$$

als das Riemann-Integral von f (siehe etwa [158]).

Satz 4.23 (Riemann-Integral und Lebesgue-Integral). Sei $f : I \rightarrow \mathbb{R}$ Riemann-integrierbar auf $I = [a, b]$. Dann ist f Lebesgue-integrierbar auf I mit Integral

$$\int_I f d\lambda = \int_a^b f(x) dx.$$

Beweis. Sei t so gewählt, dass (4.6) gilt. Nach Voraussetzung gibt es ein $n \in \mathbb{N}$ mit $|U_n^t(f)| < \infty$ und $|O_n^t(f)| < \infty$. Also ist f beschränkt. Indem wir f durch $f + \|f\|_\infty$ ersetzen, können wir annehmen, dass $f \geq 0$ gilt. Setze

$$g_n := f(b) \mathbb{1}_{\{b\}} + \sum_{i=1}^n (\inf f([t_{i-1}^n, t_i^n])) \mathbb{1}_{[t_{i-1}^n, t_i^n]},$$

$$h_n := f(b) \mathbb{1}_{\{b\}} + \sum_{i=1}^n (\sup f([t_{i-1}^n, t_i^n])) \mathbb{1}_{[t_{i-1}^n, t_i^n]}.$$

Da t^{n+1} eine Verfeinerung von t^n ist, gilt $g_n \leq g_{n+1} \leq h_{n+1} \leq h_n$. Also existieren g und h mit $g_n \uparrow g$ und $h_n \downarrow h$. Nach Konstruktion gilt $g \leq h$ und

$$\begin{aligned}\int_I g d\lambda &= \lim_{n \rightarrow \infty} \int_I g_n d\lambda = \lim_{n \rightarrow \infty} U_n^t(f) \\ &= \lim_{n \rightarrow \infty} O_n^t(f) = \lim_{n \rightarrow \infty} \int_I h_n d\lambda = \int_I h d\lambda.\end{aligned}$$

Also ist λ -fast überall $h = g$. Nach Konstruktion ist $g \leq f \leq h$, und g und h sind als Limiten von Elementarfunktionen messbar bezüglich $\mathcal{B}(I) - \mathcal{B}(\mathbb{R})$. Es folgt, dass für jedes $\alpha \in \mathbb{R}$

$$\{f \leq \alpha\} = (\{g \leq \alpha\} \cap \{g = h\}) \uplus (\{f \leq \alpha\} \cap \{g \neq h\})$$

die Vereinigung einer $\mathcal{B}(I)$ -Menge mit einer Teilmenge einer Nullmenge ist, also in $\mathcal{B}(I)^*$ (der Lebesgue'schen Vervollständigung von $\mathcal{B}(I)$) liegt. Mithin ist f messbar bezüglich $\mathcal{B}(I)^*$. Nach dem Satz über monotone Konvergenz (Satz 4.20) ist

$$\int_I f d\lambda = \lim_{n \rightarrow \infty} \int_I g_n d\lambda = \int_a^b f(x) dx. \quad \square$$

Beispiel 4.24. Sei $f : [0, 1] \rightarrow \mathbb{R}$, $x \mapsto \mathbb{1}_{\mathbb{Q}}$. Dann ist f offenbar nicht Riemann-integrierbar, weil $U_n(f) = 0$ und $O_n(f) = 1$ für jedes $n \in \mathbb{N}$. Andererseits ist f Lebesgue-integrierbar mit Integral $\int_{[0,1]} f d\lambda = 0$, denn $\mathbb{Q} \cap [0, 1]$ ist eine Nullmenge. ◇

Bemerkung 4.25. Eine uneigentlich Riemann-integrierbare Funktion f auf einem halboffenen Intervall $I = (a, b]$ oder $I = [0, \infty)$ ist nicht notwendigerweise auch Lebesgue-integrierbar. Hier wird nämlich das uneigentliche Integral $\int_0^\infty f(x) dx := \lim_{n \rightarrow \infty} \int_0^n f(x) dx$ durch eine Grenzwertprozedur definiert, die Rücksicht auf die Geometrie von \mathbb{R} nimmt. Dies tut das Lebesgue-Integral nicht. So ist die Funktion $f : [0, \infty) \rightarrow \mathbb{R}$, $x \mapsto \frac{1}{1+x} \sin(x)$ (uneigentlich) Riemann-integrierbar, jedoch nicht Lebesgue-integrierbar, weil $\int_{[0, \infty)} |f| d\lambda = \infty$ ist. ◇

Wir haben schon gesehen, dass uneigentlich Riemann-integrierbare Funktionen nicht notwendigerweise auch Lebesgue-integrierbar sind. Andererseits gibt es Lebesgue-integrierbare Funktionen, die nicht Riemann-integrierbar sind (wie etwa $\mathbb{1}_{\mathbb{Q}}$). Geometrisch lässt sich dies so interpretieren, dass das Riemann-Integral die Geometrie des Integrationsbereiches respektiert, indem es als Grenzwert von Flächen schmaler senkrechter Streifen entsteht, während das Lebesgue-Integral als Grenzwert mit flachen waagerechten Streifen gedacht werden kann. Insbesondere macht dieses Integral gar keine Annahmen an den Definitionsbereich des Integranden, weshalb es eben universeller einsetzbar ist. Um dies zu unterstreichen, bringen wir einen Satz, der uns auch ansonsten noch nützlich sein wird.

Satz 4.26. Sei $f : \Omega \rightarrow \mathbb{R}$ messbar und $f \geq 0$ fast überall. Dann gelten

$$\sum_{n=1}^{\infty} \mu(\{f \geq n\}) \leq \int f d\mu \leq \sum_{n=0}^{\infty} \mu(\{f > n\}) \quad (4.7)$$

und

$$\int f d\mu = \int_0^{\infty} \mu(\{f \geq t\}) dt. \quad (4.8)$$

Beweis. Setze $f' = \lfloor f \rfloor$ und $f'' = \lceil f \rceil$. Dann ist $f' \leq f \leq f''$ und deshalb $\int f' d\mu \leq \int f d\mu \leq \int f'' d\mu$. Nun ist

$$\begin{aligned} \int f' d\mu &= \sum_{k=1}^{\infty} \mu(\{f' = k\}) \cdot k = \sum_{k=1}^{\infty} \sum_{n=1}^k \mu(\{f' = k\}) \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mu(\{f' = k\}) \\ &= \sum_{n=1}^{\infty} \mu(\{f' \geq n\}) = \sum_{n=1}^{\infty} \mu(\{f \geq n\}). \end{aligned}$$

Analog ist

$$\int f'' d\mu = \sum_{n=1}^{\infty} \mu(\{f'' \geq n\}) = \sum_{n=1}^{\infty} \mu(\{f > n-1\}).$$

Hieraus folgt (4.7).

Gilt $g(t) := \mu(\{f \geq t\}) = \infty$ für ein $t > 0$, so sind beide Seiten in (4.8) gleich ∞ . Sei im Folgenden also $g(t) < \infty$ für alle $t > 0$.

Für $\varepsilon > 0$ und $k \in \mathbb{N}$ setze $g^\varepsilon := g \wedge g(\varepsilon)$, $f^\varepsilon := f \mathbf{1}_{\{f \geq \varepsilon\}}$ sowie $f_k^\varepsilon = 2^k f^\varepsilon$ und

$$\alpha_k^\varepsilon := 2^{-k} \sum_{n=1}^{\infty} \mu(\{f_k^\varepsilon \geq n 2^{-k}\}).$$

Dann gilt $\alpha_k^\varepsilon \xrightarrow{k \rightarrow \infty} \int_0^\infty g^\varepsilon(t) dt$. Ferner gilt nach (4.7) (mit f_k^ε statt f)

$$\begin{aligned} \alpha_k^\varepsilon &= 2^{-k} \sum_{n=1}^{\infty} \mu(\{f_k^\varepsilon \geq n\}) \leq \int f^\varepsilon d\mu \\ &\leq 2^{-k} \sum_{n=0}^{\infty} \mu(\{f_k^\varepsilon > n\}) = 2^{-k} \sum_{n=0}^{\infty} \mu(\{f^\varepsilon > n 2^{-k}\}) \leq \alpha_k^\varepsilon + 2^{-k} g(\varepsilon). \end{aligned}$$

Wegen $2^{-k} g(\varepsilon) \xrightarrow{k \rightarrow \infty} 0$ folgt $\int_0^\infty g^\varepsilon(t) dt = \int f^\varepsilon d\mu$. Wegen $f^\varepsilon \uparrow f$ und $g^\varepsilon \uparrow g$ für $\varepsilon \downarrow 0$ folgt (4.8) aus dem Satz über monotone Konvergenz. \square

Übung 4.3.1. Sei $f : [0, 1] \rightarrow \mathbb{R}$ beschränkt. Zeige: f ist genau dann (eigentlich) Riemann-integrierbar, wenn f λ -f.ü. stetig ist. ♣

Übung 4.3.2. Ist $f : [0, 1] \rightarrow \mathbb{R}$ Riemann-integrierbar, so ist f Lebesgue-messbar. Man zeige durch ein Beispiel, dass f nicht Borel-messbar sein muss. (*Hinweis:* Man verwende ohne Beweis die Existenz einer Teilmenge von $[0, 1]$, die nicht Borel-messbar ist und konstruiere hieraus eine nicht-Borel'sche Menge, deren Abschluss eine Nullmenge ist.) ♣

Übung 4.3.3. Sei $f : [0, 1] \rightarrow (0, \infty)$ Riemann integrierbar. Man zeige direkt (ohne das Lebesgue-Integral zu verwenden), dass $\int_0^1 f(x) dx > 0$ gilt. ♣

5 Momente und Gesetze der Großen Zahl

Die wichtigsten Kenngrößen für Zufallsvariablen sind Median, Erwartungswert und Varianz. Der Erwartungswert beschreibt für großes n den typischen ungefähren Wert des arithmetischen Mittels $(X_1 + \dots + X_n)/n$ von u.i.v. Zufallsvariablen (Gesetz der Großen Zahl). In Kapitel 15 werden wir sehen, wie die Varianz hingegen die typischen Abweichungen des arithmetischen Mittels vom Erwartungswert determiniert.

5.1 Momente

Im Folgenden sei $(\Omega, \mathcal{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum.

Definition 5.1. Sei X eine reelle Zufallsvariable.

(i) Ist $X \in \mathcal{L}^1(\mathbf{P})$, so heißt X **integrierbar**, und wir nennen

$$\mathbf{E}[X] := \int X d\mathbf{P}$$

den **Erwartungswert** von X . Ist $\mathbf{E}[X] = 0$, so heißt X **zentriert**. Etwas allgemeiner schreiben wir auch $\mathbf{E}[X] = \int X^- d\mathbf{P}$, falls nur X^- oder X^+ integrierbar ist.

(ii) Ist $n \in \mathbb{N}$ und $X \in \mathcal{L}^n(\mathbf{P})$, so heißen die Größen

$$m_k := \mathbf{E}[X^k], \quad M_k := \mathbf{E}[|X|^k] \quad \text{für jedes } k = 1, \dots, n,$$

die k -ten **Momente** beziehungsweise **absoluten Momente** von X .

(iii) Ist $X \in \mathcal{L}^2(\mathbf{P})$, so heißt X **quadratintegrierbar**, und wir nennen

$$\mathbf{Var}[X] := \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

die **Varianz** von X . Die Zahl $\sigma := \sqrt{\mathbf{Var}[X]}$ heißt die **Streuung** oder **Standardabweichung** von X . Formal setzen wir manchmal $\mathbf{Var}[X] = \infty$, falls $\mathbf{E}[X^2] = \infty$ ist.

(iv) Sind $X, Y \in \mathcal{L}^2(\mathbf{P})$, so definieren wir die **Kovarianz** von X und Y durch

$$\mathbf{Cov}[X, Y] := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])].$$

X und Y heißen **unkorreliert**, falls $\mathbf{Cov}[X, Y] = 0$ ist.

Bemerkung 5.2. (i) Die Definition in (ii) ist sinnvoll, denn für $X \in \mathcal{L}^n(\mathbf{P})$ ist nach Satz 4.19 $M_k < \infty$ für jedes $k = 1, \dots, n$.

(ii) Sind $X, Y \in \mathcal{L}^2(\mathbf{P})$, so ist wegen $|XY| \leq X^2 + Y^2$ auch $XY \in \mathcal{L}^1(\mathbf{P})$. Deshalb ist die Definition in (iv) sinnvoll, und es gilt

$$\mathbf{Cov}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].$$

Speziell ist $\mathbf{Var}[X] = \mathbf{Cov}[X, X]$. ◇

Wir fassen die wichtigsten Rechenregeln für Erwartungswerte als Satz zusammen. Alle aufgeführten Eigenschaften folgen direkt aus den Eigenschaften des Integrals.

Satz 5.3 (Rechenregeln für den Erwartungswert). Seien $X, Y, X_n, Z_n, n \in \mathbb{N}$, reelle integrierbare Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbf{P})$. Dann gilt

(i) Ist $\mathbf{P}_X = \mathbf{P}_Y$, so ist $\mathbf{E}[X] = \mathbf{E}[Y]$.

(ii) (Linearität) Sei $c \in \mathbb{R}$. Dann gelten $cX \in \mathcal{L}^1(\mathbf{P})$ und $X + Y \in \mathcal{L}^1(\mathbf{P})$ sowie

$$\mathbf{E}[cX] = c\mathbf{E}[X] \quad \text{und} \quad \mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y].$$

(iii) Ist $X \geq 0$ fast sicher, so gilt

$$\mathbf{E}[X] = 0 \iff X = 0 \text{ fast sicher.}$$

(iv) (Monotonie) Gilt $X \leq Y$ fast sicher, so gilt $\mathbf{E}[X] \leq \mathbf{E}[Y]$ mit Gleichheit genau dann, wenn $X = Y$ fast sicher.

(v) (Dreiecksungleichung) Es ist $|\mathbf{E}[X]| \leq \mathbf{E}[|X|]$.

(vi) Ist $X_n \geq 0$ fast sicher für jedes $n \in \mathbb{N}$, so ist $\mathbf{E}\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty} \mathbf{E}[X_n]$.

(vii) Gilt $Z_n \uparrow Z$ für ein Z , so gilt $\mathbf{E}[Z] = \lim_{n \rightarrow \infty} \mathbf{E}[Z_n] \in (-\infty, \infty]$.

Die Wahrscheinlichkeitstheorie beginnt wieder an der Stelle, wo die Unabhängigkeit ins Spiel kommt, wir also den Bereich der linearen Integrationstheorie verlassen.

Satz 5.4 (Unabhängige Zufallsvariablen sind unkorreliert).

Seien $X, Y \in \mathcal{L}^1(\mathbf{P})$ unabhängig. Dann ist $(XY) \in \mathcal{L}^1(\mathbf{P})$ und $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$. Speziell sind unabhängige Zufallsvariablen unkorreliert.

Beweis. Wir nehmen zunächst an, dass X und Y nur endlich viele Werte annehmen. Dann nimmt auch XY nur endlich viele Werte an, speziell ist offenbar $XY \in \mathcal{L}^1(\mathbf{P})$. Es folgt

$$\begin{aligned}\mathbf{E}[XY] &= \sum_{z \in \mathbb{R} \setminus \{0\}} z \mathbf{P}[XY = z] \\ &= \sum_{z \in \mathbb{R} \setminus \{0\}} \sum_{x \in \mathbb{R} \setminus \{0\}} x \frac{z}{x} \mathbf{P}[X = x, Y = z/x] \\ &= \sum_{y \in \mathbb{R} \setminus \{0\}} \sum_{x \in \mathbb{R} \setminus \{0\}} xy \mathbf{P}[X = x] \mathbf{P}[Y = y] \\ &= \left(\sum_{x \in \mathbb{R}} x \mathbf{P}[X = x] \right) \left(\sum_{y \in \mathbb{R}} y \mathbf{P}[Y = y] \right) \\ &= \mathbf{E}[X] \mathbf{E}[Y].\end{aligned}$$

Für $N \in \mathbb{N}$ sind auch die Zufallsvariablen $X_N := (2^{-N} \lfloor 2^N |X| \rfloor) \wedge N$ und $Y_N := (2^{-N} \lfloor 2^N |Y| \rfloor) \wedge N$, die nur endlich viele Werte annehmen, unabhängig, und es gilt $X_N \uparrow |X|$ sowie $Y_N \uparrow |Y|$. Nach dem Satz von der monotonen Konvergenz (Satz 4.20) ist daher

$$\begin{aligned}\mathbf{E}[|XY|] &= \lim_{N \rightarrow \infty} \mathbf{E}[X_N Y_N] = \lim_{N \rightarrow \infty} \mathbf{E}[X_N] \mathbf{E}[Y_N] \\ &= \left(\lim_{N \rightarrow \infty} \mathbf{E}[X_N] \right) \left(\lim_{N \rightarrow \infty} \mathbf{E}[Y_N] \right) = \mathbf{E}[|X|] \mathbf{E}[|Y|] < \infty.\end{aligned}$$

Also ist $XY \in \mathcal{L}^1(\mathbf{P})$. Außerdem haben wir damit den Satz gezeigt für den Fall, wo X und Y nichtnegativ sind. Daher (und weil jede der Familien $\{X^+, Y^+\}$, $\{X^-, Y^+\}$, $\{X^+, Y^-\}$ und $\{X^-, Y^-\}$ unabhängig ist) gilt

$$\begin{aligned}\mathbf{E}[XY] &= \mathbf{E}[(X^+ - X^-)(Y^+ - Y^-)] \\ &= \mathbf{E}[X^+ Y^+] - \mathbf{E}[X^- Y^+] - \mathbf{E}[X^+ Y^-] + \mathbf{E}[X^- Y^-] \\ &= \mathbf{E}[X^+] \mathbf{E}[Y^+] - \mathbf{E}[X^-] \mathbf{E}[Y^+] - \mathbf{E}[X^+] \mathbf{E}[Y^-] + \mathbf{E}[X^-] \mathbf{E}[Y^-] \\ &= \mathbf{E}[X^+ - X^-] \mathbf{E}[Y^+ - Y^-] = \mathbf{E}[X] \mathbf{E}[Y].\end{aligned} \quad \square$$

Satz 5.5 (Wald'sche Identität). Seien T, X_1, X_2, \dots unabhängige, reelle Zufallsvariablen in $\mathcal{L}^1(\mathbf{P})$. Es sei $\mathbf{P}[T \in \mathbb{N}_0] = 1$, und es seien X_1, X_2, \dots identisch verteilt. Wir setzen

$$S_T := \sum_{i=1}^T X_i.$$

Dann ist $S_T \in \mathcal{L}^1(\mathbf{P})$ und $\mathbf{E}[S_T] = \mathbf{E}[T] \mathbf{E}[X_1]$.

Beweis. Setze $S_n = \sum_{i=1}^n X_i$ für $n \in \mathbb{N}_0$. Dann ist $S_T = \sum_{n=1}^{\infty} S_n \mathbf{1}_{\{T=n\}}$. Nach Bemerkung 2.15 sind S_n und $\mathbf{1}_{\{T=n\}}$ unabhängig für jedes $n \in \mathbb{N}$ und damit unkorreliert. Es folgt (mit Hilfe der Dreiecksungleichung, siehe Satz 5.3(v))

$$\begin{aligned}\mathbf{E}[|S_T|] &= \sum_{n=1}^{\infty} \mathbf{E}[|S_n| \mathbf{1}_{\{T=n\}}] = \sum_{n=1}^{\infty} \mathbf{E}[|S_n|] \mathbf{E}[\mathbf{1}_{\{T=n\}}] \\ &\leq \sum_{n=1}^{\infty} \mathbf{E}[|X_1|] n \mathbf{P}[T=n] = \mathbf{E}[|X_1|] \mathbf{E}[T].\end{aligned}$$

Die selbe Rechnung ohne Betragstriche liefert den Rest der Behauptung. \square

Wir stellen hier ein paar einfache Eigenschaften der Varianz zusammen.

Satz 5.6. Sei $X \in \mathcal{L}^2(\mathbf{P})$. Dann gilt:

- (i) $\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] \geq 0$.
- (ii) $\mathbf{Var}[X] = 0 \iff X = \mathbf{E}[X]$ fast sicher.
- (iii) Die Abbildung $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \mathbf{E}[(X-x)^2]$ ist minimal genau in $x_0 = \mathbf{E}[X]$ mit $f(\mathbf{E}[X]) = \mathbf{Var}[X]$.

Beweis. (i) Klar nach Bemerkung 5.2(ii).

(ii) Nach Satz 5.3(iii) ist $\mathbf{E}[(X - \mathbf{E}[X])^2] = 0 \iff (X - \mathbf{E}[X])^2 = 0$ f.s.

(iii) Es ist $f(x) = \mathbf{E}[X^2] - 2x \mathbf{E}[X] + x^2 = \mathbf{Var}[X] + (x - \mathbf{E}[X])^2$. \square

Satz 5.7. Die Abbildung $\mathbf{Cov} : \mathcal{L}^2(\mathbf{P}) \times \mathcal{L}^2(\mathbf{P}) \rightarrow \mathbb{R}$ ist eine positiv semidefinite symmetrische Bilinearform, und es gilt $\mathbf{Cov}[X, Y] = 0$, falls Y fast sicher konstant ist. Ausgeschrieben heißt dies: Für $X_1, \dots, X_m, Y_1, \dots, Y_n \in \mathcal{L}^2(\mathbf{P})$ und $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_n \in \mathbb{R}$ sowie $d, e \in \mathbb{R}$ gilt

$$\mathbf{Cov} \left[d + \sum_{i=1}^m \alpha_i X_i, e + \sum_{j=1}^n \beta_j Y_j \right] = \sum_{i,j} \alpha_i \beta_j \mathbf{Cov}[X_i, Y_j]. \quad (5.1)$$

Speziell gilt die **Bienaymé-Gleichung**

$$\mathbf{Var} \left[\sum_{i=1}^m X_i \right] = \sum_{i=1}^m \mathbf{Var}[X_i] + \sum_{\substack{i,j=1 \\ i \neq j}}^m \mathbf{Cov}[X_i, X_j]. \quad (5.2)$$

Für unkorrelierte X_1, \dots, X_m gilt $\mathbf{Var}[\sum_{i=1}^m X_i] = \sum_{i=1}^m \mathbf{Var}[X_i]$.

Beweis.

$$\begin{aligned}
& \mathbf{Cov} \left[d + \sum_{i=1}^m \alpha_i X_i, e + \sum_{j=1}^n \beta_j Y_j \right] \\
&= \mathbf{E} \left[\left(\sum_{i=1}^m \alpha_i (X_i - \mathbf{E}[X_i]) \right) \left(\sum_{j=1}^n \beta_j (Y_j - \mathbf{E}[Y_j]) \right) \right] \\
&= \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j \mathbf{E}[(X_i - \mathbf{E}[X_i])(Y_j - \mathbf{E}[Y_j])] \\
&= \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j \mathbf{Cov}[X_i, Y_j]. \quad \square
\end{aligned}$$

Satz 5.8 (Cauchy-Schwarz'sche Ungleichung). Sind $X, Y \in \mathcal{L}^2(\mathbf{P})$, so gilt

$$(\mathbf{Cov}[X, Y])^2 \leq \mathbf{Var}[X] \mathbf{Var}[Y].$$

Es gilt genau dann Gleichheit, wenn es $a, b, c \in \mathbb{R}$ gibt mit $|a| + |b| + |c| > 0$ und $aX + bY + c = 0$ fast sicher.

Beweis. Die Cauchy-Schwarz'sche Ungleichung gilt für jede positiv semidefinite Bilinearform, also insbesondere für die Kovarianz. In der Notation von Varianz und Kovarianz sieht der Beweis so aus:

1. Fall: $\mathbf{Var}[Y] = 0$. Hier ist die Aussage trivialerweise richtig (mit $a = 0, b = 1$ und $c = -\mathbf{E}[Y]$).

2. Fall: $\mathbf{Var}[Y] > 0$. Sei $\theta := -\frac{\mathbf{Cov}[X, Y]}{\mathbf{Var}[Y]}$. Dann ist nach Satz 5.6(i)

$$\begin{aligned}
0 \leq \mathbf{Var}[X + \theta Y] \mathbf{Var}[Y] &= \left(\mathbf{Var}[X] + 2\theta \mathbf{Cov}[X, Y] + \theta^2 \mathbf{Var}[Y] \right) \mathbf{Var}[Y] \\
&= \mathbf{Var}[X] \mathbf{Var}[Y] - \mathbf{Cov}[X, Y]^2
\end{aligned}$$

mit Gleichheit genau dann, wenn $X + \theta Y$ fast sicher konstant ist. Wähle nun $a = 1$ und $b = \theta$ sowie $c = -\mathbf{E}[X] - b \mathbf{E}[Y]$. \square

Beispiel 5.9. (i) Es sei $p \in [0, 1]$ und $X \sim \text{Ber}_p$. Dann ist $\mathbf{E}[X^2] = \mathbf{E}[X] = \mathbf{P}[X = 1] = p$ und damit $\mathbf{Var}[X] = p(1 - p)$.

(ii) Seien $n \in \mathbb{N}$ und $p \in [0, 1]$ sowie X binomialverteilt $X \sim b_{n,p}$. Dann ist

$$\begin{aligned}\mathbf{E}[X] &= \sum_{k=0}^n k \mathbf{P}[X = k] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \cdot \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np.\end{aligned}$$

Ferner ist

$$\begin{aligned}\mathbf{E}[X(X-1)] &= \sum_{k=0}^n k(k-1) \mathbf{P}[X = k] \\ &= \sum_{k=0}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \cdot \sum_{k=1}^n (k-1) \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= n(n-1)p^2 \cdot \sum_{k=2}^n \binom{n-2}{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)} \\ &= n(n-1)p^2.\end{aligned}$$

Also ist $\mathbf{E}[X^2] = \mathbf{E}[X(X-1)] + \mathbf{E}[X] = n^2 p^2 + np(1-p)$ und damit $\mathbf{Var}[X] = np(1-p)$.

Etwas einfacher als durch die direkte Berechnung sehen wir dies ein, indem wir bemerken (siehe nach Beispiel 3.4(ii)), dass $b_{n,p} = b_{1,p}^{*n}$. Das heißt, es gilt (siehe Satz 2.31) $\mathbf{P}_X = \mathbf{P}_{Y_1+\dots+Y_n}$, wo Y_1, \dots, Y_n unabhängig sind und $Y_i \sim \text{Ber}_p$ für jedes $i = 1, \dots, n$. Es folgt

$$\begin{aligned}\mathbf{E}[X] &= n \mathbf{E}[Y_1] = np, \\ \mathbf{Var}[X] &= n \mathbf{Var}[Y_1] = np(1-p).\end{aligned}\tag{5.3}$$

(iii) Seien $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ sowie X normalverteilt $X \sim \mathcal{N}_{\mu, \sigma^2}$. Dann ist

$$\begin{aligned}\mathbf{E}[X] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/(2\sigma^2)} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x + \mu) e^{-x^2/(2\sigma^2)} dx \\ &= \mu + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-x^2/(2\sigma^2)} dx = \mu.\end{aligned}\tag{5.4}$$

Ähnlich folgt $\mathbf{Var}[X] = \mathbf{E}[X^2] - \mu^2 = \dots = \sigma^2$.

(iv) Sei $\theta > 0$ und X exponentialverteilt $X \sim \exp_\theta$. Dann ist

$$\mathbf{E}[X] = \theta \int_0^\infty x e^{-\theta x} dx = \frac{1}{\theta},$$

$$\mathbf{Var}[X] = -\theta^{-2} + \theta \int_0^\infty x^2 e^{-\theta x} dx = \theta^{-2} \left(-1 + \int_0^\infty x^2 e^{-x} dx \right) = \theta^{-2}. \quad \diamond$$

Satz 5.10 (Blackwell-Girshick). Seien T, X_1, X_2, \dots unabhängige, reelle Zufallsvariablen in $\mathcal{L}^2(\mathbf{P})$. Es sei $\mathbf{P}[T \in \mathbb{N}_0] = 1$, und es seien X_1, X_2, \dots identisch verteilt. Wir setzen

$$S_T := \sum_{i=1}^T X_i.$$

Dann ist $S_T \in \mathcal{L}^2(\mathbf{P})$ und

$$\mathbf{Var}[S_T] = \mathbf{E}[X_1]^2 \mathbf{Var}[T] + \mathbf{E}[T] \mathbf{Var}[X_1].$$

Beweis. Wir setzen $S_n = \sum_{i=1}^n X_i$ für $n \in \mathbb{N}$. Dann sind (wie beim Beweis der Wald'schen Identität) S_n und $\mathbf{1}_{\{T=n\}}$ unabhängig, also S_n^2 und $\mathbf{1}_{\{T=n\}}$ unkorreliert und damit

$$\begin{aligned} \mathbf{E}[S_T^2] &= \sum_{n=0}^{\infty} \mathbf{E}[\mathbf{1}_{\{T=n\}} S_n^2] \\ &= \sum_{n=0}^{\infty} \mathbf{E}[\mathbf{1}_{\{T=n\}}] \mathbf{E}[S_n^2] \\ &= \sum_{n=0}^{\infty} \mathbf{P}[T=n] (\mathbf{Var}[S_n] + \mathbf{E}[S_n]^2) \\ &= \sum_{n=0}^{\infty} \mathbf{P}[T=n] \left(n \mathbf{Var}[X_1] + n^2 \mathbf{E}[X_1]^2 \right) \\ &= \mathbf{E}[T] \mathbf{Var}[X_1] + \mathbf{E}[T^2] \mathbf{E}[X_1]^2. \end{aligned}$$

Nach der Wald'schen Identität (Satz 5.5) ist $\mathbf{E}[S_T] = \mathbf{E}[T] \mathbf{E}[X_1]$, also ist

$$\mathbf{Var}[S_T] = \mathbf{E}[S_T^2] - \mathbf{E}[S_T]^2 = \mathbf{E}[T] \mathbf{Var}[X_1] + (\mathbf{E}[T^2] - \mathbf{E}[T]^2) \mathbf{E}[X_1]^2.$$

Dies ist aber die Behauptung. □

Übung 5.1.1. Man zeige (mit Satz 4.15): Ist X eine integrierbare reelle Zufallsvariable, deren Verteilung \mathbf{P}_X die Dichte f (bezüglich des Lebesgue-Maßes λ) besitzt, so gilt

$$\mathbf{E}[X] = \int_{\mathbb{R}} x f(x) \lambda(dx). \quad \clubsuit$$

Übung 5.1.2. Sei $X \sim \beta_{r,s}$ eine Beta-verteilte Zufallsvariable mit Parametern $r, s > 0$ (vergleiche Beispiel 1.107(ii)). Man zeige

$$\mathbf{E}[X^n] = \prod_{k=0}^{n-1} \frac{r+k}{r+s+k} \quad \text{für jedes } n \in \mathbb{N}. \quad \clubsuit$$

Übung 5.1.3. Es seien X_1, X_2, \dots u.i.v. nichtnegative Zufallsvariablen. Man zeige mit Hilfe des Lemmas von Borel-Cantelli:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} X_n = \begin{cases} 0 \text{ f.s.,} & \text{falls } \mathbf{E}[X_1] < \infty, \\ \infty \text{ f.s.,} & \text{falls } \mathbf{E}[X_1] = \infty. \end{cases} \quad \clubsuit$$

Übung 5.1.4. Es seien X_1, X_2, \dots u.i.v. nichtnegative Zufallsvariablen. Man zeige mit Hilfe des Lemmas von Borel-Cantelli: Für jedes $c \in (0, 1)$ gilt

$$\sum_{n=1}^{\infty} e^{X_n} c^n = \begin{cases} < \infty \text{ f.s.,} & \text{falls } \mathbf{E}[X_1] < \infty, \\ = \infty \text{ f.s.,} & \text{falls } \mathbf{E}[X_1] = \infty. \end{cases} \quad \clubsuit$$

5.2 Schwaches Gesetz der Großen Zahl

Satz 5.11 (Markov'sche Ungleichung, Chebyshev'sche Ungleichung).

Sei X eine Zufallsvariable und $f : [0, \infty) \rightarrow [0, \infty)$ monoton wachsend. Dann gilt für jedes $\varepsilon > 0$ mit $f(\varepsilon) > 0$ die **Markov'sche Ungleichung**

$$\mathbf{P}[|X| \geq \varepsilon] \leq \frac{\mathbf{E}[f(|X|)]}{f(\varepsilon)}.$$

Im Spezialfall $f(x) = x^2$ erhalten wir $\mathbf{P}[|X| \geq \varepsilon] \leq \varepsilon^{-2} \mathbf{E}[X^2]$ und, falls $X \in \mathcal{L}^2(\mathbf{P})$, insbesondere die **Chebyshev'sche Ungleichung**

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq \varepsilon] \leq \varepsilon^{-2} \mathbf{Var}[X].$$

Beweis. Es gilt

$$\begin{aligned} \mathbf{E}[f(|X|)] &\geq \mathbf{E}[f(|X|) \mathbf{1}_{\{f(|X|) \geq f(\varepsilon)\}}] \\ &\geq \mathbf{E}[f(\varepsilon) \mathbf{1}_{\{f(|X|) \geq f(\varepsilon)\}}] \\ &\geq f(\varepsilon) \mathbf{P}[|X| \geq \varepsilon]. \end{aligned} \quad \square$$

Definition 5.12. Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge reeller Zufallsvariablen in $\mathcal{L}^1(\mathbf{P})$ und $\tilde{S}_n = \sum_{i=1}^n (X_i - \mathbf{E}[X_i])$.

(i) Wir sagen, $(X_n)_{n \in \mathbb{N}}$ genüge dem **schwachen Gesetz der großen Zahl**, falls

$$\lim_{n \rightarrow \infty} \mathbf{P} \left[\left| \frac{1}{n} \tilde{S}_n \right| > \varepsilon \right] = 0 \quad \text{für jedes } \varepsilon > 0.$$

(ii) Wir sagen, $(X_n)_{n \in \mathbb{N}}$ genüge dem **starken Gesetz der großen Zahl**, falls

$$\mathbf{P} \left[\limsup_{n \rightarrow \infty} \left| \frac{1}{n} \tilde{S}_n \right| = 0 \right] = 1.$$

Bemerkung 5.13. Das starke Gesetz der großen Zahl impliziert das schwache. Ist nämlich $A_n^\varepsilon := \left\{ \left| \frac{1}{n} \tilde{S}_n \right| > \varepsilon \right\}$ und $A = \left\{ \limsup_{n \rightarrow \infty} \left| \frac{1}{n} \tilde{S}_n \right| > 0 \right\}$, so gilt offenbar

$$A = \bigcup_{m \in \mathbb{N}} \limsup_{n \rightarrow \infty} A_n^{1/m},$$

also $\mathbf{P} \left[\limsup_{n \rightarrow \infty} A_n^\varepsilon \right] = 0$ für $\varepsilon > 0$. Nach dem Lemma von Fatou (Satz 4.21) ist

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbf{P} [A_n^\varepsilon] &= 1 - \liminf_{n \rightarrow \infty} \mathbf{E} [\mathbb{1}_{(A_n^\varepsilon)^c}] \\ &\leq 1 - \mathbf{E} \left[\liminf_{n \rightarrow \infty} \mathbb{1}_{(A_n^\varepsilon)^c} \right] = \mathbf{E} \left[\limsup_{n \rightarrow \infty} \mathbb{1}_{A_n^\varepsilon} \right] = 0. \quad \square \end{aligned}$$

Satz 5.14. Seien X_1, X_2, \dots unkorrelierte Zufallsvariablen in $\mathcal{L}^2(\mathbf{P})$ mit $V := \sup_{n \in \mathbb{N}} \mathbf{Var}[X_n] < \infty$. Dann genügt $(X_n)_{n \in \mathbb{N}}$ dem schwachen Gesetz der großen Zahl. Es gilt sogar für jedes $\varepsilon > 0$

$$\mathbf{P} \left[\left| \frac{1}{n} \tilde{S}_n \right| \geq \varepsilon \right] \leq \frac{V}{\varepsilon^2 n} \quad \text{für jedes } n \in \mathbb{N}. \quad (5.5)$$

Beweis. Ohne Einschränkung sei $\mathbf{E}[X_i] = 0$ für jedes $i \in \mathbb{N}$ und damit $\tilde{S}_n = X_1 + \dots + X_n$. Nach der Formel von Bienaymé (Satz 5.7) ist

$$\mathbf{Var} \left[\frac{1}{n} \tilde{S}_n \right] = n^{-2} \sum_{i=1}^n \mathbf{Var} [X_i] \leq \frac{V}{n}.$$

Nach der Chebyshev'schen Ungleichung (Satz 5.11) gilt für $\varepsilon > 0$

$$\mathbf{P} [|\tilde{S}_n/n| \geq \varepsilon] \leq \frac{V}{\varepsilon^2 n} \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Beispiel 5.15 (Weierstraß'scher Approximationssatz). Sei $f : [0, 1] \rightarrow \mathbb{R}$ eine stetige Abbildung. Nach dem Weierstraß'schen Approximationssatz existieren Polynome f_n vom Grad höchstens n , sodass

$$\|f_n - f\|_\infty \xrightarrow{n \rightarrow \infty} 0,$$

wobei wir mit $\|f\|_\infty := \sup\{|f(x)| : x \in [0, 1]\}$ die Supremumsnorm von f bezeichnen.

Wir führen hier einen probabilistischen Beweis dieser Aussage vor. Für $n \in \mathbb{N}$ sei das Polynom f_n definiert durch

$$f_n(x) := \sum_{k=0}^n f(k/n) \binom{n}{k} x^k (1-x)^{n-k} \quad \text{für } x \in [0, 1].$$

Dieses Polynom heißt **Bernstein-Polynom** der Ordnung n .

Sei $\varepsilon > 0$ fest gewählt. Da f auf $[0, 1]$ stetig ist, ist f sogar gleichmäßig stetig. Es existiert also ein $\delta > 0$, sodass

$$|f(x) - f(y)| < \varepsilon \quad \text{für alle } x, y \in [0, 1] \text{ mit } |x - y| < \delta.$$

Sei nun $p \in [0, 1]$ fest gewählt, und seien X_1, X_2, \dots unabhängige Zufallsvariablen mit $X_i \sim \text{Ber}_p$, $i \in \mathbb{N}$. Dann ist $S_n := X_1 + \dots + X_n \sim b_{n,p}$ und deshalb

$$\mathbf{E}[f(S_n/n)] = \sum_{k=0}^n f(k/n) \mathbf{P}[S_n = k] = f_n(p).$$

Wir erhalten

$$|f(S_n/n) - f(p)| \leq \varepsilon + 2\|f\|_\infty \mathbf{1}_{\{|(S_n/n) - p| \geq \delta\}}$$

und daher (nach Satz 5.14 mit $V = p(1-p) \leq \frac{1}{4}$)

$$\begin{aligned} |f_n(p) - f(p)| &\leq \mathbf{E}[|f(S_n/n) - f(p)|] \\ &\leq \varepsilon + 2\|f\|_\infty \mathbf{P}\left[\left|\frac{S_n}{n} - p\right| \geq \delta\right] \\ &\leq \varepsilon + \frac{\|f\|_\infty}{2\delta^2 n}, \end{aligned}$$

für jedes $p \in [0, 1]$. Also gilt $\|f_n - f\|_\infty \xrightarrow{n \rightarrow \infty} 0$. ◇

Übung 5.2.1 (Bernstein-Chernov Abschätzung). Seien $n \in \mathbb{N}$ und $p_1, \dots, p_n \in [0, 1]$ sowie X_1, \dots, X_n unabhängige Zufallsvariablen mit $X_i = \text{Ber}_{p_i}$ für jedes $i = 1, \dots, n$. Setze $S_n = X_1 + \dots + X_n$ und $m := \mathbf{E}[S_n]$. Man zeige, dass für jedes $\delta > 0$ die beiden folgenden Abschätzungen gelten

$$\mathbf{P}[S_n \geq (1 + \delta)m] \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^m$$

und

$$\mathbf{P}[S_n \leq (1 - \delta)m] \leq \exp\left(-\frac{\delta^2 m}{2}\right).$$

Hinweis: Verwende für S_n die Markov'sche Ungleichung mit $f(x) = e^{\lambda x}$ für gewisses $\lambda > 0$ und finde dasjenige λ , das die Abschätzung optimiert. ♣

5.3 Starkes Gesetz der Großen Zahl

Wir kommen nun zu einem starken Gesetz der großen Zahl, nämlich dem in der Form von Etemadi für identisch verteilte, paarweise unabhängige Zufallsvariablen. Es gibt viele verschiedene Formen von starken Gesetzen der großen Zahl, die unterschiedliche Voraussetzungen machen. So kann man darauf verzichten, dass die Zufallsvariablen identisch verteilt sind, wenn man stärkere Annahmen, etwa beschränkte Varianzen, macht und so weiter. Wir werden hier nicht bis in alle Tiefen gehen, sondern nur exemplarisch ein paar Aussagen vorstellen. Um die Methode für den Beweis des Satzes von Etemadi zu illustrieren, stellen wir zunächst ein Starkes Gesetz der großen Zahl unter stärkeren Annahmen vor.

Satz 5.16. Sind $X_1, X_2, \dots \in \mathcal{L}^2(\mathbf{P})$ paarweise unabhängig (das heißt, X_i und X_j sind unabhängig für alle $i, j \in \mathbb{N}$ mit $i \neq j$) und identisch verteilt, so genügt $(X_n)_{n \in \mathbb{N}}$ dem starken Gesetz der großen Zahl.

Beweis. Es sind $(X_n^+)_{n \in \mathbb{N}}$ und $(X_n^-)_{n \in \mathbb{N}}$ wieder paarweise unabhängige Familien quadratintegrierbarer Zufallsvariablen (vergleiche Bemerkung 2.15(ii)). Es reicht daher, $(X_n^+)_{n \in \mathbb{N}}$ zu betrachten. Wir nehmen also im Folgenden an, dass $X_n \geq 0$ ist fast sicher für jedes $n \in \mathbb{N}$.

Sei $S_n = X_1 + \dots + X_n$ für $n \in \mathbb{N}$. Wähle $\varepsilon > 0$. Für jedes $n \in \mathbb{N}$ setzen wir $k_n = \lfloor (1 + \varepsilon)^n \rfloor \geq \frac{1}{2}(1 + \varepsilon)^n$. Dann ist nach der Chebyshev'schen Ungleichung (Satz 5.11)

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbf{P}\left[\left|\frac{S_{k_n}}{k_n} - \mathbf{E}[X_1]\right| \geq (1 + \varepsilon)^{-n/4}\right] &\leq \sum_{n=1}^{\infty} (1 + \varepsilon)^{n/2} \mathbf{Var}[k_n^{-1} S_{k_n}] \\ &= \sum_{n=1}^{\infty} (1 + \varepsilon)^{n/2} k_n^{-1} \mathbf{Var}[X_1] \quad (5.6) \\ &\leq 2 \mathbf{Var}[X_1] \sum_{n=1}^{\infty} (1 + \varepsilon)^{-n/2} < \infty. \end{aligned}$$

Nach dem Lemma von Borel-Cantelli gibt es daher für \mathbf{P} -f.a. ω ein $n_0 = n_0(\omega)$ mit

$$\left| \frac{S_{k_n}}{k_n} - \mathbf{E}[X_1] \right| < (1 + \varepsilon)^{-n/4} \quad \text{für jedes } n \geq n_0.$$

Also gilt

$$\limsup_{n \rightarrow \infty} \left| k_n^{-1} S_{k_n} - \mathbf{E}[X_1] \right| = 0 \quad \text{fast sicher.}$$

Für hinreichend großes $n \in \mathbb{N}$ ist $k_{n+1} \leq (1 + 2\varepsilon)k_n$. Für $l \in \{k_n, \dots, k_{n+1}\}$ ist dann

$$\frac{1}{1 + 2\varepsilon} k_n^{-1} S_{k_n} \leq k_{n+1}^{-1} S_{k_n} \leq l^{-1} S_l \leq k_n^{-1} S_{k_{n+1}} \leq (1 + 2\varepsilon) k_{n+1}^{-1} S_{k_{n+1}}.$$

Wegen $1 - (1 + 2\varepsilon)^{-1} \leq 2\varepsilon$ folgt

$$\begin{aligned} \limsup_{l \rightarrow \infty} \left| l^{-1} S_l - \mathbf{E}[X_1] \right| &\leq \limsup_{n \rightarrow \infty} \left| k_n^{-1} S_{k_n} - \mathbf{E}[X_1] \right| + 2\varepsilon \limsup_{n \rightarrow \infty} k_n^{-1} S_{k_n} \\ &\leq 2\varepsilon \mathbf{E}[X_1] \quad \text{fast sicher,} \end{aligned}$$

und damit gilt das starke Gesetz der großen Zahl. \square

Die Ähnlichkeit der Varianzabschätzungen im schwachen Gesetz der großen Zahl und in (5.6) legen nahe, dass im vorangehenden Satz auf die Bedingung verzichtet werden kann, dass die Zufallsvariablen X_1, X_2, \dots identisch verteilt sind, wenn man nur fordert, dass die Varianzen beschränkt sind (siehe Übung 5.3.1).

Wir können die Bedingung in Satz 5.16 in anderer Weise abschwächen, indem wir nur Integrierbarkeit statt Quadratintegrierbarkeit der Zufallsvariablen fordern.

Satz 5.17 (Starkes Gesetz der großen Zahl von Etemadi (1981)). Es seien $X_1, X_2, \dots \in \mathcal{L}^1(\mathbf{P})$ paarweise unabhängig und identisch verteilt. Dann genügt $(X_n)_{n \in \mathbb{N}}$ dem starken Gesetz der großen Zahl.

Wir folgen der Darstellung in [44, Kapitel 2.4]. Definiere $\mu = \mathbf{E}[X_1]$ und $S_n = X_1 + \dots + X_n$. Zur Vorbereitung des Beweises stellen wir ein paar Lemmata bereit.

Lemma 5.18. Für $n \in \mathbb{N}$ seien $Y_n := X_n \mathbb{1}_{\{|X_n| \leq n\}}$ und $T_n = Y_1 + \dots + Y_n$. Die Folge $(X_n)_{n \in \mathbb{N}}$ erfüllt das starke Gesetz der großen Zahl, falls $T_n/n \xrightarrow{n \rightarrow \infty} \mu$ f.s.

Beweis. Nach Satz 4.26 ist $\sum_{n=1}^{\infty} \mathbf{P}[|X_n| > n] \leq \mathbf{E}[|X_1|] < \infty$. Nach dem Lemma von Borel-Cantelli ist daher

$$\mathbf{P}[X_n \neq Y_n \text{ für unendlich viele } n] = 0.$$

Es gibt also ein $n_0 = n_0(\omega)$ mit $X_n = Y_n$ für jedes $n \geq n_0$. Daher gilt für $n \geq n_0$

$$\frac{T_n - S_n}{n} = \frac{T_{n_0} - S_{n_0}}{n} \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Lemma 5.19. Für jedes $x \geq 0$ ist $2x \sum_{n>x} n^{-2} \leq 4$.

Beweis. Für $m \in \mathbb{N}$ ist nach dem Integralvergleichskriterium

$$\sum_{n=m}^{\infty} n^{-2} \leq m^{-2} + \int_m^{\infty} t^{-2} dt = m^{-2} + m^{-1} \leq \frac{2}{m}. \quad \square$$

Lemma 5.20. Es gilt $\sum_{n=1}^{\infty} \frac{\mathbf{E}[Y_n^2]}{n^2} \leq 4 \mathbf{E}[|X_1|]$.

Beweis. Nach Satz 4.26 ist $\mathbf{E}[Y_n^2] = \int_0^{\infty} \mathbf{P}[Y_n^2 > t] dt$. Mit der Substitution $x = \sqrt{t}$ erhalten wir

$$\mathbf{E}[Y_n^2] = \int_0^{\infty} 2x \mathbf{P}[|Y_n| > x] dx \leq \int_0^n 2x \mathbf{P}[|X_1| > x] dx.$$

Nach Lemma 5.19 gilt für $m \rightarrow \infty$

$$f_m(x) = \left(\sum_{n=1}^m n^{-2} \mathbb{1}_{\{x < n\}} \right) 2x \mathbf{P}[|X_1| > x] \uparrow f(x) \leq 4 \mathbf{P}[|X_1| > x].$$

Nach dem Satz über monotone Konvergenz können wir also Summation und Integral vertauschen und erhalten

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{\mathbf{E}[Y_n^2]}{n^2} &\leq \sum_{n=1}^{\infty} n^{-2} \int_0^{\infty} \mathbb{1}_{\{x < n\}} 2x \mathbf{P}[|X_1| > x] dx \\ &= \int_0^{\infty} \left(\sum_{n=1}^{\infty} n^{-2} \mathbb{1}_{\{x < n\}} \right) 2x \mathbf{P}[|X_1| > x] dx \\ &\leq 4 \int_0^{\infty} \mathbf{P}[|X_1| > x] dx = 4 \mathbf{E}[|X_1|]. \end{aligned} \quad \square$$

Beweis von Satz 5.17. Wie im Beweis von Satz 5.16 reicht es, den Fall $X_n \geq 0$ zu betrachten. Wähle $\varepsilon > 0$ und setze $\alpha = 1 + \varepsilon$. Für $n \in \mathbb{N}$ setzen wir $k_n = \lfloor \alpha^n \rfloor$ und haben speziell $k_n \geq \alpha^n / 2$. Es ist also für $m \in \mathbb{N}$ (mit $n_0 = \lceil \log m / \log \alpha \rceil$)

$$\sum_{n: k_n \geq m} k_n^{-2} \leq 4 \sum_{n=n_0}^{\infty} \alpha^{-2n} = 4 \alpha^{-2n_0} (1 - \alpha^{-2})^{-1} \leq 4(1 - \alpha^{-2})^{-1} m^{-2}. \quad (5.7)$$

Unser Ziel ist es, mit Hilfe von Lemma 5.20 die Abschätzung (5.6) für $(Y_n)_{n \in \mathbb{N}}$ und $(T_n)_{n \in \mathbb{N}}$ zu verfeinern. Die Chebyshev'sche Ungleichung liefert (zusammen mit (5.7)) wiederum für $\delta > 0$

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbf{P}[|T_{k_n} - \mathbf{E}[T_{k_n}]| > \delta k_n] &\leq \delta^{-2} \sum_{n=1}^{\infty} \frac{\mathbf{Var}[T_{k_n}]}{k_n^2} \\
&= \delta^{-2} \sum_{n=1}^{\infty} k_n^{-2} \sum_{m=1}^{k_n} \mathbf{Var}[Y_m] = \delta^{-2} \sum_{m=1}^{\infty} \mathbf{Var}[Y_m] \sum_{n: k_n \geq m} k_n^{-2} \\
&\leq 4(1 - \alpha^{-2})^{-1} \delta^{-2} \sum_{m=1}^{\infty} m^{-2} \mathbf{E}[Y_m^2] < \infty \text{ nach Lemma 5.20.}
\end{aligned}$$

(Im dritten Schritt durften wir die Summationsreihenfolge vertauschen, weil alle Summanden nichtnegativ sind.) Da $\delta > 0$ beliebig war, folgt (mit dem Lemma von Borel-Cantelli)

$$\lim_{n \rightarrow \infty} \frac{T_{k_n} - \mathbf{E}[T_{k_n}]}{k_n} = 0 \quad \text{fast sicher.} \quad (5.8)$$

Nach dem Satz über monotone Konvergenz (Satz 4.20) gilt

$$\mathbf{E}[Y_n] = \mathbf{E}[X_1 \mathbb{1}_{\{X_1 \leq n\}}] \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_1].$$

Also gilt $\mathbf{E}[T_{k_n}]/k_n \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_1]$ und wegen (5.8) auch $T_{k_n}/k_n \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_1]$ f.s. Wie im Beweis von Satz 5.16 gilt jetzt auch (weil $Y_n \geq 0$)

$$\lim_{l \rightarrow \infty} \frac{T_l}{l} = \mathbf{E}[X_1] \quad \text{fast sicher.}$$

Nach Lemma 5.18 folgt hieraus die Behauptung von Satz 5.17. \square

Beispiel 5.21 (Monte Carlo Integration). Betrachte eine Funktion $f : [0, 1] \rightarrow \mathbb{R}$, deren Integral $I := \int_0^1 f(x) dx$ numerisch bestimmt werden soll. Wir nehmen an, dass uns der Computer Zahlen X_1, X_2, \dots generiert, die wir als unabhängige Zufallszahlen auffassen können, die auf $[0, 1]$ gleichverteilt sind. Für $n \in \mathbb{N}$ definieren wir den Schätzwert

$$\widehat{I}_n := \frac{1}{n} \sum_{i=1}^n f(X_i).$$

Unter der Annahme, dass $f \in \mathcal{L}^1([0, 1])$ ist, liefert das starke Gesetz der großen Zahl $\widehat{I}_n \xrightarrow{n \rightarrow \infty} I$ fast sicher.

Allerdings haben wir im letzten Satz keine Aussage zur Geschwindigkeit der Konvergenz, also zur Größe $\mathbf{P}[|\widehat{I}_n - I| > \varepsilon]$ bekommen. Um genauere Schätzungen für das Integral zu bekommen, benötigen wir zusätzliche Information, etwa den Wert $V_1 := \int f^2(x) dx - I^2$, falls $f \in \mathcal{L}^2([0, 1])$ ist. (Für beschränktes f etwa lässt sich V_1 leicht nach oben abschätzen.) Dann ist nämlich $\mathbf{Var}[\widehat{I}_n] = V_1/n$, also ist nach der Chebyshev'schen Ungleichung

$$\mathbf{P}\left[|\widehat{I}_n - I| > \varepsilon n^{-1/2}\right] \leq V_1/\varepsilon^2.$$

Der Fehler ist mithin maximal von der Größenordnung $n^{-1/2}$. Der Zentrale Grenzwertsatz wird uns zeigen, dass der Fehler exakt von dieser Ordnung ist.

Wenn über f minimale Glattheitseigenschaften bekannt sind, so liefern die gängigen numerischen Verfahren sehr viel bessere Konvergenzordnungen. Die hier beschriebene **Monte Carlo Simulation** sollte also nur angewandt werden, wenn alle anderen Verfahren ungeeignet sind. Speziell ist dies der Fall, wenn statt $[0, 1]$ ein Gebiet $G \subset \mathbb{R}^d$ für sehr großes d betrachtet wird. \diamond

Definition 5.22 (Empirische Verteilungsfunktion). Seien X_1, X_2, \dots reelle Zufallsvariablen. Dann heißt die Abbildung $F_n : \mathbb{R} \rightarrow [0, 1]$, $x \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, x]}(X_i)$ die **empirische Verteilungsfunktion** von X_1, \dots, X_n .

Satz 5.23 (Glivenko-Cantelli). Seien X_1, X_2, \dots u.i.v. reelle Zufallsvariablen mit Verteilungsfunktion F , und seien F_n , $n \in \mathbb{N}$, die empirischen Verteilungsfunktionen. Dann gilt

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0 \quad \text{fast sicher.}$$

Beweis. Wähle ein $x \in \mathbb{R}$ und setze $Y_n(x) = \mathbb{1}_{(-\infty, x]}(X_n)$ und $Z_n(x) = \mathbb{1}_{(-\infty, x)}(X_n)$ für $n \in \mathbb{N}$. Wir definieren außerdem die linksseitigen Limiten $F(x-) = \lim_{y \uparrow x} F(y)$ und analog für F_n . Dann sind $(Y_n(x))_{n \in \mathbb{N}}$ und $(Z_n(x))_{n \in \mathbb{N}}$ jeweils unabhängige Familien und $\mathbf{E}[Y_n(x)] = \mathbf{P}[X_n \leq x] = F(x)$ sowie $\mathbf{E}[Z_n(x)] = \mathbf{P}[X_n < x] = F(x-)$. Nach dem starken Gesetz der großen Zahl gilt also

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n Y_i(x) \xrightarrow{n \rightarrow \infty} F(x) \quad \text{fast sicher}$$

und

$$F_n(x-) = \frac{1}{n} \sum_{i=1}^n Z_i(x) \xrightarrow{n \rightarrow \infty} F(x-) \quad \text{fast sicher.}$$

Wir setzen formal $F(-\infty) = 0$ und $F(\infty) = 1$. Wir wählen ein beliebiges $N \in \mathbb{N}$ und setzen

$$x_j := \inf \{x \in \overline{\mathbb{R}} : F(x) \geq j/N\}, \quad j = 0, \dots, N,$$

und

$$R_n := \max_{j=1, \dots, N-1} (|F_n(x_j) - F(x_j)| + |F_n(x_j-) - F(x_j-)|).$$

Nach dem Gezeigten gilt $R_n \xrightarrow{n \rightarrow \infty} 0$ fast sicher. Für $x \in (x_{j-1}, x_j)$ ist (nach Definition von x_j)

$$F_n(x) \leq F_n(x_j-) \leq F(x_j-) + R_n \leq F(x) + R_n + \frac{1}{N}$$

und

$$F_n(x) \geq F_n(x_{j-1}) \geq F(x_{j-1}) - R_n \geq F(x) - R_n - \frac{1}{N}.$$

Also gilt

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{1}{N} + \limsup_{n \rightarrow \infty} R_n = \frac{1}{N}.$$

Indem wir $N \rightarrow \infty$ gehen lassen, folgt die Behauptung. \square

Beispiel 5.24 (Satz von Shannon). Wir betrachten eine Informationsquelle, die zufällig (und unabhängig) hintereinander Zeichen X_1, X_2, \dots eines endlichen Alphabets E (also einer beliebigen endlichen Menge) ausgibt. Dabei soll p_e die Wahrscheinlichkeit für das Auftreten des Zeichens $e \in E$ sein. Formal sind also X_1, X_2, \dots u.i.v. E -wertige Zufallsvariablen mit $\mathbf{P}[X_i = e] = p_e$ für $e \in E$.

Sei für jedes $\omega \in \Omega$ und $n \in \mathbb{N}$

$$\pi_n(\omega) := \prod_{i=1}^n p_{X_i(\omega)}$$

die Wahrscheinlichkeit, dass die beobachtete Sequenz $X_1(\omega), \dots, X_n(\omega)$ auftritt. Wir setzen $Y_n(\omega) := -\log(p_{X_n(\omega)})$. Dann ist $(Y_n)_{n \in \mathbb{N}}$ u.i.v. und $\mathbf{E}[Y_n] = H(p)$, wobei

$$H(p) := - \sum_{e \in E} p_e \log(p_e)$$

die **Entropie** der Verteilung $p = (p_e)_{e \in E}$ ist. Aus dem starken Gesetz der großen Zahl folgt der Satz von Shannon:

$$-\frac{1}{n} \log \pi_n = \frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{n \rightarrow \infty} H(p) \text{ fast sicher.} \quad \diamond$$

Entropie und Quellenkodierungssatz*

Wir wollen kurz auf die Bedeutung von π_n und der Entropie eingehen. Wie groß ist die „Information“, die in einer Nachricht $X_1(\omega), \dots, X_n(\omega)$ steckt? Diese Information kann man messen durch die Länge der kürzesten Folge von Nullen und Einsen, mit der man die Nachricht kodieren kann. Wir wollen jetzt natürlich nicht für jede Nachricht eine eigene Kodierung erfinden, und auch nicht für jede mögliche Nachrichtenlänge. Stattdessen ordnen wir jedem einzelnen Zeichen $e \in E$ eine Folge von Nullen und Einsen zu, die dann aneinander gereiht die Nachricht ergeben. Die Länge $l(e)$ der Folge, die das Zeichen e kodiert, darf dabei von e abhängen. So geben wir Zeichen, die häufiger auftreten, einen kürzeren Code als den selteneren Zeichen, um einen möglichst effizienten Code zu bekommen. Beim Morse-Alphabet ist dies

so ähnlich eingerichtet („e“ und „t“, die im Englischen häufig vorkommen, haben die Codes „kurz“ und „lang“, der seltener Buchstabe „q“ hat den Code „lang-lang-kurz-lang“), allerdings besteht der Morse-Code nicht nur aus kurzen und langen Signalen, sondern auch noch aus Pausenzeichen, die das Ende eines Buchstabens signalisieren. Wenn wir nur Nullen und Einsen verwenden dürfen, haben wir keine solchen Pausenzeichen und müssen den Code so anlegen, dass der Code eines Zeichens nicht gleichzeitig der Anfang eines Codes eines anderen Zeichens ist. Wir dürfen also nicht etwa ein Zeichen mit 0110 kodieren und ein anderes mit 011011. Ein Code, der diese Bedingung erfüllt, heißt ein **binärer Präfixcode**. Wir bezeichnen mit $c(e) \in \{0, 1\}^{l(e)}$ den Code von e , wobei $l(e)$ die Länge ist. Wir können die Codes aller Zeichen in einem Baum darstellen.

Wir wollen nun einen Code $C = (c(e), e \in E)$ herstellen, der effizient ist in dem Sinne, dass die erwartete Länge des Codes (für ein zufälliges Zeichen)

$$L_p(C) := \sum_{e \in E} p_e l(e)$$

möglichst klein ist.

Wir konstruieren zunächst einen Code und zeigen dann, dass dieser fast optimal ist. Als ersten Schritt nummerieren wir $E = \{e_1, \dots, e_N\}$ so, dass $p_{e_1} \geq p_{e_2} \geq \dots \geq p_{e_N}$ gilt. Wir definieren $\ell(e) \in \mathbb{N}$ für jedes $e \in E$ durch

$$2^{-\ell(e)} \leq p_e < 2^{-\ell(e)+1}.$$

Setze $\tilde{p}_e = 2^{-\ell(e)}$ für jedes $e \in E$ und $\tilde{q}_k = \sum_{l < k} \tilde{p}_{e_l}$ für $k = 1, \dots, N$.

Nach Konstruktion gilt $\ell(e_l) \leq \ell(e_k)$ für alle $l \leq k$, also hat die Binärdarstellung von \tilde{q}_k maximal $\ell(e_k)$ Stellen:

$$\tilde{q}_k = \sum_{i=1}^{\ell(e_k)} c_i(e_k) 2^{-i},$$

wobei die Zahlen $c_1(e_k), \dots, c_{\ell(e_k)}(e_k) \in \{0, 1\}$ eindeutig festgelegt sind.

Offenbar ist $\tilde{q}_l \geq \tilde{q}_k + 2^{-\ell(e_k)}$ für jedes $l > k$, also ist

$$(c_1(e_k), \dots, c_{\ell(e_k)}(e_k)) \neq (c_1(e_l), \dots, c_{\ell(e_k)}(e_l)) \quad \text{für alle } l > k.$$

Daher ist $C = (c(e), e \in E)$ ein Präfixcode.

Wir schreiben für jedes $b > 0$ und $x > 0$ $\log_b(x) := \frac{\log(x)}{\log(b)}$ für den Logarithmus von x zur Basis b . Nach Konstruktion ist $-\log_2(p_e) \leq \ell(e) \leq 1 - \log_2(p_e)$. Also ist die erwartete Länge

$$-\sum_{e \in E} p_e \log_2(p_e) \leq L_p(C) \leq 1 - \sum_{e \in E} p_e \log_2(p_e).$$

Die Länge dieses Codes für die ersten n Zeichen unserer zufälligen Quelle ist also ungefähr $-\sum_{k=1}^n \log_2(p_{X_k(\omega)}) = -\log_2 \pi_n(\omega)$, womit wir den Anknüpfungspunkt zum Satz von Shannon haben. Dieser trifft also eine Aussage über die Länge eines Binärcodes, der zur Übertragung einer langen Nachricht gebraucht wird.

Ist nun der oben angegebene Code optimal, oder gibt es Codes mit geringerer erwarteter Länge? Antwort gibt der Quellenkodierungssatz, den wir hier mit einer Definition und einem Lemma vorbereiten.

Definition 5.25 (Entropie). Sei $p = (p_e)_{e \in E}$ eine Wahrscheinlichkeitsverteilung auf einer höchstens abzählbaren Menge E . Für $b > 0$ definieren wir

$$H_b(p) := - \sum_{e \in E} p_e \log_b(p_e),$$

wobei wir $0 \log_b(0) := 0$ festlegen. Wir nennen $H(p) := H_e(p)$ (e die Euler'sche Zahl) die **Entropie** und $H_2(p)$ die **binäre Entropie** von p .

Man beachte, dass nur für endliches E die Entropie stets endlich ist.

Lemma 5.26 (Entropie-Ungleichung). Seien b und p wie oben. Ferner sei q eine Sub-Wahrscheinlichkeitsverteilung, also $q_e \geq 0$ für jedes $e \in E$ und $\sum_{e \in E} q_e \leq 1$. Dann gilt

$$H_b(p) \leq - \sum_{e \in E} p_e \log_b(q_e) \quad (5.9)$$

mit Gleichheit genau dann, wenn $H_b(p) = \infty$ oder $q = p$.

Beweis. Ohne Einschränkung können wir mit $b = e$, also mit dem natürlichen Logarithmus rechnen. Es gilt $\log(1 + x) \leq x$ für $x > -1$ mit Gleichheit genau dann, wenn $x = 0$ ist. Ist in (5.9) die linke oder die rechte Seite endlich, so können wir die rechte von der linken Seite abziehen und erhalten

$$\begin{aligned} H(p) + \sum_{e \in E} p_e \log(q_e) &= \sum_{e: p_e > 0} p_e \log(q_e/p_e) \\ &= \sum_{e: p_e > 0} p_e \log \left(1 + \frac{q_e - p_e}{p_e} \right) \\ &\leq \sum_{e: p_e > 0} p_e \frac{q_e - p_e}{p_e} = \sum_{e \in E} (q_e - p_e) \leq 0. \end{aligned}$$

Ist $q \neq p$, so ist $q_e \neq p_e$ für ein $e \in E$ mit $p_e > 0$. Ist dies nun der Fall, so gilt strikte Ungleichheit, falls $H(p) < \infty$. \square

Satz 5.27 (Quellenkodierungssatz). Sei $p = (p_e)_{e \in E}$ eine Wahrscheinlichkeitsverteilung auf dem endlichen Alphabet E . Für jeden binären Präfixcode $C = (c(e), e \in E)$ gilt $L_p(C) \geq H_2(p)$, und es gibt einen solchen Code C mit $L_p(C) \leq H_2(p) + 1$.

Beweis. Der zweite Teil des Satzes wurde durch die obige Konstruktion schon gezeigt. Sei nun also ein Präfixcode gegeben. Sei $L = \max_{e \in E} l(e)$. Für $e \in E$ sei $C_L(e) = \{c \in \{0, 1\}^L : c_k = c_k(e) \text{ für } k \leq l(e)\}$ die Menge aller Binärfolgen der Länge L , die wie $c(e)$ beginnen. Da wir einen Präfixcode vorliegen haben, sind die $C_L(e), e \in E$, paarweise disjunkt und $\bigcup_{e \in E} C_L(e) \subset \{0, 1\}^L$. Setzen wir also $q_e := 2^{-l(e)}$, so ist (beachte: $\#C_L(e) = 2^{L-l(e)}$)

$$\sum_{e \in E} q_e = 2^{-L} \sum_{e \in E} \#C_L(e) \leq 1.$$

Nach Lemma 5.26 gilt $L_p(C) = \sum_{e \in E} p_e l(e) = - \sum_{e \in E} p_e \log_2(q_e) \geq H_2(p)$. \square

Übung 5.3.1. Man zeige die folgende Verbesserung von Satz 5.16: Sind $X_1, X_2, \dots \in \mathcal{L}^2(\mathbf{P})$ paarweise unabhängig mit beschränkten Varianzen, so genügt $(X_n)_{n \in \mathbb{N}}$ dem starken Gesetz der großen Zahl. \clubsuit

Übung 5.3.2. Man zeige: Ist $(X_n)_{n \in \mathbb{N}}$ eine unabhängige Folge identisch verteilter Zufallsvariablen mit $\frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{n \rightarrow \infty} Y$ fast sicher für eine Zufallsvariable Y , so ist $X_1 \in \mathcal{L}^1(\mathbf{P})$ und $Y = \mathbf{E}[X_1]$ fast sicher.

Hinweis: Man zeige zunächst

$$\mathbf{P}[|X_n| > n \text{ für unendlich viele } n] = 0 \iff X_1 \in \mathcal{L}^1(\mathbf{P}). \quad \clubsuit$$

Übung 5.3.3. Sei E endlich und p ein Wahrscheinlichkeitsvektor auf E . Man zeige, dass die Entropie $H(p)$ minimal ist (nämlich Null), falls $p = \delta_e$ für ein $e \in E$ und maximal (nämlich $\log(\#E)$), falls p die Gleichverteilung auf E ist. \clubsuit

Übung 5.3.4 (Subadditivität der Entropie). Für $i = 1, 2$ sei E^i endlich und p^i ein Wahrscheinlichkeitsvektor auf E^i . Sei p ein Wahrscheinlichkeitsvektor auf $E^1 \times E^2$ mit Marginalverteilungen p^1 und p^2 . Das heißt

$$\sum_{e^2 \in E^2} p_{(e^1, e^2)} = p_{e^1}^1 \quad \text{und} \quad \sum_{f^1 \in E^1} p_{(f^1, f^2)} = p_{f^2}^2 \quad \text{für alle } e^1 \in E^1, f^2 \in E^2.$$

Zeige: $H(p) \leq H(p^1) + H(p^2)$. \clubsuit

Übung 5.3.5. Sei $b \in \{2, 3, 4, \dots\}$. Ein b -adischer Präfixcode ist ähnlich definiert wie ein binärer Präfixcode, jedoch sind jetzt als Zeichen die Zahlen $0, 1, \dots, b-1$ zugelassen. Man zeige, dass die Aussage des Quellenkodierungssatzes sinngemäß für b -adische Präfixcodes gilt, mit $H_b(p)$ statt $H_2(p)$. \clubsuit

Übung 5.3.6. Es soll die Effizienz des Morse-Alphabets geprüft werden. Dazu benötigen wir eine Tabelle der Morse-Codes sowie der Häufigkeiten von Buchstaben, etwa in typischen deutschen Texten (die tatsächliche Häufigkeit variiert

natürlich je nach Text). Die folgenden Häufigkeiten wurden [14, Seite 10] entnommen.

Buchstabe	Morsecode	Häufigkeit	Buchstabe	Morsecode	Häufigkeit
A	. -	0.0651	N	- .	0.0978
B	- . . .	0.0189	O	---	0.0251
C	- . - .	0.0306	P	. -- .	0.0079
D	- ..	0.0508	Q	-- . -	0.0002
E	.	0.1740	R	. - .	0.07
F	. . - .	0.0166	S	0.0727
G	-- .	0.0301	T	-	0.0615
H	0.0476	U	. . -	0.0435
I	.. .	0.0755	V -	0.0067
J	. ---	0.0027	W	. --	0.0189
K	- . -	0.0121	X	- . . -	0.0003
L	. - . .	0.0344	Y	- . --	0.0004
M	--	0.0253	Z	-- . .	0.0113

Beim Morsecode steht ' .' für ein kurzes Signal und ' -' für ein langes Signal. Jeder Buchstabe wird mit einem Pausenzeichen beendet. Der Morsecode ist also ein ternärer Präfixcode.

Man bestimme die mittlere Codelänge eines Zeichens und vergleiche sie mit der Entropie H_3 , um die Effizienz des Morsecodes zu bestimmen. ♣

5.4 Konvergenzrate im starken GGZ

Im schwachen Gesetz der großen Zahl hatten wir auch eine Aussage über die Konvergenzgeschwindigkeit gemacht (Satz 5.14), im starken Gesetz der großen Zahl hingegen nicht. Da wir hier nur erste Momente der Zufallsvariablen gefordert hatten, können wir auch keine brauchbaren allgemein gültigen Aussagen erwarten. Nehmen wir hingegen höhere Momente an, so bekommen wir nützliche Aussagen zur Konvergenzgeschwindigkeit.

Das Herzstück des schwachen Gesetzes der großen Zahl ist die Chebyshev'sche Ungleichung. Hier geben wir eine schärfere Ungleichung an, die mit der gleichen oberen Schranke nun das Maximum über alle Partialsummen bis zu einer bestimmten Grenze abschätzt.

Satz 5.28 (Kolmogorov'sche Ungleichung). Seien $n \in \mathbb{N}$ und X_1, X_2, \dots, X_n unabhängige Zufallsvariablen mit $\mathbf{E}[X_i] = 0$ und $\mathbf{Var}[X_i] < \infty$ für $i = 1, \dots, n$. Ferner sei $S_k = X_1 + \dots + X_k$ für $k = 1, \dots, n$. Dann gilt für jedes $t > 0$

$$\mathbf{P}[\max\{S_k : k = 1, \dots, n\} \geq t] \leq \frac{\mathbf{Var}[S_n]}{t^2 + \mathbf{Var}[S_n]} \quad (5.10)$$

sowie die Kolmogorov'sche Ungleichung

$$\mathbf{P}[\max\{|S_k| : k = 1, \dots, n\} \geq t] \leq t^{-2} \mathbf{Var}[S_n]. \quad (5.11)$$

In Satz 11.2 werden wir mit der Doob'schen Ungleichung eine Verallgemeinerung der Kolmogorov'schen Ungleichung kennen lernen.

Beweis. Wir führen eine Zerlegung durch nach dem ersten Zeitpunkt τ , zu dem die Partialsummen den Wert t überschreiten. Sei also

$$\tau := \min \{k \in \{1, \dots, n\} : S_k \geq t\}$$

und $A_k = \{\tau = k\}$ für $k = 1, \dots, n$ sowie

$$A = \bigcup_{k=1}^n A_k = \{\max\{S_k : k = 1, \dots, n\} \geq t\}.$$

Sei $c \geq 0$. Die Zufallsvariable $(S_k + c) \mathbb{1}_{A_k}$ ist messbar bezüglich $\sigma(X_1, \dots, X_k)$, und $S_n - S_k$ ist messbar bezüglich $\sigma(X_{k+1}, \dots, X_n)$. Nach Satz 2.26 sind die beiden Zufallsvariablen unabhängig, und es gilt

$$\mathbf{E}[(S_k + c) \mathbb{1}_{A_k} (S_n - S_k)] = \mathbf{E}[(S_k + c) \mathbb{1}_{A_k}] \mathbf{E}[S_n - S_k] = 0.$$

Offenbar sind die Ereignisse A_1, \dots, A_n paarweise disjunkt, also $\sum_{k=1}^n \mathbb{1}_{A_k} = \mathbb{1}_A \leq 1$. Wir erhalten so

$$\begin{aligned} \mathbf{Var}[S_n] + c^2 &= \mathbf{E}[(S_n + c)^2] \\ &\geq \mathbf{E}\left[\sum_{k=1}^n (S_k + c)^2 \mathbb{1}_{A_k}\right] = \sum_{k=1}^n \mathbf{E}[(S_k + c)^2 \mathbb{1}_{A_k}] \\ &= \sum_{k=1}^n \mathbf{E}\left[((S_k + c)^2 + 2(S_k + c)(S_n - S_k) + (S_n - S_k)^2) \mathbb{1}_{A_k}\right] \quad (5.12) \\ &= \sum_{k=1}^n \mathbf{E}[(S_k + c)^2 \mathbb{1}_{A_k}] + \sum_{k=1}^n \mathbf{E}[(S_n - S_k)^2 \mathbb{1}_{A_k}] \\ &\geq \sum_{k=1}^n \mathbf{E}[(S_k + c)^2 \mathbb{1}_{A_k}]. \end{aligned}$$

Wegen $c \geq 0$ ist $(S_k + c)^2 \mathbb{1}_{A_k} \geq (t + c)^2 \mathbb{1}_{A_k}$, also können wir (5.12) fortsetzen durch

$$\geq \sum_{k=1}^n \mathbf{E} [(t+c)^2 \mathbb{1}_{A_k}] = (t+c)^2 \mathbf{P}[A].$$

Für $c = \mathbf{Var}[S_n]/t \geq 0$ erhalten wir

$$\mathbf{P}[A] \leq \frac{\mathbf{Var}[S_n] + c^2}{(t+c)^2} = \frac{c(t+c)}{(t+c)^2} = \frac{tc}{t^2 + tc} = \frac{\mathbf{Var}[S_n]}{t^2 + \mathbf{Var}[S_n]}.$$

Damit ist (5.10) gezeigt. Um (5.11) zu zeigen, wähle $\bar{\tau} := \min \{k \in \{1, \dots, n\} : |S_k| \geq t\}$ und $\bar{A}_k = \{\bar{\tau} = k\}$ sowie $\bar{A} = \{\bar{\tau} \leq n\}$. Die obige Fortsetzung von (5.12) mit $c > 0$ ist jetzt nicht zulässig. Wenn wir aber $c = 0$ wählen, gilt $S_k^2 \mathbb{1}_{\bar{A}_k} \geq t^2 \mathbb{1}_{\bar{A}_k}$. Mit der selben Rechnung wie in (5.12) erhalten wir $\mathbf{P}[\bar{A}] \leq t^{-2} \mathbf{Var}[S_n]$. \square

Wir folgern aus der Kolmogorov'schen Ungleichung eine erste Verschärfung des starken Gesetzes der großen Zahl.

Satz 5.29. Seien X_1, X_2, \dots unabhängige Zufallsvariablen mit $\mathbf{E}[X_n] = 0$ für jedes $n \in \mathbb{N}$ und $V := \sup\{\mathbf{Var}[X_n] : n \in \mathbb{N}\} < \infty$. Dann gilt für jedes $\varepsilon > 0$

$$\limsup_{n \rightarrow \infty} \frac{|S_n|}{n^{1/2}(\log(n))^{(1/2)+\varepsilon}} = 0 \quad \text{fast sicher.}$$

Beweis. Setze $k_n = 2^n$ und $l(n) = n^{1/2}(\log(n))^{(1/2)+\varepsilon}$ für $n \in \mathbb{N}$. Es gilt $l(k_{n+1})/l(k_n) \xrightarrow{n \rightarrow \infty} \sqrt{2}$. Daher gilt für hinreichend großes n und für $k \in \mathbb{N}$ mit $k_{n-1} \leq k \leq k_n$, dass $|S_k|/l(k) \leq 2|S_k|/l(k_n)$. Also reicht es, für $\delta > 0$ zu zeigen, dass

$$\limsup_{n \rightarrow \infty} l(k_n)^{-1} \max\{|S_k| : k \leq k_n\} \leq \delta \quad \text{fast sicher.} \quad (5.13)$$

Für $\delta > 0$ und $n \in \mathbb{N}$ setze $A_n^\delta := \{\max\{|S_k| : k \leq k_n\} > \delta l(k_n)\}$. Die Kolmogorov'sche Ungleichung liefert

$$\sum_{n=1}^{\infty} \mathbf{P}[A_n^\delta] \leq \sum_{n=1}^{\infty} \delta^{-2} (l(k_n))^{-2} V k_n = \frac{V}{\delta^2 (\log 2)^{1+2\varepsilon}} \sum_{n=1}^{\infty} n^{-1-2\varepsilon} < \infty.$$

Das Borel-Cantelli Lemma liefert nun $\mathbf{P}[\limsup_{n \rightarrow \infty} A_n^\delta] = 0$, also (5.13). \square

Wir werden in Kapitel 22 sehen, dass für unabhängige, identisch verteilte, quadrat-integrierbare, zentrierte Zufallsvariablen X_1, X_2, \dots die folgende Verschärfung gilt

$$\limsup_{n \rightarrow \infty} \frac{|S_n|}{\sqrt{2n \mathbf{Var}[X_1] \log(\log(n))}} = 1 \quad \text{fast sicher.}$$

Die Konvergenzrate ist also in diesem Fall genau bekannt. Sind die X_1, X_2, \dots nicht unabhängig, sondern nur paarweise unabhängig, so verschlechtert sich die Konvergenzrate, wenngleich nicht drastisch: Wir geben hier ohne Beweis einen Satz an, den Rademacher 1922 [151] und Menshov 1923 [123] unabhängig voneinander gefunden haben.

Satz 5.30 (Rademacher–Menshov). Seien X_1, X_2, \dots unkorrelierte quadratintegrierbare zentrierte Zufallsvariablen und $(a_n)_{n \in \mathbb{N}}$ eine wachsende Folge nichtnegativer Zahlen mit

$$\sum_{n=1}^{\infty} (\log n)^2 a_n^{-2} \mathbf{Var}[X_n] < \infty. \quad (5.14)$$

Dann gilt $\limsup_{n \rightarrow \infty} \left| a_n^{-1} \sum_{k=1}^n X_k \right| = 0$ fast sicher.

Beweis. Siehe etwa [137]. □

Bemerkung 5.31. Die Bedingung (5.14) ist scharf in dem Sinne, dass es für jede wachsende Folge $(a_n)_{n \in \mathbb{N}}$ mit $\sum_{n=1}^{\infty} a_n^{-2} (\log n)^2 = \infty$ eine Folge paarweise unabhängiger, quadratintegrierbarer, zentrierter Zufallsvariablen X_1, X_2, \dots mit $\mathbf{Var}[X_n] = 1$ für jedes $n \in \mathbb{N}$ gibt, sodass

$$\limsup_{n \rightarrow \infty} \left| a_n^{-1} \sum_{k=1}^n X_k \right| = \infty \text{ fast sicher.}$$

Siehe [27]. Dort wird ein Beispiel von Tandori [173] für so genannte orthogonale Reihen weiter entwickelt. Siehe auch [127]. ◇

Für Zufallsvariablen mit unendlicher Varianz werden Aussagen über die Konvergenzgeschwindigkeit naturgemäß schwächer. Es gilt beispielsweise (siehe [11]):

Satz 5.32 (Baum und Katz (1965)). Sei $\gamma > 1$, und seien X_1, X_2, \dots u.i.v. sowie $S_n = X_1 + \dots + X_n$ für $n \in \mathbb{N}$. Dann gilt

$$\sum_{n=1}^{\infty} n^{\gamma-2} \mathbf{P}[|S_n|/n > \varepsilon] < \infty \text{ für jedes } \varepsilon > 0 \iff \mathbf{E}[|X_1|^\gamma] < \infty \text{ und } \mathbf{E}[X_1] = 0.$$

Übung 5.4.1. Seien X_1, \dots, X_n unabhängige, reelle Zufallsvariablen und $S_k = X_1 + \dots + X_k$ für jedes $k = 1, \dots, n$. Man zeige, dass für jedes $t > 0$ die Ungleichung von Etemadi gilt:

$$\mathbf{P}\left[\max_{k=1, \dots, n} |S_k| \geq t \right] \leq 3 \max_{k=1, \dots, n} \mathbf{P}[|S_k| \geq t/3]. \quad \clubsuit$$

5.5 Der Poissonprozess

Wir wollen ein Modell für die Anzahl der Klicks entwickeln, die ein Geigerzähler in einem (Zeit-)Intervall $I = (a, b]$ macht. Die Anzahl der Klicks soll dabei

- zufällig sein und unabhängig für disjunkte Intervalle,
- zeitlich homogen in dem Sinne, dass die Anzahl der Klicks in $I = (a, b]$ die selbe Verteilung hat wie die Anzahl der Klicks in $c + I = (a + c, b + c]$,
- einen Erwartungswert besitzen,
- keine Doppelpunkte aufweisen: der Zähler macht zu jedem Zeitpunkt nur einen Klick.

Wir formalisieren diese Forderungen, indem wir die Notation einführen:

$$\mathcal{I} := \{(a, b] : a, b \in [0, \infty), a \leq b\},$$

$$\ell((a, b]) := b - a \quad (\text{die Länge des Intervalls } I = (a, b]).$$

Für $I \in \mathcal{I}$ sei N_I die Anzahl der Klicks nach Zeitpunkt a und nicht später als b . Speziell setzen wir $N_t := N_{(0,t]}$ für die Gesamtzahl aller Klicks bis zur Zeit t . Die obigen Forderungen lassen sich nun übersetzen zu: $(N_I, I \in \mathcal{I})$ ist eine Familie von Zufallsvariablen mit Werten in \mathbb{N}_0 mit den Eigenschaften

- (P1) $N_{I \cup J} = N_I + N_J$, falls $I \cap J = \emptyset$ und $I \cup J \in \mathcal{I}$ ist.
- (P2) Die Verteilung von N_I hängt nur von der Länge von I ab: $\mathbf{P}_{N_I} = \mathbf{P}_{N_J}$ für alle $I, J \in \mathcal{I}$ mit $\ell(I) = \ell(J)$.
- (P3) Ist $\mathcal{J} \subset \mathcal{I}$ mit $I \cap J = \emptyset$ für alle $I, J \in \mathcal{J}$ mit $I \neq J$, so ist $(N_J, J \in \mathcal{J})$ eine unabhängige Familie.
- (P4) Für jedes $I \in \mathcal{I}$ gilt $\mathbf{E}[N_I] < \infty$.
- (P5) Es gilt $\limsup_{\varepsilon \downarrow 0} \varepsilon^{-1} \mathbf{P}[N_\varepsilon \geq 2] = 0$.

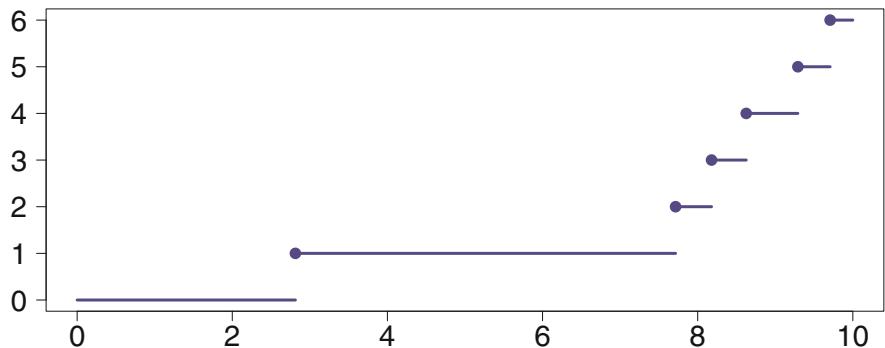


Abb. 5.1. Simulation eines Poissonprozesses mit Rate $\alpha = 0.5$.

Die Bedeutung von (P5) erklärt sich durch die folgende Rechnung. Setzen wir $\lambda := \limsup_{\varepsilon \downarrow 0} \varepsilon^{-1} \mathbf{P}[N_\varepsilon \geq 2]$, so ist für jedes $n \in \mathbb{N}$ und $\varepsilon > 0$

$$\mathbf{P}[N_{2^{-n}} \geq 2] \geq [2^{-n}/\varepsilon] \mathbf{P}[N_\varepsilon \geq 2] - [2^{-n}/\varepsilon]^2 \mathbf{P}[N_\varepsilon \geq 2]^2.$$

Also gilt (indem man ε in geeigneter Weise nach 0 gehen lässt)

$$2^n \mathbf{P}[N_{2^{-n}} \geq 2] \geq \lambda - 2^{-n} \lambda^2 \xrightarrow{n \rightarrow \infty} \lambda.$$

Aus der Tatsache, dass $(1 - a_k/k)^k \xrightarrow{k \rightarrow \infty} e^{-a}$ gilt, falls $a_k \xrightarrow{k \rightarrow \infty} a$, folgern wir nun

$$\begin{aligned} \mathbf{P}[\text{es gibt einen Doppelklick in } (0, 1)] &= \lim_{n \rightarrow \infty} \mathbf{P}\left[\bigcup_{k=0}^{2^n-1} \{N_{(k2^{-n}, (k+1)2^{-n}]} \geq 2\}\right] \\ &= 1 - \lim_{n \rightarrow \infty} \mathbf{P}\left[\bigcap_{k=0}^{2^n-1} \{N_{(k2^{-n}, (k+1)2^{-n}]} \leq 1\}\right] \\ &= 1 - \lim_{n \rightarrow \infty} \prod_{k=0}^{2^n-1} \mathbf{P}[N_{(k2^{-n}, (k+1)2^{-n}]} \leq 1] \\ &= 1 - \lim_{n \rightarrow \infty} (1 - \mathbf{P}[N_{2^{-n}} \geq 2])^{2^n} = 1 - e^{-\lambda}. \end{aligned}$$

Wir müssen also $\lambda = 0$ fordern; dies ist aber gerade (P5).

Der nächste Satz zeigt, dass die Bedingungen (P1) – (P5) die Zufallsvariablen $(N_I, I \in \mathcal{I})$ eindeutig charakterisieren und zwar als Poissonprozess.

Definition 5.33 (Poissonprozess). Eine Familie $(N_t, t \geq 0)$ von \mathbb{N}_0 -wertigen Zufallsvariablen heißt **Poissonprozess mit Intensität $\alpha \geq 0$** , falls $N_0 = 0$ und:

(i) Für jedes $n \in \mathbb{N}$ und je $n + 1$ Zahlen $0 = t_0 < t_1 < \dots < t_n$ ist die Familie $(N_{t_i} - N_{t_{i-1}}, i = 1, \dots, n)$ unabhängig.

(ii) Für $t > s \geq 0$ ist $N_t - N_s$ Poisson-verteilt mit Parameter $\alpha(t - s)$, also

$$\mathbf{P}[N_t - N_s = k] = e^{-\alpha(t-s)} \frac{(\alpha(t-s))^k}{k!} \quad \text{für jedes } k \in \mathbb{N}_0.$$

Die Existenz eines Poissonprozesses ist an dieser Stelle noch nicht gesichert. Darauf kommen wir in Satz 5.36 zurück.

Satz 5.34. Erfüllt $(N_I, I \in \mathcal{I})$ die Bedingungen (P1) – (P5), so ist $(N_{(0,t]}, t \geq 0)$ ein Poissonprozess mit Intensität $\alpha := \mathbf{E}[N_{(0,1]}]$. Ist umgekehrt $(N_t, t \geq 0)$ ein Poissonprozess, so erfüllt $(N_t - N_s, (s, t] \in \mathcal{I})$ die Bedingungen (P1) – (P5).

Beweis. Sei zunächst $(N_t, t \geq 0)$ ein Poissonprozess mit Intensität $\alpha \geq 0$. Per Konstruktion gilt (P1). Für $I = (a, b]$ ist offenbar $\mathbf{P}_{N_I} = \text{Poi}_{\alpha(b-a)} = \text{Poi}_{\alpha\ell(I)}$. Also gilt (P2). Wegen (i) gilt (P3). Offenbar ist $\mathbf{E}[N_I] = \alpha\ell(I) < \infty$, also gilt (P4). Schließlich ist $\mathbf{P}[N_\varepsilon \geq 2] = 1 - e^{-\alpha\varepsilon} - \alpha\varepsilon$, also

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \mathbf{P}[N_\varepsilon \geq 2] = \lim_{\varepsilon \downarrow 0} \frac{1 - e^{-\alpha\varepsilon}}{\varepsilon} - \alpha = 0.$$

Also gilt auch (P5).

Erfüllt nun $(N_I, I \in \mathcal{I})$ die Bedingungen (P1) – (P5). Setze $\alpha(t) := \mathbf{E}[N_t]$. Dann ist (wegen (P1) und (P2))

$$\alpha(s+t) = \mathbf{E}[N_{(0,s]} + N_{(s,s+t]}] = \mathbf{E}[N_{(0,s]}] + \mathbf{E}[N_{(0,t]}] = \alpha(s) + \alpha(t).$$

Da $t \mapsto \alpha(t)$ monoton wachsend ist, folgt hieraus sogar $\alpha(t) = t\alpha(1)$ für jedes $t \geq 0$. Wir setzen $\alpha := \alpha(1)$ und erhalten $\mathbf{E}[N_I] = \alpha\ell(I)$. Wir müssen nur noch zeigen, dass $\mathbf{P}_{N_t} = \text{Poi}_{\alpha t}$ gilt. Um den Satz über die Poissonapproximation (Satz 3.7) zu verwenden, zerlegen wir für festes $n \in \mathbb{N}$, das Intervall $(0, t]$ in 2^n disjunkte gleich lange Intervalle

$$I^n(k) := ((k-1)2^{-n}t, k2^{-n}t], \quad k = 1, \dots, 2^n,$$

und setzen $X^n(k) := N_{I^n(k)}$ sowie

$$\overline{X}^n(k) := \begin{cases} 1, & \text{falls } X^n(k) \geq 1, \\ 0, & \text{sonst.} \end{cases}$$

Nach den Annahmen (P2) und (P3) sind $(X^n(k), k = 1, \dots, 2^n)$ unabhängig und identisch verteilt. Daher ist auch $(\overline{X}^n(k), k = 1, \dots, 2^n)$ unabhängig und identisch verteilt, nämlich $\overline{X}^n(k) \sim \text{Ber}_{p_n}$, wobei $p_n = \mathbf{P}[N_{2^{-n}t} \geq 1]$.

Schließlich setzen wir $N_t^n := \sum_{k=1}^{2^n} \overline{X}^n(k)$. Dann ist $N_t^n \sim b_{2^n, p_n}$. Offenbar ist $N_t^{n+1} - N_t^n \geq 0$. Nun gilt nach (P5)

$$\mathbf{P}[N_t \neq N_t^n] \leq \sum_{k=1}^{2^n} \mathbf{P}[X^n(k) \geq 2] = 2^n \mathbf{P}[N_{2^{-n}t} \geq 2] \xrightarrow{n \rightarrow \infty} 0. \quad (5.15)$$

Also ist $\mathbf{P}\left[N_t = \lim_{n \rightarrow \infty} N_t^n\right] = 1$. Nach dem Satz über monotone Konvergenz gilt

$$\alpha t = \mathbf{E}[N_t] = \lim_{n \rightarrow \infty} \mathbf{E}[N_t^n] = \lim_{n \rightarrow \infty} p_n 2^n.$$

Nach dem Satz über Poisson-Approximation (Satz 3.7) gilt daher für jedes $l \in \mathbb{N}_0$

$$\mathbf{P}[N_t = l] = \lim_{n \rightarrow \infty} \mathbf{P}[N_t^n = l] = \text{Poi}_{\alpha t}(\{l\}).$$

Also ist $\mathbf{P}_{N_t} = \text{Poi}_{\alpha t}$. □

Bislang steht noch der Nachweis aus, dass es überhaupt Poissonprozesse gibt. Wir geben hier ein allgemeines, zweistufiges Konstruktionsprinzip an, das später in Kapitel 24 in ähnlicher Form auch für die Konstruktion von Poissonprozessen auf allgemeineren Räumen (statt der Zeitmenge $[0, \infty)$) verwendet wird. Im ersten Schritt wird die Anzahl der Sprünge im Intervall $[0, 1]$ bestimmt. Im zweiten Schritt werden diese Sprünge unabhängig und uniform im Intervall $[0, 1]$ verteilt. Genau genommen wird so der Poissonprozess nur mit der Zeitmenge $[0, 1]$ konstruiert, aber es ist klar, dass wir für das Intervall $(1, 2]$ diese Konstruktion einfach unabhängig noch einmal ausführen können, dann für das Intervall $(2, 3]$ und so weiter. Wir beschränken uns hier also auf die Konstruktion des Poissonprozesses mit Zeitmenge $[0, 1]$ und verweisen auf Übung 5.5.1.

Sei $\alpha > 0$ gegeben und L eine Poi_α -verteilte Zufallsvariable. Seien ferner X_1, X_2, \dots unabhängige Zufallsvariablen, die uniform im Einheitsintervall verteilt sind: $X_k \sim \mathcal{U}_{(0,1)}$ für jedes k . Wir nehmen an, dass $\{L, X_1, X_2, \dots\}$ eine unabhängige Familie von Zufallsvariablen ist. Wir definieren dann $N = (N_t)_{t \in [0,1]}$ durch

$$N_t := \sum_{l=1}^L \mathbb{1}_{(0,t]}(X_l) \quad \text{für } t \in [0, 1]. \quad (5.16)$$

Satz 5.35. *Die in (5.16) definierte Familie N von Zufallsvariablen ist ein Poissonprozess mit Intensität α und Zeitmenge $[0, 1]$.*

Beweis. Wir müssen zeigen, dass die Zuwächse von N in endlich vielen paarweise disjunkten Intervallen unabhängig und Poisson-verteilt sind. Sei also $m \in \mathbb{N}$ gegeben und $0 = t_0 < t_1 < \dots < t_m = 1$. Wir verwenden die Abkürzungen $p_i := t_i - t_{i-1}$ und $\lambda_i = \alpha \cdot (t_i - t_{i-1})$ und zeigen

$$(N_{t_i} - N_{t_{i-1}})_{i=1, \dots, m} \text{ ist unabhängig} \quad (5.17)$$

und

$$N_{t_i} - N_{t_{i-1}} \sim \text{Poi}_{\lambda_i} \quad \text{für alle } i = 1, \dots, m. \quad (5.18)$$

Dies ist gleichwertig damit, dass für jede Wahl $k_1, \dots, k_m \in \mathbb{N}_0$ gilt:

$$\mathbf{P}[N_{t_i} - N_{t_{i-1}} = k_i \text{ für jedes } i = 1, \dots, m] = \prod_{i=1}^m \left(e^{-\lambda_i} \frac{\lambda_i^{k_i}}{k_i!} \right). \quad (5.19)$$

Wir schreiben

$$M_{n,i} := \#\{l \leq n : t_{i-1} < X_l \leq t_i\} = \sum_{l=1}^n \mathbb{1}_{(t_{i-1}, t_i]}(X_l).$$

Nach Übung 2.2.3 ist $(M_{n,1}, \dots, M_{n,m})$ multinomialverteilt mit Parametern n und $p = (p_1, \dots, p_m)$. Nehmen wir jetzt also $n := k_1 + \dots + k_m$ an, so gilt

$$\mathbf{P}[M_{n,1} = k_1, \dots, M_{n,m} = k_m] = \frac{n!}{k_1! \cdots k_m!} p_1^{k_1} \cdots p_m^{k_m}.$$

Um (5.19) zu zeigen, beachte man, dass im Ereignis von (5.19) notwendigerweise $L = n$ gilt und dass L und $(M_{n,1}, \dots, M_{n,m})$ unabhängig sind. Daher ist

$$\begin{aligned} & \mathbf{P}[N_{t_i} - N_{t_{i-1}} = k_i \text{ für } i = 1, \dots, m] \\ &= \mathbf{P}[\{N_{t_i} - N_{t_{i-1}} = k_i \text{ für } i = 1, \dots, m\} \cap \{L = n\}] \\ &= \mathbf{P}[\{M_{n,1} = k_1, \dots, M_{n,m} = k_m\} \cap \{L = n\}] \\ &= \mathbf{P}[M_{n,1} = k_1, \dots, M_{n,m} = k_m] \cdot \mathbf{P}[L = n] \\ &= \frac{n!}{k_1! \cdots k_m!} p_1^{k_1} \cdots p_m^{k_m} e^{-\alpha} \frac{\alpha^n}{n!} \\ &= \prod_{i=1}^m \left(e^{-\lambda_i} \frac{\lambda_i^{k_i}}{k_i!} \right). \end{aligned} \quad \square$$

Zum Abschluss dieses Abschnitts stellen wir eine weitere, instruktive Konstruktion des Poissonprozesses vor, die auf den Wartezeiten zwischen den Klicks basiert, oder formal zwischen den Unstetigkeitsstellen der Abbildung $t \mapsto N_t(\omega)$. Wie groß ist die Wahrscheinlichkeit, dass wir zur Zeit s auf den nächsten Klick des Zählers länger als t Zeiteinheiten warten müssen? Wenn wir die Klicks als Poissonprozess mit Intensität α modellieren, ist diese Wahrscheinlichkeit

$$\mathbf{P}[N_{(s,s+t]} = 0] = e^{-\alpha t}.$$

Mithin ist die Wartezeit auf den nächsten Klick exponentialverteilt mit Parameter α . Außerdem sollten die Wartezeiten unabhängig voneinander sein. Wir nehmen nun die Wartezeiten als Startpunkt der Betrachtung und konstruieren hieraus den Poissonprozess.

Sei W_1, W_2, \dots eine unabhängige Familie von exponentialverteilten Zufallsvariablen mit Parameter $\alpha > 0$, also $\mathbf{P}[W_n > x] = e^{-\alpha x}$. Wir setzen

$$T_n := \sum_{k=1}^n W_k$$

und interpretieren W_n als die Wartezeit zwischen dem $(n-1)$ -ten und dem n -ten Klick. T_n ist der Zeitpunkt des n -ten Klicks.

In Anlehnung an diese Intuition definieren wir

$$N_t := \max\{n \in \mathbb{N} : T_n \leq t\}$$

als die Anzahl der Klicks bis zur Zeit t . Es ist dann

$$\{N_t = k\} = \{T_k \leq t < T_{k+1}\}.$$

Speziell ist also N_t eine Zufallsvariable.

Satz 5.36. Die Familie $(N_t, t \geq 0)$ ist ein Poissonprozess mit Intensität α .

Beweis. (Wir folgen dem Beweis in [65, Satz 3.34].) Wir müssen zeigen, dass für jedes $n \in \mathbb{N}$ und jede Folge $0 = t_0 < t_1 < \dots < t_n$ gilt: $(N_{t_i} - N_{t_{i-1}}, i = 1, \dots, n)$ ist unabhängig und $N_{t_i} - N_{t_{i-1}} \sim \text{Poi}_{\alpha(t_i - t_{i-1})}$. Wir wissen zwar, dass es nicht ausreicht, dies nur für $n = 2$ zu zeigen. Allerdings wird der Schreibaufwand für $n \geq 3$ extrem groß, und das Prinzip, wie man den Beweis für allgemeines $n \in \mathbb{N}$ führt, wird klar, wenn man $n = 2$ untersucht hat. Daher beschränken wir uns hier auf den Fall $n = 2$.

Wir zeigen also, für $0 < s < t$ und $l, k \in \mathbb{N}_0$

$$\mathbf{P}[N_s = k, N_t - N_s = l] = \left(e^{-\alpha s} \frac{(\alpha s)^k}{k!} \right) \left(e^{-\alpha(t-s)} \frac{(\alpha(t-s))^l}{l!} \right). \quad (5.20)$$

Hieraus folgt, dass N_s und $(N_t - N_s)$ unabhängig sind. Außerdem folgt, indem wir über $k \in \mathbb{N}_0$ summieren, dass $N_t - N_s \sim \text{Poi}_{\alpha(t-s)}$.

Nach Korollar 2.22 hat die Verteilung $\mathbf{P}_{(W_1, \dots, W_{k+l+1})}$ die Dichte

$$x \mapsto \alpha^{k+l+1} e^{-\alpha S_{k+l+1}(x)},$$

wobei $S_n(x) := x_1 + \dots + x_n$. Es reicht nun, $l \geq 1$ zu betrachten, da wir den ($l = 0$)-Term durch die Normierung des W-Maßes erhalten. Sei also $l \geq 1$. Wir erhalten

$$\begin{aligned} \mathbf{P}[N_s = k, N_t - N_s = l] &= \mathbf{P}[T_k \leq s < T_{k+1}, T_{k+l} \leq t < T_{k+l+1}] \\ &= \int_0^\infty \cdots \int_0^\infty dx_1 \cdots dx_{k+l+1} \\ &\quad \alpha^{k+l+1} e^{-\alpha S_{k+l+1}(x)} \mathbb{1}_{\{S_k(x) \leq s < S_{k+1}(x)\}} \mathbb{1}_{\{S_{k+l}(x) \leq t < S_{k+l+1}(x)\}}. \end{aligned}$$

Wir integrieren nun sukzessive, mit x_{k+l+1} beginnend. Im ersten Schritt substituieren wir $z = S_{k+l+1}(x)$ und erhalten

$$\int_0^\infty dx_{k+l+1} \alpha e^{-\alpha S_{k+l+1}(x)} \mathbb{1}_{\{S_{k+l+1}(x) > t\}} = \int_t^\infty dz \alpha e^{-\alpha z} = e^{-\alpha t}.$$

Nun halten wir x_1, \dots, x_k fest und erhalten für die restlichen Variablen durch die Substitution $y_1 = S_{k+1}(x) - s, y_2 = x_{k+2}, \dots, y_l = x_{k+l}$

$$\begin{aligned} &\int_0^\infty \cdots \int_0^\infty dx_{k+1} \cdots dx_{k+l} \mathbb{1}_{\{s < S_{k+1}(x) \leq S_{k+l} \leq t\}} \\ &= \int_0^\infty \cdots \int_0^\infty dy_1 \cdots dy_l \mathbb{1}_{\{y_1 + \dots + y_l \leq t-s\}} = \frac{(t-s)^l}{l!}. \end{aligned}$$

(Dies erhält man zum Beispiel per Induktion über l .) Wir integrieren nun über die verbleibenden Variablen x_1, \dots, x_k und erhalten

$$\int_0^\infty \cdots \int_0^\infty dx_1 \cdots dx_k \mathbb{1}_{\{S_k(x) \leq s\}} = \frac{s^k}{k!}.$$

Insgesamt erhalten wir

$$\mathbf{P}[N_s = k, N_t - N_s = l] = e^{-\alpha t} \alpha^{k+l} \frac{s^k}{k!} \frac{(t-s)^l}{l!},$$

also (5.20). \square

Übung 5.5.1. Seien $L_n, X_k^n, k, n \in \mathbb{N}$ unabhängige Zufallsvariablen mit $L_n \sim \text{Poi}_\alpha$ und $X_k^n \sim \mathcal{U}_{(n-1, n]}$ (Gleichverteilung auf $(n-1, n]$) für alle $k, n \in \mathbb{N}$. Setze

$$N_t := \#\{(k, n) \in \mathbb{N}^2 : k \leq L_n \text{ und } X_k^n \leq t\}.$$

Man zeige, dass $(N_t)_{t \geq 0}$ ein Poissonprozess mit Intensität α ist. \clubsuit

Übung 5.5.2. Sei $T > 0$ und seien X_1, X_2, \dots unabhängige auf $[0, 1]$ gleichverteilte Zufallsvariablen. Sei

$$N := \max \{n \in \mathbb{N}_0 : X_1 + \dots + X_n \leq T\}.$$

Bestimme $\mathbf{E}[N]$. \clubsuit

6 Konvergenzsätze

Im starken und schwachen Gesetz der großen Zahl hatten wir implizit schon die Begriffe von fast sicherer und stochastischer Konvergenz von Folgen von Zufallsvariablen kennen gelernt und gesehen, dass die fast sichere die stochastische Konvergenz impliziert. In diesem Kapitel definieren wir die Begriffe von fast sicherer und stochastischer Konvergenz sowie Konvergenz im Mittel von Folgen messbarer Abbildungen und setzen sie in Beziehung zueinander. Eine Schlüsselrolle kommt dabei dem Konzept der gleichgradigen Integrierbarkeit zu.

6.1 Fast-überall- und stochastische Konvergenz

Im Folgenden ist $(\Omega, \mathcal{A}, \mu)$ stets ein σ -endlicher Maßraum. Wir definieren zunächst Fast-überall-Konvergenz und stochastische Konvergenz in metrischen Räumen und vergleichen dann beide Konzepte miteinander. Hierfür benötigen wir zunächst zwei Lemmata, die sicherstellen, dass die Abstandsfunktion zweier messbarer Abbildungen wieder messbar ist. Sei im Folgenden (E, d) ein separabler, metrischer Raum mit Borel'scher σ -Algebra $\mathcal{B}(E)$. „Separabel“ heißt dabei bekanntlich, dass es eine abzählbare, dichte Teilmenge gibt. Für $x \in E$ und $r > 0$ bezeichnen wir mit $B_r(x) = \{y \in E : d(x, y) < r\}$ die offene Kugel um x mit Radius r .

Lemma 6.1. *Seien $f, g : \Omega \rightarrow E$ messbar bezüglich $\mathcal{A} - \mathcal{B}(E)$. Dann ist die Abbildung $H : \Omega \rightarrow [0, \infty)$, $\omega \mapsto d(f(\omega), g(\omega))$ messbar bezüglich $\mathcal{A} - \mathcal{B}([0, \infty))$.*

Beweis. Sei $F \subset E$ abzählbar und dicht. Nach der Dreiecksungleichung ist $d(x, z) + d(z, y) \geq d(x, y)$ für alle $x, y \in E$ und für $z \in F$. Sei $(z_n)_{n \in \mathbb{N}}$ eine Folge in F mit $z_n \xrightarrow{n \rightarrow \infty} x$. Da d stetig ist, gilt $d(x, z_n) + d(z_n, y) \xrightarrow{n \rightarrow \infty} d(x, y)$. Insgesamt haben wir $\inf_{z \in F} (d(x, z) + d(z, y)) = d(x, y)$. Da $x \mapsto d(x, z)$ stetig ist, also messbar, sind auch $f_z, g_z : \Omega \rightarrow [0, \infty)$ mit $f_z(\omega) = d(f(\omega), z)$ und $g_z(\omega) = d(g(\omega), z)$ messbar, also auch $f_z + g_z$ sowie $H = \inf_{z \in F} (f_z + g_z)$.

(Etwas systematischer lässt sich der Satz zeigen, wenn wir benutzen, dass (f, g) messbar ist bezüglich $\mathcal{A} - \mathcal{B}(E \times E)$ (dies folgt aus Satz 14.8) und $d : E \times E \rightarrow [0, \infty)$ stetig also messbar bezüglich $\mathcal{B}(E \times E) - \mathcal{B}([0, \infty))$. Als Verknüpfung messbarer Abbildungen ist dann auch $\omega \mapsto d(f(\omega), g(\omega))$ messbar.) \square

Seien $f, f_1, f_2, \dots : \Omega \rightarrow E$ messbar bezüglich $\mathcal{A} - \mathcal{B}(E)$.

Definition 6.2. Wir sagen: $(f_n)_{n \in \mathbb{N}}$ konvergiert gegen f

(i) **μ -stochastisch** (oder **dem Maße nach**), in Formeln $f_n \xrightarrow{\text{stoch}} f$, wenn für jedes $A \in \mathcal{A}$ mit $\mu(A) < \infty$ und für jedes $\varepsilon > 0$ gilt, dass

$$\mu(\{d(f, f_n) > \varepsilon\} \cap A) \xrightarrow{n \rightarrow \infty} 0.$$

(ii) **μ -fast überall**, in Formeln $f_n \xrightarrow{\text{f.ü.}} f$, wenn es eine μ -Nullmenge $N \in \mathcal{A}$ gibt, sodass für jedes $\omega \in \Omega \setminus N$ gilt, dass

$$d(f(\omega), f_n(\omega)) \xrightarrow{n \rightarrow \infty} 0.$$

Ist μ ein W-Maß, so sagen wir in diesem Fall auch, dass $(f_n)_{n \in \mathbb{N}}$ **fast sicher** konvergiert und schreiben $f_n \xrightarrow{\text{f.s.}} f$. Gelegentlich werden wir die Hinweise „fast überall“ und „fast sicher“ auch weglassen.

Bemerkung 6.3. Seien $A_1, A_2, \dots \in \mathcal{A}$ mit $A_N \uparrow \Omega$ und $\mu(A_N) < \infty$ für jedes $N \in \mathbb{N}$. Dann ist F.ü.-Konvergenz äquivalent zur F.ü.-Konvergenz auf jedem A_N . \diamond

Bemerkung 6.4. Fast-überall-Konvergenz impliziert die stochastische: Sei zu $\varepsilon > 0$

$$D_n(\varepsilon) = \{d(f, f_m) > \varepsilon \text{ für ein } m \geq n\}.$$

Dann gilt $D(\varepsilon) := \bigcap_{n=1}^{\infty} D_n(\varepsilon) \subset N$, wobei N die Nullmenge aus der Definition der F.ü.-Konvergenz ist. Die σ -Stetigkeit von oben von μ impliziert

$$\mu(D_n(\varepsilon) \cap A) \xrightarrow{n \rightarrow \infty} \mu(D(\varepsilon) \cap A) = 0$$

für jedes $A \in \mathcal{A}$ mit $\mu(A) < \infty$. \diamond

Bemerkung 6.5. Stochastische oder Fast-überall-Konvergenz legen den Grenzwert eindeutig fest bis auf Gleichheit fast überall. In der Tat: Sei $f_n \xrightarrow{\text{stoch}} f$ und $f_n \xrightarrow{\text{stoch}} g$. Seien $A_1, A_2, \dots \in \mathcal{A}$ mit $A_n \uparrow \Omega$ und $\mu(A_n) < \infty$ für jedes $n \in \mathbb{N}$. Dann ist (wegen $d(f, g) \leq d(f, f_n) + d(g, f_n)$) für jedes $m \in \mathbb{N}$ und $\varepsilon > 0$

$$\begin{aligned} \mu(A_m \cap \{d(f, g) > \varepsilon\}) \\ \leq \mu(A_m \cap \{d(f, f_n) > \varepsilon/2\}) + \mu(A_m \cap \{d(g, f_n) > \varepsilon/2\}) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Also ist $\mu(\{d(f, g) > 0\}) = 0$. \diamond

Bemerkung 6.6. Im Allgemeinen impliziert stochastische Konvergenz nicht F.ü.-Konvergenz. In der Tat: Sei $(X_n)_{n \in \mathbb{N}}$ eine unabhängige Familie von Zufallsvariablen mit $X_n \sim \text{Ber}_{1/n}$. Dann gilt $X_n \xrightarrow{\text{stoch}} 0$, jedoch ist nach dem Lemma von Borel-Cantelli $\limsup_{n \rightarrow \infty} X_n = 1$ fast sicher. \diamond

Satz 6.7. Seien $A_1, A_2, \dots \in \mathcal{A}$ mit $A_N \uparrow \Omega$ und $\mu(A_N) < \infty$ für jedes $N \in \mathbb{N}$. Für messbare $f, g : \Omega \rightarrow E$ setze

$$\tilde{d}(f, g) := \sum_{N=1}^{\infty} \frac{2^{-N}}{1 + \mu(A_N)} \int_{A_N} (1 \wedge d(f(\omega), g(\omega))) \mu(d\omega). \quad (6.1)$$

Dann ist \tilde{d} eine Metrik, die die stochastische Konvergenz erzeugt: Sind f, f_1, f_2, \dots messbar, so gilt

$$f_n \xrightarrow{\text{stoch}} f \iff \tilde{d}(f, f_n) \xrightarrow{n \rightarrow \infty} 0.$$

Beweis. Für $N \in \mathbb{N}$ setze

$$\tilde{d}_N(f, g) := \int_{A_N} (1 \wedge d(f(\omega), g(\omega))) \mu(d\omega).$$

Genau dann gilt $\tilde{d}(f, f_n) \xrightarrow{n \rightarrow \infty} 0$, wenn $\tilde{d}_N(f, f_n) \xrightarrow{n \rightarrow \infty} 0$ für jedes $N \in \mathbb{N}$.

„ \Rightarrow “ Es gelte $f_n \xrightarrow{\text{stoch}} f$. Dann ist für jedes $\varepsilon \in (0, 1)$

$$\tilde{d}_N(f, f_n) \leq \mu(A_N \cap \{d(f, f_n) > \varepsilon\}) + \varepsilon \mu(A_N) \xrightarrow{n \rightarrow \infty} \varepsilon \mu(A_N).$$

Da $\varepsilon \in (0, 1)$ beliebig war, gilt $\tilde{d}_N(f, f_n) \xrightarrow{n \rightarrow \infty} 0$.

„ \Leftarrow “ Es gelte $\tilde{d}(f, f_n) \xrightarrow{n \rightarrow \infty} 0$. Sei $B \in \mathcal{A}$ mit $\mu(B) < \infty$. Wähle $\delta > 0$ und $N \in \mathbb{N}$ so groß, dass $\mu(B \setminus A_N) < \delta$. Dann ist für $\varepsilon \in (0, 1)$

$$\begin{aligned} \mu(B \cap \{d(f, f_n) > \varepsilon\}) &\leq \delta + \mu(A_N \cap \{d(f, f_n) > \varepsilon\}) \\ &\leq \delta + \varepsilon^{-1} \tilde{d}_N(f, f_n) \xrightarrow{n \rightarrow \infty} \delta. \end{aligned}$$

Da $\delta > 0$ beliebig war, folgt $\mu(B \cap \{d(f, f_n) > \varepsilon\}) \xrightarrow{n \rightarrow \infty} 0$, also $f_n \xrightarrow{\text{stoch}} f$. \square

Wir betrachten nun den wichtigen Fall $E = \mathbb{R}$ mit der euklidischen Metrik. Hier haben wir durch das Integral einen weiteren Konvergenzbegriff zur Verfügung.

Definition 6.8 (Konvergenz im Mittel). Seien $f, f_1, f_2, \dots \in \mathcal{L}^1(\mu)$. Wir sagen $(f_n)_{n \in \mathbb{N}}$ konvergiere **im Mittel** gegen f , in Formeln

$$f_n \xrightarrow{L^1} f,$$

falls $\|f_n - f\|_1 \xrightarrow{n \rightarrow \infty} 0$.

Bemerkung 6.9. Gilt $f_n \xrightarrow{L^1} f$, so gilt insbesondere $\int f_n d\mu \xrightarrow{n \rightarrow \infty} \int f d\mu$. \diamond

Bemerkung 6.10. Gilt $f_n \xrightarrow{L^1} f$ und $f_n \xrightarrow{L^1} g$, so ist $f = g$ fast überall. In der Tat ist nach der Dreiecksungleichung $\|f - g\|_1 \leq \|f_n - f\|_1 + \|f_n - g\|_1 \xrightarrow{n \rightarrow \infty} 0$. \diamond

Bemerkung 6.11. L^1 -Konvergenz und F.ü.-Konvergenz implizieren jeweils stochastische Konvergenz. Alle anderen Implikationen sind im Allgemeinen falsch. \diamond

Satz 6.12 (Schnelle Konvergenz). Sei (E, d) ein separabler, metrischer Raum. Damit die Folge $(f_n)_{n \in \mathbb{N}}$ messbarer Abbildungen $\Omega \rightarrow E$ fast überall konvergiert, ist hinreichend, dass eine der folgenden Bedingungen gilt.

(i) Es gilt $E = \mathbb{R}$, es gibt ein $p \in [1, \infty)$ mit $f_n \in \mathcal{L}^p(\mu)$ für jedes $n \in \mathbb{N}$, und es gibt ein $f \in \mathcal{L}^p(\mu)$ mit $\sum_{n=1}^{\infty} \|f_n - f\|_p < \infty$.

(ii) Es gibt ein messbares f mit $\sum_{n=1}^{\infty} \mu(A \cap \{d(f, f_n) > \varepsilon\}) < \infty$ für jedes $\varepsilon > 0$ und für jedes $A \in \mathcal{A}$ mit $\mu(A) < \infty$.

In beiden Fällen gilt $f_n \xrightarrow{n \rightarrow \infty} f$ fast überall.

(iii) E ist vollständig, und es gibt eine summierbare Folge $(\varepsilon_n)_{n \in \mathbb{N}}$, sodass

$$\sum_{n=1}^{\infty} \mu(A \cap \{d(f_n, f_{n+1}) > \varepsilon_n\}) < \infty \quad \text{für jedes } A \in \mathcal{A} \text{ mit } \mu(A) < \infty.$$

Beweis. Offenbar impliziert (i) schon (ii), denn nach der Markov'schen Ungleichung ist $\mu(\{|f - f_n| > \varepsilon\}) \leq \varepsilon^{-p} \|f - f_n\|_p^p$.

Nach Bemerkung 6.3 reicht es, den Fall $\mu(\Omega) < \infty$ zu betrachten.

Gelte nun (ii). Sei $B_n(\varepsilon) = \{d(f, f_n) > \varepsilon\}$ und $B(\varepsilon) = \limsup_{n \rightarrow \infty} B_n(\varepsilon)$. Das Lemma von Borel-Cantelli liefert $\mu(B(\varepsilon)) = 0$. Sei $N = \bigcup_{n=1}^{\infty} B(1/n)$. Dann gilt $\mu(N) = 0$ und $f_n(\omega) \xrightarrow{n \rightarrow \infty} f(\omega)$ für jedes $\omega \in \Omega \setminus N$.

Gelte (iii). Sei $B_n = \{d(f_n, f_{n+1}) > \varepsilon_n\}$ und $B = \limsup_{n \rightarrow \infty} B_n$. Dann ist $\mu(B) = 0$, und für jedes $\omega \in \Omega \setminus B$ ist $(f_n(\omega))_{n \in \mathbb{N}}$ eine Cauchy-Folge in E . Da E vollständig ist, existiert der Limes $f(\omega) := \lim_{n \rightarrow \infty} f_n(\omega)$. Für $\omega \in B$ setze $f(\omega) = 0$. \square

Korollar 6.13. Sei (E, d) ein separabler metrischer Raum. Seien f, f_1, f_2, \dots messbare Abbildungen $\Omega \rightarrow E$. Dann sind die beiden folgenden Aussagen äquivalent.

(i) $f_n \xrightarrow{n \rightarrow \infty} f$ stochastisch,

(ii) Zu jeder Teilfolge von $(f_n)_{n \in \mathbb{N}}$ existiert eine gegen f fast überall konvergente Teilfolge.

Beweis. „(ii) \Rightarrow (i)“ Wir nehmen an, dass (i) nicht gilt. Sei \tilde{d} eine Metrik, die die stochastische Konvergenz bestimmt (siehe Satz 6.7). Dann gibt es ein $\varepsilon > 0$ und eine Teilfolge $(f_{n_k})_{k \in \mathbb{N}}$ mit $\tilde{d}(f_{n_k}, f) > \varepsilon$ für jedes $k \in \mathbb{N}$. Offenbar konvergiert keine Teilfolge von $(f_{n_k})_{k \in \mathbb{N}}$ stochastisch gegen f , also auch nicht f.ü.

„(i) \Rightarrow (ii)“ Gelte nun (i). Seien $A_1, A_2, \dots \in \mathcal{A}$ mit $A_N \uparrow \Omega$ und $\mu(A_N) < \infty$ für jedes $N \in \mathbb{N}$. Sei $n_k \uparrow \infty$ beliebig. Wegen $f_{n_k} \xrightarrow{\text{stoch}} f$ für $k \rightarrow \infty$, können wir eine Teilfolge $(f_{n_{k_l}})_{l \in \mathbb{N}}$ wählen, sodass

$$\mu\left(A_l \cap \left(d(f, f_{n_{k_l}}) > \frac{1}{l}\right)\right) < 2^{-l} \quad \text{für jedes } l \in \mathbb{N}.$$

Also gilt für jedes $N \in \mathbb{N}$

$$\sum_{l=1}^{\infty} \mu\left(A_N \cap \left(d(f, f_{n_{k_l}}) > \frac{1}{l}\right)\right) \leq N \mu(A_N) + \sum_{l=N+1}^{\infty} 2^{-l} < \infty.$$

Nach Satz 6.12(ii) konvergiert $(f_{n_{k_l}})_{l \in \mathbb{N}}$ fast überall auf A_N gegen f . Nach Bemerkung 6.3 konvergiert $(f_{n_{k_l}})_{l \in \mathbb{N}}$ daher fast überall auf Ω gegen f . \square

Korollar 6.14. Ist $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, bei dem stochastische und F.ü.-Konvergenz nicht zusammenfallen, so gibt es keine Topologie auf der Menge der messbaren Abbildungen $\Omega \rightarrow E$, die die F.ü.-Konvergenz erzeugt.

Beweis. Wir nehmen an, dass es eine Topologie gibt, die die F.ü.-Konvergenz erzeugt. Seien f, f_1, f_2, \dots messbare Abbildungen mit der Eigenschaft, dass $f_n \xrightarrow{\text{stoch}} f$, jedoch nicht $f_n \xrightarrow{n \rightarrow \infty} f$ fast überall. Sei nun U eine offene Menge, die f enthält, für die jedoch $f_n \notin U$ für unendlich viele $n \in \mathbb{N}$ gilt. Sei also $(f_{n_k})_{k \in \mathbb{N}}$ eine Teilfolge mit $f_{n_k} \notin U$ für jedes $k \in \mathbb{N}$. Wegen $f_{n_k} \xrightarrow{k \rightarrow \infty} f$ stochastisch, gibt es nach Korollar 6.13 wiederum eine Teilfolge $(f_{n_{k_l}})_{l \in \mathbb{N}}$ von $(f_{n_k})_{k \in \mathbb{N}}$ mit $f_{n_{k_l}} \xrightarrow{l \rightarrow \infty} f$ fast überall. Es ist dann aber $f_{n_{k_l}} \in U$ für alle bis auf endlich viele l , was einen Widerspruch darstellt. \square

Korollar 6.15. Sei (E, d) ein separabler, vollständiger metrischer Raum. Es sei $(f_n)_{n \in \mathbb{N}}$ eine stochastische Cauchy-Folge in E , das heißt, für jedes $A \in \mathcal{A}$ mit $\mu(A) < \infty$ und jedes $\varepsilon > 0$ gilt

$$\mu\left(A \cap \{d(f_m, f_n) > \varepsilon\}\right) \rightarrow 0 \quad \text{für } m, n \rightarrow \infty.$$

Dann konvergiert $(f_n)_{n \in \mathbb{N}}$ stochastisch.

Beweis. Ohne Einschränkung kann $\mu(\Omega) < \infty$ angenommen werden. Wähle eine Teilfolge $(f_{n_k})_{k \in \mathbb{N}}$, sodass

$$\mu\left(\{d(f_n, f_{n_k}) > 2^{-k}\}\right) < 2^{-k} \quad \text{für jedes } n \geq n_k.$$

Nach Satz 6.12(iii) gibt es ein f mit $f_{n_k} \xrightarrow{k \rightarrow \infty} f$ fast überall, also insbesondere

$$\mu(\{d(f_{n_k}, f) > \varepsilon/2\}) \xrightarrow{k \rightarrow \infty} 0 \quad \text{für jedes } \varepsilon > 0.$$

Nun ist aber

$$\mu(\{d(f_n, f) > \varepsilon\}) \leq \mu(\{d(f_{n_k}, f_n) > \varepsilon/2\}) + \mu(\{d(f_{n_k}, f) > \varepsilon/2\}).$$

Ist k so groß, dass $2^{-k} < \varepsilon/2$ ist und $n \geq n_k$, so ist der erste Summand kleiner als 2^{-k} , also gilt

$$\mu(\{d(f_n, f) > \varepsilon\}) \xrightarrow{n \rightarrow \infty} 0,$$

das heißt, es gilt $f_n \xrightarrow{\text{stoch}} f$. □

Übung 6.1.1. Man zeige: Ist Ω höchstens abzählbar, so folgt aus stochastischer Konvergenz schon F.ü.-Konvergenz. ♣

Übung 6.1.2. Man gebe jeweils ein Beispiel an für eine Folge, die

- (i) in L^1 konvergiert, aber nicht fast überall,
- (ii) fast überall konvergiert, aber nicht in L^1 .



Übung 6.1.3. (Satz von Egorov (1911)) Sei $(\Omega, \mathcal{A}, \mu)$ ein endlicher Maßraum, und seien f_1, f_2, \dots messbare Funktionen, die fast überall gegen ein f konvergieren. Man zeige: Zu jedem $\varepsilon > 0$ gibt es eine Menge $A \in \mathcal{A}$ mit $\mu(\Omega \setminus A) < \varepsilon$ und $\sup_{\omega \in A} |f_n(\omega) - f(\omega)| \xrightarrow{n \rightarrow \infty} 0$. ♣

Übung 6.1.4. Man zeige: Sind X_1, X_2, \dots unabhängige, quadratintegrierbare Zufallsvariablen mit $E[X_i] = 0$ für jedes $i \in \mathbb{N}$ und $\sum_{i=1}^{\infty} \text{Var}[X_i] < \infty$, dann existiert ein quadratintegrierbares X mit $X = \lim_{n \rightarrow \infty} \sum_{i=1}^n X_i$ fast sicher. ♣

6.2 Gleichgradige Integrierbarkeit

Wie bekommt man nun die Konvergenz im Mittel aus der F.ü.-Konvergenz? Das notwendige Additiv ist die gleichgradige Integrierbarkeit.

Definition 6.16. Eine Familie $\mathcal{F} \subset \mathcal{L}^1(\mu)$ heißt **gleichgradig integrierbar**, falls

$$\inf_{0 \leq g \in \mathcal{L}^1(\mu)} \sup_{f \in \mathcal{F}} \int (|f| - g)^+ d\mu = 0. \quad (6.2)$$

Satz 6.17. Die Familie $\mathcal{F} \subset \mathcal{L}^1(\mu)$ ist genau dann gleichgradig integrierbar, wenn

$$\inf_{0 \leq \tilde{g} \in \mathcal{L}^1(\mu)} \sup_{f \in \mathcal{F}} \int_{\{|f| > \tilde{g}\}} |f| d\mu = 0. \quad (6.3)$$

Ist $\mu(\Omega) < \infty$, so ist die gleichgradige Integrierbarkeit äquivalent zu jeder der beiden folgenden Bedingungen

$$(i) \inf_{a \in [0, \infty)} \sup_{f \in \mathcal{F}} \int (|f| - a)^+ d\mu = 0,$$

$$(ii) \inf_{a \in [0, \infty)} \sup_{f \in \mathcal{F}} \int_{\{|f| > a\}} |f| d\mu = 0.$$

Beweis. Offenbar gilt $(|f| - g)^+ \leq |f| \cdot \mathbb{1}_{\{|f| > g\}}$, also impliziert (6.3) die gleichgradige Integrierbarkeit.

Gelte nun (6.2). Für jedes $\varepsilon > 0$ sei $g_\varepsilon \in \mathcal{L}^1(\mu)$ so gewählt, dass

$$\sup_{f \in \mathcal{F}} \int (|f| - g_\varepsilon)^+ d\mu \leq \varepsilon. \quad (6.4)$$

Setze $\tilde{g}_\varepsilon = 2g_{\varepsilon/2}$. Dann ist für $f \in \mathcal{F}$

$$\int_{\{|f| > \tilde{g}_\varepsilon\}} |f| d\mu \leq \int_{\{|f| > \tilde{g}_\varepsilon\}} (|f| - g_{\varepsilon/2})^+ d\mu + \int_{\{|f| > \tilde{g}_\varepsilon\}} g_{\varepsilon/2} d\mu.$$

Per Konstruktion ist $\int_{\{|f| > \tilde{g}_\varepsilon\}} (|f| - g_{\varepsilon/2})^+ d\mu \leq \varepsilon/2$ und

$$g_{\varepsilon/2} \mathbb{1}_{\{|f| > \tilde{g}_\varepsilon\}} \leq (|f| - g_{\varepsilon/2})^+ \mathbb{1}_{\{|f| > \tilde{g}_\varepsilon\}},$$

also auch

$$\int_{\{|f| > \tilde{g}_\varepsilon\}} g_{\varepsilon/2} d\mu \leq \int_{\{|f| > \tilde{g}_\varepsilon\}} (|f| - g_{\varepsilon/2})^+ d\mu \leq \varepsilon/2.$$

Insgesamt haben wir also

$$\sup_{f \in \mathcal{F}} \int_{\{|f| > \tilde{g}_\varepsilon\}} |f| d\mu \leq \varepsilon. \quad (6.5)$$

Offenbar impliziert (ii) schon (i), und (i) impliziert die gleichgradige Integrierbarkeit von \mathcal{F} , falls $\mu(\Omega) < \infty$, denn das Infimum wird hier ja über die kleinere Menge der konstanten Funktionen gebildet. Wir müssen noch zeigen, dass gleichgradige Integrierbarkeit (ii) impliziert. Sei also \mathcal{F} gleichgradig integrierbar (jedoch

nicht notwendigerweise $\mu(\Omega) < \infty$. Zu gegebenem $\varepsilon > 0$ (und g_ε und \tilde{g}_ε wie oben) wählen wir a_ε so, dass $\int_{\{\tilde{g}_{\varepsilon/2} > a_\varepsilon\}} \tilde{g}_{\varepsilon/2} d\mu < \frac{\varepsilon}{2}$. Dann ist

$$\int_{\{|f| > a_\varepsilon\}} |f| d\mu \leq \int_{\{|f| > \tilde{g}_{\varepsilon/2}\}} |f| d\mu + \int_{\{\tilde{g}_{\varepsilon/2} > a_\varepsilon\}} \tilde{g}_{\varepsilon/2} d\mu < \varepsilon. \quad \square$$

Satz 6.18. (i) Ist $\mathcal{F} \subset \mathcal{L}^1(\mu)$ eine endliche Menge, so ist \mathcal{F} gleichgradig integrierbar.

(ii) Sind $\mathcal{F}, \mathcal{G} \subset \mathcal{L}^1(\mu)$ gleichgradig integrierbar, dann sind auch $(f + g : f \in \mathcal{F}, g \in \mathcal{G})$ und $(f - g : f \in \mathcal{F}, g \in \mathcal{G})$ sowie $\{|f| : f \in \mathcal{F}\}$ gleichgradig integrierbar.

(iii) Ist \mathcal{F} gleichgradig integrierbar und existiert zu jedem $g \in \mathcal{G}$ ein $f \in \mathcal{F}$ mit $|g| \leq |f|$, so ist auch \mathcal{G} gleichgradig integrierbar.

Beweis. Der einfache Beweis verbleibt zur Übung. \square

Der folgende Satz beschreibt ein sehr gut anwendbares Kriterium für gleichgradige Integrierbarkeit. Wir werden diesen Satz an vielen Stellen einsetzen.

Satz 6.19. Für endliches μ ist $\mathcal{F} \subset \mathcal{L}^1(\mu)$ genau dann gleichgradig integrierbar, wenn es eine Funktion $H : [0, \infty) \rightarrow [0, \infty)$ gibt mit $\lim_{x \rightarrow \infty} H(x)/x = \infty$ und

$$\sup_{f \in \mathcal{F}} \int H(|f|) d\mu < \infty.$$

H kann sogar monoton wachsend und konvex gewählt werden.

Beweis. „ \Leftarrow “ Es existiere H mit den angegebenen Eigenschaften. Dann gilt $K_a := \inf_{x \geq a} \frac{H(x)}{x} \uparrow \infty$, wenn $a \uparrow \infty$. Also ist für $a > 0$

$$\begin{aligned} \sup_{f \in \mathcal{F}} \int_{\{|f| \geq a\}} |f| d\mu &\leq \frac{1}{K_a} \sup_{f \in \mathcal{F}} \int_{\{|f| \geq a\}} H(|f|) d\mu \\ &\leq \frac{1}{K_a} \sup_{f \in \mathcal{F}} \int H(|f|) d\mu \xrightarrow{a \rightarrow \infty} 0. \end{aligned}$$

„ \Rightarrow “ Sei \mathcal{F} gleichgradig integrierbar. Da $\mu(\Omega) < \infty$ gilt, gibt es (nach Satz 6.17) eine Folge $a_n \uparrow \infty$ mit

$$\sup_{f \in \mathcal{F}} \int (|f| - a_n)^+ d\mu < 2^{-n}.$$

Wir setzen

$$H(x) = \sum_{n=1}^{\infty} (x - a_n)^+ \quad \text{für jedes } x \geq 0.$$

Dann ist H als Summe konvexer Funktionen konvex. Ferner gilt für jedes $n \in \mathbb{N}$ und $x \geq 2a_n$, dass $H(x)/x \geq \sum_{k=1}^n (1 - a_k/x)^+ \geq n/2$, also gilt $H(x)/x \uparrow \infty$. Schließlich ist nach dem Satz über monotone Konvergenz für jedes $f \in \mathcal{F}$

$$\int H(|f(\omega)|) \mu(d\omega) = \sum_{n=1}^{\infty} \int (|f| - a_n)^+ d\mu \leq \sum_{n=1}^{\infty} 2^{-n} = 1. \quad \square$$

Zur Notation $\|\cdot\|_p$ erinnere man sich an Definition 4.16.

Definition 6.20. Sei $p \in [1, \infty]$. Eine Familie $\mathcal{F} \subset \mathcal{L}^p(\mu)$ heißt beschränkt in $\mathcal{L}^p(\mu)$, falls $\sup\{\|f\|_p : f \in \mathcal{F}\} < \infty$ gilt.

Korollar 6.21. Ist $\mu(\Omega) < \infty$ und $p > 1$ sowie \mathcal{F} beschränkt in $\mathcal{L}^p(\mu)$, dann ist \mathcal{F} gleichgradig integrierbar.

Beweis. Wende Satz 6.19 an mit der konvexen Abbildung $H(x) = x^p$. \square

Korollar 6.22. Ist $(X_i)_{i \in I}$ eine Familie von Zufallsvariablen mit

$$\sup\{|\mathbf{E}[X_i]| : i \in I\} < \infty \quad \text{und} \quad \sup\{\mathbf{Var}[X_i] : i \in I\} < \infty,$$

dann ist $(X_i)_{i \in I}$ gleichgradig integrierbar.

Beweis. Dies folgt aus Korollar 6.21 mit $p = 2$, denn $\mathbf{E}[X_i^2] = \mathbf{E}[X_i]^2 + \mathbf{Var}[X_i]$ ist in $i \in I$ beschränkt. \square

Lemma 6.23. Es existiert eine Abbildung $h \in \mathcal{L}^1(\mu)$ mit $h > 0$ fast überall.

Beweis. Seien $A_1, A_2, \dots \in \mathcal{A}$ mit $A_n \uparrow \Omega$ und $\mu(A_n) < \infty$ für $n \in \mathbb{N}$. Setze

$$h = \sum_{n=1}^{\infty} 2^{-n} (1 + \mu(A_n))^{-1} \mathbb{1}_{A_n}.$$

Dann ist $h > 0$ fast überall und $\int h d\mu = \sum_{n=1}^{\infty} 2^{-n} \frac{\mu(A_n)}{1 + \mu(A_n)} \leq 1$. \square

Satz 6.24. Eine Familie $\mathcal{F} \subset \mathcal{L}^1(\mu)$ ist genau dann gleichgradig integrierbar, wenn die beiden folgenden Bedingungen erfüllt sind.

$$(i) C := \sup_{f \in \mathcal{F}} \int |f| d\mu < \infty.$$

(ii) Es gibt eine Funktion $0 \leq h \in \mathcal{L}^1(\mu)$, sodass für jedes $\varepsilon > 0$ ein $\delta(\varepsilon) > 0$ existiert mit

$$\sup_{f \in \mathcal{F}} \int_A |f| d\mu \leq \varepsilon \quad \text{für jedes } A \in \mathcal{A} \text{ mit } \int_A h d\mu < \delta(\varepsilon).$$

Ist $\mu(\Omega) < \infty$, so ist (ii) gleichwertig mit (iii):

(iii) Für jedes $\varepsilon > 0$ gibt es ein $\delta(\varepsilon) > 0$, sodass

$$\sup_{f \in \mathcal{F}} \int_A |f| d\mu \leq \varepsilon \quad \text{für jedes } A \in \mathcal{A} \text{ mit } \mu(A) < \delta(\varepsilon).$$

Beweis. „ \implies “ Sei \mathcal{F} gleichgradig integrierbar. Sei $h \in \mathcal{L}^1(\mu)$ mit $h > 0$ fast überall. Sei $\varepsilon > 0$ und $\tilde{g}_{\varepsilon/3}$ eine $\varepsilon/3$ -Schranke für \mathcal{F} (wie in (6.5)). Wegen $\{\tilde{g}_{\varepsilon/3} \geq \alpha h\} \downarrow \emptyset$ für $\alpha \rightarrow \infty$, gilt für hinreichend großes $\alpha = \alpha(\varepsilon)$

$$\int_{\{\tilde{g}_{\varepsilon/3} \geq \alpha h\}} \tilde{g}_{\varepsilon/3} d\mu < \frac{\varepsilon}{3}.$$

Mit $\delta(\varepsilon) := \frac{\varepsilon}{3\alpha(\varepsilon)}$ gilt dann für jedes $A \in \mathcal{A}$ mit $\int_A h d\mu < \delta(\varepsilon)$ und jedes $f \in \mathcal{F}$

$$\begin{aligned} \int_A |f| d\mu &\leq \int_{\{|f| > \tilde{g}_{\varepsilon/3}\}} |f| d\mu + \int_A \tilde{g}_{\varepsilon/3} d\mu \\ &\leq \frac{\varepsilon}{3} + \alpha \int_A h d\mu + \int_{\{\tilde{g}_{\varepsilon/3} \geq \alpha h\}} \tilde{g}_{\varepsilon/3} d\mu \leq \varepsilon. \end{aligned}$$

Damit ist (ii) gezeigt. Setzen wir in die Rechnung $A = \Omega$ ein, so erhalten wir

$$\int |f| d\mu \leq \frac{2\varepsilon}{3} + \alpha \int h d\mu < \infty.$$

Damit ist auch (i) gezeigt.

„ \Leftarrow “ Es gelten (i) und (ii). Sei $\varepsilon > 0$. Wähle h und $\delta(\varepsilon) > 0$ wie in (ii) und C wie in (i). Setze $\tilde{h} = \frac{C}{\delta(\varepsilon)}h$. Dann ist

$$\int_{\{|f| > \tilde{h}\}} h d\mu = \frac{\delta(\varepsilon)}{C} \int_{\{|f| > \tilde{h}\}} \tilde{h} d\mu \leq \frac{\delta(\varepsilon)}{C} \int |f| d\mu \leq \delta(\varepsilon),$$

also nach Voraussetzung $\int_{\{|f| > \tilde{h}\}} |f| d\mu < \varepsilon$.

„(ii) \Rightarrow (iii)“ Es gelte (ii). Sei $\varepsilon > 0$ und $\delta = \delta(\varepsilon)$ wie in (ii) gewählt. Sei $K < \infty$ so groß gewählt, dass $\int_{\{h \geq K\}} h d\mu < \delta/2$. Für $A \in \mathcal{A}$ mit $\mu(A) < \delta/(2K)$ ist dann

$$\int_A h d\mu \leq K\mu(A) + \int_{\{h \geq K\}} h d\mu < \delta,$$

also $\int_A |f| d\mu \leq \varepsilon$ für jedes $f \in \mathcal{F}$.

„(iii) \Rightarrow (ii)“ Gelten (iii) und $\mu(\Omega) < \infty$, so leistet $h \equiv 1$ das Gewünschte. \square

Wir kommen zum Hauptsatz dieses Abschnittes.

Satz 6.25. Sei $\{f_n : n \in \mathbb{N}\} \subset \mathcal{L}^1(\mu)$. Die folgenden Aussagen sind äquivalent.

- (i) Es gibt ein $f \in \mathcal{L}^1(\mu)$ mit $f_n \xrightarrow{n \rightarrow \infty} f$ in L^1 .
- (ii) $(f_n)_{n \in \mathbb{N}}$ ist $\mathcal{L}^1(\mu)$ -Cauchy-Folge, also $\|f_n - f_m\|_1 \rightarrow 0$ für $m, n \rightarrow \infty$.
- (iii) $(f_n)_{n \in \mathbb{N}}$ ist gleichgradig integrierbar, und es gibt eine messbare Abbildung f mit $f_n \xrightarrow{\text{stoch}} f$.

Die Limiten in (i) und (iii) stimmen überein.

Beweis. „(i) \Rightarrow (ii)“ Dies ist klar.

„(ii) \Rightarrow (iii)“ Für jedes $\varepsilon > 0$ gibt es ein $n_\varepsilon \in \mathbb{N}$, sodass $\|f_n - f_{n_\varepsilon}\|_1 < \varepsilon$ für jedes $n \geq n_\varepsilon$ gilt. Also ist für jedes $n \geq n_\varepsilon$ auch $\|(|f_n| - |f_{n_\varepsilon}|)^+\|_1 < \varepsilon$. Mithin ist $g_\varepsilon = \max\{|f_1|, \dots, |f_{n_\varepsilon}|\}$ eine ε -Schranke für $(f_n)_{n \in \mathbb{N}}$ (wie in (6.4)). Für $\varepsilon > 0$ ist

$$\mu(\{|f_m - f_n| > \varepsilon\}) \leq \varepsilon^{-1} \|f_m - f_n\|_1 \rightarrow 0 \quad \text{für } m, n \rightarrow \infty.$$

Deshalb ist $(f_n)_{n \in \mathbb{N}}$ auch eine stochastische Cauchy-Folge, also stochastisch konvergent nach Korollar 6.15.

„(iii) \Rightarrow (i)“ Sei f der stochastische Grenzwert der Folge $(f_n)_{n \in \mathbb{N}}$. Wir nehmen an, dass $(f_n)_{n \in \mathbb{N}}$ nicht in L^1 gegen f konvergiert. Dann gibt es ein $\varepsilon > 0$ und eine Teilfolge $(f_{n_k})_{k \in \mathbb{N}}$ mit

$$\|f - f_{n_k}\|_1 > 2\varepsilon \quad \text{für jedes } k \in \mathbb{N}, \tag{6.6}$$

wobei wir $\|f - f_{n_k}\|_1 = \infty$ setzen, falls $f \notin \mathcal{L}^1(\mu)$ ist. Nach Korollar 6.13 gibt es eine Teilfolge $(f'_{n'_k})_{k \in \mathbb{N}}$ von $(f_{n_k})_{k \in \mathbb{N}}$ mit $f'_{n'_k} \xrightarrow{k \rightarrow \infty} f$ fast überall. Nach dem Lemma von Fatou (Satz 4.21) mit 0 als Minorante gilt daher

$$\int |f| d\mu \leq \liminf_{k \rightarrow \infty} \int |f_{n'_k}| d\mu < \infty.$$

Also ist $f \in \mathcal{L}^1(\mu)$. Nach Satz 6.18(ii) (mit $\mathcal{G} = \{f\}$) ist $(f - f_{n'_k})_{k \in \mathbb{N}}$ gleichgradig integrierbar, also gibt es ein $0 \leq g \in \mathcal{L}^1(\mu)$, sodass $\int (|f - f_{n'_k}| - g)^+ d\mu < \varepsilon$. Setze

$$g_k = |f_{n'_k} - f| \wedge g \quad \text{für jedes } k \in \mathbb{N}.$$

Dann gilt $g_k \xrightarrow{k \rightarrow \infty} 0$ fast überall und $g - g_k \geq 0$. Nach dem Lemma von Fatou ist

$$\begin{aligned} \limsup_{k \rightarrow \infty} \int g_k d\mu &= \int g d\mu - \liminf_{k \rightarrow \infty} \int (g - g_k) d\mu \\ &\leq \int g d\mu - \int \left(\lim_{k \rightarrow \infty} (g - g_k) \right) d\mu = 0. \end{aligned}$$

Wegen $|f - f_{n'_k}| = (|f - f_{n'_k}| - g)^+ + g_k$ ist also

$$\limsup_{k \rightarrow \infty} \|f - f_{n'_k}\|_1 \leq \limsup_{k \rightarrow \infty} \int (|f - f_{n'_k}| - g)^+ d\mu + \limsup_{k \rightarrow \infty} \int g_k d\mu \leq \varepsilon,$$

im Widerspruch zu (6.6). \square

Korollar 6.26 (Lebesgue'scher Konvergenzsatz, majorisierte Konvergenz). Sei f messbar und $(f_n)_{n \in \mathbb{N}}$ eine Folge in $\mathcal{L}^1(\mu)$ mit $f_n \xrightarrow{n \rightarrow \infty} f$ stochastisch. Es existiere eine integrierbare Majorante $0 \leq g \in \mathcal{L}^1(\mu)$ mit $|f_n| \leq g$ fast überall für jedes $n \in \mathbb{N}$. Dann gilt $f \in \mathcal{L}^1(\mu)$ und $f_n \xrightarrow{n \rightarrow \infty} f$ in L^1 , also insbesondere $\int f_n d\mu \xrightarrow{n \rightarrow \infty} \int f d\mu$.

Beweis. Das folgt aus Satz 6.25, weil die Majorante die gleichgradige Integrierbarkeit der Folge $(f_n)_{n \in \mathbb{N}}$ sichert. \square

Übung 6.2.1. Sei $H \in \mathcal{L}^1(\mu)$ mit $H > 0$ μ -f.ü. (siehe Lemma 6.23) und (E, d) ein separabler metrischer Raum. Man zeige:

(i) Durch

$$d_H(f, g) := \int (1 \wedge d(f(\omega), g(\omega))) H(\omega) \mu(d\omega)$$

wird eine Metrik definiert, die die stochastische Konvergenz erzeugt.

(ii) Ist (E, d) vollständig, so ist d_H vollständig. \clubsuit

6.3 Vertauschung von Integral und Ableitung

Wir wollen untersuchen, wie sich Eigenschaften wie Stetigkeit und Differenzierbarkeit von Zweiparameterfunktionen unter Integration nach einer Variablen erhalten.

Satz 6.27 (Stetigkeitslemma). Sei (E, d) ein metrischer Raum, $x_0 \in E$ und $f : \Omega \times E \rightarrow \mathbb{R}$ eine Abbildung mit den Eigenschaften

- (i) für jedes $x \in E$ ist die Abbildung $\omega \mapsto f(\omega, x)$ in $\mathcal{L}^1(\mu)$,
- (ii) für fast alle $\omega \in \Omega$ ist die Abbildung $x \mapsto f(\omega, x)$ stetig im Punkte x_0 ,
- (iii) es gibt ein $h \in \mathcal{L}^1(\mu)$, $h \geq 0$, mit $|f(\cdot, x)| \leq h$ μ -f.ü. für alle $x \in E$.

Dann ist die Abbildung $F : E \rightarrow \mathbb{R}$, $x \mapsto \int f(\omega, x) \mu(d\omega)$ stetig in x_0 .

Beweis. Sei $(x_n)_{n \in \mathbb{N}}$ eine Folge in E mit $\lim_{n \rightarrow \infty} x_n = x_0$. Setze $f_n = f(\cdot, x_n)$. Nach Voraussetzung ist $|f_n| \leq h$ und $f_n \xrightarrow{n \rightarrow \infty} f(\cdot, x_0)$ fast überall. Nach dem Satz von der majorisierten Konvergenz (Korollar 6.26) ist

$$F(x_n) = \int f_n d\mu \xrightarrow{n \rightarrow \infty} \int f(\cdot, x_0) d\mu = F(x_0).$$

Also ist F stetig in x_0 . □

Satz 6.28 (Differentiationslemma). Sei $I \subset \mathbb{R}$ ein nichttriviales, offenes Intervall und $f : \Omega \times I \rightarrow \mathbb{R}$ eine Abbildung mit den Eigenschaften

- (i) für jedes $x \in I$ ist $(\omega \mapsto f(\omega, x)) \in \mathcal{L}^1(\mu)$,
- (ii) für fast alle $\omega \in \Omega$ ist $I \rightarrow \mathbb{R}$, $x \mapsto f(\omega, x)$ differenzierbar, wobei wir die Ableitung mit f' bezeichnen,
- (iii) es gibt ein $h \in \mathcal{L}^1(\mu)$, $h \geq 0$, mit $|f'(\cdot, x)| \leq h$ μ -f.ü. für alle $x \in I$.

Dann gilt: Für jedes $x \in I$ ist $f'(\cdot, x) \in \mathcal{L}^1(\mu)$. Die Funktion $F : x \mapsto \int f(\omega, x) \mu(d\omega)$ ist differenzierbar mit Ableitung

$$F'(x) = \int f'(\omega, x) \mu(d\omega).$$

Beweis. Sei $x_0 \in I$ und $(x_n)_{n \in \mathbb{N}}$ eine Folge in I mit $x_n \neq x_0$ für jedes $n \in \mathbb{N}$ sowie $\lim_{n \rightarrow \infty} x_n = x_0$. Wir zeigen, dass entlang der Folge $(x_n)_{n \in \mathbb{N}}$ die Differenzenquotienten konvergieren. Setze

$$g_n(\omega) = \frac{f(\omega, x_n) - f(\omega, x_0)}{x_n - x_0} \quad \text{für jedes } \omega \in \Omega.$$

Nach Voraussetzung (ii) gilt

$$g_n \xrightarrow{n \rightarrow \infty} f'(\cdot, x_0) \quad \mu\text{-fast überall.}$$

Nach dem Zwischenwertsatz der Differentialrechnung existiert zu jedem $n \in \mathbb{N}$ und fast jedem $\omega \in \Omega$ ein $y_n(\omega) \in I$ mit $g_n(\omega) = f'(\omega, y_n(\omega))$. Speziell ist $|g_n| \leq h$ fast überall für jedes $n \in \mathbb{N}$. Nach dem Satz von der majorisierten Konvergenz (Korollar 6.26) ist also die Grenzfunktion $f'(\cdot, x_0)$ in $\mathcal{L}^1(\mu)$ und

$$\lim_{n \rightarrow \infty} \frac{F(x_n) - F(x_0)}{x_n - x_0} = \lim_{n \rightarrow \infty} \int g_n(\omega) \mu(d\omega) = \int f'(\omega, x_0) \mu(d\omega). \quad \square$$

Beispiel 6.29 (Laplace-Transformation). Sei X eine nichtnegative Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbf{P})$. In der Notation von Satz 6.28 sei $I = [0, \infty)$ und $f(x, \lambda) = e^{-\lambda x}$ für $\lambda \in I$. Dann ist

$$F(\lambda) = \mathbf{E}[e^{-\lambda X}]$$

in $(0, \infty)$ unendlich oft differenzierbar. Die ersten Ableitungen sind $F'(\lambda) = -\mathbf{E}[X e^{-\lambda X}]$ und $F''(\lambda) = \mathbf{E}[(X^2)e^{-\lambda X}]$. Sukzessive erhalten wir die n -te Ableitung $F^{(n)}(\lambda) = \mathbf{E}[(-X)^n e^{-\lambda X}]$. Es gilt (monotone Konvergenz)

$$\mathbf{E}[X] = -\lim_{\lambda \downarrow 0} F'(\lambda) \tag{6.7}$$

und

$$\mathbf{E}[X^n] = (-1)^n \lim_{\lambda \downarrow 0} F^{(n)}(\lambda) \quad \text{für jedes } n \in \mathbb{N}. \tag{6.8}$$

In der Tat: Für $\varepsilon > 0$ und $I = (\varepsilon, \infty)$ ist

$$\sup_{x \geq 0, \lambda \in I} \left| \frac{d}{d\lambda} f(x, \lambda) \right| = \sup_{x \geq 0, \lambda \in I} x e^{-\lambda x} = \varepsilon^{-1} e^{-1} < \infty.$$

Damit erfüllt F die Voraussetzungen von Satz 6.28. Iterativ erhalten wir die Aussage für $F^{(n)}$, denn

$$\left| \frac{d^n}{d\lambda^n} f(x, \lambda) \right| \leq (n/\varepsilon)^n e^{-n} < \infty \quad \text{für } x \geq 0 \text{ und } \lambda \geq \varepsilon. \quad \diamond$$

Übung 6.3.1. Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbf{P})$ und

$$\Lambda(t) := \log(\mathbf{E}[e^{tX}]) \quad \text{für jedes } t \in \mathbb{R}.$$

Man zeige, dass $D := \{t \in \mathbb{R} : \Lambda(t) < \infty\}$ ein nichtleeres Intervall ist, und dass Λ im Inneren von D unendlich oft differenzierbar ist. ♣

7 L^p -Räume und Satz von Radon-Nikodym

In diesem Kapitel wollen wir die Räume der Funktionen untersuchen, deren p -te Potenz integrierbar ist. Wir leiten in Abschnitt 7.2 zunächst wichtige Ungleichungen her (Hölder, Minkowski, Jensen) und untersuchen dann in Abschnitt 7.3 den Fall $p = 2$, wo wir Hilberträume vorliegen haben, im Detail. Neben den genannten Ungleichungen sind die wichtigsten Ergebnisse für die Stochastik der Zerlegungssatz von Lebesgue sowie der Satz von Radon-Nikodym in Abschnitt 7.4. Der Leser mag beim ersten Lesen die anderen, eher analytisch als stochastisch ausgerichteten, Teile dieses Kapitels überschlagen.

7.1 Definitionen

Im Folgenden ist $(\Omega, \mathcal{A}, \mu)$ stets ein σ -endlicher Maßraum. In Definition 4.16 hatten wir für messbares $f : \Omega \rightarrow \overline{\mathbb{R}}$ definiert

$$\|f\|_p := \left(\int |f|^p d\mu \right)^{1/p} \quad \text{für } p \in [1, \infty),$$

und

$$\|f\|_\infty := \inf \{K \geq 0 : \mu(|f| > K) = 0\}.$$

Ferner hatten wir die Räume definiert, wo diese Ausdrücke endlich sind

$$\mathcal{L}^p(\Omega, \mathcal{A}, \mu) = \mathcal{L}^p(\mathcal{A}, \mu) = \mathcal{L}^p(\mu) = \{f : \Omega \rightarrow \overline{\mathbb{R}} \text{ ist messbar und } \|f\|_p < \infty\}.$$

Wir hatten gesehen, dass $\|\cdot\|_1$ eine Pseudonorm auf $\mathcal{L}^1(\mu)$ ist. Unser erstes Ziel ist es hier, $\|\cdot\|_p$ zu einer echten Norm zu machen, und zwar für jedes $p \in [1, \infty]$. Abgesehen davon, dass die Dreiecksungleichung noch zu zeigen ist, müssen wir zu diesem Zwecke auch den Raum verändern, denn es gilt nur

$$\|f - g\|_p = 0 \iff f = g \quad \mu\text{-f.ü.}$$

Bei einer echten (also nicht nur Pseudo-)Norm muss aus der linken Seite schon Gleichheit (nicht nur f.ü.) von f und g gelten. Wir sehen daher f und g als äquivalent an, falls $f = g$ fast überall. Sei also

$$\mathcal{N} = \{f \text{ ist messbar und } f = 0 \quad \mu\text{-f.ü.}\}.$$

Für jedes $p \in [1, \infty]$ ist \mathcal{N} ein Untervektorraum von $\mathcal{L}^p(\mu)$. Wir können also formal den Quotientenraum bilden. Dies ist das Standardverfahren, um aus einer Pseudonorm eine Norm zu machen.

Definition 7.1 (Quotientenraum). Für jedes $p \in [1, \infty]$ definieren wir

$$L^p(\Omega, \mathcal{A}, \mu) := \mathcal{L}^p(\Omega, \mathcal{A}, \mu)/\mathcal{N} = \{\bar{f} := f + \mathcal{N} : f \in \mathcal{L}^p(\mu)\}.$$

Für $\bar{f} \in L^p(\mu)$ setzen wir $\|\bar{f}\|_p = \|f\|_p$ für ein $f \in \bar{f}$ und $\int \bar{f} d\mu = \int f d\mu$, falls dieser Ausdruck für f definiert ist.

Man beachte, dass $\|\bar{f}\|_p$ nicht von der Wahl des Repräsentanten $f \in \bar{f}$ abhängt.

Wir wollen jetzt zunächst die Konvergenz bezüglich $\|\cdot\|_p$ untersuchen und erweitern dazu den entsprechenden Satz (Satz 6.25) über die Konvergenz bezüglich $\|\cdot\|_1$.

Definition 7.2. Seien $p \in [1, \infty]$ und $f, f_1, f_2, \dots \in \mathcal{L}^p(\mu)$. Falls $\|f_n - f\|_p \xrightarrow{n \rightarrow \infty} 0$ gilt, so sagen wir, dass $(f_n)_{n \in \mathbb{N}}$ im **p-ten Mittel** gegen f konvergiere und schreiben $f_n \xrightarrow{L^p} f$.

Satz 7.3. Seien $p \in [1, \infty]$ und $f_1, f_2, \dots \in \mathcal{L}^p(\mu)$. Dann sind äquivalent:

- (i) Es gibt ein $f \in \mathcal{L}^p(\mu)$ mit $f_n \xrightarrow{L^p} f$.
- (ii) $(f_n)_{n \in \mathbb{N}}$ ist eine Cauchy-Folge in $\mathcal{L}^p(\mu)$.

Ist $p < \infty$ so sind (i) und (ii) zudem äquivalent zu

- (iii) $(|f_n|^p)_{n \in \mathbb{N}}$ ist gleichgradig integrierbar, und es gibt ein messbares f mit $f_n \xrightarrow{\text{stoch.}} f$.

Die Limiten in (i) und (iii) stimmen überein.

Beweis. Für $p = \infty$ ist die Äquivalenz von (i) und (ii) eine einfache Schlussfolgerung aus der Dreiecksungleichung.

Sei nun $p \in [1, \infty)$. Der Beweis verläuft ähnlich wie der Beweis von Satz 6.25.

„(i) \implies (ii)“ Für $x, y \in \mathbb{R}$ ist $|x + y|^p \leq 2^p (|x|^p + |y|^p)$. Also ist

$$\|f_m - f_n\|_p^p \leq 2^p (\|f_m - f\|_p^p + \|f_n - f\|_p^p) \xrightarrow{n \rightarrow \infty} 0 \quad \text{für } m, n \rightarrow \infty.$$

„(ii) \implies (iii)“ Das geht genauso wie im Beweis von Satz 6.25.

„(iii) \implies (i)“ Wegen $|f_n|^p \xrightarrow{n \rightarrow \infty} |f|^p$ stoch., gilt $|f|^p \in \mathcal{L}^1(\mu)$ nach Satz 6.25, und damit $f \in \mathcal{L}^p(\mu)$. Setze $g_n = |f_n - f|^p$ für $n \in \mathbb{N}$. Dann ist $g_n \xrightarrow{n \rightarrow \infty} 0$

stochastisch, und $(g_n)_{n \in \mathbb{N}}$ ist gleichgradig integrierbar, da $g_n \leq 2^p (|f_n|^p + |f|^p)$. Also gilt (nach Satz 6.25) $\|f_n - f\|_p^p = \|g_n\|_1 \xrightarrow{n \rightarrow \infty} 0$. \square

Übung 7.1.1. Seien $(X_i)_{i \in \mathbb{N}}$ unabhängige, quadratintegrierbare Zufallsvariablen mit $E[X_i] = 0$ für jedes $i \in \mathbb{N}$.

- (i) Man zeige: Gilt $\sum_{i=1}^{\infty} \text{Var}[X_i] < \infty$, so existiert eine reelle Zufallsvariable X mit $\sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} X$ fast sicher.
- (ii) Gilt in (i) auch die Umkehrung? ♣

Übung 7.1.2. Sei $f : \Omega \rightarrow \mathbb{R}$ messbar. Zeige:

- (i) Gilt $\int |f|^p d\mu < \infty$ für ein $p \in (0, \infty)$, so gilt $\|f\|_p \xrightarrow{p \rightarrow \infty} \|f\|_\infty$.
- (ii) Auf die Integrierbarkeitsbedingung in (i) kann nicht verzichtet werden. ♣

Übung 7.1.3. Sei $p \in (1, \infty)$, $f \in L^p(\lambda)$, wobei λ das Lebesgue-Maß auf \mathbb{R} ist, und $T : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto x + 1$. Man zeige:

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \xrightarrow{n \rightarrow \infty} 0 \quad \text{in } L^p(\lambda). \quad \clubsuit$$

7.2 Ungleichungen und Satz von Fischer-Riesz

Wir wollen eine der wichtigsten Ungleichungen der Wahrscheinlichkeitstheorie, die Jensen'sche Ungleichung für konvexe Funktionen, herleiten. Aus dieser kann man die Hölder'sche Ungleichung und die Minkowski'sche Ungleichung folgern, die uns die Dreiecksungleichung für $\|\cdot\|_p$ liefern sowie den Dualraum zu bestimmen helfen. Allerdings geben wir hier direkte (und einfache) Beweise für die beiden letztgenannten Ungleichungen.

Bevor wir zur Jensen'schen Ungleichung kommen, wiederholen wir kurz Grundätzliches zur Konvexität von Mengen und Funktionen.

Definition 7.4. Eine Teilmenge G eines Vektorraums (beziehungsweise eines affin-linearen Raums) heißt **konvex**, falls für je zwei Punkte $x, y \in G$ und jedes $\lambda \in [0, 1]$ auch $\lambda x + (1 - \lambda)y \in G$ ist.

- Beispiele 7.5.**
- (i) Die konvexen Teilmengen von \mathbb{R} sind die Intervalle.
 - (ii) Ein linearer Unterraum eines Vektorraums ist konvex.
 - (iii) Die Menge aller W-Maße auf einem Messraum ist eine konvexe Menge. \diamond

Definition 7.6. Sei G eine konvexe Menge. Eine Abbildung $\varphi : G \rightarrow \mathbb{R}$ heißt **konvex**, falls für je zwei Punkte $x, y \in G$ und jedes $\lambda \in [0, 1]$ gilt

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda \varphi(x) + (1 - \lambda) \varphi(y).$$

φ heißt **konkav**, falls $(-\varphi)$ konvex ist.

Ist $I \subset \mathbb{R}$ ein Intervall und $\varphi : I \rightarrow \mathbb{R}$ stetig und im Inneren I° zweimal stetig differenzierbar mit zweiter Ableitung φ'' , so ist φ genau dann konvex, wenn $\varphi''(x) \geq 0$ ist für alle $x \in I^\circ$. Anders ausgedrückt: Die erste Ableitung φ' einer konvexen Funktion ist eine monoton wachsende Funktion. Wir werden im nächsten Satz sehen, dass dies auch dann noch gilt, wenn φ nicht zweimal stetig differenzierbar ist, wenn wir zur rechtsseitigen Ableitung $D^+ \varphi$ übergehen (oder zur linksseitigen), von der wir zeigen, dass sie immer existiert.

Satz 7.7. Sei $I \subset \mathbb{R}$ ein Intervall mit Innerem I° sowie $\varphi : I \rightarrow \mathbb{R}$ eine konvexe Abbildung. Dann gilt:

(i) φ ist stetig in I° und insbesondere messbar bezüglich $\mathcal{B}(I)$.

(ii) Für $x \in I^\circ$ definiere die Funktion der Differenzenquotienten

$$g_x(y) := \frac{\varphi(y) - \varphi(x)}{y - x} \quad \text{für } y \in I \setminus \{x\}.$$

Dann ist g_x monoton wachsend, und es existieren die links- und rechtsseitigen Ableitungen

$$D^- \varphi(x) := \lim_{y \uparrow x} g_x(y) = \sup\{g_x(y) : y < x\}$$

und

$$D^+ \varphi(x) := \lim_{y \downarrow x} g_x(y) = \inf\{g_x(y) : y > x\}.$$

(iii) Für $x \in I^\circ$ gilt $D^- \varphi(x) \leq D^+ \varphi(x)$ und

$$\varphi(x) + (y - x)t \leq \varphi(y) \quad \text{für jedes } y \in I \quad \iff \quad t \in [D^- \varphi(x), D^+ \varphi(x)].$$

$D^- \varphi(x)$ und $D^+ \varphi(x)$ sind also die minimale und die maximale Tangentensteigung in x .

(iv) Die Abbildungen $x \mapsto D^- \varphi(x)$ und $x \mapsto D^+ \varphi(x)$ sind monoton wachsend. $x \mapsto D^- \varphi(x)$ ist linksstetig und $x \mapsto D^+ \varphi(x)$ ist rechtsstetig. Es gilt $D^- \varphi(x) = D^+ \varphi(x)$ in allen Stetigkeitspunkten von $D^- \varphi$ und $D^+ \varphi$.

(v) φ ist genau dann in x differenzierbar, wenn $D^- \varphi(x) = D^+ \varphi(x)$ ist. In diesem Fall ist die Ableitung $\varphi'(x) = D^+ \varphi(x)$.

(vi) φ ist fast überall differenzierbar, und es gilt $\varphi(b) - \varphi(a) = \int_a^b D^+ \varphi(x) dx$ für $a, b \in I^\circ$.

Beweis. (i) Sei $x \in I^\circ$. Wir nehmen an, dass $\liminf_{n \rightarrow \infty} \varphi(x - 1/n) \leq \varphi(x) - \varepsilon$ für ein $\varepsilon > 0$ gilt. Da φ konvex ist, gilt

$$\varphi(y) \geq \varphi(x) + n(y - x)(\varphi(x) - \varphi(x - 1/n)) \quad \text{für jedes } y > x \text{ und } n \in \mathbb{N}.$$

Zusammen mit der obigen Annahme folgt $\varphi(y) = \infty$ für jedes $y > x$. Mithin war die Annahme falsch. Die analoge Überlegung für die rechte Seite liefert die Stetigkeit von φ in x .

(ii) Die Monotonie folgt aus der Konvexität. Die anderen Aussagen sind klar.

(iii) Aufgrund der Monotonie von g_x gilt $D^- \varphi(x) \leq D^+ \varphi(x)$. Per Konstruktion ist $\varphi(x) + (y - x)t \leq \varphi(y)$ für alle $y < x$ genau dann, wenn $t \geq D^- \varphi(x)$ ist und für alle $y > x$ genau dann, wenn $t \leq D^+ \varphi(x)$ ist.

(iv) Für $\varepsilon > 0$ ist aufgrund der Konvexität $x \mapsto g_x(x + \varepsilon)$ monoton wachsend und nach (i) stetig. Als Infimum monotoner, stetiger Funktionen ist $x \mapsto D^+ \varphi(x)$ monoton wachsend und rechtsstetig. Analog folgt die Aussage für $D^- \varphi$. Da $x \mapsto g_x(y)$ monoton ist, folgt $D^+ \varphi(x') \geq D^- \varphi(x') \geq D^+ \varphi(x)$ für $x' > x$. Ist $D^+ \varphi$ stetig in x , so ist $D^- \varphi(x) = D^+ \varphi(x)$.

(v) Dies ist klar, da $D^- \varphi$ und $D^+ \varphi$ die Limiten der linksseitigen und rechtsseitigen Sekantensteigungsfolgen sind.

(vi) Für $\varepsilon > 0$ sei $A_\varepsilon = \{x \in I : D^+ \varphi(x) \geq \varepsilon + \lim_{y \uparrow x} D^+ \varphi(y)\}$ die Menge der Unstetigkeitsstellen der Höhe mindestens ε . Für je zwei Punkte $a, b \in I$ mit $a < b$ ist $\#(A_\varepsilon \cap (a, b)) \leq \varepsilon^{-1}(D^+ \varphi(b) - D^+ \varphi(a))$, also $A_\varepsilon \cap (a, b)$ eine endliche Menge. Daher ist A_ε höchstens abzählbar. Also ist auch $A = \bigcup_{n=1}^{\infty} A_{1/n}$ höchstens abzählbar und damit eine Nullmenge. Nach (iv) und (v) ist φ differenzierbar in $I^\circ \setminus A$ mit Ableitung $D^+ \varphi$. \square

Ist I ein Intervall, so heißt eine Abbildung $g : I \rightarrow \mathbb{R}$ affin linear, falls es Zahlen $a, b \in \mathbb{R}$ gibt mit $g(x) = ax + b$ für alle $x \in I$. Ist $\varphi : I \rightarrow \mathbb{R}$ eine Abbildung, so schreiben wir

$$L(\varphi) := \{g : I \rightarrow \mathbb{R} \text{ ist affin linear und } g \leq \varphi\}.$$

Wir schreiben kurz $\sup L(\varphi)$ für die Abbildung $x \mapsto \sup\{f(x) : f \in L(\varphi)\}$.

Korollar 7.8. Sei $I \subset \mathbb{R}$ ein offenes Intervall und $\varphi : I \rightarrow \mathbb{R}$ eine Abbildung. Dann sind äquivalent

(i) φ ist konvex.

(ii) Zu jedem $x_0 \in I$ existiert ein $g \in L(\varphi)$ mit $g(x_0) = \varphi(x_0)$.

(iii) $L(\varphi)$ ist nicht leer und $\varphi = \sup L(\varphi)$.

(iv) Es existiert eine Folge $(g_n)_{n \in \mathbb{N}}$ in $L(\varphi)$ mit $\varphi = \lim_{n \rightarrow \infty} \max\{g_1, \dots, g_n\}$.

Beweis. „(ii) \implies (iii) \iff (iv)“ Dies ist klar.

„(iii) \implies (i)“ Das Supremum konvexer Funktionen ist konvex, und jede affin lineare Funktion ist konvex. Also ist $\sup L(\varphi)$ konvex, falls $L(\varphi) \neq \emptyset$.

„(i) \implies (ii)“ Nach Satz 7.7(iii) ist für jedes $x_0 \in I$ die Abbildung $x \mapsto \varphi(x_0) + (x - x_0)D^+\varphi(x_0)$ in $L(\varphi)$. \square

Satz 7.9 (Jensen'sche Ungleichung). Sei $I \subset \mathbb{R}$ ein Intervall und X eine Zufallsvariable mit Werten in I und $\mathbf{E}[|X|] < \infty$. Ist φ konvex, dann gilt $\mathbf{E}[\varphi(X)^-] < \infty$ und

$$\mathbf{E}[\varphi(X)] \geq \varphi(\mathbf{E}[X]).$$

Beweis. Da nach Korollar 7.8(iii) $L(\varphi) \neq \emptyset$ ist, können wir $a, b \in \mathbb{R}$ so wählen, dass $ax + b \leq \varphi(x)$ gilt für alle $x \in I$. Es ist dann

$$\mathbf{E}[\varphi(X)^-] \leq \mathbf{E}[(aX + b)^-] \leq |b| + |a| \cdot \mathbf{E}[|X|] < \infty.$$

Wir unterscheiden die Fälle, wo $\mathbf{E}[X]$ im Inneren I° oder am Rand ∂I liegt.

1. Fall Ist $\mathbf{E}[X] \in I^\circ$, so sei $t^+ := D^+\varphi(\mathbf{E}[X])$ die maximale Tangentensteigung von φ in $\mathbf{E}[X]$. Dann ist $\varphi(x) \geq t^+ \cdot (x - \mathbf{E}[X]) + \varphi(\mathbf{E}[X])$ für jedes $x \in I$, also

$$\mathbf{E}[\varphi(X)] \geq t^+ \mathbf{E}[X - \mathbf{E}[X]] + \mathbf{E}[\varphi(\mathbf{E}[X])] = \varphi(\mathbf{E}[X]).$$

2. Fall Ist $\mathbf{E}[X] \in \partial I$, so ist $X = \mathbf{E}[X]$ f.s., also $\mathbf{E}[\varphi(X)] = \mathbf{E}[\varphi(\mathbf{E}[X])] = \varphi(\mathbf{E}[X])$. \square

Die Jensen'sche Ungleichung lässt sich auf den \mathbb{R}^n ausweiten. Hierfür benötigen wir eine Darstellung konvexer Funktionen mehrerer Veränderlicher als Supremum von affin linearen Funktionen. Dabei heißt eine Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$ affin linear, wenn es ein $a \in \mathbb{R}^n$ und ein $b \in \mathbb{R}$ gibt mit $g(x) = \langle a, x \rangle + b$ für jedes x . Hierbei bezeichnet $\langle \cdot, \cdot \rangle$ das gewöhnliche Skalarprodukt auf \mathbb{R}^n .

Satz 7.10. Sei $G \subset \mathbb{R}^n$ offen und konvex und $\varphi : G \rightarrow \mathbb{R}$ eine Abbildung. Dann gilt Korollar 7.8 sinngemäß mit $I = G$. Ist φ konvex, so ist φ stetig und insbesondere messbar. Ist φ zweimal stetig differenzierbar, so ist φ genau dann konvex, wenn die Hesse-Matrix positiv semidefinit ist.

Beweis. Da wir die Aussagen nur für den Beweis der mehrdimensionalen Jensen'schen Ungleichung benötigen, die aber im weiteren Verlaufe keine tragende Bedeutung hat, geben wir nur die Literatur an: Im Buch von Rockafellar [155] folgt die Stetigkeit aus Theorem 10.1, die Aussagen von Korollar 7.8 aus Theorem 12.1 und Theorem 18.8. Die Aussage über die Hesse-Matrix steht in Theorem 4.5. \square

Satz 7.11 (Jensen'sche Ungleichung im \mathbb{R}^n). Sei $G \subset \mathbb{R}^n$ konvex, und seien X_1, \dots, X_n integrierbare reelle Zufallsvariablen mit $\mathbf{P}[(X_1, \dots, X_n) \in G] = 1$. Sei ferner $\varphi : G \rightarrow \mathbb{R}$ konvex. Dann ist $\mathbf{E}[\varphi(X_1, \dots, X_n)^-] < \infty$ und

$$\mathbf{E}[\varphi(X_1, \dots, X_n)] \geq \varphi(\mathbf{E}[X_1], \dots, \mathbf{E}[X_n]).$$

Beweis. Wir betrachten zunächst den Fall, wo G offen ist. Die Argumentation läuft hier ähnlich wie beim Beweis von Satz 7.9.

Sei $g \in L(\varphi)$ mit $g(\mathbf{E}[X_1], \dots, \mathbf{E}[X_n]) = \varphi(\mathbf{E}[X_1], \dots, \mathbf{E}[X_n])$. Da $g \leq \varphi$ linear ist, folgt

$$\mathbf{E}[\varphi(X_1, \dots, X_n)] \geq \mathbf{E}[g(X_1, \dots, X_n)] = g(\mathbf{E}[X_1], \dots, \mathbf{E}[X_n]).$$

Die Integrierbarkeit von $\varphi(X_1, \dots, X_n)^-$ folgt völlig analog wie im eindimensionalen Fall. Sei jetzt der allgemeine Fall betrachtet, das heißt derjenige, wo G nicht notwendigerweise offen ist. Hier ist das Problem, wenn $(\mathbf{E}[X_1], \dots, \mathbf{E}[X_n]) \in \partial G$ liegt, etwas kniffliger als im eindimensionalen Fall, weil ∂G flache Stücke haben kann, die aber selbst notwendigerweise wieder konvex sind. Man kann also nicht schließen, dass (X_1, \dots, X_n) fast sicher gleich dem Erwartungswert ist. Wir skizzieren nur das Argument: Zunächst kann man nur folgern, dass (X_1, \dots, X_n) fast sicher in einem solchen flachen Stück liegt. Dieses ist dann notwendigerweise von Dimension kleiner als n (oder Null, falls das Stück schon ein Punkt ist). Jetzt muss man φ auf das flache Stück einschränken und sich iterativ in der Dimension herunter arbeiten. Die Details finden sich beispielsweise in [42, Theorem 10.2.6]. \square

Beispiel 7.12. Sei X eine reelle Zufallsvariable mit $\mathbf{E}[X^2] < \infty$, $I = \mathbb{R}$ und $\varphi(x) = x^2$. Aus der Jensen'schen Ungleichung folgt

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 \geq 0. \quad \diamond$$

Beispiel 7.13. $G = [0, \infty) \times [0, \infty)$, und $\alpha \in (0, 1)$ sowie $\varphi(x, y) = x^\alpha y^{1-\alpha}$. φ ist konkav (Übung!), daher gilt für nichtnegative Zufallsvariablen X und Y mit endlicher Erwartung (nach Satz 7.11)

$$\mathbf{E}[X^\alpha Y^{1-\alpha}] \leq (\mathbf{E}[X])^\alpha (\mathbf{E}[Y])^{1-\alpha}. \quad \diamond$$

Beispiel 7.14. Seien G , X und Y wie in Beispiel 7.13. Sei $p \in (1, \infty)$. Dann ist $\psi(x, y) = (x^{1/p} + y^{1/p})^p$ konkav. Daher gilt (nach Satz 7.11)

$$(\mathbf{E}[X]^{1/p} + \mathbf{E}[Y]^{1/p})^p \geq \mathbf{E}[(X^{1/p} + Y^{1/p})^p]. \quad \diamond$$

Wir kommen nun zu den beiden weiteren wichtigen Ungleichungen, der Hölder'schen Ungleichung und der Minkowski'schen Ungleichung. Zur Vorbereitung bringen wir ein Lemma.

Lemma 7.15 (Young'sche Ungleichung). Für $p, q \in (1, \infty)$ mit $\frac{1}{p} + \frac{1}{q} = 1$ und für $x, y \in [0, \infty)$ gilt

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}. \quad (7.1)$$

Beweis. Wir halten $y \in [0, \infty)$ fest und definieren $f(x) := \frac{x^p}{p} + \frac{y^q}{q} - xy$ für $x \in [0, \infty)$. f ist zweimal stetig differenzierbar in $(0, \infty)$ mit Ableitungen $f'(x) = x^{p-1} - y$ und $f''(x) = (p-1)x^{p-2}$. Speziell ist f strikt konvex und besitzt daher eine eindeutige Minimalstelle bei $x_0 = y^{1/(p-1)}$. Nach Voraussetzung ist $q = \frac{p}{p-1}$, also $x_0^p = y^q$ und daher

$$f(x_0) = \left(\frac{1}{p} + \frac{1}{q}\right)y^q - y^{1/(p-1)}y = 0. \quad \square$$

Satz 7.16 (Hölder'sche Ungleichung). Seien $p, q \in [1, \infty]$ mit $\frac{1}{p} + \frac{1}{q} = 1$ und $f \in \mathcal{L}^p(\mu)$, $g \in \mathcal{L}^q(\mu)$. Dann gilt $(fg) \in \mathcal{L}^1(\mu)$ und

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q.$$

Beweis. Die Fälle $p = 1$ und $p = \infty$ sind trivial. Sei also nun $p \in (1, \infty)$ und $f \in \mathcal{L}^p(\mu)$ und $g \in \mathcal{L}^q(\mu)$ nicht fast überall Null. Indem wir zu $f/\|f\|_p$ und $g/\|g\|_q$ übergehen, können wir $\|f\|_p = \|g\|_q = 1$ annehmen. Nach Lemma 7.15 ist

$$\begin{aligned} \|fg\|_1 &= \int |f| \cdot |g| d\mu \leq \frac{1}{p} \int |f|^p d\mu + \frac{1}{q} \int |g|^q d\mu \\ &= \frac{1}{p} + \frac{1}{q} = 1 = \|f\|_p \cdot \|g\|_q. \end{aligned} \quad \square$$

Satz 7.17 (Minkowski'sche Ungleichung). Für $p \in [1, \infty]$ und $f, g \in \mathcal{L}^p(\mu)$ gilt

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \quad (7.2)$$

Beweis. Der Fall $p = \infty$ ist wiederum trivial. Sei also $p \in [1, \infty)$. Die linke Seite in (7.2) wird nicht kleiner, wenn wir f und g durch $|f|$ und $|g|$ ersetzen. Wir können also ohne Einschränkung annehmen, dass $f \geq 0$, $g \geq 0$ und $\|f + g\|_p > 0$ gelten.

Nun ist $(f + g)^p \leq 2^p(f^p \vee g^p) \leq 2^p(f^p + g^p)$, also ist $f + g \in \mathcal{L}^p(\mu)$. Mit Hilfe der Hölder'schen Ungleichung, angewandt auf $f \cdot (f + g)^{p-1}$ und auf $g \cdot (f + g)^{p-1}$, erhalten wir

$$\begin{aligned} \|f + g\|_p^p &= \int (f + g)^p d\mu = \int f(f + g)^{p-1} d\mu + \int g(f + g)^{p-1} d\mu \\ &\leq \|f\|_p \cdot \|(f + g)^{p-1}\|_q + \|g\|_p \cdot \|(f + g)^{p-1}\|_q \\ &= (\|f\|_p + \|g\|_p) \cdot \|f + g\|_p^{p-1}, \end{aligned}$$

wobei wir im letzten Schritt ausgenutzt haben, dass $p - p/q = 1$ ist. Teilen wir nun beide Seiten durch $\|f + g\|_p^{p-1}$, so folgt (7.2). \square

Wir haben in Satz 7.17 die Dreiecksungleichung gezeigt und damit, dass $\|\cdot\|_p$ eine Norm ist. In Satz 7.3 wurde hingegen gezeigt, dass diese Norm vollständig ist (jede Cauchy-Folge konvergiert). Ein vollständiger normierter Vektorraum heißt **Banachraum**. Wir haben also den folgenden Satz gezeigt:

Satz 7.18 (Fischer-Riesz). Für $p \in [1, \infty]$ ist $(L^p(\mu), \|\cdot\|_p)$ ein Banachraum.

Übung 7.2.1. Zeige die Hölder'sche Ungleichung mit Hilfe der Jensen'schen Ungleichung mit der Funktion aus Beispiel 7.13. \clubsuit

Übung 7.2.2. Zeige die Minkowski'sche Ungleichung mit Hilfe der Jensen'schen Ungleichung mit der Funktion aus Beispiel 7.14. \clubsuit

Übung 7.2.3. Sei X eine reelle Zufallsvariable und $p, q \in (1, \infty)$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Zeige: X ist genau dann in $L^p(\mathbf{P})$, wenn es ein $C < \infty$ gibt, sodass $|\mathbf{E}[XY]| \leq C \|Y\|_q$ für jede beschränkte Zufallsvariable Y . \clubsuit

7.3 Hilberträume

In diesem Abschnitt betrachten wir nur den Fall $p = 2$. Ziel ist der Darstellungssatz von Riesz-Fréchet für stetige lineare Funktionale auf Hilberträumen. Dieser Satz stellt einen Baustein für einen funktionalanalytischen Beweis des Satzes von Radon-Nikodym dar.

Definition 7.19. Sei V ein reeller Vektorraum. Eine Abbildung $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ heißt **Skalarprodukt**, falls

- (i) (Linearität) $\langle x, \alpha y + z \rangle = \alpha \langle x, y \rangle + \langle x, z \rangle$ für alle $x, y, z \in V$ und $\alpha \in \mathbb{R}$.
- (ii) (Symmetrie) $\langle x, y \rangle = \langle y, x \rangle$ für alle $x, y \in V$.
- (iii) (Positive Definitheit) $\langle x, x \rangle > 0$ für jedes $x \in V \setminus \{0\}$.

Gelten lediglich (i), (ii) und $\langle x, x \rangle \geq 0$ für jedes x , so heißt $\langle \cdot, \cdot \rangle$ eine positiv semidefinite symmetrische Bilinearform.

Ist $\langle \cdot, \cdot \rangle$ ein Skalarprodukt, so heißt $(V, \langle \cdot, \cdot \rangle)$ ein (reeller) **Hilbertraum**, falls die durch $\|x\| := \langle x, x \rangle^{1/2}$ definierte Norm vollständig ist, falls also $(V, \|\cdot\|)$ ein Banachraum ist.

Definition 7.20. Für $f, g \in \mathcal{L}^2(\mu)$ definieren wir

$$\langle f, g \rangle := \int fg d\mu.$$

Für $\bar{f}, \bar{g} \in L^2(\mu)$ definieren wir $\langle \bar{f}, \bar{g} \rangle := \langle f, g \rangle$, wobei $f \in \bar{f}$ und $g \in \bar{g}$.

Man beachte, dass diese Definition unabhängig von der Wahl der Repräsentanten f und g ist.

Satz 7.21. $\langle \cdot, \cdot \rangle$ ist ein Skalarprodukt auf $L^2(\mu)$ und eine positiv semidefinite symmetrische Bilinearform auf $\mathcal{L}^2(\mu)$. Es gilt $\|f\|_2 = \langle f, f \rangle^{1/2}$.

Beweis. Übung! □

Als Korollar zu Satz 7.18 erhalten wir:

Korollar 7.22. $(L^2(\mu), \langle \cdot, \cdot \rangle)$ ist ein reeller Hilbertraum.

Lemma 7.23. Ist $\langle \cdot, \cdot \rangle$ eine positiv semidefinite Bilinearform auf dem reellen Vektorraum V , so ist $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ stetig (bezüglich der Produkttopologie der Topologie auf V , die von der Pseudometrik $d(x, y) = \langle x - y, x - y \rangle^{1/2}$ erzeugt wird).

Beweis. Klar. □

Definition 7.24 (Orthogonales Komplement). Sei V ein reeller Vektorraum mit Skalarprodukt $\langle \cdot, \cdot \rangle$. Ist $W \subset V$, so bezeichnen wir den Untervektorraum

$$W^\perp := \{v \in V : \langle v, w \rangle = 0 \text{ für alle } w \in W\}$$

als das orthogonale Komplement von W .

Satz 7.25 (Orthogonale Zerlegung). Sei $(V, \langle \cdot, \cdot \rangle)$ ein Hilbertraum und $W \subset V$ ein abgeschlossener linearer Unterraum. Für jedes $x \in V$ existiert eine eindeutige Darstellung $x = y + z$, wobei $y \in W$ und $z \in W^\perp$ ist.

Beweis. Sei $x \in V$ und $c := \inf\{\|x - w\| : w \in W\}$. Sei ferner $(w_n)_{n \in \mathbb{N}}$ eine Folge in W mit $\|x - w_n\| \xrightarrow{n \rightarrow \infty} c$. Die Parallelogrammgleichung ergibt

$$\|w_m - w_n\|^2 = 2\|w_m - x\|^2 + 2\|w_n - x\|^2 - 4\left\|\frac{1}{2}(w_m + w_n) - x\right\|^2.$$

Da W linear ist, ist $(w_m + w_n)/2 \in W$, also $\|\frac{1}{2}(w_m + w_n) - x\| \geq c$. Also ist $(w_n)_{n \in \mathbb{N}}$ eine Cauchy-Folge: $\|w_m - w_n\| \rightarrow 0$, falls $m, n \rightarrow \infty$.

Da V vollständig ist und W abgeschlossen, ist auch W vollständig, also gibt es ein $y \in W$ mit $w_n \xrightarrow{n \rightarrow \infty} y$. Setze nun $z := x - y$. Dann ist $\|z\| = \lim_{n \rightarrow \infty} \|w_n - x\| = c$ aufgrund der Stetigkeit der Norm (Lemma 7.23).

Betrachte ein beliebiges $w \in W \setminus \{0\}$. Wir setzen $\varrho := -\langle z, w \rangle / \|w\|^2$ und erhalten $y + \varrho w \in W$, also

$$c^2 \leq \|x - (y + \varrho w)\|^2 = \|z\|^2 + \varrho^2 \|w\|^2 + 2\varrho \langle z, w \rangle = c^2 - \varrho^2 \|w\|^2.$$

Folglich ist $\langle z, w \rangle = 0$ für alle $w \in W$ und damit $z \in W^\perp$.

Die Eindeutigkeit der Darstellung klar: Ist $x = y' + z'$ eine weitere orthogonale Zerlegung, so ist $y - y' \in W$ und $z - z' \in W^\perp$ sowie $y - y' + z - z' = 0$, also ist

$$\begin{aligned} 0 &= \|y - y' + z - z'\|^2 = \|y - y'\|^2 + \|z - z'\|^2 + 2\langle y - y', z - z' \rangle \\ &= \|y - y'\|^2 + \|z - z'\|^2. \end{aligned}$$

Es folgt $y = y'$ und $z = z'$. □

Satz 7.26 (Darstellungssatz von Riesz-Fréchet). Sei $(V, \langle \cdot, \cdot \rangle)$ ein Hilbertraum und $F : V \rightarrow \mathbb{R}$ eine Abbildung. Dann sind äquivalent:

- (i) F ist stetig und linear.
- (ii) Es gibt ein $f \in V$ mit $F(x) = \langle x, f \rangle$ für alle $x \in V$.

Das Element $f \in V$ in (ii) ist eindeutig bestimmt.

Beweis. „(ii) \implies (i)“ Für jedes $f \in V$ ist per Definition des Skalarprodukts die Abbildung $x \mapsto \langle x, f \rangle$ linear. Nach Lemma 7.23 ist diese Abbildung auch stetig.

„(i) \implies (ii)“ Ist $F \equiv 0$, so wähle $f = 0$. Sei nun F nicht identisch Null. Da F stetig ist, ist der Kern $W := F^{-1}(\{0\})$ ein abgeschlossener echter linearer Unterraum von V . Sei $v \in V \setminus W$, und sei $v = y + z$ für $y \in W$ und $z \in W^\perp$ die orthogonale Zerlegung von v . Dann ist $z \neq 0$, und $F(z) = F(v) - F(y) = F(v) \neq 0$, und wir können $u := z/F(z) \in W^\perp$ definieren. Offenbar ist $F(u) = 1$, und für jedes $x \in V$ ist $F(x - F(x)u) = F(x) - F(x)F(u) = 0$, also $x - F(x)u \in W$ und damit $\langle x - F(x)u, u \rangle = 0$. Folglich ist $F(x) = \langle x, u \rangle / \|u\|^2$. Setzen wir nun $f := u / \|u\|^2$, so ist $F(x) = \langle x, f \rangle$ für alle $x \in V$.

„Eindeutigkeit“ Sei $\langle x, f \rangle = \langle x, g \rangle$ für alle $x \in V$. Setzen wir $x = f - g$, so erhalten wir $0 = \langle f - g, f - g \rangle$, also $f = g$. □

Wir werden den Darstellungssatz im folgenden Abschnitt für den Raum $\mathcal{L}^2(\mu)$ brauchen statt für den Hilbertraum $L^2(\mu)$. Mit ein bisschen *abstract nonsense* lässt sich aber der vorangehende Satz auf diese Situation anwenden. Wir erinnern daran, dass $\mathcal{N} = \{f \in \mathcal{L}^2(\mu) : \langle f, f \rangle = 0\}$ der Unterraum der Funktionen ist, die fast überall Null sind, und $L^2(\mu) = \mathcal{L}^2(\mu) / \mathcal{N}$ der Quotientenraum. Dies ist ein Spezialfall der

Situation, wo $(V, \langle \cdot, \cdot \rangle)$ ein linearer Raum mit vollständiger positiv semidefiniter symmetrischer Bilinearform ist. In diesem Fall ist $\mathcal{N} := \{v \in V : \langle v, v \rangle = 0\}$ und $V_0 = V/\mathcal{N} := \{f + \mathcal{N} : f \in V\}$. Wir schreiben $\langle v + \mathcal{N}, w + \mathcal{N} \rangle_0 := \langle v, w \rangle$ und erhalten so einen Hilbertraum $(V_0, \langle \cdot, \cdot \rangle_0)$.

Korollar 7.27. *Sei $(V, \langle \cdot, \cdot \rangle)$ ein linearer Vektorraum mit vollständiger positiv semidefiniter symmetrischer Bilinearform. Die Abbildung $F : V \rightarrow \mathbb{R}$ ist genau dann stetig und linear, wenn es ein $f \in V$ gibt mit $F(x) = \langle x, f \rangle$ für alle $x \in V$.*

Beweis. Die eine Implikation ist trivial. Sei also F stetig und linear. Dann ist $F(0) = 0$, weil F linear ist, und für jedes $v \in \mathcal{N}$ ist $F(v) = F(0) = 0$, weil F stetig ist (klar: v liegt in jeder offenen Umgebung von 0, also muss F in v denselben Wert annehmen wie in 0). Also induziert F eine stetige lineare Abbildung $F_0 : V_0 \rightarrow \mathbb{R}$ durch $F_0(x + \mathcal{N}) = F(x)$. Nach Satz 7.26 existiert ein $f + \mathcal{N} \in V_0$ mit $F_0(x + \mathcal{N}) = \langle x + \mathcal{N}, f + \mathcal{N} \rangle_0$ für jedes $x + \mathcal{N} \in V_0$. Nach Definition von F_0 und $\langle \cdot, \cdot \rangle_0$ ist nun aber $F(x) = \langle x, f \rangle$ für jedes $x \in V$. \square

Korollar 7.28. *Die Abbildung $F : \mathcal{L}^2(\mu) \rightarrow \mathbb{R}$ ist genau dann stetig und linear, wenn es ein $f \in \mathcal{L}^2(\mu)$ gibt mit $F(g) = \int g f d\mu$ für alle $g \in \mathcal{L}^2(\mu)$.*

Beweis. Der Raum $\mathcal{L}^2(\mu)$ erfüllt die Bedingungen des vorangehenden Korollars. \square

Übung 7.3.1 (Fourier-Reihen). Für $n \in \mathbb{N}_0$ seien $S_n, C_n : [0, 1] \rightarrow [0, 1]$ definiert durch $S_n(x) = \sin(2\pi n x)$, $C_n(x) = \cos(2\pi n x)$. Für zwei quadratisch summierbare reelle Folgen $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}_0}$ sei $h_{a,b} := b_0 + \sum_{n=1}^{\infty} (a_n S_n + b_n C_n)$. Ferner sei W der Vektorraum aller solcher $h_{a,b}$.

Man zeige:

- (i) Die Funktionen $C_0, S_n, C_n, n \in \mathbb{N}$ sind ein Orthogonalsystem in $L^2([0, 1], \lambda)$.
- (ii) Die Reihe, die $h_{a,b}$ definiert, konvergiert in $L^2([0, 1], \lambda)$.
- (iii) W ist ein abgeschlossener linearer Unterraum von $L^2([0, 1], \lambda)$.
- (iv) Für jedes $f \in \mathcal{L}^2([0, 1], \lambda)$ existieren eindeutige quadratisch summierbare Folgen $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}_0}$ mit $f = h_{a,b}$. Es gilt $\|f\|_2^2 = b_0^2 + \sum_{n=1}^{\infty} (a_n^2 + b_n^2)$.

Hinweis: Man zeige (iv) zunächst für Treppenfunktionen (siehe Übung 4.2.6). \clubsuit

7.4 Lebesgue'scher Zerlegungssatz

In diesem Abschnitt benutzen wir die eben gewonnenen Aussagen über Hilberträume, um ein Maß zu zerlegen in einen singulären und einen absolutstetigen Anteil bezüglich eines zweiten Maßes. Für den absolutstetigen Anteil zeigen wir, dass er

eine Dichte besitzt. Seien μ und ν Maße auf (Ω, \mathcal{A}) . Nach Definition 4.13 heißt eine messbare Funktion $f : \Omega \rightarrow [0, \infty)$ eine **Dichte** von ν bezüglich μ , falls

$$\nu(A) := \int f \mathbb{1}_A d\mu \quad \text{für jedes } A \in \mathcal{A}. \quad (7.3)$$

Andererseits definiert für jedes messbare $f : \Omega \rightarrow [0, \infty)$ Gleichung (7.3) ein Maß ν auf (Ω, \mathcal{A}) . Wir schreiben in diesem Fall auch

$$\nu = f\mu \quad \text{und} \quad f = \frac{d\nu}{d\mu}. \quad (7.4)$$

Beispielsweise hat die Normalverteilung $\nu = \mathcal{N}_{0,1}$ die Dichte $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ bezüglich des Lebesgue-Maßes $\mu = \lambda$ auf \mathbb{R} .

Ist $g : \Omega \rightarrow [0, \infty]$ messbar, so gilt (nach Satz 4.15)

$$\int g d\nu = \int gf d\mu. \quad (7.5)$$

Wir erhalten so, dass genau dann $g \in \mathcal{L}^1(\nu)$ ist, wenn $gf \in \mathcal{L}^1(\mu)$ gilt, und in diesem Fall ist (7.5) erfüllt.

Gilt $\nu = f\mu$, so ist offenbar $\nu(A) = 0$ für jedes $A \in \mathcal{A}$ mit $\mu(A) = 0$. In gewissem Sinne komplementär ist die Situation beispielsweise bei der Poissonverteilung $\mu = \text{Poi}_\varrho$ mit Parameter $\varrho > 0$ und $\nu = \mathcal{N}_{0,1}$. Hier ist $\mathbb{N}_0 \subset \mathbb{R}$ eine ν -Nullmenge mit $\mu(\mathbb{R} \setminus \mathbb{N}_0) = 0$. Wir sagen, dass ν **singulär** zu μ ist.

Das Ziel dieses Kapitels ist es, im allgemeinen Fall zu zeigen, dass ein beliebiges σ -endliches Maß ν auf einem Messraum (Ω, \mathcal{A}) zerlegt werden kann in einen Teil, der singulär zum σ -endlichen Maß μ ist, und einen Teil, der eine Dichte bezüglich μ hat (Lebesgue'scher Zerlegungssatz, Satz 7.33).

Satz 7.29 (Eindeutigkeit der Dichte). *Sei ν σ -endlich. Sind f_1 und f_2 Dichten von ν bezüglich μ , so gilt $f_1 = f_2$ μ -fast überall. Speziell ist die Dichtefunktion $\frac{d\nu}{d\mu}$ eindeutig bis auf Gleichheit μ -fast überall.*

Beweis. Sei $E_n \uparrow \Omega$ mit $\nu(E_n) < \infty$, $n \in \mathbb{N}$. Sei $A_n = E_n \cap \{f_1 > f_2\}$ für $n \in \mathbb{N}$. Dann ist $\nu(A_n) < \infty$, also

$$0 = \nu(A_n) - \nu(A_n) = \int_{A_n} (f_1 - f_2) d\mu.$$

Nach Satz 4.8(i) gilt $f_2 \mathbb{1}_{A_n} = f_1 \mathbb{1}_{A_n}$ μ -f.ü. Da $f_1 > f_2$ auf A_n gilt, erhalten wir $\mu(A_n) = 0$ und damit auch $\mu(\{f_1 > f_2\}) = \mu(\bigcup_{n \in \mathbb{N}} A_n) = 0$. Analog folgt $\mu(\{f_1 < f_2\}) = 0$, also $f_1 = f_2$ μ -f.ü. \square

Definition 7.30. Seien μ und ν zwei Maße auf (Ω, \mathcal{A}) .

(i) ν heißt **absolutstetig** bezüglich μ (kurz $\nu \ll \mu$), falls

$$\nu(A) = 0 \quad \text{für jedes } A \in \mathcal{A} \text{ mit } \mu(A) = 0. \quad (7.6)$$

Die Maße μ und ν heißen **äquivalent** (kurz $\mu \approx \nu$), falls $\nu \ll \mu$ und $\mu \ll \nu$.

(ii) μ heißt **singulär** zu ν (kurz $\mu \perp \nu$), falls es ein $A \in \mathcal{A}$ gibt mit $\mu(A) = 0$ und $\nu(\Omega \setminus A) = 0$.

Bemerkung 7.31. Offenbar gilt $\mu \perp \nu \iff \nu \perp \mu$. \diamond

Beispiel 7.32. (i) Sei μ ein Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Dichte f bezüglich des Lebesgue-Maßes λ . Dann ist für $A \in \mathcal{A}$ mit $\lambda(A) = 0$ schon $\mu(A) = \int_A f d\lambda = 0$, also $\mu \ll \lambda$. Ist λ -fast überall $f > 0$, so ist $\mu(A) = \int_A f d\lambda > 0$ falls $\lambda(A) > 0$, also $\mu \approx \lambda$. Ist $\lambda(\{f = 0\}) > 0$, so ist (wegen $\mu(\{f = 0\}) = 0$) $\lambda \not\ll \mu$.

(ii) Betrachte die Bernoulli-Verteilungen Ber_p und Ber_q für $p, q \in [0, 1]$. Ist $p \in (0, 1)$, so gilt $\text{Ber}_q \ll \text{Ber}_p$. Ist $p \in \{0, 1\}$, so ist $\text{Ber}_q \ll \text{Ber}_p$ genau dann, wenn $p = q$, und $\text{Ber}_q \perp \text{Ber}_p$ genau dann, wenn $q = 1 - p$.

(iii) Betrachte die Poisson-Verteilungen Poi_α und Poi_β für $\alpha, \beta \geq 0$. Es ist genau dann $\text{Poi}_\alpha \ll \text{Poi}_\beta$, wenn $\beta > 0$ oder $\alpha = 0$.

(iv) Betrachte die unendlichen Produktmaße (siehe Satz 1.64) $(\text{Ber}_p)^{\otimes \mathbb{N}}$ und $(\text{Ber}_q)^{\otimes \mathbb{N}}$ auf $\Omega = \{0, 1\}^{\mathbb{N}}$. Dann ist $(\text{Ber}_p)^{\otimes \mathbb{N}} \perp (\text{Ber}_q)^{\otimes \mathbb{N}}$, falls $p \neq q$. In der Tat: Sei $X_n((\omega_1, \omega_2, \dots)) = \omega_n$ für jedes $n \in \mathbb{N}$ die Projektion von Ω auf die n -te Koordinate. Dann ist $(X_n)_{n \in \mathbb{N}}$ unabhängig und Bernoulli-verteilt (siehe Beispiel 2.18) mit Parameter r unter $(\text{Ber}_r)^{\otimes \mathbb{N}}$. Nach dem starken Gesetz der großen Zahl gibt es also für $r \in \{p, q\}$ eine messbare Menge $A_r \subset \Omega$ mit $(\text{Ber}_r)^{\otimes \mathbb{N}}(\Omega \setminus A_r) = 0$ und

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i(\omega) = r \quad \text{für jedes } \omega \in A_r.$$

Speziell ist also $A_p \cap A_q = \emptyset$, falls $p \neq q$, und damit $(\text{Ber}_p)^{\otimes \mathbb{N}} \perp (\text{Ber}_q)^{\otimes \mathbb{N}}$. \diamond

Satz 7.33 (Zerlegungssatz von Lebesgue). Seien μ und ν σ -endliche Maße auf (Ω, \mathcal{A}) . Dann lässt sich ν auf eindeutige Weise zerlegen in den (bezüglich μ) absolutstetigen Anteil ν_a und den singulären Anteil ν_s :

$$\nu = \nu_a + \nu_s, \quad \text{wobei } \nu_a \ll \mu \text{ und } \nu_s \perp \mu.$$

ν_a hat eine Dichte bezüglich μ , und $\frac{d\nu_a}{d\mu}$ ist \mathcal{A} -messbar und μ -f.ü. endlich.

Korollar 7.34 (Satz von Radon-Nikodym). Seien μ und ν σ -endliche Maße auf (Ω, \mathcal{A}) . Dann gilt

$$\nu \text{ hat eine Dichte bezüglich } \mu \iff \nu \ll \mu.$$

In diesem Fall ist $\frac{d\nu}{d\mu}$ \mathcal{A} -messbar und μ -f.ü. endlich. $\frac{d\nu}{d\mu}$ heißt **Radon-Nikodym-Ableitung** von ν nach μ .

Beweis. Die eine Richtung ist trivial. Sei also $\nu \ll \mu$. Mit Satz 7.33 bekommen wir, dass $\nu = \nu_a$ eine Dichte bezüglich μ hat. \square

Beweis (Satz 7.33). Die Idee geht auf v. Neumann zurück, wir folgen der Darstellung in [42].

Wir können uns durch die üblichen Ausschöpfungsargumente auf den Fall beschränken wo μ und ν endlich sind. Nach Satz 4.19 ist die kanonische Inklusion $i : \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu) \hookrightarrow \mathcal{L}^1(\Omega, \mathcal{A}, \mu + \nu)$ stetig. Wegen $\nu \leq \mu + \nu$ ist also auch die Linearform $\mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu) \rightarrow \mathbb{R}$, $h \mapsto \int h d\nu$ stetig. Nach dem Satz von Riesz-Fréchet (hier: Korollar 7.28) existiert daher ein $g \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu)$ mit

$$\int h d\nu = \int hg d(\mu + \nu) \quad \text{für jedes } h \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu), \quad (7.7)$$

oder äquivalent dazu

$$\int f(1-g) d(\mu + \nu) = \int f d\mu \quad \text{für jedes } f \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu). \quad (7.8)$$

Wählen wir in (7.7) speziell $h = \mathbb{1}_{\{g < 0\}}$, so folgt, dass $(\mu + \nu)$ -fast überall $g \geq 0$ gilt. Analog folgt mit $f = \mathbb{1}_{\{g > 1\}}$, in (7.8) dass $(\mu + \nu)$ -fast überall $g \leq 1$ gilt, also ist $0 \leq g \leq 1$.

Sei nun $f \geq 0$ messbar, und seien $0 \leq f_1 \leq f_2 \leq \dots$ Funktionen in $\mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu)$ mit $f_n \uparrow f$. Nach dem Satz von der monotonen Konvergenz (angewandt auf das Maß $(1-g)(\mu + \nu)$), dem Maß mit Dichte $(1-g)$ bezüglich $\mu + \nu$ erhalten wir, dass (7.8) für alle messbaren $f \geq 0$ gilt. Analog folgt die Gültigkeit von (7.7) für alle messbaren $h \geq 0$.

Sei $E := g^{-1}(\{1\})$. Setzen wir $f = \mathbb{1}_E$ in (7.8) ein, so erhalten wir $\mu(E) = 0$. Wir definieren jetzt zwei Maße ν_a und ν_s für $A \in \mathcal{A}$ durch

$$\nu_a(A) := \nu(A \setminus E) \quad \text{und} \quad \nu_s(A) := \nu(A \cap E).$$

Offenbar gilt $\nu = \nu_a + \nu_s$ und $\nu_s(\Omega \setminus E) = 0$, also $\nu_s \perp \mu$. Ist nun $A \cap E = \emptyset$ und $\mu(A) = 0$, so ist $\int \mathbb{1}_A d\mu = 0$, also nach (7.8) auch $\int_A (1-g) d(\mu + \nu) = 0$. Andererseits ist $1-g > 0$ auf A , also $\mu(A) + \nu(A) = 0$ und damit $\nu_a(A) = \nu(A) = 0$. Ist allgemeiner B messbar mit $\mu(B) = 0$, so ist $\mu(B \setminus E) = 0$, also nach dem

Gezeigten $\nu_a(B) = \nu_a(B \setminus E) = 0$. Folglich ist $\nu_a \ll \mu$ und $\nu = \nu_a + \nu_s$ die gewünschte Zerlegung.

Um die Dichte von ν_a bezüglich μ zu erhalten, setzen wir $f := \frac{g}{1-g} \mathbb{1}_{\Omega \setminus E}$. Für jedes $A \in \mathcal{A}$ ist nun nach (7.8) und (7.7) mit $h = \mathbb{1}_{A \setminus E}$

$$\int_A f d\mu = \int_{A \cap E^c} g d(\mu + \nu) = \nu(A \setminus E) = \nu_a(A).$$

Also ist $f = \frac{d\nu_a}{d\mu}$. □

Übung 7.4.1. Wir definieren eine Abbildung $F : (0, 1] \rightarrow (0, 1]$ an der Stelle $x \in (0, 1]$ mit nicht abbrechender Binärdarstellung $x = (0, x_1 x_2 x_3 \dots) := \sum_{n=1}^{\infty} x_n 2^{-n}$ durch

$$F(x) = (0, x_1 x_1 x_2 x_2 x_3 x_3 \dots) = \sum_{n=1}^{\infty} 3 x_n 4^{-n}.$$

Sei U eine in $(0, 1]$ gleichverteilte Zufallsvariable und $\mu := \mathbf{P}_{U \circ F^{-1}}$ die Verteilung von $F(U)$.

Man zeige, dass μ eine stetige Verteilungsfunktion hat, und dass μ singulär zum Lebesgue-Maß $\lambda|_{(0,1]}$ ist. ♣

Übung 7.4.2. Sei $n \in \mathbb{N}$ und $p, q \in [0, 1]$. Unter welchen Bedingungen gilt für die Binomialverteilungen $b_{n,p} \ll b_{n,q}$? Man bestimme die Radon-Nikodym Ableitung $\frac{db_{n,p}}{db_{n,q}}$. ♣

7.5 Ergänzung: Signierte Maße

In diesem Abschnitt bringen wir die Zerlegungssätze für signierte Maße (Hahn, Jordan) und liefern einen alternativen Beweis für den Lebesgue'schen Zerlegungssatz.

Definition 7.35. Seien μ und ν zwei Maße auf (Ω, \mathcal{A}) . ν heißt **totalstetig** bezüglich μ , falls es für jedes $\varepsilon > 0$ ein $\delta > 0$ gibt, sodass für jedes $A \in \mathcal{A}$ gilt

$$\mu(A) < \delta \quad \text{impliziert} \quad \nu(A) < \varepsilon. \quad (7.9)$$

Bemerkung 7.36. Die Definition der Totalstetigkeit ist sehr ähnlich der der gleichgradigen Integrierbarkeit (siehe Satz 6.24(iii)), jedenfalls für endliches μ . Auf diesen Zusammenhang kommen wir im Rahmen des Martingalkonvergenzsatzes zurück, der einen alternativen Beweis für den Satz von Radon-Nikodym (Korollar 7.34) liefert. ◇

Satz 7.37. Seien μ und ν Maße auf (Ω, \mathcal{A}) . Ist ν totalstetig bezüglich μ , dann ist $\nu \ll \mu$. Ist $\nu(\Omega) < \infty$, so gilt auch die umgekehrte Implikation.

Beweis. „ \Rightarrow “ Sei ν totalstetig bezüglich μ . Sei $A \in \mathcal{A}$ mit $\mu(A) = 0$. Für jedes $\varepsilon > 0$ ist nach Voraussetzung $\nu(A) < \varepsilon$, also $\nu(A) = 0$ und damit $\nu \ll \mu$.

„ \Leftarrow “ Sei ν endlich, aber nicht totalstetig bezüglich μ . Dann existiert ein $\varepsilon > 0$ und Mengen $A_n \in \mathcal{A}$ mit $\mu(A_n) < 2^{-n}$, aber $\nu(A_n) \geq \varepsilon$ für jedes $n \in \mathbb{N}$. Setze $A := \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$. Dann ist

$$\mu(A) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \mu(A_k) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} 2^{-k} = 0.$$

Da ν endlich ist und stetig von oben (Satz 1.36), gilt

$$\nu(A) = \lim_{n \rightarrow \infty} \nu\left(\bigcup_{k=n}^{\infty} A_k\right) \geq \inf_{n \in \mathbb{N}} \nu(A_n) \geq \varepsilon > 0.$$

Also ist $\nu \not\ll \mu$. □

Beispiel 7.38. Die Endlichkeitsannahme ist für die Umkehrung im vorigen Satz essenziell. Sei beispielsweise $\mu = \mathcal{N}_{0,1}$ die Standardnormalverteilung auf \mathbb{R} und ν das Lebesgue-Maß auf \mathbb{R} . Dann hat ν bezüglich μ die Dichte $f(x) = \sqrt{2\pi} e^{-x^2/2}$. Speziell gilt $\nu \ll \mu$. Andererseits gilt $\mu([n, \infty)) \xrightarrow{n \rightarrow \infty} 0$ und $\nu([n, \infty)) = \infty$ für jedes $n \in \mathbb{N}$. Mithin ist ν nicht totalstetig bezüglich μ . ◇

Beispiel 7.39. Sei (Ω, \mathcal{A}) ein Messraum, und seien μ und ν endliche Maße auf (Ω, \mathcal{A}) . Mit \mathcal{Z} bezeichnen wir die Menge der endlichen Zerlegungen von Ω in disjunkte, messbare Mengen. Das heißt, $Z \in \mathcal{Z}$ ist eine endliche Teilmenge von \mathcal{A} so, dass die Mengen $C \in Z$ paarweise disjunkt sind und $\bigcup_{C \in Z} C = \Omega$ für jedes Z . Für $Z \in \mathcal{Z}$ definieren wir eine Funktion $f_Z : \Omega \rightarrow \mathbb{R}$ durch

$$f_Z(\omega) = \sum_{C \in Z: \mu(C) > 0} \frac{\nu(C)}{\mu(C)} \mathbb{1}_C(\omega).$$

Wir zeigen, dass die folgenden drei Aussagen äquivalent sind:

- (i) Die Familie $(f_Z : Z \in \mathcal{Z})$ ist gleichgradig integrierbar in $\mathcal{L}^1(\mu)$ und $\int f_Z d\mu = \nu(\Omega)$ für jedes $Z \in \mathcal{Z}$.
- (ii) Es gilt $\nu \ll \mu$.
- (iii) ν ist totalstetig bezüglich μ .

Die Äquivalenz von (ii) und (iii) wurde im vorigen Satz bewiesen. Gilt (ii), so ist für jedes $Z \in \mathcal{Z}$

$$\int f_Z d\mu = \sum_{C \in Z: \mu(C) > 0} \nu(C) = \nu(\Omega),$$

weil $\nu(C) = 0$ ist für diejenigen C , die in der Summe nicht auftauchen. Sei nun $\varepsilon > 0$ gegeben. Da (iii) aus (ii) folgt, gibt es ein $\delta' > 0$, sodass $\nu(A) < \varepsilon/2$ ist für jedes $A \in \mathcal{A}$ mit $\mu(A) \leq \delta'$. Sei $K := \nu(\Omega)/\delta'$ und $\delta < \varepsilon/(2K)$. Dann ist

$$\mu \left(\bigcup_{C \in Z: K\mu(C) \leq \nu(C)} C \right) = \sum_{C \in Z: K\mu(C) \leq \nu(C)} \mu(C) \leq \frac{1}{K} \nu(\Omega) = \delta',$$

also

$$\sum_{C \in Z: K\mu(C) \leq \nu(C)} \nu(C) = \nu \left(\bigcup_{C \in Z: K\mu(C) \leq \nu(C)} C \right) < \frac{\varepsilon}{2}.$$

Es folgt schließlich für $A \in \mathcal{A}$ mit $\mu(A) < \delta$

$$\begin{aligned} \int_A f_Z d\mu &= \sum_{C \in Z: \mu(C) > 0} \mu(A \cap C) \frac{\nu(C)}{\mu(C)} \\ &= \sum_{0 < K\mu(C) \leq \nu(C)} \mu(A \cap C) \frac{\nu(C)}{\mu(C)} + \sum_{K\mu(C) > \nu(C)} \mu(A \cap C) \frac{\nu(C)}{\mu(C)} \\ &\leq \frac{\varepsilon}{2} + \sum_{K\mu(C) > \nu(C)} K \mu(A \cap C) \leq \frac{\varepsilon}{2} + K \mu(A) < \varepsilon. \end{aligned}$$

Also ist $(f_Z, Z \in \mathcal{Z})$ gleichgradig integrierbar nach Satz 6.24(iii).

Gelte nun (i). Ist $\mu = 0$, so ist $\int f d\mu = 0$ für jedes f , also $\nu(\Omega) = 0$ und damit $\nu \ll \mu$. Sei also $\mu \neq 0$. Sei $A \in \mathcal{A}$ mit $\mu(A) = 0$. Dann ist $Z = \{A, A^c\} \in \mathcal{Z}$ und $f_Z = \mathbb{1}_{A^c} \nu(A^c)/\mu(A^c)$. Nach Voraussetzung ist $\nu(\Omega) = \int f_Z d\mu = \nu(A^c)$, also $\nu(A) = 0$ und damit $\nu \ll \mu$. \diamond

Definition 7.40 (Ladungsverteilung, signiertes Maß). Eine Mengenfunktion $\varphi : \mathcal{A} \rightarrow \mathbb{R}$ heißt **signiertes Maß** oder **Ladungsverteilung** auf (Ω, \mathcal{A}) , falls sie σ -additiv ist, falls also für jede Folge paarweise disjunkter Mengen $A_1, A_2, \dots \in \mathcal{A}$ gilt, dass

$$\varphi \left(\biguplus_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \varphi(A_n). \quad (7.10)$$

Die Menge aller Ladungsverteilungen bezeichnen wir mit $\mathcal{LV} = \mathcal{LV}(\Omega, \mathcal{A})$.

Bemerkung 7.41. (i) Ist φ ein signiertes Maß, so liegt in (7.10) automatisch schon absolute Konvergenz vor. Tatsächlich ändert sich ja der Wert der linken Seite nicht, wenn wir die Mengen A_1, A_2, \dots umnummerieren. Damit dies für die rechte Seite auch gilt, muss nach dem Weierstraß'schen Umordnungssatz die Reihe absolut konvergieren. Speziell gilt für jede Folge $(A_n)_{n \in \mathbb{N}}$ paarweise disjunkter Mengen $\lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} |\varphi(A_k)| = 0$.

(ii) Ist $\varphi \in \mathcal{LV}$, so ist $\varphi(\emptyset) = 0$, da $\mathbb{R} \ni \nu(\emptyset) = \sum_{n \in \mathbb{N}} \nu(\emptyset)$.

(iii) $\varphi \in \mathcal{LV}$ ist im Allgemeinen nicht σ -subadditiv. \diamond

Beispiel 7.42. Sind μ^+, μ^- endliche Maße, so ist $\varphi := \mu^+ - \mu^- \in \mathcal{LV}$. Wir werden sehen, dass jedes signierte Maß eine solche Darstellung besitzt. \diamond

Satz 7.43 (Zerlegungssatz von Hahn). Sei φ ein signiertes Maß. Dann gibt es eine Menge $\Omega^+ \in \mathcal{A}$ mit $\varphi(A) \geq 0$ für jedes $A \in \mathcal{A}$, $A \subset \Omega^+$ und $\varphi(A) \leq 0$ für jedes $A \in \mathcal{A}$, $A \subset \Omega^- := \Omega \setminus \Omega^+$. Eine solche Darstellung $\Omega = \Omega^- \uplus \Omega^+$ wird auch Hahn-Zerlegung von Ω (bezüglich φ) genannt.

Beweis. Sei $\alpha := \sup \{ \varphi(A) : A \in \mathcal{A} \}$. Wir müssen zeigen, dass φ das Maximum α tatsächlich annimmt, dass es also ein $\Omega^+ \in \mathcal{A}$ gibt mit $\varphi(\Omega^+) = \alpha$. Dann ist nämlich $\alpha \in \mathbb{R}$, und für $A \subset \Omega^+$, $A \in \mathcal{A}$ gilt

$$\alpha \geq \varphi(\Omega^+ \setminus A) = \varphi(\Omega^+) - \varphi(A) = \alpha - \varphi(A),$$

also $\varphi(A) \geq 0$. Für $A \subset \Omega^-$, $A \in \mathcal{A}$ ist $\varphi(A) \leq 0$, denn

$$\alpha \geq \varphi(\Omega^+ \cup A) = \varphi(\Omega^+) + \varphi(A) = \alpha + \varphi(A).$$

Wir konstruieren nun Ω^+ mit $\varphi(\Omega^+) = \alpha$. Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge in \mathcal{A} mit $\alpha = \lim_{n \rightarrow \infty} \varphi(A_n)$. Setze $A := \bigcup_{n=1}^{\infty} A_n$. Da jedes A_n noch „Anteile mit negativer Masse“ enthalten kann, können wir nicht einfach $\Omega^+ = A$ wählen. Vielmehr müssen wir Schicht für Schicht die negativen Anteile abfischen.

Setze $A_n^0 := A_n$ und $A_n^1 := A \setminus A_n$ sowie

$$\mathcal{P}_n := \left\{ \bigcap_{i=1}^n A_i^{s(i)} : s \in \{0, 1\}^n \right\}$$

die Partition von A , die von A_1, \dots, A_n erzeugt wird. Offensichtlich gilt für $B, C \in \mathcal{P}_n$ entweder $B = C$ oder $B \cap C = \emptyset$. Außerdem gilt $A_n = \biguplus_{\substack{B \in \mathcal{P}_n \\ B \subset A_n}} B$. Setze

$$\mathcal{P}_n^- := \{B \in \mathcal{P}_n : \varphi(B) < 0\}, \quad \mathcal{P}_n^+ := \mathcal{P}_n \setminus \mathcal{P}_n^-,$$

und

$$C_n := \bigcup_{B \in \mathcal{P}_n^+} B.$$

Wegen der endlichen Additivität von φ ist

$$\varphi(A_n) = \sum_{\substack{B \in \mathcal{P}_n \\ B \subset A_n}} \varphi(B) \leq \sum_{\substack{B \in \mathcal{P}_n^+ \\ B \subset A_n}} \varphi(B) \leq \sum_{B \in \mathcal{P}_n^+} \varphi(B) = \varphi(C_n).$$

Für $m \leq n$ setze $E_m^n = C_m \cup \dots \cup C_n$. Für $m < n$ ist $E_m^n \setminus E_m^{n-1} \subset C_n$, also

$$E_m^n \setminus E_m^{n-1} = \biguplus_{\substack{B \in \mathcal{P}_n^+ \\ B \subset E_m^n \setminus E_m^{n-1}}} B.$$

Speziell ist $\varphi(E_m^n \setminus E_m^{n-1}) \geq 0$. Für $E_m := \bigcup_{n \geq m} C_n$ gilt außerdem $E_m^n \uparrow E_m$ ($n \rightarrow \infty$) und

$$\begin{aligned} \varphi(A_m) &\leq \varphi(C_m) = \varphi(E_m^m) \leq \varphi(E_m^m) + \sum_{n=m+1}^{\infty} \varphi(E_m^n \setminus E_m^{n-1}) \\ &= \varphi\left(E_m^m \cup \bigcup_{n=m+1}^{\infty} (E_m^n \setminus E_m^{n-1})\right) = \varphi\left(\bigcup_{n=m}^{\infty} E_m^n\right) = \varphi(E_m). \end{aligned}$$

Wir setzen jetzt $\Omega^+ = \bigcap_{m=1}^{\infty} E_m$, also $E_m \downarrow \Omega^+$. Dann ist

$$\begin{aligned} \varphi(E_m) &= \varphi\left(\Omega^+ \uplus \biguplus_{n \geq m} (E_n \setminus E_{n+1})\right) \\ &= \varphi(\Omega^+) + \sum_{n=m}^{\infty} \varphi(E_n \setminus E_{n+1}) \xrightarrow{m \rightarrow \infty} \varphi(\Omega^+), \end{aligned}$$

wobei wir im letzten Schritt Bemerkung 7.41(i) ausgenutzt haben. Insgesamt ist

$$\alpha = \lim_{m \rightarrow \infty} \varphi(A_m) \leq \lim_{m \rightarrow \infty} \varphi(E_m) = \varphi(\Omega^+).$$

Per Definition ist aber $\alpha \geq \varphi(\Omega^+)$, also $\alpha = \varphi(\Omega^+)$, was zu zeigen war. \square

Korollar 7.44 (Zerlegungssatz von Jordan). Sei $\varphi \in \mathcal{LV}(\Omega, \mathcal{A})$ ein signiertes Maß. Dann gibt es eindeutig bestimmte endliche Maße φ^+, φ^- mit $\varphi = \varphi^+ - \varphi^-$ und $\varphi^+ \perp \varphi^-$.

Beweis. Sei $\Omega = \Omega^+ \uplus \Omega^-$ die Hahn-Zerlegung. Setze $\varphi^+(A) := \varphi(A \cap \Omega^+)$ und $\varphi^-(A) := -\varphi(A \cap \Omega^-)$.

Die Eindeutigkeit der Zerlegung ist trivial. \square

Korollar 7.45. Sei $\varphi \in \mathcal{LV}(\Omega, \mathcal{A})$ und $\varphi = \varphi^+ - \varphi^-$ die Jordan-Zerlegung von φ sowie $\Omega = \Omega^+ \uplus \Omega^-$ die Hahn-Zerlegung von Ω . Dann definiert

$$\begin{aligned}\|\varphi\|_{TV} &:= \sup \{\varphi(A) - \varphi(\Omega \setminus A) : A \in \mathcal{A}\} \\ &= \varphi(\Omega^+) - \varphi(\Omega^-) \\ &= \varphi^+(\Omega) + \varphi^-(\Omega)\end{aligned}$$

eine Norm auf $\mathcal{LV}(\Omega, \mathcal{A})$, die so genannte **Totalvariationsnorm**.

Beweis. Zu zeigen ist nur die Dreiecksungleichung. Seien $\varphi_1, \varphi_2 \in \mathcal{LV}$. Sei $\Omega = \Omega^+ \uplus \Omega^-$ die Hahn-Zerlegung bezüglich $\varphi := \varphi_1 + \varphi_2$ und $\Omega = \Omega_i^+ \uplus \Omega_i^-$ die bezüglich φ_i , $i = 1, 2$. Dann gilt

$$\begin{aligned}\|\varphi_1 + \varphi_2\|_{TV} &= \varphi_1(\Omega^+) - \varphi_1(\Omega^-) + \varphi_2(\Omega^+) - \varphi_2(\Omega^-) \\ &\leq \varphi_1(\Omega_1^+) - \varphi_1(\Omega_1^-) + \varphi_2(\Omega_2^+) - \varphi_2(\Omega_2^-) \\ &= \|\varphi_1\|_{TV} + \|\varphi_2\|_{TV}.\end{aligned}\quad \square$$

Wir wollen jetzt einen alternativen Beweis des Zerlegungssatzes von Lebesgue (Satz 7.33) angeben und bereiten dies mit einem Lemma vor.

Lemma 7.46. Seien μ, ν endliche Maße auf (Ω, \mathcal{A}) , die nicht singulär zueinander sind, kurz: $\mu \not\perp \nu$. Dann gibt es ein $A \in \mathcal{A}$ mit $\mu(A) > 0$ und ein $\varepsilon > 0$ mit

$$\varepsilon \mu(E) \leq \nu(E) \quad \text{für jedes } E \in \mathcal{A} \text{ mit } E \subset A.$$

Beweis. Für $n \in \mathbb{N}$ sei $\Omega = \Omega_n^+ \uplus \Omega_n^-$ eine Hahn-Zerlegung zu $(\nu - \frac{1}{n}\mu) \in \mathcal{LV}$. Setze $M := \bigcap_{n \in \mathbb{N}} \Omega_n^-$. Offenbar ist $(\nu - \frac{1}{n}\mu)(M) \leq 0$, also $\nu(M) \leq \frac{1}{n}\mu(M)$ für jedes $n \in \mathbb{N}$ und deshalb $\nu(M) = 0$. Wegen $\mu \not\perp \nu$ folgt $\mu(\Omega \setminus M) = \mu(\bigcup_{n \in \mathbb{N}} \Omega_n^+) > 0$, also $\mu(\Omega_{n_0}^+) > 0$ für ein $n_0 \in \mathbb{N}$. Setze $A := \Omega_{n_0}^+$ und $\varepsilon := \frac{1}{n_0}$. Damit ist dann $\mu(A) > 0$ und $(\nu - \varepsilon\mu)(E) \geq 0$ für jedes $E \subset A$, $E \in \mathcal{A}$. \square

Alternativer Beweis von Satz 7.33. Wir zeigen hier nur die Existenz der Zerlegung. Indem wir eine geeignete Folge $\Omega_n \uparrow \Omega$ betrachten, können wir annehmen, dass ν schon endlich ist. Betrachte die Menge der Funktionen

$$\mathcal{G} := \left\{ g : \Omega \rightarrow [0, \infty] : g \text{ ist messbar und } \int_A g \, d\mu \leq \nu(A) \text{ für alle } A \in \mathcal{A} \right\}$$

und setze

$$\gamma := \sup \left\{ \int g \, d\mu : g \in \mathcal{G} \right\}.$$

Unser Ziel ist es, ein maximales Element f in \mathcal{G} zu konstruieren (also eines mit $\int f \, d\mu = \gamma$), das dann die gesuchte Dichte von ν_a ist.

Offenbar ist $0 \in \mathcal{G}$, also $\mathcal{G} \neq \emptyset$. Weiter gilt

$$f, g \in \mathcal{G} \quad \text{impliziert} \quad f \vee g \in \mathcal{G}. \quad (7.11)$$

Mit $E := \{f \geq g\}$ ist nämlich für $A \in \mathcal{A}$

$$\int_A (f \vee g) d\mu = \int_{A \cap E} f d\mu + \int_{A \setminus E} g d\mu \leq \nu(A \cap E) + \nu(A \setminus E) = \nu(A).$$

Wähle eine Folge $(g_n)_{n \in \mathbb{N}}$ in \mathcal{G} mit $\int g_n d\mu \xrightarrow{n \rightarrow \infty} \gamma$ und setze $f_n = g_1 \vee \dots \vee g_n$. Wegen (7.11) ist $f_n \in \mathcal{G}$. Der Satz von der monotonen Konvergenz liefert für $f := \sup\{f_n : n \in \mathbb{N}\}$

$$\int_A f d\mu = \sup_{n \in \mathbb{N}} \int_A f_n d\mu \leq \nu(A) \quad \text{für jedes } A \in \mathcal{A},$$

(das heißt $f \in \mathcal{G}$) und weiter

$$\int f d\mu = \sup_{n \in \mathbb{N}} \int f_n d\mu \geq \sup_{n \in \mathbb{N}} \int g_n d\mu = \gamma,$$

also $\int f d\mu = \gamma \leq \nu(\Omega)$. Wir definieren nun für jedes $A \in \mathcal{A}$

$$\nu_a(A) := \int_A f d\mu, \quad \nu_s(A) := \nu(A) - \nu_a(A).$$

Nach Konstruktion ist nun $\nu_a \ll \mu$ ein endliches Maß mit Dichte f bezüglich μ . Wegen $f \in \mathcal{G}$ ist $\nu_s(A) = \nu(A) - \int_A f d\mu \geq 0$ für jedes $A \in \mathcal{A}$, also ist auch ν_s ein endliches Maß. Es bleibt zu zeigen, dass $\nu_s \perp \mu$.

An dieser Stelle benutzen wir Lemma 7.46. Wir nehmen an, dass $\nu_s \not\perp \mu$ gälte. Dann gäbe es ein $\varepsilon > 0$ und ein $A \in \mathcal{A}$ mit $\mu(A) > 0$ so, dass $\varepsilon\mu(E) \leq \nu_s(E)$ für jedes $E \subset A$, $E \in \mathcal{A}$. Für $B \in \mathcal{A}$ wäre dann

$$\begin{aligned} \int_B (f + \varepsilon \mathbf{1}_A) d\mu &= \int_B f d\mu + \varepsilon\mu(A \cap B) \\ &\leq \nu_a(B) + \nu_s(A \cap B) \leq \nu_a(B) + \nu_s(B) = \nu(B). \end{aligned}$$

Mit anderen Worten: $(f + \varepsilon \mathbf{1}_A) \in \mathcal{G}$ und damit $\int (f + \varepsilon \mathbf{1}_A) d\mu = \gamma + \varepsilon\mu(A) > \gamma$, was im Widerspruch zur Definition von γ steht. Also ist tatsächlich $\nu_s \perp \mu$. \square

Übung 7.5.1. Sei μ ein σ -endliches Maß auf (Ω, \mathcal{A}) und φ ein signiertes Maß auf (Ω, \mathcal{A}) . Man zeige, dass, analog zum Satz von Radon-Nikodym, die beiden folgenden Aussagen äquivalent sind:

- (i) Für jedes $A \in \mathcal{A}$ mit $\mu(A) = 0$ ist $\varphi(A) = 0$.
- (ii) Es gibt ein $f \in \mathcal{L}^1(\mu)$ mit $\varphi = f\mu$, also $\int_A f d\mu = \varphi(A)$ für jedes $A \in \mathcal{A}$. \clubsuit

Übung 7.5.2. Seien μ, ν, α endliche Maße auf (Ω, \mathcal{A}) mit $\nu \ll \mu \ll \alpha$.

(i) Zeige, dass die Kettenregel für die Radon-Nikodym-Ableitung gilt:

$$\frac{d\nu}{d\alpha} = \frac{d\nu}{d\mu} \frac{d\mu}{d\alpha} \quad \text{a-f.ü.}$$

(ii) Zeige, dass $f := \frac{d\nu}{d(\mu+\nu)}$ existiert und dass μ -f.ü. $\frac{d\nu}{d\mu} = \frac{f}{1-f}$ gilt. ♣

7.6 Ergänzung: Dualräume

Nach dem Darstellungssatz von Riesz-Fréchet (Satz 7.26) hat jede stetige Linearform $F : L^2(\mu) \rightarrow \mathbb{R}$ eine Darstellung $F(g) = \langle f, g \rangle$ für ein $f \in L^2(\mu)$. Andererseits ist für jedes $f \in L^2(\mu)$ die Abbildung $L^2(\mu) \rightarrow \mathbb{R}$, $g \mapsto \langle f, g \rangle$ stetig und linear. Daher ist $L^2(\mu)$ in kanonischer Weise isomorph zu seinem topologischen Dualraum $(L^2(\mu))'$. Dieser ist allgemein wie folgt definiert.

Definition 7.47 (Dualraum). Sei $(V, \|\cdot\|)$ ein Banachraum. Der **Dualraum** V' von V ist definiert durch

$$V' := \{F : V \rightarrow \mathbb{R} \text{ ist stetig und linear}\}.$$

Für $F \in V'$ setzen wir $\|F\|' := \sup\{|F(f)| : \|f\| = 1\}$.

Bemerkung 7.48. Da F stetig ist, existiert für jedes $\delta > 0$ ein $\varepsilon > 0$, sodass $|F(f)| < \delta$ gilt für jedes $f \in V$ mit $\|f\| < \varepsilon$. Also ist $\|F\|' \leq \delta/\varepsilon < \infty$. ◊

Wir sind nun an dem Fall $V = L^p(\mu)$ für $p \in [1, \infty]$ interessiert und schreiben $\|F\|'_p$ für die Norm von $F \in V'$. Betrachten wir speziell $V = L^2(\mu)$, so ist nach der Cauchy-Schwarz'schen Ungleichung $\|F\|'_2 = \|f\|_2$. Dies lässt sich verallgemeinern:

Lemma 7.49. Seien $p, q \in [1, \infty]$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann ist die kanonische Abbildung

$$\begin{aligned} \kappa : L^q(\mu) &\rightarrow (L^p(\mu))' \\ \kappa(f)(g) &= \int fg \, d\mu \quad \text{für } f \in L^q(\mu), g \in L^p(\mu) \end{aligned}$$

eine Isometrie, das heißt $\|\kappa(f)\|'_p = \|f\|_q$.

Beweis. Wir zeigen die Gleichheit, indem wir beide Ungleichungen zeigen.

„≤“ Dies folgt direkt aus der Hölder'schen Ungleichung.

„ \geq “ Für jedes zulässige Paar p, q und $f \in \mathcal{L}^q(\mu)$, $g \in \mathcal{L}^p(\mu)$ gilt per Definition der Operatornorm $\|\kappa(f)\|'_p \|g\|_p \geq |\int f g d\mu|$. Wir definieren die Vorzeichenfunktion $\text{sign}(x) = \mathbb{1}_{(0,\infty)}(x) - \mathbb{1}_{(-\infty,0)}(x)$. Indem wir g durch $\tilde{g} := |g| \text{ sign}(f)$ ersetzen (beachte $\|\tilde{g}\|_p = \|g\|_p$), erhalten wir

$$\|\kappa(f)\|'_p \|g\|_p \geq \left| \int f \tilde{g} d\mu \right| = \|fg\|_1. \quad (7.12)$$

Sei zunächst $q = 1$ und $f \in \mathcal{L}^1(\mu)$. Mit $g \equiv 1 \in \mathcal{L}^\infty(\mu)$ in (7.12) folgt $\|\kappa(f)\|'_\infty \geq \|f\|_1$.

Sei nun $q \in (1, \infty)$. Wähle $g = |f|^{q-1}$. Wegen $\frac{q-1}{q} = \frac{1}{p}$ ist dann

$$\|\kappa(f)\|'_p \cdot \|g\|_p \geq \|fg\|_1 = \||f|^q\|_1 = \|f\|_q^q = \|f\|_q \cdot \|f\|_q^{q-1} = \|f\|_q \cdot \|g\|_p.$$

Sei schließlich $q = \infty$. Ohne Einschränkung sei $\|f\|_\infty \in (0, \infty)$. Sei $\varepsilon > 0$. Dann existiert ein $A_\varepsilon \in \mathcal{A}$ mit $0 < \mu(A_\varepsilon) < \infty$, sodass

$$A_\varepsilon \subset \{|f| > (1 - \varepsilon)\|f\|_\infty\}.$$

Setzen wir $g = \frac{1}{\mu(A_\varepsilon)} \mathbb{1}_{A_\varepsilon}$, so ist $\|g\|_1 = 1$ und $\|\kappa(f)\|'_1 \geq \|fg\|_1 \geq (1 - \varepsilon)\|f\|_\infty$. \square

Satz 7.50. Für $p \in [1, \infty)$ und $\frac{1}{p} + \frac{1}{q} = 1$ ist $L^q(\mu)$ isomorph zu $(L^p(\mu))'$ vermöge der Isometrie κ .

Beweis. Für den Beweis greifen wir zurück auf den Satz von Radon-Nikodym (Korollar 7.34). Allerdings skizzieren wir den Beweis nur, weil wir die Theorie der signierten Maße und Inhalte nicht vertiefen wollen. Ein signierter Inhalt ν ist eine additive Mengenfunktion, die sich als Differenz $\nu = \nu^+ - \nu^-$ zweier endlicher Inhalte darstellen lässt, also auch negative Werte annehmen kann. (Diese Begriffsbildung ist analog zu der des signierten Maßes, das sich ja als Differenz zweier Maße darstellen lässt.)

Da κ eine Isometrie ist, ist κ insbesondere injektiv. Wir müssen also nur noch zeigen, dass κ surjektiv ist. Sei $F \in (L^p(\mu))'$. Dann ist $\nu(A) = F(\mathbb{1}_A)$ ein signierter Inhalt auf \mathcal{A} , und es gilt

$$|\nu(A)| \leq \|F\|'_p (\mu(A))^{1/p}.$$

Da μ \emptyset -stetig ist, ist also auch ν \emptyset -stetig und daher ein signiertes Maß auf \mathcal{A} . Es gilt sogar $\nu \ll \mu$. Nach dem Satz von Radon-Nikodym (Korollar 7.34) (angewandt auf die Maße ν^- und ν^+ , vergleiche Übung 7.5.1) besitzt ν eine Dichte bezüglich μ , also eine messbare Funktion f mit $\nu = f\mu$.

Sei $\mathbb{E}_f := \{g : g \text{ ist Elementarfunktion mit } \mu(g \neq 0) < \infty\}$ und $\mathbb{E}_f^+ := \{g \in \mathbb{E}_f : g \geq 0\}$. Dann ist für $g \in \mathbb{E}_f$

$$F(g) = \int gf \, d\mu. \quad (7.13)$$

Um zu zeigen, dass (7.13) für alle $g \in \mathcal{L}^p(\mu)$ gilt, müssen wir zunächst zeigen, dass $f \in \mathcal{L}^q(\mu)$ liegt. Wir unterscheiden zwei Fälle.

Fall 1: $p = 1$. Für jedes $\alpha > 0$ ist

$$\begin{aligned} \mu(\{|f| > \alpha\}) &\leq \frac{1}{\alpha} \nu(\{|f| > \alpha\}) \\ &= \frac{1}{\alpha} F(\mathbb{1}_{\{|f| > \alpha\}}) \leq \frac{1}{\alpha} \|F\|'_1 \cdot \|\mathbb{1}_{\{|f| > \alpha\}}\|_1 = \frac{1}{\alpha} \|F\|'_1 \cdot \mu(\{|f| > \alpha\}). \end{aligned}$$

Es folgt $\mu(\{|f| > \alpha\}) = 0$, falls $\alpha > \|F\|'_1$, also $\|f\|_\infty \leq \|F\|'_1 < \infty$.

Fall 2: $p \in (1, \infty)$. Nach Satz 1.96 (und weil μ σ -endlich ist) existieren $g_1, g_2, \dots \in \mathbb{E}_f^+$ so, dass $g_n \uparrow |f|$ μ -f.ü. Setzen wir $h_n = \text{sign}(f)(g_n)^{q-1} \in \mathbb{E}_f$, so gilt

$$\begin{aligned} \|g_n\|_q^q &\leq \int h_n f \, d\mu = F(h_n) \\ &\leq \|F\|'_p \cdot \|h_n\|_p = \|F\|'_p \cdot (\|g_n\|_q)^{q-1}, \end{aligned}$$

also ist $\|g_n\|_q \leq \|F\|'_p$. Monotone Konvergenz (Satz 4.20) liefert nun $\|f\|_q \leq \|F\|'_p < \infty$ also $f \in \mathcal{L}^q(\mu)$.

Daher ist die Abbildung $\tilde{F} : g \mapsto \int gf \, d\mu$ in $(L^p(\mu))'$ und $\tilde{F}(g) = F(g)$ für jedes $g \in \mathbb{E}_f$. Da \tilde{F} stetig ist und $\mathbb{E}_f \subset L^p(\mu)$ dicht liegt, gilt schon $\tilde{F} = F$. \square

Bemerkung 7.51. Die Aussage von Satz 7.50 ist für $p = \infty$ im Allgemeinen falsch. (Für endliches \mathcal{A} ist die Aussage trivialerweise auch für $p = \infty$ richtig.) Sei beispielsweise $\Omega = \mathbb{N}$, $\mathcal{A} = 2^\Omega$ und μ das Zählmaß. Wir betrachten also Folgenräume $\ell^p = L^p(\mathbb{N}, 2^\mathbb{N}, \mu)$. Für den Unterraum $\ell^K \subset \ell^\infty$ der konvergenten Folgen ist $F : \ell^K \rightarrow \mathbb{R}$, $(a_n)_{n \in \mathbb{N}} \mapsto \lim_{n \rightarrow \infty} a_n$ ein stetiges lineares Funktional. Nach den Hahn-Banach Sätzen der Funktionalanalysis (siehe etwa [80] oder [179]) kann F zu einem stetigen linearen Funktional auf ℓ^∞ fortgesetzt werden. Offenbar gibt es jedoch kein $(b_n)_{n \in \mathbb{N}} \in \ell^1$ mit $F((a_n)_{n \in \mathbb{N}}) = \sum_{m=1}^{\infty} a_m b_m$. \diamond

Übung 7.6.1. Man zeige, dass $\mathbb{E}_f \subset L^p(\mu)$ dicht liegt, falls $p \in [1, \infty)$. \clubsuit

8 Bedingte Erwartungen

Wenn über den Ausgang eines Zufallsexperimentes eine Teilinformation vorhanden ist, ändern sich die Wahrscheinlichkeiten für die möglichen Ereignisse. Das Konzept der bedingten Wahrscheinlichkeiten und bedingten Erwartungen formalisiert den zugehörigen Kalkül.

8.1 Elementare bedingte Wahrscheinlichkeiten

Beispiel 8.1. Wir werfen einen fairen sechsseitigen Würfel und betrachten die Ereignisse

$$\begin{aligned} A &:= \{\text{Augenzahl ungerade}\}, \\ B &:= \{\text{Augenzahl Drei oder kleiner}\}. \end{aligned}$$

Offenbar ist $\mathbf{P}[A] = \frac{1}{2}$ und $\mathbf{P}[B] = \frac{1}{2}$. Wie groß ist aber die Wahrscheinlichkeit, dass A eintritt, wenn wir schon wissen, dass B eintritt?

Wir modellieren das Experiment auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$, wobei $\Omega = \{1, \dots, 6\}$, $\mathcal{A} = 2^\Omega$ und \mathbf{P} die Gleichverteilung auf Ω ist. Dann ist

$$A = \{1, 3, 5\} \quad \text{und} \quad B = \{1, 2, 3\}.$$

Wenn wir nur wissen, dass B eingetreten ist, liegt es nahe, auf $\{1, 2, 3\}$ die Gleichverteilung zu vermuten. Wir definieren also auf $(B, 2^B)$ ein neues W-Maß \mathbf{P}_B durch

$$\mathbf{P}_B[C] = \frac{\#C}{\#B} \quad \text{für } C \subset B.$$

Indem wir Punkten in $\Omega \setminus B$ die Wahrscheinlichkeit Null geben (die können ja nicht eingetreten sein, wenn B eingetreten ist), können wir \mathbf{P}_B auf Ω fortsetzen durch

$$\mathbf{P}[C|B] := \mathbf{P}_B[C \cap B] = \frac{\#(C \cap B)}{\#B} \quad \text{für } C \subset \Omega.$$

$$\text{So erhalten wir } \mathbf{P}[A|B] = \frac{\#\{1, 3\}}{\#\{1, 2, 3\}} = \frac{2}{3}.$$

◇

Durch das Beispiel motiviert treffen wir die folgende Definition.

Definition 8.2 (Bedingte Wahrscheinlichkeit). Sei $(\Omega, \mathcal{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und $B \in \mathcal{A}$. Dann definieren wir die **bedingte Wahrscheinlichkeit gegeben B** für jedes $A \in \mathcal{A}$ durch

$$\mathbf{P}[A|B] = \begin{cases} \frac{\mathbf{P}[A \cap B]}{\mathbf{P}[B]}, & \text{falls } \mathbf{P}[B] > 0, \\ 0, & \text{sonst.} \end{cases} \quad (8.1)$$

Bemerkung 8.3. Die genaue Festsetzung in (8.1) für den Fall $\mathbf{P}[B] = 0$ ist willkürlich und unerheblich. \diamond

Satz 8.4. Ist $\mathbf{P}[B] > 0$, so ist $\mathbf{P}[\cdot | B]$ ein W-Maß auf (Ω, \mathcal{A}) .

Beweis. Trivial! \square

Satz 8.5. Seien $A, B \in \mathcal{A}$ mit $\mathbf{P}[A], \mathbf{P}[B] > 0$. Dann gilt

$$A, B \text{ sind unabhängig} \iff \mathbf{P}[A|B] = \mathbf{P}[A] \iff \mathbf{P}[B|A] = \mathbf{P}[B].$$

Beweis. Trivial! \square

Satz 8.6 (Formel von der totalen Wahrscheinlichkeit).

Sei I eine höchstens abzählbare Menge und $(B_i)_{i \in I}$ paarweise disjunkte Mengen mit $\mathbf{P}[\bigcup_{i \in I} B_i] = 1$. Dann gilt für jedes $A \in \mathcal{A}$

$$\mathbf{P}[A] = \sum_{i \in I} \mathbf{P}[A|B_i] \mathbf{P}[B_i]. \quad (8.2)$$

Beweis. Wegen der σ -Additivität von \mathbf{P} ist

$$\mathbf{P}[A] = \mathbf{P}\left[\bigcup_{i \in I} (A \cap B_i)\right] = \sum_{i \in I} \mathbf{P}[A \cap B_i] = \sum_{i \in I} \mathbf{P}[A|B_i] \mathbf{P}[B_i]. \quad \square$$

Satz 8.7 (Bayes'sche Formel). Sei I eine höchstens abzählbare Menge sowie $(B_i)_{i \in I}$ paarweise disjunkte Mengen mit $\mathbf{P}[\bigcup_{i \in I} B_i] = 1$. Dann gilt für jedes $A \in \mathcal{A}$ mit $\mathbf{P}[A] > 0$ und jedes $k \in I$

$$\mathbf{P}[B_k | A] = \frac{\mathbf{P}[A|B_k] \mathbf{P}[B_k]}{\sum_{i \in I} \mathbf{P}[A|B_i] \mathbf{P}[B_i]}. \quad (8.3)$$

Beweis. Es gilt

$$\mathbf{P}[B_k | A] = \frac{\mathbf{P}[B_k \cap A]}{\mathbf{P}[A]} = \frac{\mathbf{P}[A | B_k] \mathbf{P}[B_k]}{\mathbf{P}[A]}.$$

Setze jetzt (8.2) für $\mathbf{P}[A]$ ein. \square

Beispiel 8.8. Bei der Produktion gewisser elektronischer Bauteile sind 2% der Ware defekt. Ein schnelles Testverfahren erkennt ein defektes Bauteil mit Wahrscheinlichkeit 95%, meldet aber bei 10% der intakten Bauteile falschen Alarm.

Mit welcher Wahrscheinlichkeit ist ein als defekt erkanntes Bauteil wirklich defekt?

Wir formalisieren die obige Beschreibung. Seien

$$\begin{aligned} A &:= \{\text{Bauteil wird als defekt deklariert}\}, \\ B &:= \{\text{Bauteil ist defekt}\}, \end{aligned}$$

sowie

$$\mathbf{P}[B] = 0.02, \quad \mathbf{P}[B^c] = 0.98,$$

$$\mathbf{P}[A | B] = 0.95, \quad \mathbf{P}[A | B^c] = 0.1.$$

Die Bayes'sche Formel liefert nun

$$\begin{aligned} \mathbf{P}[B | A] &= \frac{\mathbf{P}[A | B] \mathbf{P}[B]}{\mathbf{P}[A | B] \mathbf{P}[B] + \mathbf{P}[A | B^c] \mathbf{P}[B^c]} \\ &= \frac{0.95 \cdot 0.02}{0.95 \cdot 0.02 + 0.1 \cdot 0.98} = \frac{19}{117} \approx 0.162. \end{aligned}$$

Andererseits ist die Wahrscheinlichkeit, dass ein nicht als defekt erkanntes Bauteil dennoch defekt ist

$$\mathbf{P}[B | A^c] = \frac{0.05 \cdot 0.02}{0.05 \cdot 0.02 + 0.9 \cdot 0.98} = \frac{1}{883} \approx 0.00113. \quad \diamond$$

Sei nun $X \in \mathcal{L}^1(\mathbf{P})$. Ist $A \in \mathcal{A}$, so ist offenbar auch $\mathbb{1}_A X \in \mathcal{L}^1(\mathbf{P})$, und wir setzen

$$\mathbf{E}[X; A] := \mathbf{E}[\mathbb{1}_A X]. \quad (8.4)$$

Ist $\mathbf{P}[A] > 0$, so ist $\mathbf{P}[\cdot | A]$ ein W-Maß. Wegen $\mathbb{1}_A X \in \mathcal{L}^1(\mathbf{P})$ ist auch $X \in \mathcal{L}^1(\mathbf{P}[\cdot | A])$. Also können wir den Erwartungswert von X bezüglich $\mathbf{P}[\cdot | A]$ definieren.

Definition 8.9. Sei $X \in \mathcal{L}^1(\mathbf{P})$ und $A \in \mathcal{A}$. Dann setzen wir

$$\mathbf{E}[X | A] := \int X(\omega) \mathbf{P}[d\omega | A] = \begin{cases} \frac{\mathbf{E}[\mathbb{1}_A X]}{\mathbf{P}[A]}, & \text{falls } \mathbf{P}[A] > 0, \\ 0, & \text{sonst.} \end{cases} \quad (8.5)$$

Offenbar ist $\mathbf{P}[B|A] = \mathbf{E}[\mathbb{1}_B|A]$ für jedes $B \in \mathcal{A}$.

Wir betrachten nun die Situation, die wir bei der Formel von der totalen Wahrscheinlichkeit untersucht hatten. Sei also I eine höchstens abzählbare Menge, und seien $(B_i)_{i \in I}$ paarweise disjunkte Ereignisse mit $\biguplus_{i \in I} B_i = \Omega$. Wir definieren $\mathcal{F} := \sigma(B_i, i \in I)$. Für $X \in \mathcal{L}^1(\mathbf{P})$ definieren wir eine Abbildung $\mathbf{E}[X|\mathcal{F}] : \Omega \rightarrow \mathbb{R}$ durch

$$\mathbf{E}[X|\mathcal{F}](\omega) = \mathbf{E}[X|B_i] \iff B_i \ni \omega. \quad (8.6)$$

Lemma 8.10. Die Abbildung $\mathbf{E}[X|\mathcal{F}]$ hat die folgenden Eigenschaften:

(i) $\mathbf{E}[X|\mathcal{F}]$ ist \mathcal{F} -messbar;

(ii) $\mathbf{E}[X|\mathcal{F}] \in \mathcal{L}^1(\mathbf{P})$, und für jedes $A \in \mathcal{F}$ gilt $\int_A \mathbf{E}[X|\mathcal{F}] d\mathbf{P} = \int_A X d\mathbf{P}$.

Beweis. (i) Sei f die Abbildung $f : \Omega \rightarrow I$, mit

$$f(\omega) = i \iff B_i \ni \omega.$$

Ferner sei $g : I \rightarrow \mathbb{R}$, $i \mapsto \mathbf{E}[X|B_i]$. Da I diskret ist, ist g messbar. Da f messbar ist bezüglich \mathcal{F} , ist auch $\mathbf{E}[X|\mathcal{F}] = g \circ f$ messbar bezüglich \mathcal{F} .

(ii) Sei $A \in \mathcal{F}$ und $J \subset I$ mit $A = \bigcup_{j \in J} B_j$. Sei $J' := \{i \in J : \mathbf{P}[B_i] > 0\}$. Dann ist

$$\int_A \mathbf{E}[X|\mathcal{F}] d\mathbf{P} = \sum_{i \in J'} \mathbf{P}[B_i] \mathbf{E}[X|B_i] = \sum_{i \in J'} \mathbf{E}[\mathbb{1}_{B_i} X] = \int_A X d\mathbf{P}. \quad \square$$

Übung 8.1.1 (Gedächtnislosigkeit der Exponentialverteilung). Sei $X > 0$ eine strikt positive Zufallsvariable und $\theta > 0$. Man zeige: Genau dann ist X exponentiellverteilt, wenn

$$\mathbf{P}[X > t + s | X > s] = \mathbf{P}[X > t] \quad \text{für alle } s, t \geq 0.$$

Insbesondere gilt für $\theta > 0$: Genau dann ist $X \sim \exp_\theta$, wenn $\mathbf{P}[X > t + s | X > s] = e^{-\theta t}$ für alle $s, t \geq 0$ gilt. ♣

Übung 8.1.2. Ein Theater mit n nummerierten Sitzen ist für heute Abend ausgebucht und jede der n wartenden Personen hat eine gültige Platzkarte. Nacheinander betreten die Gäste den Saal und nehmen Platz. Da die erste Person zerstreut ist, wählt sie jedoch zufällig einen Platz aus. Jede folgende Person setzt sich auf ihren reservierten Platz, falls der frei ist, und wählt sonst zufällig einen der freien Plätze aus.

(i) Wie groß ist die Wahrscheinlichkeit, dass die letzte Person ihren reservierten Platz bekommt?

(ii) Wie groß ist die Wahrscheinlichkeit, dass die k -te Person ihren reservierten Platz bekommt? ♣

8.2 Bedingte Erwartungen

Wir nehmen an, dass X eine uniform auf $[0, 1]$ verteilte Zufallsvariable ist, und dass bei Kenntnis des Wertes $X = x$ die Zufallsvariablen Y_1, \dots, Y_n unabhängig und Ber_x -verteilt sind. Mit unserem Apparat können wir bisher bedingte Wahrscheinlichkeiten vom Typ $\mathbf{P}[\cdot | X \in [a, b]]$, $a < b$, ausrechnen. Wie sieht es aber aus mit $\mathbf{P}[Y_1 = \dots = Y_n = 1 | X = x]$? Intuitiv sollte dies x^n sein. Wir brauchen einen Begriff der bedingten Wahrscheinlichkeit, der auch für Ereignisse mit Wahrscheinlichkeit Null in konsistenter Weise unserer Intuition entspricht. Wir werden (im nächsten Abschnitt) sehen, dass dies im vorliegenden Beispiel mit Hilfe von Übergangskernen möglich ist. Zunächst aber betrachten wir die allgemeine Situation.

Sei im Folgenden stets $\mathcal{F} \subset \mathcal{A}$ eine Unter- σ -Algebra und $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$. In Anlehnung an Lemma 8.10 treffen wir die folgende Definition.

Definition 8.11 (Bedingte Erwartung). Eine Zufallsvariable Y heißt **bedingte Erwartung** von X gegeben \mathcal{F} , symbolisch $\mathbf{E}[X | \mathcal{F}] := Y$, falls gilt:

- (i) Y ist \mathcal{F} -messbar.
- (ii) Für jedes $A \in \mathcal{F}$ gilt $\mathbf{E}[X \mathbb{1}_A] = \mathbf{E}[Y \mathbb{1}_A]$.

Für $B \in \mathcal{A}$ heißt $\mathbf{P}[B | \mathcal{F}] := \mathbf{E}[\mathbb{1}_B | \mathcal{F}]$ die **bedingte Wahrscheinlichkeit** von B gegeben \mathcal{F} .

Satz 8.12. $\mathbf{E}[X | \mathcal{F}]$ existiert und ist eindeutig (bis auf Gleichheit fast sicher).

Da bedingte Erwartungen nur bis auf Gleichheit f.s. definiert sind, sind alle Gleichheiten mit bedingten Erwartungen immer nur als Gleichheiten f.s. zu verstehen, auch wenn nicht explizit darauf hingewiesen wird.

Beweis. Eindeutigkeit. Seien Y und Y' Zufallsvariablen, die (i) und (ii) erfüllen. Setze $A = \{Y > Y'\} \in \mathcal{F}$. Dann ist nach Bedingung (ii)

$$0 = \mathbf{E}[Y \mathbb{1}_A] - \mathbf{E}[Y' \mathbb{1}_A] = \mathbf{E}[(Y - Y') \mathbb{1}_A].$$

Wegen $(Y - Y') \mathbb{1}_A \geq 0$, ist dann $\mathbf{P}[A] = 0$, also $Y \leq Y'$ fast sicher. Analog folgt $Y \geq Y'$ fast sicher.

Existenz. Seien $X^+ = X \vee 0$ und $X^- = X^+ - X$. Durch

$$Q^\pm(A) := \mathbf{E}[X^\pm \mathbb{1}_A] \quad \text{für jedes } A \in \mathcal{F},$$

werden zwei endliche Maße auf (Ω, \mathcal{F}) definiert. Offenbar ist $Q^\pm \ll \mathbf{P}$, also liefert der Satz von Radon-Nikodym (Korollar 7.34) die Existenz von \mathcal{F} -messbaren Dichten Y^\pm , sodass

$$Q^\pm(A) = \int_A Y^\pm d\mathbf{P} = \mathbf{E}[Y^\pm \mathbb{1}_A].$$

Setze nun $Y = Y^+ - Y^-$. □

Definition 8.13. Ist Y eine Zufallsvariable und $X \in \mathcal{L}^1(\mathbf{P})$, so definieren wir $\mathbf{E}[X|Y] := \mathbf{E}[X|\sigma(Y)]$.

Satz 8.14 (Eigenschaften der bedingten Erwartung). Seien $(\Omega, \mathcal{A}, \mathbf{P})$ und X wie oben sowie $\mathcal{G} \subset \mathcal{F} \subset \mathcal{A}$ σ -Algebren. Ferner sei $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$. Dann gilt:

(i) (**Linearität**) $\mathbf{E}[\lambda X + Y | \mathcal{F}] = \lambda \mathbf{E}[X | \mathcal{F}] + \mathbf{E}[Y | \mathcal{F}]$.

(ii) (**Monotonie**) Ist $X \geq Y$ f.s., so ist $\mathbf{E}[X | \mathcal{F}] \geq \mathbf{E}[Y | \mathcal{F}]$.

(iii) Ist $\mathbf{E}[|XY|] < \infty$ und ist Y messbar bezüglich \mathcal{F} , dann ist

$$\mathbf{E}[XY | \mathcal{F}] = Y \mathbf{E}[X | \mathcal{F}] \quad \text{und} \quad \mathbf{E}[Y | \mathcal{F}] = \mathbf{E}[Y | Y] = Y.$$

(iv) (**Turmeigenschaft**) $\mathbf{E}[\mathbf{E}[X | \mathcal{F}] | \mathcal{G}] = \mathbf{E}[\mathbf{E}[X | \mathcal{G}] | \mathcal{F}] = \mathbf{E}[X | \mathcal{G}]$.

(v) (**Dreiecksungleichung**) $\mathbf{E}[|X| | \mathcal{F}] \geq |\mathbf{E}[X | \mathcal{F}]|$.

(vi) (**Unabhängigkeit**) Sind $\sigma(X)$ und \mathcal{F} unabhängig, so ist $\mathbf{E}[X | \mathcal{F}] = \mathbf{E}[X]$.

(vii) Gilt $\mathbf{P}[A] \in \{0, 1\}$ für jedes $A \in \mathcal{F}$, so ist $\mathbf{E}[X | \mathcal{F}] = \mathbf{E}[X]$.

(viii) (**Majorisierte Konvergenz**) Ist $Y \in \mathcal{L}^1(\mathbf{P})$, $Y \geq 0$ und ist $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen mit $|X_n| \leq Y$ für $n \in \mathbb{N}$ sowie $X_n \xrightarrow{n \rightarrow \infty} X$ f.s., so gilt

$$\lim_{n \rightarrow \infty} \mathbf{E}[X_n | \mathcal{F}] = \mathbf{E}[X | \mathcal{F}] \quad \text{f.s. und in } L^1(\mathbf{P}). \quad (8.7)$$

Beweis. (i) Die rechte Seite ist \mathcal{F} -messbar, und für $A \in \mathcal{F}$ ist

$$\begin{aligned} \mathbf{E}[\mathbb{1}_A (\lambda \mathbf{E}[X | \mathcal{F}] + \mathbf{E}[Y | \mathcal{F}])] &= \lambda \mathbf{E}[\mathbb{1}_A \mathbf{E}[X | \mathcal{F}]] + \mathbf{E}[\mathbb{1}_A \mathbf{E}[Y | \mathcal{F}]] \\ &= \lambda \mathbf{E}[\mathbb{1}_A X] + \mathbf{E}[\mathbb{1}_A Y] \\ &= \mathbf{E}[\mathbb{1}_A (\lambda X + Y)]. \end{aligned}$$

(ii) Sei $A = \{\mathbf{E}[X | \mathcal{F}] < \mathbf{E}[Y | \mathcal{F}]\} \in \mathcal{F}$. Wegen $X \geq Y$ ist $\mathbf{E}[\mathbb{1}_A (X - Y)] \geq 0$, also $\mathbf{P}[A] = 0$.

(iii) Sei zunächst $X \geq 0$ und $Y \geq 0$. Für $n \in \mathbb{N}$ setze $Y_n = 2^{-n} \lfloor 2^n Y \rfloor$. Dann ist $Y_n \uparrow Y$ sowie $Y_n \mathbf{E}[X | \mathcal{F}] \uparrow Y \mathbf{E}[X | \mathcal{F}]$ (da $\mathbf{E}[X | \mathcal{F}] \geq 0$ nach (ii)). Es gilt nach dem Satz von der monotonen Konvergenz (Lemma 4.6(ii))

$$\mathbf{E}[\mathbb{1}_A Y_n \mathbf{E}[X | \mathcal{F}]] \xrightarrow{n \rightarrow \infty} \mathbf{E}[\mathbb{1}_A Y \mathbf{E}[X | \mathcal{F}]].$$

Andererseits ist

$$\begin{aligned}\mathbf{E}[\mathbb{1}_A Y_n \mathbf{E}[X|\mathcal{F}]] &= \sum_{k=1}^{\infty} \mathbf{E}[\mathbb{1}_A \mathbb{1}_{\{Y_n=k2^{-n}\}} k 2^{-n} \mathbf{E}[X|\mathcal{F}]] \\ &= \sum_{k=1}^{\infty} \mathbf{E}[\mathbb{1}_A \mathbb{1}_{\{Y_n=k2^{-n}\}} k 2^{-n} X] \\ &= \mathbf{E}[\mathbb{1}_A Y_n X] \xrightarrow{n \rightarrow \infty} \mathbf{E}[\mathbb{1}_A Y X].\end{aligned}$$

Also gilt $\mathbf{E}[\mathbb{1}_A Y \mathbf{E}[X|\mathcal{F}]] = \mathbf{E}[\mathbb{1}_A Y X]$. Im allgemeinen Fall schreiben wir $X = X^+ - X^-$ und $Y = Y^+ - Y^-$ und nutzen die Linearität der bedingten Erwartung aus.

(iv) Die zweite Gleichung folgt aus (iii) mit $Y = \mathbf{E}[X|\mathcal{G}]$ und $X = 1$. Sei nun $A \in \mathcal{G}$. Dann ist insbesondere auch $A \in \mathcal{F}$, also

$$\mathbf{E}[\mathbb{1}_A \mathbf{E}[\mathbf{E}[X|\mathcal{F}]|\mathcal{G}]] = \mathbf{E}[\mathbb{1}_A \mathbf{E}[X|\mathcal{F}]] = \mathbf{E}[\mathbb{1}_A X] = \mathbf{E}[\mathbb{1}_A \mathbf{E}[X|\mathcal{G}]].$$

(v) Das folgt aus (i) und (ii) mit $X = X^+ - X^-$.

(vi) Trivialerweise ist $\mathbf{E}[X]$ messbar bezüglich \mathcal{F} . Sei $A \in \mathcal{F}$. Dann sind X und $\mathbb{1}_A$ unabhängig, also ist $\mathbf{E}[\mathbf{E}[X|\mathcal{F}] \mathbb{1}_A] = \mathbf{E}[X \mathbb{1}_A] = \mathbf{E}[X] \mathbf{E}[\mathbb{1}_A]$.

(vii) Für jedes $A \in \mathcal{F}$ und $B \in \mathcal{A}$ gilt $\mathbf{P}[A \cap B] = 0$, falls $\mathbf{P}[A] = 0$ ist, und $\mathbf{P}[A \cap B] = \mathbf{P}[B]$, falls $\mathbf{P}[A] = 1$ ist. Also ist \mathcal{F} von \mathcal{A} unabhängig und damit auch von jeder Teil- σ -Algebra von \mathcal{A} . Speziell ist \mathcal{F} von $\sigma(X)$ unabhängig. Die Aussage folgt also aus (vi).

(viii) Sei $|X_n| \leq Y$ für jedes $n \in \mathbb{N}$ und $X_n \xrightarrow{n \rightarrow \infty} X$ fast sicher. Setze $Z_n := \sup_{k \geq n} |X_k - X|$. Dann ist $0 \leq Z_n \leq 2Y$ und $Z_n \xrightarrow{\text{f.s.}} 0$. Nach Korollar 6.26 (majorisierte Konvergenz) gilt $\mathbf{E}[Z_n] \xrightarrow{n \rightarrow \infty} 0$, also nach der Dreiecksungleichung

$$\mathbf{E}[|\mathbf{E}[X_n|\mathcal{F}] - \mathbf{E}[X|\mathcal{F}]|] \leq \mathbf{E}[\mathbf{E}[|X_n - X| |\mathcal{F}]] = \mathbf{E}[|X_n - X|] \leq \mathbf{E}[Z_n] \xrightarrow{n \rightarrow \infty} 0.$$

Dies ist aber die $L^1(\mathbf{P})$ -Konvergenz in (8.7). Da $(Z_n)_{n \in \mathbb{N}}$ monoton fallend ist, gilt dies nach (ii) auch für $(\mathbf{E}[Z_n|\mathcal{F}])_{n \in \mathbb{N}}$. Daher existiert $Z := \lim_{n \rightarrow \infty} \mathbf{E}[Z_n|\mathcal{F}]$. Nach dem Lemma von Fatou ist

$$\mathbf{E}[Z] \leq \lim_{n \rightarrow \infty} \mathbf{E}[\mathbf{E}[Z_n|\mathcal{F}]] = \lim_{n \rightarrow \infty} \mathbf{E}[Z_n] = 0,$$

also ist $Z = 0$ und damit $\mathbf{E}[Z_n|\mathcal{F}] \xrightarrow{n \rightarrow \infty} 0$ fast sicher. Nach (v) ist aber

$$|\mathbf{E}[X_n|\mathcal{F}] - \mathbf{E}[X|\mathcal{F}]| \leq \mathbf{E}[Z_n|\mathcal{F}]. \quad \square$$

Bemerkung 8.15. Intuitiv ist $\mathbf{E}[X|\mathcal{F}]$ die beste Vorhersage, die wir für den Wert von X machen können, wenn uns die Information aus der σ -Algebra \mathcal{F} zur Verfügung steht. Ist beispielsweise $\sigma(X) \subset \mathcal{F}$, kennen wir also X schon, dann ist

$\mathbf{E}[X|\mathcal{F}] = X$, wie in (iii) gezeigt. Am anderen Ende der Skala ist der Fall, wo X und \mathcal{F} unabhängig sind, wir also durch Kenntnis von \mathcal{F} keine Information über X gewinnen. Hier ist die beste Vorhersage für X der Erwartungswert selber, also $\mathbf{E}[X] = \mathbf{E}[X|\mathcal{F}]$ wie in (vi) gezeigt.

Was heißt dabei aber eigentlich genau „beste Vorhersage“? Wir wollen dies für quadratintegrierbare Zufallsvariablen X als diejenige \mathcal{F} -messbare Zufallsvariable verstehen, die den L^2 -Abstand zu X minimiert. Dass dies die bedingte Erwartung tatsächlich tut, ist der Inhalt des folgenden Korollars. \diamond

Bemerkung 8.16. Für eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ mit $X^- \in \mathcal{L}^1(\mathbf{P})$ können wir den bedingten Erwartungswert als monotonen Limes definieren

$$\mathbf{E}[X|\mathcal{F}] := \lim_{n \rightarrow \infty} \mathbf{E}[X_n|\mathcal{F}],$$

wobei $-X^- \leq X_1$ und $X_n \uparrow X$ gilt. Mit Hilfe der Monotonie (Satz 8.14(ii)) zeigt man leicht, dass der Grenzwert nicht von der Wahl der Folge (X_n) abhängt und die Bedingungen von Definition 8.11 erfüllt. Analog verfährt man für X mit $X^+ \in \mathcal{L}^1(\mathbf{P})$. Auch für diese Verallgemeinerung des bedingten Erwartungswertes gilt $\mathbf{E}[X|\mathcal{F}] \leq \mathbf{E}[Y|\mathcal{F}]$ f.s., falls $Y \geq X$ f.s. (siehe Übung 8.2.1). \diamond

Korollar 8.17 (Bedingte Erwartung als Projektion). Sei $\mathcal{F} \subset \mathcal{A}$ eine σ -Algebra und X eine Zufallsvariable mit $\mathbf{E}[X^2] < \infty$. Dann ist $\mathbf{E}[X|\mathcal{F}]$ die orthogonale Projektion von X auf $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$. Es gilt also für jedes \mathcal{F} -messbare Y mit $\mathbf{E}[Y^2] < \infty$

$$\mathbf{E}[(X - Y)^2] \geq \mathbf{E}[(X - \mathbf{E}[X|\mathcal{F}])^2]$$

mit Gleichheit genau dann, wenn $Y = \mathbf{E}[X|\mathcal{F}]$.

Beweis. Wir nehmen zunächst an, dass $\mathbf{E}[\mathbf{E}[X|\mathcal{F}]^2] < \infty$ gilt. (In Satz 8.20 werden wir sehen, dass sogar $\mathbf{E}[\mathbf{E}[X|\mathcal{F}]^2] \leq \mathbf{E}[X^2]$ gilt, aber wir wollen den Beweis hier etwas elementarer halten.) Sei Y messbar bezüglich \mathcal{F} und $\mathbf{E}[Y^2] < \infty$. Dann ist nach der Cauchy-Schwarz Ungleichung $\mathbf{E}[|XY|] < \infty$, also (mit der Turmeigenschaft) $\mathbf{E}[XY] = \mathbf{E}[\mathbf{E}[X|\mathcal{F}]Y]$. Entsprechend gilt $\mathbf{E}[XE[X|\mathcal{F}]] = \mathbf{E}[\mathbf{E}[XE[X|\mathcal{F}]]|\mathcal{F}] = \mathbf{E}[\mathbf{E}[X|\mathcal{F}]^2]$, also

$$\begin{aligned} \mathbf{E}[(X - Y)^2] - \mathbf{E}[(X - \mathbf{E}[X|\mathcal{F}])^2] \\ &= \mathbf{E}[X^2 - 2XY + Y^2 - X^2 + 2X\mathbf{E}[X|\mathcal{F}] - \mathbf{E}[X|\mathcal{F}]^2] \\ &= \mathbf{E}[Y^2 - 2Y\mathbf{E}[X|\mathcal{F}] + \mathbf{E}[X|\mathcal{F}]^2] \\ &= \mathbf{E}[(Y - \mathbf{E}[X|\mathcal{F}])^2] \geq 0. \end{aligned}$$

Für den Fall, wo $\mathbf{E}[\mathbf{E}[X|\mathcal{F}]^2] < \infty$ gilt, sind wir fertig. Es reicht also zu zeigen, dass diese Bedingung stets erfüllt ist, wenn $\mathbf{E}[X^2] < \infty$ ist. Für $N \in \mathbb{N}$

betrachten wir die abgeschnittenen Zufallsvariablen $|X| \wedge N$, für die offenbar $\mathbf{E}[\mathbf{E}[|X| \wedge N | \mathcal{F}]^2] \leq N^2$ gilt. Nach dem schon Gezeigten (mit $|X| \wedge N$ statt X und mit $Y = 0 \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$) und mit Hilfe der Ungleichung $a^2 \leq 2(a - b)^2 + 2b^2$, $a, b \in \mathbb{R}$, folgern wir

$$\begin{aligned}\mathbf{E}[\mathbf{E}[|X| \wedge N | \mathcal{F}]^2] &\leq 2\mathbf{E}\left[\left((|X| \wedge N) - \mathbf{E}[|X| \wedge N | \mathcal{F}]\right)^2\right] + 2\mathbf{E}[(|X| \wedge N)^2] \\ &\leq 4\mathbf{E}[|X| \wedge N]^2 \leq 4\mathbf{E}[X^2].\end{aligned}$$

Nach Satz 8.14(ii) und (viii) gilt $\mathbf{E}[|X| \wedge N | \mathcal{F}] \uparrow \mathbf{E}[|X| | \mathcal{F}]$ für $N \rightarrow \infty$. Nach der Dreiecksungleichung (Satz 8.14(v)) und dem Satz über monotone Konvergenz (Satz 4.20) folgt

$$\mathbf{E}[\mathbf{E}[X | \mathcal{F}]^2] \leq \mathbf{E}[\mathbf{E}[|X| | \mathcal{F}]^2] = \lim_{N \rightarrow \infty} \mathbf{E}[\mathbf{E}[|X| \wedge N | \mathcal{F}]^2] \leq 4\mathbf{E}[X^2] < \infty,$$

und der Beweis ist vollständig. \square

Beispiel 8.18. Seien $X, Y \in \mathcal{L}^1(\mathbf{P})$ unabhängig. Dann ist

$$\mathbf{E}[X + Y | Y] = \mathbf{E}[X | Y] + \mathbf{E}[Y | Y] = \mathbf{E}[X] + Y. \quad \diamond$$

Beispiel 8.19. Seien X_1, \dots, X_N unabhängig mit $\mathbf{E}[X_i] = 0$, $i = 1, \dots, N$. Setze $\mathcal{F}_n := \sigma(X_1, \dots, X_n)$ und $S_n := X_1 + \dots + X_n$ für $n = 1, \dots, N$. Dann ist für $n \geq m$

$$\begin{aligned}\mathbf{E}[S_n | \mathcal{F}_m] &= \mathbf{E}[X_1 | \mathcal{F}_m] + \dots + \mathbf{E}[X_n | \mathcal{F}_m] \\ &= X_1 + \dots + X_m + \mathbf{E}[X_{m+1}] + \dots + \mathbf{E}[X_n] \\ &= S_m.\end{aligned}$$

Nach Satz 8.14(iv) ist wegen $\sigma(S_m) \subset \mathcal{F}_m$ auch

$$\mathbf{E}[S_n | S_m] = \mathbf{E}[\mathbf{E}[S_n | \mathcal{F}_m] | S_m] = \mathbf{E}[S_m | S_m] = S_m. \quad \diamond$$

Wir kommen nun zur Jensen'schen Ungleichung für bedingte Erwartungen.

Satz 8.20 (Jensen'sche Ungleichung). Sei $I \subset \mathbb{R}$ ein Intervall, und sei $\varphi : I \rightarrow \mathbb{R}$ konvex und X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbf{P})$ mit Werten in I . Ferner sei $\mathbf{E}[|X|] < \infty$ und $\mathcal{F} \subset \mathcal{A}$ eine σ -Algebra. Dann gilt

$$\infty \geq \mathbf{E}[\varphi(X) | \mathcal{F}] \geq \varphi(\mathbf{E}[X | \mathcal{F}]).$$

Beweis. Für die Existenz von $\mathbf{E}[\varphi(X) | \mathcal{F}]$ mit Werten in $(-\infty, \infty]$, beachte, dass $\varphi(X)^- \in \mathcal{L}^1(\mathbf{P})$ und siehe Bemerkung 8.16. Nach Übung 8.2.2 ist $\mathbf{E}[X | \mathcal{F}] \in I$ fast sicher, also ist $\varphi(\mathbf{E}[X | \mathcal{F}])$ wohldefiniert.

(Man erinnere sich der Definition 1.68 zur Sprechweise „fast sicher auf A “.) Auf dem Ereignis $\{\mathbf{E}[X|\mathcal{F}]\}$ ist ein Randpunkt von I ist $X = \mathbf{E}[X|\mathcal{F}]$ fast sicher, und die Aussage ist trivial. In der Tat: Ohne Einschränkung sei 0 der linke Randpunkt von I und $A := \{\mathbf{E}[X|\mathcal{F}] = 0\}$. Da X Werte in $I \subset [0, \infty)$ annimmt, ist $0 \leq \mathbf{E}[X \mathbf{1}_A] = \mathbf{E}[\mathbf{E}[X|\mathcal{F}] \mathbf{1}_A] = 0$, also ist $X \mathbf{1}_A = 0$. Der Fall eines rechten Randpunktes geht analog.

Sei also nun das Ereignis $B := \{\mathbf{E}[X|\mathcal{F}] \text{ ist innerer Punkt von } I\}$ betrachtet.

Für jeden inneren Punkt $x \in I$ sei $D^+ \varphi(x)$ die maximale Tangentensteigung von φ in x , also der maximale Wert t mit $\varphi(y) \geq (y - x)t + \varphi(x)$ für alle $y \in I$ (siehe Satz 7.7). Dann gibt es für jedes $x \in I^\circ$ eine \mathbf{P} -Nullmenge N_x , sodass für jedes $\omega \in B \setminus N_x$ gilt, dass

$$\begin{aligned}\mathbf{E}[\varphi(X)|\mathcal{F}](\omega) &\geq \varphi(x) + \mathbf{E}[D^+ \varphi(x)(X - x)|\mathcal{F}](\omega) \\ &= \varphi(x) + D^+ \varphi(x)(\mathbf{E}[X|\mathcal{F}](\omega) - x) =: \psi_\omega(x).\end{aligned}\quad (8.8)$$

Sei $V := \mathbb{Q} \cap I^\circ$. Dann ist auch $N := \bigcup_{x \in V} N_x$ eine \mathbf{P} -Nullmenge, und für jedes $\omega \in B \setminus N$ und jedes $x \in V$ gilt (8.8).

Die Abbildung $x \mapsto D^+ \varphi(x)$ ist rechtsstetig (nach Satz 7.7(iv)), also ist $x \mapsto \psi_\omega(x)$ rechtsstetig. Daher gilt für jedes $\omega \in B \setminus N$

$$\begin{aligned}\varphi(\mathbf{E}[X|\mathcal{F}](\omega)) &= \psi_\omega(\mathbf{E}[X|\mathcal{F}](\omega)) \\ &\leq \sup_{x \in I^\circ} \psi_\omega(x) = \sup_{x \in V} \psi_\omega(x) \leq \mathbf{E}[\varphi(X)|\mathcal{F}](\omega).\end{aligned}\quad \square$$

Korollar 8.21. Sei $p \in [1, \infty]$ und $\mathcal{F} \subset \mathcal{A}$ eine Teil- σ -Algebra. Dann ist die Abbildung $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbf{P}) \rightarrow \mathcal{L}^p(\Omega, \mathcal{F}, \mathbf{P})$, $X \mapsto \mathbf{E}[X|\mathcal{F}]$ eine Kontraktion (das heißt: $\|\mathbf{E}[X|\mathcal{F}]\|_p \leq \|X\|_p$) und damit insbesondere stetig. Es gilt also für $X, X_1, X_2, \dots \in \mathcal{L}^p(\Omega, \mathcal{A}, \mathbf{P})$ mit $\|X_n - X\|_p \xrightarrow{n \rightarrow \infty} 0$ auch

$$\|\mathbf{E}[X_n|\mathcal{F}] - \mathbf{E}[X|\mathcal{F}]\|_p \xrightarrow{n \rightarrow \infty} 0.$$

Beweis. Für $p \in [1, \infty)$ benutze die Jensen'sche Ungleichung mit $\varphi(x) = |x|^p$. Für $p = \infty$ beachte, dass $|\mathbf{E}[X|\mathcal{F}]| \leq \mathbf{E}[|X||\mathcal{F}] \leq \mathbf{E}[|X|_\infty|\mathcal{F}] = \|X\|_\infty$. \square

Korollar 8.22. Ist $(X_i, i \in I)$ gleichgradig integrierbar und $(\mathcal{F}_j, j \in J)$ eine Familie von Teil- σ -Algebren von \mathcal{A} sowie $X_{i,j} := \mathbf{E}[X_i|\mathcal{F}_j]$, dann ist $(X_{i,j}, (i, j) \in I \times J)$ gleichgradig integrierbar. Insbesondere ist für $X \in \mathcal{L}^1(\mathbf{P})$ die Familie $(\mathbf{E}[X|\mathcal{F}_j], j \in J)$ gleichgradig integrierbar.

Beweis. Nach Satz 6.19 existiert eine wachsende, konvexe Funktion f mit der Eigenschaft $f(x)/x \rightarrow \infty$, $x \rightarrow \infty$ und $L := \sup_{i \in I} \mathbf{E}[f(|X_i|)] < \infty$. Dann ist $x \mapsto f(|x|)$ konvex, also nach der Jensen'schen Ungleichung

$$\mathbf{E}[f(|X_{i,j}|)] = \mathbf{E}[f(|\mathbf{E}[X_i|\mathcal{F}_j]|)] \leq L < \infty.$$

Nach Satz 6.19 ist daher $(X_{i,j}, (i, j) \in I \times J)$ gleichgradig integrierbar. \square

Beispiel 8.23. Seien μ und ν endliche Maße mit $\nu \ll \mu$. Sei $f = d\nu/d\mu$ die Radon-Nikodym-Ableitung, und sei $I = \{\mathcal{F} \subset \mathcal{A} : \mathcal{F} \text{ ist eine } \sigma\text{-Algebra}\}$. Betrachte die auf \mathcal{F} eingeschränkten Maße $\mu|_{\mathcal{F}}$ und $\nu|_{\mathcal{F}}$. Dann ist $\nu|_{\mathcal{F}} \ll \mu|_{\mathcal{F}}$ (klar, denn in \mathcal{F} gibt es ja weniger μ -Nullmengen), also existiert die Radon-Nikodym-Ableitung $f_{\mathcal{F}} := d\nu|_{\mathcal{F}}/d\mu|_{\mathcal{F}}$. Dann ist $(f_{\mathcal{F}} : \mathcal{F} \in I)$ gleichgradig integrierbar (bezüglich μ). (Für endliche σ -Algebren \mathcal{F} wurde dies schon in Beispiel 7.39 gezeigt.) In der Tat: Sei $\mathbf{P} = \mu/\mu(\Omega)$ und $\mathbf{Q} = \nu/\mu(\Omega)$. Dann ist $f_{\mathcal{F}} = d\mathbf{Q}|_{\mathcal{F}}/d\mathbf{P}|_{\mathcal{F}}$. Für jedes $F \in \mathcal{F}$ ist also $\mathbf{E}[f_{\mathcal{F}} \mathbf{1}_F] = \int_F f_{\mathcal{F}} d\mathbf{P} = \mathbf{Q}(F) = \int_F f d\mathbf{P} = \mathbf{E}[f \mathbf{1}_F]$, also $f_{\mathcal{F}} = \mathbf{E}[f | \mathcal{F}]$. Nach dem vorangehenden Korollar ist $(f_{\mathcal{F}} : \mathcal{F} \in I)$ gleichgradig integrierbar bezüglich \mathbf{P} und damit auch bezüglich μ . \diamond

Übung 8.2.1. Man zeige die Aussagen von Bemerkung 8.16. \clubsuit

Übung 8.2.2. Sei $I \subset \mathbb{R}$ ein beliebiges Intervall und $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ eine Zufallsvariable mit $X \in I$ fast sicher. Man zeige für $\mathcal{F} \subset \mathcal{A}$, dass $\mathbf{E}[X | \mathcal{F}] \in I$ fast sicher gilt. Ist dies auch noch richtig, wenn nur $X^- \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ gilt? \clubsuit

Übung 8.2.3 (Bayes'sche Formel). Seien $A \in \mathcal{A}$ und $B \in \mathcal{F} \subset \mathcal{A}$. Man zeige

$$\mathbf{P}[B | A] = \frac{\int_B \mathbf{P}[A | \mathcal{F}] d\mathbf{P}}{\int \mathbf{P}[A | \mathcal{F}] d\mathbf{P}}.$$

Wird \mathcal{F} von paarweise disjunkten Mengen B_1, B_2, \dots erzeugt, so ist dies gerade die Bayes'sche Formel aus Satz 8.7. \clubsuit

Übung 8.2.4. Man zeige durch ein Beispiel, dass $\mathbf{E}[\mathbf{E}[X | \mathcal{F}] | \mathcal{G}] \neq \mathbf{E}[\mathbf{E}[X | \mathcal{G}] | \mathcal{F}]$ gelten kann. \clubsuit

Übung 8.2.5. Man zeige die bedingte Markov'sche Ungleichung: Für monoton wachsendes $f : [0, \infty) \rightarrow [0, \infty)$ und $\varepsilon > 0$ mit $f(\varepsilon) > 0$ ist

$$\mathbf{P}[|X| \geq \varepsilon | \mathcal{F}] \leq \frac{\mathbf{E}[f(|X|) | \mathcal{F}]}{f(\varepsilon)}. \quad \clubsuit$$

Übung 8.2.6. Man zeige die bedingte Cauchy-Schwarz'sche Ungleichung: Für quadratintegrierbare Zufallsvariablen X, Y gilt

$$\mathbf{E}[XY | \mathcal{F}]^2 \leq \mathbf{E}[X^2 | \mathcal{F}] \mathbf{E}[Y^2 | \mathcal{F}]. \quad \clubsuit$$

Übung 8.2.7. Seien X_1, \dots, X_n integrierbar, unabhängig und identisch verteilt. Sei $S_n = X_1 + \dots + X_n$. Zeige:

$$\mathbf{E}[X_i | S_n] = \frac{1}{n} S_n \quad \text{für jedes } i = 1, \dots, n. \quad \clubsuit$$

Übung 8.2.8. Seien X_1 und X_2 unabhängig und exponentialverteilt mit Parameter $\theta > 0$. Man bestimme $\mathbf{E}[X_1 \wedge X_2 | X_1]$.



Übung 8.2.9. Seien X und Y reelle Zufallsvariablen mit gemeinsamer Dichte f , und sei $h : \mathbb{R} \rightarrow \mathbb{R}$ messbar mit $\mathbf{E}[|h(X)|] < \infty$. Es bezeichne λ das Lebesgue-Maß auf \mathbb{R} .

(i) Zeige, dass fast sicher gilt:

$$\mathbf{E}[h(X)|Y] = \frac{\int h(x)f(x, Y) \lambda(dx)}{\int f(x, Y) \lambda(dx)}.$$

(ii) Seien speziell X und Y unabhängig und \exp_θ -verteilt für ein $\theta > 0$. Bestimme $\mathbf{E}[X|X+Y]$ und $\mathbf{P}[X \leq x|X+Y]$ für $x \geq 0$.



8.3 Reguläre Version der bedingten Verteilung

Ist X eine Zufallsvariable mit Werten in einem Messraum (E, \mathcal{E}) , so können wir mit unserem Apparat bisher für festes $A \in \mathcal{A}$ die bedingte Wahrscheinlichkeit $\mathbf{P}[A|X]$ angeben. Können wir die Situation aber auch so einrichten, dass wir für jedes $x \in E$ ein W-Maß $\mathbf{P}[\cdot|X=x]$ angeben können, sodass für jedes $A \in \mathcal{A}$ gilt $\mathbf{P}[A|X] = \mathbf{P}[A|X=x]$ auf $\{X=x\}$?

Wir sind beispielsweise an einem zweistufigen Zufallsexperiment interessiert: Im ersten Schritt wird eine Münze in zufälliger Weise so gefälscht, dass sie die Erfolgswahrscheinlichkeit X hat. Danach werden unabhängige Würfe Y_1, \dots, Y_n mit dieser Münze durchgeführt. Die „bedingte“ Verteilung von (Y_1, \dots, Y_n) gegeben $\{X=x\}$ sollte also $(\text{Ber}_x)^{\otimes n}$ sein.

Sei X wie oben und Z eine $\sigma(X)$ -messbare, reelle Zufallsvariable. Nach dem Faktorisierungslemma (Korollar 1.97 mit $f = X$ und $g = Z$) existiert eine Abbildung $\varphi : E \rightarrow \mathbb{R}$ mit

$$\varphi \text{ ist } \mathcal{E} - \mathcal{B}(\mathbb{R})\text{-messbar} \quad \text{und} \quad \varphi(X) = Z. \quad (8.9)$$

Ist X surjektiv, so ist φ eindeutig festgelegt. Wir schreiben dann $Z \circ X^{-1} := \varphi$ (auch wenn die Umkehrabbildung X^{-1} selber nicht existiert).

Definition 8.24. Sei $Y \in \mathcal{L}^1(\mathbf{P})$ und $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$. Dann definieren wir die bedingte Erwartung von Y gegeben $X = x$ durch $\mathbf{E}[Y|X=x] := \varphi(x)$, wobei φ die Funktion aus (8.9) ist mit $Z = \mathbf{E}[Y|X]$.

Wir setzen analog $\mathbf{P}[A|X=x] = \mathbf{E}[\mathbb{1}_A | X=x]$ für $A \in \mathcal{A}$.

Für eine Menge $B \in \mathcal{A}$ mit $\mathbf{P}[B] > 0$ ist die bedingte Wahrscheinlichkeit $\mathbf{P}[\cdot|B]$ ein W-Maß. Gilt das Gleiche für $\mathbf{P}[\cdot|X=x]$? Der Fall liegt hier komplizierter, da

wir für jedes $A \in \mathcal{A}$ den Ausdruck $\mathbf{P}[A|X = x]$ für x nur bis auf eine Ausnahmemenge, die allerdings von A abhängt, definiert haben. Wenn wir die σ -Algebra \mathcal{A} nun durch abzählbar viele A genügend gut approximieren können, besteht Hoffnung, dass die Ausnahmemengen sich zu einer Nullmenge vereinigen. Wir fassen zunächst die Begriffe genauer und zeigen dann das angedeutete Ergebnis.

Definition 8.25 (Übergangskern, Markovkern). Sind $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$ Messräume, so heißt $\kappa : \Omega_1 \times \mathcal{A}_2 \rightarrow [0, \infty]$ ein (σ -)endlicher **Übergangskern** (von Ω_1 nach Ω_2), falls

- (i) $\omega_1 \mapsto \kappa(\omega_1, A_2)$ ist \mathcal{A}_1 -messbar für jedes $A_2 \in \mathcal{A}_2$.
- (ii) $A_2 \mapsto \kappa(\omega_1, A_2)$ ist ein (σ -)endliches Maß auf $(\Omega_2, \mathcal{A}_2)$ für jedes $\omega_1 \in \Omega_1$.

Ist das Maß in (ii) ein W-Maß für jedes $\omega_1 \in \Omega_1$, so heißt κ **stochastischer Kern** oder **Markovkern**. Wird in (ii) zusätzlich $\kappa(\omega_1, \Omega_2) \leq 1$ für jedes $\omega_1 \in \Omega_1$ gefordert, so heißt κ **sub-Markov'sch** oder **substochastisch**.

Bemerkung 8.26. Es reicht, in Definition 8.25 die Eigenschaft (i) nur für Mengen A_2 aus einem schnittstabilen Erzeuger \mathcal{E} von \mathcal{A}_2 , der Ω_2 oder eine Folge $E_n \uparrow \Omega_2$ enthält, zu fordern. Es ist dann nämlich

$$\mathcal{D} := \{A_2 \in \mathcal{A}_2 : \omega_1 \mapsto \kappa(\omega_1, A_2) \text{ ist } \mathcal{A}_1\text{-messbar}\}$$

ein Dynkin-System (Übung!). Wegen $\mathcal{E} \subset \mathcal{D}$ ist (Satz 1.19) $\mathcal{D} = \sigma(\mathcal{E}) = \mathcal{A}_2$. \diamond

Beispiel 8.27. (i) Sind $(\Omega_1, \mathcal{A}_1)$ und $(\Omega_2, \mathcal{A}_2)$ diskrete Messräume, so liefert jede Matrix $(K_{ij})_{\substack{i \in \Omega_1 \\ j \in \Omega_2}}$ mit nichtnegativen Einträgen und endlichen Zeilensummen

$$K_i := \sum_{j \in \Omega_2} K_{ij} < \infty \quad \text{für } i \in \Omega_1,$$

einen endlichen Übergangskern von Ω_1 nach Ω_2 vermöge $\kappa(i, A) = \sum_{j \in A} K_{ij}$. Der Kern ist stochastisch, falls $K_i = 1$ für jedes $i \in \Omega_1$ und substochastisch, falls $K_i \leq 1$ für jedes $i \in \Omega_1$.

(ii) Ist μ_2 ein endliches Maß auf Ω_2 , dann ist $\kappa(\omega_1, \cdot) \equiv \mu_2$ ein endlicher Übergangskern.

(iii) $\kappa(x, \cdot) = \text{Poi}_x$ ist ein stochastischer Kern von $[0, \infty)$ nach \mathbb{N}_0 (beachte: für jedes $A \subset \mathbb{N}_0$ ist $x \mapsto \text{Poi}_x(A)$ stetig, also insbesondere messbar).

(iv) Sei μ eine Verteilung auf \mathbb{R}^n und X eine Zufallsvariable mit $\mathbf{P}_X = \mu$. Dann definiert $\kappa(x, \cdot) = \mathbf{P}[X + x \in \cdot] = \delta_x * \mu$ einen stochastischen Kern von \mathbb{R}^n nach \mathbb{R}^n . In der Tat: Die Mengen $(-\infty, y]$, $y \in \mathbb{R}^n$, bilden einen schnittstabilen Erzeuger von $\mathcal{B}(\mathbb{R}^n)$ und $x \mapsto \kappa(x, (-\infty, y]) = \mu((-\infty, y - x])$ ist linksstetig, also messbar. Nach Bemerkung 8.26 ist daher $x \mapsto \kappa(x, A)$ messbar für jedes $A \in \mathcal{B}(\mathbb{R}^n)$. \diamond

Definition 8.28. Sei Y eine Zufallsvariable mit Werten in einem Messraum (E, \mathcal{E}) und $\mathcal{F} \subset \mathcal{A}$ eine Unter- σ -Algebra. Ein stochastischer Kern $\kappa_{Y, \mathcal{F}}$ von (Ω, \mathcal{F}) nach (E, \mathcal{E}) heißt **reguläre Version der bedingten Verteilung** von Y gegeben \mathcal{F} , falls $\kappa_{Y, \mathcal{F}}(\omega, B) = \mathbf{P}[\{Y \in B\} | \mathcal{F}](\omega)$ für \mathbf{P} -f.a. $\omega \in \Omega$ und jedes $B \in \mathcal{E}$, also falls

$$\int \mathbb{1}_B(Y) \mathbb{1}_A d\mathbf{P} = \int \kappa_{Y, \mathcal{F}}(\cdot, B) \mathbb{1}_A d\mathbf{P} \quad \text{für alle } A \in \mathcal{F}, B \in \mathcal{E}. \quad (8.10)$$

Sei speziell $\mathcal{F} = \sigma(X)$ für eine Zufallsvariable X (in einem beliebigen Messraum (E', \mathcal{E}')). Dann heißt der stochastische Kern

$$(x, A) \mapsto \kappa_{Y, X}(x, A) = \mathbf{P}[\{Y \in A\} | X = x] = \kappa_{Y, \sigma(X)}(X^{-1}(x), A)$$

(die Funktion aus dem Faktorisierungslemma mit beliebiger Festsetzung für $x \notin X(\Omega)$) eine reguläre Version der bedingten Verteilung von Y gegeben X .

Satz 8.29 (Reguläre bedingte Verteilungen in \mathbb{R}). Ist $Y : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ reellwertig, dann existiert eine reguläre Version $\kappa_{Y, \mathcal{F}}$ der bedingten Verteilung $\mathbf{P}[\{Y \in \cdot\} | \mathcal{F}]$.

Beweis. Die Strategie besteht darin, eine messbare Version der Verteilungsfunktion der bedingten Verteilung von Y zu konstruieren, indem diese zunächst für rationale Werte festgelegt wird (bis auf eine Nullmenge) und dann auf die reellen Zahlen fortgesetzt wird.

Für $r \in \mathbb{Q}$ sei $F(r, \cdot)$ eine Version der bedingten Wahrscheinlichkeit $\mathbf{P}[Y \in (-\infty, r] | \mathcal{F}]$. Für $r \leq s$ ist offenbar $\mathbb{1}_{\{Y \in (-\infty, r]\}} \leq \mathbb{1}_{\{Y \in (-\infty, s]\}}$, also gibt es nach Satz 8.14(ii) (Monotonie der bedingten Erwartung) eine Nullmenge $A_{r,s} \in \mathcal{F}$ mit

$$F(r, \omega) \leq F(s, \omega) \quad \text{für jedes } \omega \in \Omega \setminus A_{r,s}. \quad (8.11)$$

Nach Satz 8.14(viii) (majorisierte Konvergenz) gibt es Nullmengen $(B_r)_{r \in \mathbb{Q}} \in \mathcal{F}$ und $C \in \mathcal{F}$, sodass

$$\lim_{n \rightarrow \infty} F\left(r + \frac{1}{n}, \omega\right) = F(r, \omega) \quad \text{für jedes } \omega \in \Omega \setminus B_r \quad (8.12)$$

sowie

$$\inf_{n \in \mathbb{N}} F(-n, \omega) = 0 \quad \text{und} \quad \sup_{n \in \mathbb{N}} F(n, \omega) = 1 \quad \text{für jedes } \omega \in \Omega \setminus C. \quad (8.13)$$

Setze $N := (\bigcup_{r,s \in \mathbb{Q}} A_{r,s}) \cup (\bigcup_{r \in \mathbb{Q}} B_r) \cup C$. Für $\omega \in \Omega \setminus N$ definieren wir

$$\tilde{F}(z, \omega) := \inf \{F(r, \omega) : r \in \mathbb{Q}, r > z\} \quad \text{für alle } z \in \mathbb{R}.$$

Per Konstruktion ist $\tilde{F}(\cdot, \omega)$ monoton wachsend und rechtsstetig. Aus (8.11) und (8.12) folgt

$$\tilde{F}(z, \omega) = F(z, \omega) \quad \text{für jedes } z \in \mathbb{Q} \text{ und } \omega \in \Omega \setminus N. \quad (8.14)$$

Wegen (8.13) ist $\tilde{F}(\cdot, \omega)$ also eine Verteilungsfunktion für jedes $\omega \in \Omega \setminus N$. Für $\omega \in N$ setze $\tilde{F}(\cdot, \omega) = F_0$, wobei F_0 eine beliebige fest gewählte Verteilungsfunktion ist.

Für jedes $\omega \in \Omega$ definieren wir $\kappa(\omega, \cdot)$ als das durch die Verteilungsfunktion $\tilde{F}(\cdot, \omega)$ definierte W-Maß auf (Ω, \mathcal{A}) . Für $r \in \mathbb{Q}$ und $B = (-\infty, r]$ ist dann (nach (8.14))

$$\omega \mapsto \kappa(\omega, B) = F(r, \omega) \mathbb{1}_{N^c}(\omega) + F_0(r) \mathbb{1}_N(\omega) \quad (8.15)$$

\mathcal{F} -messbar. Nun ist $\{(-\infty, r], r \in \mathbb{Q}\}$ ein schnittstabiler Erzeuger von $\mathcal{B}(\mathbb{R})$. Nach Bemerkung 8.26 gilt die Messbarkeit also für jedes $B \in \mathcal{B}(\mathbb{R})$, und damit ist κ als stochastischer Kern erkannt.

Wir müssen noch zeigen, dass κ eine Version der bedingten Verteilung ist. Nach (8.15) und der Definition von F ist für $A \in \mathcal{F}$, $r \in \mathbb{Q}$ und $B = (-\infty, r]$

$$\int_A \kappa(\omega, B) \mathbf{P}[d\omega] = \int_A \mathbf{P}[Y \in B | \mathcal{F}] d\mathbf{P} = \mathbf{P}[A \cap \{Y \in B\}].$$

Als Funktion von B sind beide Seiten endliche Maße auf $\mathcal{B}(\mathbb{R})$, die auf dem schnittstabilen Erzeuger $\{(-\infty, r], r \in \mathbb{Q}\}$ übereinstimmen. Nach dem Eindeutigkeitssatz (Lemma 1.42) gilt daher für jedes $B \in \mathcal{B}(\mathbb{R})$ Gleichheit und damit \mathbf{P} -fast sicher $\kappa(\cdot, B) = \mathbf{P}[Y \in B | \mathcal{F}]$, also $\kappa = \kappa_{Y, \mathcal{F}}$. \square

Beispiel 8.30. Seien Z_1, Z_2 unabhängig und Poisson-verteilt mit den Parametern $\lambda_1, \lambda_2 \geq 0$. Dann kann man zeigen (Übung!), dass (mit $Y = Z_1$ und $X = Z_1 + Z_2$)

$$\mathbf{P}[Z_1 = k | Z_1 + Z_2 = n] = b_{n,p}(k) \quad \text{für } k = 0, \dots, n,$$

wobei $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$ ist. \diamond

Dieses Beispiel ließ sich aber im Grunde genommen auch noch mit elementaren Mitteln bearbeiten. Die volle Stärke des Ergebnisses nutzen wir in den folgenden Beispielen aus.

Beispiel 8.31. Seien X und Y reelle Zufallsvariablen mit gemeinsamer Dichtefunktion f (bezüglich des Lebesgue-Maßes λ^2 auf \mathbb{R}^2). Für $x \in \mathbb{R}$ setzen wir

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \lambda(dy).$$

Offenbar ist $f_X(x) > 0$ für \mathbf{P}_X -f.a. $x \in \mathbb{R}$ und f_X^{-1} ist die Dichte des absolutstetigen Anteils des Lebesgue-Maßes λ bezüglich \mathbf{P}_X . Die reguläre Version der bedingten Verteilung von Y gegeben X hat die Dichte

$$\frac{\mathbf{P}[Y \in dy | X = x]}{dy} = f_{Y|X}(x, y) := \frac{f(x, y)}{f_X(x)} \text{ für } \mathbf{P}_X[dx]\text{-f.a. } x \in \mathbb{R}. \quad (8.16)$$

In der Tat ist nach dem Satz von Fubini (siehe Satz 14.16) die Abbildung $x \mapsto \int_B f_{Y|X}(x, y) \lambda(dy)$ messbar für jedes $B \in \mathcal{B}(\mathbb{R})$, und für $A, B \in \mathcal{B}(\mathbb{R})$ gilt

$$\begin{aligned} & \int_A \mathbf{P}[X \in dx] \int_B f_{Y|X}(x, y) \lambda(dy) \\ &= \int_A \mathbf{P}[X \in dx] f_X(x)^{-1} \int_B f(x, y) \lambda(dy) \\ &= \int_A \lambda(dx) \int_B f(x, y) \lambda(dy) \\ &= \int_{A \times B} f d\lambda^2 = \mathbf{P}[X \in A, Y \in B]. \end{aligned} \quad \diamond$$

Beispiel 8.32. Seien $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ und Z_1, Z_2 unabhängig und $\mathcal{N}_{\mu_i, \sigma_i^2}$ verteilt ($i = 1, 2$). Dann existiert eine reguläre Version der bedingten Verteilung

$$\mathbf{P}[Z_1 \in \cdot | Z_1 + Z_2 = x] \quad \text{für } x \in \mathbb{R}.$$

Setzen wir $X = Z_1 + Z_2$ und $Y = Z_1$, so ist $(X, Y) \sim \mathcal{N}_{\mu, \Sigma}$ bivariat normalverteilt mit Kovarianzmatrix $\Sigma := \begin{pmatrix} \sigma_1^2 + \sigma_2^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 \end{pmatrix}$ und mit $\mu := \begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 \end{pmatrix}$. Wegen

$$\Sigma^{-1} = (\sigma_1^2 \sigma_2^2)^{-1} \begin{pmatrix} \sigma_1^2 & -\sigma_1^2 \\ -\sigma_1^2 & \sigma_1^2 + \sigma_2^2 \end{pmatrix} = (\sigma_1^2 \sigma_2^2)^{-1} B^T B,$$

wo $B = \begin{pmatrix} \sigma_1 & -\sigma_1 \\ 0 & \sigma_2 \end{pmatrix}$ ist, hat (X, Y) die Dichte (siehe Beispiel 1.105(ix))

$$\begin{aligned} f(x, y) &= \det(2\pi \Sigma)^{-1/2} \exp \left(-\frac{1}{2\sigma_1^2 \sigma_2^2} \left\| B \begin{pmatrix} x - (\mu_1 + \mu_2) \\ y - \mu_1 \end{pmatrix} \right\|^2 \right) \\ &= (4\pi^2 \sigma_1^2 \sigma_2^2)^{-1/2} \exp \left(-\frac{\sigma_1^2(y - (x - \mu_2))^2 + \sigma_2^2(y - \mu_1)^2}{2\sigma_1^2 \sigma_2^2} \right) \\ &= C_x \exp(- (y - \mu_x)^2 / 2\sigma_x^2), \end{aligned}$$

wobei C_x eine Normalisierungskonstante ist und

$$\mu_x = \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} (x - \mu_1 - \mu_2) \quad \text{und} \quad \sigma_x^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Nach (8.16) hat $\mathbf{P}[Z_1 \in \cdot | Z_1 + Z_2 = x]$ die Dichte

$$y \mapsto f_{Y|X}(x, y) = \frac{C_x}{f_X(x)} \exp \left(-\frac{(y - \mu_x)^2}{2\sigma_x^2} \right),$$

also ist

$$\mathbf{P}[Z_1 \in \cdot | Z_1 + Z_2 = x] = \mathcal{N}_{\mu_x, \sigma_x^2} \text{ für fast alle } x \in \mathbb{R}. \quad \diamond$$

Beispiel 8.33. Seien X und Y unabhängige, reelle Zufallsvariablen. Dann ist für \mathbf{P}_X -fast alle $x \in \mathbb{R}$

$$\mathbf{P}[X + Y \in \cdot | X = x] = \delta_x * \mathbf{P}_Y. \quad \diamond$$

Die Situation ist noch nicht vollends zufriedenstellend, da wir die sehr starke Annahme gemacht haben, dass Y reellwertig ist. Ursprünglich waren wir aber auch an einer Situation interessiert, wo Y Werte in \mathbb{R}^n annimmt, oder sogar in allgemeineren Räumen. Wir dehnen nun das Ergebnis auf eine größere Klasse von Wertebereichen von Y aus.

Definition 8.34. Zwei Messräume (E, \mathcal{E}) und (E', \mathcal{E}') heißen **isomorph**, falls es eine bijektive Abbildung $\varphi : E \rightarrow E'$ gibt, sodass φ messbar ist bezüglich $\mathcal{E}-\mathcal{E}'$ und die Umkehrabbildung φ^{-1} messbar ist bezüglich $\mathcal{E}'-\mathcal{E}$. Wir nennen dann φ einen **Messraum-Isomorphismus**. Sind zudem μ und μ' Maße auf (E, \mathcal{E}) und (E', \mathcal{E}') und gilt $\mu' = \mu \circ \varphi^{-1}$, so ist φ ein **Maßraum-Isomorphismus**, und die Maßräume (E, \mathcal{E}, μ) und (E', \mathcal{E}', μ') heißen **isomorph**.

Definition 8.35. Ein Messraum (E, \mathcal{E}) heißt **Borel'scher Raum**, falls es eine Borel'sche Menge $B \in \mathcal{B}(\mathbb{R})$ gibt, sodass (E, \mathcal{E}) und $(B, \mathcal{B}(B))$ isomorph sind.

Ein separabler topologischer Raum, dessen Topologie durch eine vollständige Metrik erzeugt wird, heißt **polnischer Raum**. Wichtige Beispiele hierfür sind \mathbb{R}^d , \mathbb{Z}^d , $\mathbb{R}^\mathbb{N}$, und $(C([0, 1]), \|\cdot\|_\infty)$. Abgeschlossene Teilmengen von polnischen Räumen sind ebenfalls polnisch. Wir kommen auf polnische Räume im Zusammenhang mit der Konvergenz von Maßen in Kapitel 13 zurück. Ohne Beweis bringen wir das folgende topologische Ergebnis (siehe etwa [42, Theorem 13.1.1]).

Satz 8.36. Ist E ein polnischer Raum mit Borel'scher σ -Algebra \mathcal{E} , dann ist (E, \mathcal{E}) ein Borel'scher Raum.

Satz 8.37 (Reguläre bedingte Verteilungen). Sei $\mathcal{F} \subset \mathcal{A}$ eine Unter- σ -Algebra. Sei Y eine Zufallsvariable mit Werten in einem Borel'schen Raum (E, \mathcal{E}) (also zum Beispiel E polnisch, $E = \mathbb{R}^d$, $E = \mathbb{R}^\infty$, $E = C([0, 1])$ usw.). Dann existiert eine reguläre Version $\kappa_{Y, \mathcal{F}}$ der bedingten Verteilung $\mathbf{P}[\{Y \in \cdot\} | \mathcal{F}]$.

Beweis. Sei $B \in \mathcal{B}(\mathbb{R})$ und $\varphi : E \rightarrow B$ ein Messraum-Isomorphismus. Mit Satz 8.29 erhalten wir eine reguläre bedingte Verteilung $\kappa_{Y', \mathcal{F}}$ der reellen Zufallsvariablen $Y' = \varphi \circ Y$. Wir setzen nun $\kappa_{Y, \mathcal{F}}(\omega, A) = \kappa_{Y', \mathcal{F}}(\omega, \varphi(A))$ für $A \in \mathcal{E}$. \square

Abschließend greifen wir das eingangs betrachtete Beispiel wieder auf und können nun die dort in Anführungszeichen gemachte Aussage formal hinschreiben. Sei also

X uniform auf $[0, 1]$ verteilt, und gegeben den Wert $X = x$ seien (Y_1, \dots, Y_n) unabhängig Ber_x -verteilt. Wir setzen $Y = (Y_1, \dots, Y_n)$. Nach Satz 8.37 (mit $E = \{0, 1\}^n \subset \mathbb{R}^n$) existiert die reguläre bedingte Verteilung

$$\kappa_{Y,X}(x, \cdot) = \mathbf{P}[Y \in \cdot | X = x] \quad \text{für } x \in [0, 1].$$

In der Tat ist für fast alle $x \in [0, 1]$

$$\mathbf{P}[Y \in \cdot | X = x] = (\text{Ber}_x)^{\otimes n}.$$

Satz 8.38. Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbf{P})$ mit Werten in einem Borel'schen Raum (E, \mathcal{E}) . Sei $\mathcal{F} \subset \mathcal{A}$ eine σ -Algebra und $\kappa_{X,\mathcal{F}}$ eine reguläre Version der bedingten Verteilung von X gegeben \mathcal{F} . Sei ferner $f : E \rightarrow \mathbb{R}$ messbar und $\mathbf{E}[|f(X)|] < \infty$. Dann ist

$$\mathbf{E}[f(X)|\mathcal{F}](\omega) = \int f(x) \kappa_{X,\mathcal{F}}(\omega, dx) \quad \text{für } \mathbf{P}\text{-fast alle } \omega. \quad (8.17)$$

Beweis. Wir rechnen nach, dass die rechte Seite in (8.17) die Eigenschaften des bedingten Erwartungswertes hat.

Es reicht, den Fall $f \geq 0$ zu betrachten. Indem wir f durch Elementarfunktionen approximieren, sehen wir, dass die rechte Seite in (8.17) messbar bezüglich \mathcal{F} ist (siehe Lemma 14.20 für ein formales Argument). Nach Satz 1.96 existieren dann Mengen $A_1, A_2, \dots \in \mathcal{E}$ und Zahlen $\alpha_1, \alpha_2, \dots \geq 0$ mit

$$g_n := \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i} \xrightarrow{n \rightarrow \infty} f.$$

Für jedes $n \in \mathbb{N}$ und $B \in \mathcal{F}$ ist nun

$$\begin{aligned} \mathbf{E}[g_n(X) \mathbb{1}_B] &= \sum_{i=1}^n \alpha_i \mathbf{P}[\{X \in A_i\} \cap B] \\ &= \sum_{i=1}^n \alpha_i \int_B \mathbf{P}[\{X \in A_i\} | \mathcal{F}] \mathbf{P}[d\omega] \\ &= \sum_{i=1}^n \alpha_i \int_B \kappa_{X,\mathcal{F}}(\omega, A_i) \mathbf{P}[d\omega] \\ &= \int_B \sum_{i=1}^n \alpha_i \kappa_{X,\mathcal{F}}(\omega, A_i) \mathbf{P}[d\omega] \\ &= \int_B \left(\int g_n(x) \kappa_{X,\mathcal{F}}(\omega, dx) \right) \mathbf{P}[d\omega]. \end{aligned}$$

Nach dem Satz von der monotonen Konvergenz konvergiert für fast jedes ω das innere Integral gegen $\int f(x) \kappa_{X,\mathcal{F}}(\omega, dx)$. Erneute Anwendung des Satzes von der monotonen Konvergenz liefert

$$\mathbf{E}[f(X) \mathbb{1}_B] = \lim_{n \rightarrow \infty} \mathbf{E}[g_n(X) \mathbb{1}_B] = \int_B \int f(x) \kappa_{X,\mathcal{F}}(\omega, dx) \mathbf{P}[d\omega]. \quad \square$$

Übung 8.3.1. Sei (E, \mathcal{E}) ein Borel'scher Raum und μ ein atomloses Maß (das heißt, $\mu(\{x\}) = 0$ für jedes $x \in E$). Man zeige: Für jedes $A \in \mathcal{E}$ und jedes $n \in \mathbb{N}$ existieren paarweise disjunkte Mengen $A_1, \dots, A_n \in \mathcal{E}$ mit $\biguplus_{k=1}^n A_k = A$ und $\mu(A_k) = \mu(A)/n$ für jedes $k = 1, \dots, n$. \clubsuit

Übung 8.3.2. Seien $p, q \in (1, \infty)$ mit $\frac{1}{p} + \frac{1}{q} = 1$, und seien $X \in \mathcal{L}^p(\mathbf{P})$ und $Y \in \mathcal{L}^q(\mu)$. Sei $\mathcal{F} \subset \mathcal{A}$ eine σ -Algebra. Man zeige mit Hilfe des vorangehenden Satzes die bedingte Version der Hölder'schen Ungleichung:

$$\mathbf{E}[|XY| \mid \mathcal{F}] \leq \mathbf{E}[|X|^p \mid \mathcal{F}]^{1/p} \mathbf{E}[|Y|^q \mid \mathcal{F}]^{1/q} \text{ fast sicher.} \quad \clubsuit$$

Übung 8.3.3. Sei (X, Y) uniform verteilt auf $B := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ beziehungsweise auf $[-1, 1]^2$.

- (i) Man bestimme jeweils die bedingte Verteilung von Y gegeben $X = x$.
- (ii) Sei $R := \sqrt{X^2 + Y^2}$ und $\Theta = \arctan(Y/X)$. Man bestimme jeweils die bedingte Verteilung von Θ gegeben $R = r$. \clubsuit

Übung 8.3.4. Sei $A \subset \mathbb{R}^n$ eine Borel-messbare Menge mit endlichem Lebesgue-Maß $\lambda(A) \in (0, \infty)$, und sei $B \subset A$ messbar mit $\lambda(B) > 0$. Zeige: Ist X uniform verteilt (siehe Beispiel 1.75) auf A , so ist die bedingte Verteilung von X gegeben $\{X \in B\}$ die uniforme Verteilung auf B . \clubsuit

Übung 8.3.5 (Borel'sches Paradoxon). Wir wollen die Erde als Kugel ansehen und betrachten einen zufälligen uniform auf der Erdoberfläche verteilten Punkt X . Wir wollen die Koordinaten von X durch die geografische Länge Θ und Breite Φ angeben. Allerdings soll, entgegen der üblichen Konvention, Θ die Werte in $[0, \pi]$ annehmen und Φ in $[-\pi, \pi]$. Damit wird für festes Θ ein kompletter Großkreis beschrieben, wenn Φ seinen Wertebereich durchläuft. Ist nun Φ gegeben Θ uniform verteilt auf $[-\pi, \pi]$? Man sollte annehmen, dass jeder Punkt auf dem Großkreis gleich wahrscheinlich ist. Dies ist jedoch nicht der Fall! Der etwas „aufgedickte“ Großkreis, mit Längen zwischen Θ und $\Theta + \varepsilon$ (für kleines ε) ist am Äquator dicker als an den Polen. Lassen wir $\varepsilon \rightarrow 0$ gehen, so sollten wir, zumindest intuitiv, die bedingten Wahrscheinlichkeiten erhalten.

- (i) Man zeige: $\mathbf{P}[\{\Phi \in \cdot\} \mid \Theta = \theta]$ hat für fast alle θ die Dichte $\frac{1}{4}|\cos(\phi)|$ für $\phi \in [-\pi, \pi]$.
- (ii) Man zeige: $\mathbf{P}[\{\Theta \in \cdot\} \mid \Phi = \phi] = \mathcal{U}_{[0, \pi]}$ für fast alle ϕ .

Hinweis: Man zeige, dass Θ und Φ unabhängig sind und bestimme die Verteilungen von Θ und Φ . \clubsuit

Übung 8.3.6 (Verwerfungsmethode zur Erzeugung von Zufallsvariablen). Sei E höchstens abzählbar und P und Q Wahrscheinlichkeitsmaße auf E . Es gebe ein $c > 0$ mit $f(e) := \frac{Q(\{e\})}{P(\{e\})} \leq c$ für jedes $e \in E$ mit $P(\{e\}) > 0$. Seien X_1, X_2, \dots unabhängige Zufallsvariablen mit Verteilung P und U_1, U_2, \dots davon unabhängige u.i.v. Zufallsvariablen, die uniform auf $[0, 1]$ verteilt sind. Wähle N als die (zufällige) kleinste natürliche Zahl n , sodass $U_n \leq f(X_n)/c$, und setze $Y := X_N$.

Man zeige: Y hat die Verteilung Q .

Anmerkung: Dieses Verfahren zur Erzeugung einer Zufallsvariable mit einer gewünschten Verteilung Q wird auch Verwerfungsmethode (*rejection sampling*) genannt, denn man kann es so interpretieren: Die Zufallsvariable X_1 ist ein Vorschlag für den möglichen Wert von Y . Dieser Vorschlag wird mit Wahrscheinlichkeit $f(X_1)/c$ angenommen, ansonsten wird X_2 betrachtet und so weiter. ♣

Übung 8.3.7. Sei E ein polnischer Raum und $P, Q \in \mathcal{M}_1(\mathbb{R})$ sowie $c > 0$ mit $f := \frac{dQ}{dP} \leq c$ P -fast sicher. Zeige die analoge Aussage zu Übung 8.3.6. ♣

Übung 8.3.8. Man zeige, dass $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ isomorph sind und folgere, dass jede Borel'sche Menge $B \in \mathcal{B}(\mathbb{R}^n)$ (zusammen mit der Spur- σ -Algebra) ein Borel'scher Raum ist. ♣

9 Martingale

Einer der wichtigsten Begriffe der modernen Wahrscheinlichkeitstheorie ist das Martingal, das die Idee eines fairen Spiels formalisiert. In diesem Kapitel wird der Begriffsapparat für die Beschreibung allgemeiner stochastischer Prozesse aufgebaut. Danach werden Martingale und das diskrete stochastische Integral eingeführt und auf ein Modell der Finanzmathematik angewandt.

9.1 Prozesse, Filtrationen, Stoppzeiten

Wir führen die grundlegenden technischen Begriffe für die Behandlung stochastischer Prozesse, darunter Martingale, ein. Um die Begriffe später in einem anderen Kontext weiter verwenden zu können, streben wir eine größere Allgemeinheit an als für die Behandlung von Martingalen notwendig wäre.

Im Folgenden sei stets (E, τ) ein polnischer Raum mit Borel'scher σ -Algebra \mathcal{E} . Weiter sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein W-Raum und $I \subset \mathbb{R}$ beliebig. Meistens interessieren uns die Fälle $I = \mathbb{N}_0$, $I = \mathbb{Z}$, $I = [0, \infty)$ und I ein Intervall.

Definition 9.1 (Stochastischer Prozess). Sei $I \subset \mathbb{R}$. Eine Familie von Zufallsvariablen $X = (X_t, t \in I)$ (auf $(\Omega, \mathcal{F}, \mathbf{P})$) mit Werten in (E, \mathcal{E}) heißt **stochastischer Prozess** mit Zeitbereich I und Zustandsraum E .

Bemerkung 9.2. Etwas allgemeiner werden manchmal auch beliebig indizierte Familien von Zufallsvariablen stochastischer Prozess genannt. Beispielsweise ist dies beim Poisson'schen Punktprozess aus Kapitel 24 der Fall. ◇

Bemerkung 9.3. Oftmals werden wir (gewissen Traditionen folgend) einen stochastischen Prozess auch als $X = (X_t)_{t \in I}$ schreiben, wenn wir weniger den Aspekt betonen wollen, dass X eine Familie von Zufallsvariablen ist, sondern den zeitlichen Verlauf der Beobachtungen stärker gewichten. Formal sollen beide Objekte identisch sein. ◇

Beispiel 9.4. Sei $I = \mathbb{N}_0$ und $(Y_n, n \in \mathbb{N})$ eine Familie von u.i.v. Zufallsvariablen auf einem W-Raum $(\Omega, \mathcal{F}, \mathbf{P})$, mit $\mathbf{P}[Y_n = 1] = 1 - \mathbf{P}[Y_n = -1] = \frac{1}{2}$. Setze

$E = \mathbb{Z}$ (mit der diskreten Topologie) und

$$X_t = \sum_{n=1}^t Y_n \quad \text{für jedes } t \in \mathbb{N}_0.$$

$(X_t, t \in \mathbb{N}_0)$ heißt **symmetrische einfache Irrfahrt** auf \mathbb{Z} . \diamond

Beispiel 9.5. Der Poissonprozess $X = (X_t)_{t \geq 0}$ mit Intensität $\alpha > 0$ (siehe Kapitel 5.5) ist ein stochastischer Prozess mit Wertebereich \mathbb{N}_0 . \diamond

Wir führen weitere Begriffe ein:

Definition 9.6. Ist X eine Zufallsvariable (oder ein stochastischer Prozess), so schreiben wir auch $\mathcal{L}[X] = \mathbf{P}_X$ für die Verteilung von X . Ist $\mathcal{G} \subset \mathcal{F}$ eine σ -Algebra, so schreiben wir $\mathcal{L}[X | \mathcal{G}]$ für eine reguläre Version der bedingten Verteilung von X gegeben \mathcal{G} .

Definition 9.7. Ein stochastischer Prozess $X = (X_t)_{t \in I}$ mit Werten in E heißt

- (i) **reellwertig**, falls $E = \mathbb{R}$,
- (ii) Prozess mit **unabhängigen Zuwächsen**, falls X reellwertig ist und für jedes $n \in \mathbb{N}$ und alle $t_0, \dots, t_n \in I$ mit $t_0 < t_1 < \dots < t_n$ gilt

$$(X_{t_i} - X_{t_{i-1}})_{i=1, \dots, n} \text{ ist unabhängig,}$$

- (iii) **Gauß'scher Prozess**, falls X reellwertig ist und für $n \in \mathbb{N}$ und $t_1, \dots, t_n \in I$ gilt

$$(X_{t_1}, \dots, X_{t_n}) \text{ ist } n\text{-dimensional normalverteilt,}$$

- (iv) **integrierbar** (beziehungsweise **quadratintegrierbar**), falls X reellwertig ist und $\mathbf{E}[|X_t|] < \infty$ (beziehungsweise $\mathbf{E}[(X_t)^2] < \infty$) für jedes $t \in I$ gilt.

Sei nun zusätzlich $I \subset \mathbb{R}$ abgeschlossen unter Addition. Dann heißt X

- (v) **stationär**, falls $\mathcal{L}[(X_{s+t})_{t \in I}] = \mathcal{L}[(X_t)_{t \in I}]$ für jedes $s \in I$,
- (vi) Prozess mit **stationären Zuwächsen**, falls X reellwertig ist und $\mathcal{L}[X_{s+t+r} - X_{t+r}] = \mathcal{L}[X_{s+r} - X_r]$ für alle $r, s, t \in I$. (Ist $0 \in I$, so reicht es, $r = 0$ zu betrachten.)

Beispiel 9.8. (i) Der Poissonprozess mit Intensität θ sowie die Irrfahrt auf \mathbb{Z} sind Prozesse mit stationären, unabhängigen Zuwächsen.

(ii) Sind $X_t, t \in I$, u.i.v. Zufallsvariablen, so ist $(X_t)_{t \in I}$ stationär.

(iii) Sei $(X_n)_{n \in \mathbb{Z}}$ reellwertig und stationär, und seien $k \in \mathbb{N}$ und $c_0, \dots, c_k \in \mathbb{R}$. Dann definiert

$$Y_n := \sum_{i=0}^k c_i X_{n-i}$$

einen stationären Prozess $Y = (Y_n)_{n \in \mathbb{Z}}$. Gilt $c_0, \dots, c_k \geq 0$ und $c_0 + \dots + c_k = 1$, so wird Y das **gleitende Mittel** von X (mit Gewichten c_0, \dots, c_k) genannt. \diamond

Die beiden folgenden Definitionen sind auch für allgemeinere halbgeordnete Mengen I sinnvoll, wir beschränken uns jedoch weiterhin auf den Fall $I \subset \mathbb{R}$.

Definition 9.9 (Filtration). Eine Familie $\mathbb{F} = (\mathcal{F}_t, t \in I)$ von σ -Algebren mit $\mathcal{F}_t \subset \mathcal{F}$ für jedes $t \in I$, heißt **Filtration**, falls $\mathcal{F}_s \subset \mathcal{F}_t$ für alle $s, t \in I$ mit $s \leq t$.

Definition 9.10 (adaptiert). Ein stochastischer Prozess $X = (X_t, t \in I)$ heißt **adaptiert** an die Filtration \mathbb{F} , falls X_t bezüglich \mathcal{F}_t messbar ist für jedes $t \in I$. Gilt $\mathcal{F}_t = \sigma(X_s, s \leq t)$ für jedes $t \in I$, so schreiben wir $\mathbb{F} = \sigma(X)$ und nennen \mathbb{F} die von X erzeugte **Filtration**.

Bemerkung 9.11. Offenbar ist ein stochastischer Prozess stets an seine erzeugte Filtration adaptiert. Die erzeugte Filtration ist die „kleinste“ Filtration, an die ein Prozess adaptiert ist. \diamond

Definition 9.12 (vorhersagbar / previsibel). Ein stochastischer Prozess $X = (X_n, n \in \mathbb{N}_0)$ heißt **vorhersagbar** (oder **previsibel**) bezüglich der Filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N}_0)$, falls X_0 konstant ist und für jedes $n \in \mathbb{N}$ gilt:

$$X_n \text{ ist } \mathcal{F}_{n-1}\text{-messbar.}$$

Beispiel 9.13. Seien $I = \mathbb{N}_0$, und seien Y_1, Y_2, \dots reelle Zufallsvariablen sowie $X_n := \sum_{m=1}^n Y_m$. Setze

$$\mathcal{F}_0 = \{\emptyset, \Omega\} \quad \text{und} \quad \mathcal{F}_n = \sigma(Y_1, \dots, Y_n) \quad \text{für } n \in \mathbb{N}.$$

Dann ist $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N}_0) = \sigma(Y) = (\sigma(Y_n)_{n \in \mathbb{N}})$ die von $Y = (Y_n)_{n \in \mathbb{N}}$ erzeugte Filtration, und X ist an \mathbb{F} adaptiert, also ist $\sigma(X) \subset \mathbb{F}$. Offenbar ist (Y_1, \dots, Y_n) messbar bezüglich $\sigma(X_1, \dots, X_n)$, also $\sigma(Y) \subset \sigma(X)$, und daher gilt auch $\mathbb{F} = \sigma(X)$.

Sei nun $\tilde{X}_n := \sum_{m=1}^n \mathbb{1}_{[0, \infty)}(Y_m)$. Dann ist auch \tilde{X} an \mathbb{F} adaptiert, jedoch ist im Allgemeinen $\mathbb{F} \not\supseteq \sigma(\tilde{X})$. \diamond

Beispiel 9.14. Sei $I = \mathbb{N}_0$, und seien D_1, D_2, \dots unabhängig und identisch verteilt mit $\mathbf{P}[D_i = -1] = \mathbf{P}[D_i = 1] = \frac{1}{2}$ für jedes $i \in \mathbb{N}$. Setze $D = (D_i)_{i \in \mathbb{N}}$ und

$\mathbb{F} = \sigma(D)$. Wir interpretieren D_i als das Ergebnis einer Wette, die uns pro Spielschein einen Gewinn oder Verlust von einer Geldeinheit bringt. Vor jedem Spiel entscheiden wir, wie viele Spielscheine wir einsetzen wollen. Die Anzahl H_n der in der n -ten Runde eingesetzten Spielscheine darf nur von den Ergebnissen der bisherigen Spiele abhängen, nicht aber von D_n und auch nicht von einem D_m für $m > n$. Mit anderen Worten: Es muss eine Funktion $F_n : \{-1, 1\}^{n-1} \rightarrow \mathbb{N}$ geben mit $H_n = F_n(D_1, \dots, D_{n-1})$. (Für das Petersburger Spiel (Beispiel 4.22) gilt beispielsweise $F_n(x_1, \dots, x_{n-1}) = 2^{n-1} \mathbb{1}_{\{x_1=x_2=\dots=x_{n-1}=0\}}$.) Damit ist H dann vorhersagbar. Andererseits besitzt jedes vorhersagbare H die Gestalt $H_n = F_n(D_1, \dots, D_{n-1})$, $n \in \mathbb{N}$, für gewisse Funktionen $F_n : \{-1, 1\}^{n-1} \rightarrow \mathbb{N}$, kommt also als Spielstrategie in Betracht. \diamond

Definition 9.15 (Stoppzeit). Eine Zufallsvariable τ mit Werten in $I \cup \{\infty\}$ heißt **Stoppzeit** (bezüglich \mathbb{F}), falls für jedes $t \in I$ gilt, dass

$$\{\tau \leq t\} \in \mathcal{F}_t.$$

Die Idee hinter dieser Definition ist, dass \mathcal{F}_t den Kenntnisstand eines Beobachters zur Zeit t wiedergibt. Der Wahrheitsgehalt der Aussage $\{\tau \leq t\}$ kann also aufgrund der Beobachtungen bis zur Zeit t bestimmt werden.

Satz 9.16. Ist I abzählbar, so ist τ genau dann eine Stoppzeit, wenn $\{\tau = t\} \in \mathcal{F}_t$ für jedes $t \in I$ gilt.

Beweis. Übung! \square

Beispiel 9.17. Seien $I = \mathbb{N}_0$ (oder allgemeiner $I \subset [0, \infty)$ rechts-diskret in dem Sinne, dass $t < \inf I \cap (t, \infty)$ für jedes $t \geq 0$, und damit insbesondere abzählbar) und $K \subset \mathbb{R}$ messbar sowie X ein reeller, adaptierter stochastischer Prozess. Wir betrachten den Zeitpunkt, zu dem X erstmals in K ist:

$$\tau_K := \inf\{t \in I : X_t \in K\}.$$

Intuitiv ist klar, dass τ_K eine Stoppzeit ist, denn ob $\{\tau_K \leq t\}$ eintritt oder nicht, können wir aufgrund der Beobachtungen von X bis zur Zeit t entscheiden. Formal können wir argumentieren, indem wir bemerken, dass $\{X_s \in K\} \in \mathcal{F}_s \subset \mathcal{F}_t$ für $s \leq t$ gilt. Also ist auch die abzählbare Vereinigung dieser Mengen wieder in \mathcal{F}_t :

$$\{\tau_K \leq t\} = \bigcup_{s \in I \cap [0, t]} \{X_s \in K\} \in \mathcal{F}_t.$$

Betrachte nun den zufälligen Zeitpunkt $\tilde{\tau} := \sup\{t \in I : X_t \in K\}$ des letzten Aufenthalts von X in K . Zu fester Zeit t können wir aufgrund der bisherigen Beobachtungen nicht entscheiden, ob X bereits das letzte Mal in K war. Hierzu bedürfte es der Prophetie. Also ist $\tilde{\tau}$ im Allgemeinen keine Stoppzeit. \diamond

Lemma 9.18. Sei $I \subset [0, \infty)$ abgeschlossen unter Addition, und seien σ und τ Stoppzeiten. Dann gilt:

- (i) $\sigma \vee \tau$ und $\sigma \wedge \tau$ sind Stoppzeiten.
- (ii) Gilt $\sigma, \tau \geq 0$, dann ist auch $\sigma + \tau$ eine Stoppzeit.
- (iii) Ist $s \geq 0$, dann ist $\tau + s$ eine Stoppzeit, jedoch im Allgemeinen nicht $\tau - s$.

Bevor wir zum einfachen formalen Beweis kommen, wollen wir festhalten, dass insbesondere (i) und (iii) Eigenschaften sind, die wir von Stoppzeiten erwarten können: Bei (i) ist die Interpretation klar. Für (iii) beachte man, dass $\tau - s$ um s in die Zukunft blickt (denn $\{\tau - s \leq t\} \in \mathcal{F}_{t+s}$), während $\tau + s$ um s in die Vergangenheit schaut. Stoppzeiten ist aber nur der Blick in die Vergangenheit erlaubt.

Beweis. (i) Für jedes $t \in I$ ist $\{\sigma \vee \tau \leq t\} = \{\sigma \leq t\} \cap \{\tau \leq t\} \in \mathcal{F}_t$ und $\{\sigma \wedge \tau \leq t\} = \{\sigma \leq t\} \cup \{\tau \leq t\} \in \mathcal{F}_t$.

(ii) Sei $t \in I$. Nach (i) sind $\tau \wedge t$ und $\sigma \wedge t$ Stoppzeiten für jedes $t \in I$. Speziell ist für jedes $s \leq t$ dann $\{\tau \wedge t \leq s\} \in \mathcal{F}_s \subset \mathcal{F}_t$. Andererseits ist für $s > t$ stets $\tau \wedge t \leq s$. Also sind $\tau' := (\tau \wedge t) + \mathbb{1}_{\{\tau>t\}}$ und $\sigma' := (\sigma \wedge t) + \mathbb{1}_{\{\sigma>t\}}$ messbar bezüglich \mathcal{F}_t und damit auch $\tau' + \sigma'$. Es folgt $\{\tau + \sigma \leq t\} = \{\tau' + \sigma' \leq t\} \in \mathcal{F}_t$.

(iii) Für $\tau + s$ folgt dies aus (ii) (mit der Stoppzeit $\sigma \equiv s$). Für $\tau - s$ beachte man, dass in der Definition der Stoppzeit für jedes $t \in I$ lediglich gefordert wird, dass $\{\tau - s \leq t\} = \{\tau \leq t + s\} \in \mathcal{F}_{t+s}$. Im Allgemeinen ist aber \mathcal{F}_{t+s} eine echte Obermenge von \mathcal{F}_t , also $\tau - s$ keine Stoppzeit. \square

Definition 9.19. Ist τ eine Stoppzeit, so heißt

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ für jedes } t \in I\}$$

die σ -Algebra der τ -Vergangenheit.

Beispiel 9.20. Sei $I = \mathbb{N}_0$ (oder rechts-diskret, vergleiche Beispiel 9.17), X ein adaptierter, reellwertiger stochastischer Prozess und $K \in \mathbb{R}$. Definiere die Stoppzeit $\tau = \inf\{t : X_t \geq K\}$ des ersten Eintretens in $[K, \infty)$. Betrachte die Ereignisse $A = \{\sup\{X_t : t \in I\} > K - 5\}$ und $B = \{\sup\{X_t : t \in I\} > K + 5\}$.

Für jedes $t \in I$ ist $\{\tau \leq t\} \subset A$, also ist $A \cap \{\tau \leq t\} = \{\tau \leq t\} \in \mathcal{F}_t$. Es folgt $A \in \mathcal{F}_\tau$. Jedoch ist im Allgemeinen $B \notin \mathcal{F}_\tau$, denn wir können bis zur Zeit τ eben nicht entscheiden, ob X auch die Hürde $K + 5$ noch nehmen wird oder nicht. \diamond

Lemma 9.21. Sind σ und τ Stoppzeiten mit $\sigma \leq \tau$, so gilt $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

Beweis. Sei $A \in \mathcal{F}_\sigma$ und $t \in I$. Dann ist $A \cap \{\sigma \leq t\} \in \mathcal{F}_t$. Da τ eine Stoppzeit ist, ist auch $\{\tau \leq t\} \in \mathcal{F}_t$. Wegen $\sigma \leq \tau$ ist also

$$A \cap \{\tau \leq t\} = (A \cap \{\sigma \leq t\}) \cap \{\tau \leq t\} \in \mathcal{F}_t. \quad \square$$

Definition 9.22. Ist $\tau < \infty$ Stoppzeit, dann setzen wir $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$.

Lemma 9.23. Ist I höchstens abzählbar und X adaptiert sowie $\tau < \infty$ eine Stoppzeit, dann ist X_τ messbar bezüglich \mathcal{F}_τ .

Beweis. Sei A messbar und $t \in I$. Für jedes $s \leq t$ ist dann $\{\tau = s\} \cap X_s^{-1}(A) \in \mathcal{F}_s \subset \mathcal{F}_t$. Also ist

$$X_\tau^{-1}(A) \cap \{\tau \leq t\} = \bigcup_{\substack{s \in I \\ s \leq t}} \left(\{\tau = s\} \cap X_s^{-1}(A) \right) \in \mathcal{F}_t. \quad \square$$

Für überabzählbares I und festes ω ist die Abbildung $I \rightarrow E$, $t \mapsto X_t(\omega)$ im Allgemeinen nicht messbar, also ist auch die Zusammensetzung X_τ nicht immer messbar. Hier brauchen wir Zusatzannahmen an die Regularität der Pfade $t \mapsto X_t(\omega)$, zum Beispiel Rechtsstetigkeit. Auf diesen Punkt kommen wir in Kapitel 21 zurück und belassen es hier einstweilen bei der Warnung.

9.2 Martingale

Als ein faires Spiel mit mehreren Runden würden wir sicherlich ohne Zögern die sukzessive Auszahlung von Gewinnen Y_1, Y_2, \dots bezeichnen, falls Y_1, Y_2, \dots u.i.v. mit $\mathbf{E}[Y_1] = 0$ sind. In diesem Fall ist der Prozess X der Partialsummen $X_n = Y_1 + \dots + Y_n$ integrierbar und $\mathbf{E}[X_n | \mathcal{F}_m] = X_m$, wenn $m < n$ ist (wobei $\mathbb{F} = \sigma(X)$). Wir wollen diese Gleichung als Definitionsgleichung für ein faires Spiel, das wir im Folgenden ein Martingal nennen, heranziehen und so die Bedingungen der Unabhängigkeit und der identisch verteilten Zuwächse erheblich abschwächen. Die Tragweite der folgenden Begriffsbildung wird sich erst nach und nach herausstellen.

Definition 9.24. Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, $I \subset \mathbb{R}$, und \mathbb{F} eine Filtration. Sei $X = (X_t)_{t \in I}$ ein reellwertiger, adaptierter stochastischer Prozess mit $\mathbf{E}[|X_t|] < \infty$ für jedes $t \in I$. X heißt (bezüglich \mathbb{F}) ein

Martingal, falls $\mathbf{E}[X_t | \mathcal{F}_s] = X_s$ für alle $s, t \in I$ mit $t > s$,

Submartingal, falls $\mathbf{E}[X_t | \mathcal{F}_s] \geq X_s$ für alle $s, t \in I$ mit $t > s$,

Supermartingal, falls $\mathbf{E}[X_t | \mathcal{F}_s] \leq X_s$ für alle $s, t \in I$ mit $t > s$.

Bemerkung 9.25. Offenbar ist für Martingale $t \mapsto \mathbf{E}[X_t]$ konstant, für Submartingale monoton wachsend und für Supermartingale monoton fallend. \diamond

Bemerkung 9.26. Die Etymologie des Begriffs *Martingale* ist nicht völlig geklärt. Das französische *la martingale* (ursprünglich provenzalisch *martegalo*) nach der

Stadt *Martiques*) bedeutet im Reitsport einen „beim Spring- und Geländereiten verwendeten Hilfszügel“ als Teil des Zaumzeugs ([25]). Manchmal wird die verzweigte Form, insbesondere des *Jagdmartingals* (französisch: *la martingale à anneaux*, englisch: *running martingale*), als sinnbildlich für die Verdoppelungsstrategie im Petersburger Spiel angesehen. Eben diese Verdoppelungsstrategie ist die zweite Bedeutung von *la martingale*. Von hier aus scheint eine Bedeutungsverschiebung hin zum mathematischen Begriff durchaus möglich. Eine andere Herleitung geht, statt vom Aussehen, von der Funktion des Zaumzeugs aus und nennt das Bestreben einer Spielstrategie, den Zufall *im Zaume* zu halten. So wird der Begriff des Martingals zunächst auf Spielstrategien im Allgemeinen, dann auf die Verdoppelungsstrategie im Speziellen übertragen. ◇

Bemerkung 9.27. Ist $I = \mathbb{N}$, $I = \mathbb{N}_0$ oder $I = \mathbb{Z}$, so reicht es, jeweils nur $t = s + 1$ zu betrachten, denn nach der Turmeigenschaft der bedingten Erwartung (Satz 8.14(iv)) ist

$$\mathbf{E}[X_{s+2} | \mathcal{F}_s] = \mathbf{E}[\mathbf{E}[X_{s+2} | \mathcal{F}_{s+1}] | \mathcal{F}_s],$$

und wenn die definierende Gleichung (beziehungsweise Ungleichung) in einem Zeitschritt gilt, dann zieht sie sich durch in den zweiten Zeitschritt und so fort. ◇

Bemerkung 9.28. Geben wir die Filtration \mathbb{F} nicht explizit an, so nehmen wir stillschweigend an, dass \mathbb{F} die von X erzeugte Filtration $\mathcal{F}_t = \sigma(X_s, s \leq t)$ ist. ◇

Bemerkung 9.29. Sind \mathbb{F} und \mathbb{F}' Filtrationen mit $\mathcal{F}_t \subset \mathcal{F}'_t$ für jedes t , und ist X an \mathbb{F} adaptiert und ein \mathbb{F}' -(Sub-, Super-)Martingal, dann ist X auch ein (Sub-, Super-)Martingal bezüglich der kleineren Filtration \mathbb{F} . Es gilt nämlich für $s < t$ und den Fall, wo X ein Submartingal ist,

$$\mathbf{E}[X_t | \mathcal{F}_s] = \mathbf{E}[\mathbf{E}[X_t | \mathcal{F}'_s] | \mathcal{F}_s] \geq \mathbf{E}[X_s | \mathcal{F}_s] = X_s.$$

Speziell ist ein \mathbb{F} -(Sub-, Super-)Martingal X also auch stets ein (Sub-, Super-)Martingal bezüglich der von X erzeugten Filtration $\sigma(X)$. ◇

Beispiel 9.30. Seien Y_1, \dots, Y_N unabhängige Zufallsvariablen mit $\mathbf{E}[Y_t] = 0$ für jedes $t = 1, \dots, N$. Setze $\mathcal{F}_t := \sigma(Y_1, \dots, Y_t)$ und $X_t := \sum_{s=1}^t Y_s$. Dann ist X adaptiert und integrierbar, und für $r > s$ ist $\mathbf{E}[Y_r | \mathcal{F}_s] = 0$. Also ist für $t > s$

$$\mathbf{E}[X_t | \mathcal{F}_s] = \mathbf{E}[X_s | \mathcal{F}_s] + \mathbf{E}[X_t - X_s | \mathcal{F}_s] = X_s + \sum_{r=s+1}^t \mathbf{E}[Y_r | \mathcal{F}_s] = X_s.$$

Es folgt, dass X ein \mathbb{F} -Martingal ist.

Analog ist X ein Submartingal, falls $\mathbf{E}[Y_t] \geq 0$ für jedes t gilt beziehungsweise ein Supermartingal, falls $\mathbf{E}[Y_t] \leq 0$ für jedes t gilt. ◇

Beispiel 9.31. Wir betrachten die Situation des vorangehenden Beispiels, jedoch mit $\mathbf{E}[Y_t] = 1$ und $X_t = \prod_{s=1}^t Y_s$ für $t \in \mathbb{N}_0$. Nach Satz 5.4 ist $Y_1 \cdot Y_2$ integrierbar. Iterativ erhalten wir $\mathbf{E}[|X_t|] < \infty$ für jedes $t \in \mathbb{N}_0$. Offenbar ist X an \mathbb{F} adaptiert, und für $s \in \mathbb{N}_0$ gilt

$$\mathbf{E}[X_{s+1} | \mathcal{F}_s] = \mathbf{E}[X_s Y_{s+1} | \mathcal{F}_s] = X_s \mathbf{E}[Y_{s+1} | \mathcal{F}_s] = X_s.$$

Also ist X ein \mathbb{F} -Martingal. \diamond

Satz 9.32. (i) *X ist genau dann ein Supermartingal, wenn $(-X)$ ein Submartingal ist.*

(ii) *Seien X und Y Martingale und $a, b \in \mathbb{R}$. Dann ist $(aX + bY)$ ein Martingal.*

(iii) *Seien X und Y Supermartingale und $a, b \geq 0$. Dann ist $(aX + bY)$ ein Supermartingal.*

(iv) *Seien X und Y Supermartingale. Dann ist $Z := X \wedge Y = (\min(X_t, Y_t))_{t \in I}$ ein Supermartingal.*

(v) *Ist $(X_t)_{t \in \mathbb{N}_0}$ ein Supermartingal und $\mathbf{E}[X_T] \geq \mathbf{E}[X_0]$ für ein $T \in \mathbb{N}_0$, dann ist $(X_t)_{t \in \{0, \dots, T\}}$ ein Martingal. Gibt es eine Folge $T_N \rightarrow \infty$ mit $\mathbf{E}[X_{T_N}] \geq \mathbf{E}[X_0]$, dann ist X ein Martingal.*

Beweis. (i), (ii) und (iii) Dies ist klar.

(iv) Wegen $|Z_t| \leq |X_t| + |Y_t|$ ist $\mathbf{E}[|Z_t|] < \infty$ für jedes $t \in I$. Wegen der Monotonie der bedingten Erwartung (Satz 8.14(ii)) ist $\mathbf{E}[Z_t | \mathcal{F}_s] \leq \mathbf{E}[X_t | \mathcal{F}_s] \leq X_s$ für $t > s$ und $\mathbf{E}[Z_t | \mathcal{F}_s] \leq \mathbf{E}[Y_t | \mathcal{F}_s] \leq Y_s$, also $\mathbf{E}[Z_t | \mathcal{F}_s] \leq X_s \wedge Y_s = Z_s$.

(v) Für $t \leq T$ setze $Y_t := \mathbf{E}[X_T | \mathcal{F}_t]$. Dann ist Y ein Martingal und $Y_t \leq X_t$. Daher ist

$$\mathbf{E}[X_0] \leq \mathbf{E}[X_T] = \mathbf{E}[Y_T] = \mathbf{E}[Y_t] \leq \mathbf{E}[X_t] \leq \mathbf{E}[X_0].$$

(Die erste Ungleichung gilt hierbei nach Voraussetzung.) Es folgt $Y_t = X_t$ fast sicher für jedes t , und daher ist $(X_t)_{t \in \{0, \dots, T\}}$ ein Martingal.

Sei $T_N \rightarrow \infty$ mit $\mathbf{E}[X_{T_N}] \geq \mathbf{E}[X_0]$ für jedes $N \in \mathbb{N}$. Dann gibt es für $t > s \geq 0$ ein $N \in \mathbb{N}$ mit $T_N > t$. Daher ist $\mathbf{E}[X_t | \mathcal{F}_s] = \mathbf{E}[X_s]$, also X ein Martingal. \square

Bemerkung 9.33. Viele Aussagen über Supermartingale gelten *mutatis mutandis* auch für Submartingale. So gilt im vorangehenden Satz Aussage (i) mit vertauschten Rollen, Aussage (iv) gilt für Submartingale, wenn das Minimum durch ein Maximum ersetzt wird, und so fort. Wir geben die Aussagen nicht stets sowohl für Submartingale wie für Supermartingale an, sondern wählen *pars pro toto* einen Fall aus. Man beachte aber, dass die Aussagen, die explizit über Martingale gemacht werden, nicht ohne weiteres auf Sub- oder Supermartingale übertragen werden können (vergleiche etwa (ii) im vorangehenden Satz). \diamond

Korollar 9.34. Sei X ein Submartingal und $a \in \mathbb{R}$. Dann ist $(X - a)^+$ ein Submartingal.

Beweis. Offenbar sind 0 und $Y = X - a$ Submartingale. Nach (iv) ist daher auch $(X - a)^+ = Y \vee 0$ ein Submartingal. \square

Satz 9.35. Sei X ein Martingal und $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine konvexe Funktion.

(i) Ist

$$\mathbf{E}[\varphi(X_t)^+] < \infty \quad \text{für jedes } t \in I, \quad (9.1)$$

dann ist $(\varphi(X_t))_{t \in I}$ ein Submartingal.

(ii) Ist $t^* := \sup(I) \in I$, so impliziert $\mathbf{E}[\varphi(X_{t^*})^+] < \infty$ schon (9.1).

(iii) Ist speziell $p \geq 1$ und $\mathbf{E}[|X_t|^p] < \infty$ für jedes $t \in I$, dann ist $(|X_t|^p)_{t \in I}$ ein Submartingal.

Beweis. (i) Es ist stets $\mathbf{E}[\varphi(X_t)^-] < \infty$ (Satz 7.9), also nach Voraussetzung $\mathbf{E}[|\varphi(X_t)|] < \infty$ für jedes $t \in I$. Die Jensen'sche Ungleichung (Satz 8.20) liefert für $t > s$

$$\mathbf{E}[\varphi(X_t) | \mathcal{F}_s] \geq \varphi(\mathbf{E}[X_t | \mathcal{F}_s]) = \varphi(X_s).$$

(ii) Da φ konvex ist, ist auch $x \mapsto \varphi(x)^+$ konvex. Weiter ist nach Voraussetzung $\mathbf{E}[\varphi(X_{t^*})^+] < \infty$, also gilt nach der Jensen'schen Ungleichung für jedes $t \in I$:

$$\mathbf{E}[\varphi(X_t)^+] = \mathbf{E}[\varphi(\mathbf{E}[X_{t^*} | \mathcal{F}_t])^+] \leq \mathbf{E}[\mathbf{E}[\varphi(X_{t^*})^+ | \mathcal{F}_t]] = \mathbf{E}[\varphi(X_{t^*})^+] < \infty.$$

(iii) Dies ist klar, weil $x \mapsto |x|^p$ konvex ist. \square

Beispiel 9.36. (Siehe Beispiel 9.4.) Die symmetrische einfache Irrfahrt X auf \mathbb{Z} ist ein quadratintegrierbares Martingal. Also ist $(X_n^2)_{n \in \mathbb{N}_0}$ ein Submartingal. \diamond

Übung 9.2.1. Sei Y eine Zufallsvariable mit $\mathbf{E}[|Y|] < \infty$ und \mathbb{F} eine Filtration sowie

$$X_t := \mathbf{E}[Y | \mathcal{F}_t] \quad \text{für jedes } t \in I.$$

Man zeige, dass X ein \mathbb{F} -Martingal ist. \clubsuit

Übung 9.2.2. Sei $(X_n)_{n \in \mathbb{N}_0}$ ein vorhersagbares \mathbb{F} -Martingal. Man zeige, dass dann für jedes $n \in \mathbb{N}_0$ fast sicher $X_n = X_0$ gilt. \clubsuit

Übung 9.2.3. Man zeige, dass die Aussage von Satz 9.35 auch gilt, wenn X nur ein Submartingal, φ jedoch zusätzlich monoton wachsend ist. Man zeige durch ein Beispiel, dass hier auf die Monotonie im Allgemeinen nicht verzichtet werden kann. (Vergleiche Korollar 9.34.) \clubsuit

Übung 9.2.4 (Ungleichung von Azuma). Man zeige:

- (i) Ist X eine Zufallsvariable mit $|X| \leq 1$ f.s., so gibt es eine Zufallsvariable Y mit Werten in $\{-1, +1\}$ und mit $\mathbf{E}[Y|X] = X$.
- (ii) Für X wie in (i) mit $\mathbf{E}[X] = 0$ folgere man (mit Hilfe der Jensen'schen Ungleichung)

$$\mathbf{E}[e^{\lambda X}] \leq \cosh(\lambda) \leq e^{\lambda^2/2} \quad \text{für alle } \lambda \in \mathbb{R}.$$

- (iii) Ist $(M_n)_{n \in \mathbb{N}_0}$ ein Martingal mit $M_0 = 0$, und gibt es eine Folge $(c_k)_{k \in \mathbb{N}}$ nichtnegativer Zahlen mit $|M_n - M_{n-1}| \leq c_n$ f.s. für jedes $n \in \mathbb{N}$, so gilt

$$\mathbf{E}[e^{\lambda M_n}] \leq \exp\left(\frac{1}{2}\lambda^2 \sum_{k=1}^n c_k^2\right).$$

- (iv) Unter den Bedingungen von (iii) gilt die **Azuma'sche Ungleichung**

$$\mathbf{P}[|M_n| \geq \lambda] \leq 2 \exp\left(-\frac{\lambda^2}{2 \sum_{k=1}^n c_k^2}\right) \quad \text{für alle } \lambda \geq 0.$$

Hinweis: Verwende die Markov'sche Ungleichung für $f(x) = e^{\gamma x}$ und wähle γ optimal. ♣

9.3 Diskretes stochastisches Integral

Bisher haben wir das Martingal als Partialsummenprozess eines fairen Spiels kennen gelernt. Dies kann beispielsweise auch der Kurs X einer Aktie sein, die zu diskreten Zeitpunkten $m = 0, 1, \dots$ an einer Börse gehandelt wird. Bei dieser Interpretation ist es besonders evident, dass es natürlich ist, neue stochastische Prozesse zu generieren, indem man Anlagestrategien H für die entsprechende Aktie betrachtet. Die Wertentwicklung des neuen Prozesses $H \cdot X$ entspricht dann der eines Portefeuilles, bei dem in jedem Zeitintervall $(m-1, m]$ genau H_m Aktien gehalten werden. Dies führt zu der folgenden Begriffsbildung.

Definition 9.37 (Diskretes Stochastisches Integral). Sei $(X_n)_{n \in \mathbb{N}_0}$ ein reeller, \mathbb{F} -adaptierter Prozess und $(H_n)_{n \in \mathbb{N}}$ reellwertig und \mathbb{F} -vorhersagbar. Wir definieren den stochastischen Prozess $H \cdot X$ durch

$$(H \cdot X)_n := \sum_{m=1}^n H_m (X_m - X_{m-1}) \quad \text{für } n \in \mathbb{N}_0, \tag{9.2}$$

und nennen $H \cdot X$ das **diskrete stochastische Integral** von H bezüglich X . Ist X ein Martingal, so nennen wir $H \cdot X$ auch die **Martingaltransformierte** von X .

Bemerkung 9.38. Offenbar ist $H \cdot X$ adaptiert an \mathbb{F} . ◊

Sei X ein (möglicherweise unfaires) Spiel, wobei $X_n - X_{n-1}$ den Spielgewinn pro Spielschein in der n -ten Runde bezeichnet. Wir interpretieren H_n als die Anzahl der Spielscheine, die für das n -te Spiel eingesetzt werden, und verstehen H als **Spielstrategie**. Offenbar muss die Entscheidung, wie groß H_n sein soll, zur Zeit $n-1$, also vor der Bekanntgabe des Ergebnisses X_n fallen. Das heißt, H muss vorhersagbar sein.

Ist nun X ein faires Spiel, also ein Martingal, und ist H **lokal beschränkt** (das heißt, jedes H_n ist beschränkt), dann ist (wegen $\mathbf{E}[X_{n+1} - X_n | \mathcal{F}_n] = 0$)

$$\begin{aligned}\mathbf{E}[(H \cdot X)_{n+1} | \mathcal{F}_n] &= \mathbf{E}[(H \cdot X)_n + H_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n] \\ &= (H \cdot X)_n + H_{n+1} \mathbf{E}[X_{n+1} - X_n | \mathcal{F}_n] \\ &= (H \cdot X)_n.\end{aligned}$$

Also ist $H \cdot X$ ein Martingal. Im folgenden Satz zeigen wir, dass auch die Umkehrung gilt, also X ein Martingal ist, wenn für hinreichend viele vorhersagbare Prozesse das stochastische Integral ein Martingal ist.

Satz 9.39 (Stabilitätssatz für Stochastische Integrale).

Sei $(X_n)_{n \in \mathbb{N}_0}$ ein adaptierter, reeller stochastischer Prozess mit $\mathbf{E}[|X_0|] < \infty$.

- (i) X ist genau dann ein Martingal, wenn für jeden lokal beschränkten, vorhersagbaren Prozess H das stochastische Integral $H \cdot X$ ein Martingal ist.
- (ii) X ist genau dann ein Submartingal (Supermartingal), wenn $H \cdot X$ ein Submartingal (Supermartingal) ist für jedes lokal beschränkte, vorhersagbare $H \geq 0$.

Beweis. (i) „ \Rightarrow “ Dies hat die obige Diskussion schon gezeigt.

„ \Leftarrow “ Wähle $n_0 \in \mathbb{N}$. Setze $H_n = \mathbb{1}_{\{n=n_0\}}$. Dann ist $(H \cdot X)_{n_0-1} = 0$, also

$$0 = \mathbf{E}[(H \cdot X)_{n_0} | \mathcal{F}_{n_0-1}] = \mathbf{E}[X_{n_0} | \mathcal{F}_{n_0-1}] - X_{n_0-1}.$$

(ii) Dies geht analog wie in (i). □

Der vorangehende Satz sagt uns insbesondere, dass wir keine (lokal beschränkte) Spielstrategie finden können, die aus einem Martingal (oder schlimmer: einem Supermartingal) ein Submartingal mache. Genau dies wird einem aber natürlich durch diverse Aufforderungen zum so genannten „Systemlotto“ und Ähnlichem nahe gelegt.

Beispiel 9.40 (Petersburger Spiel). Wir führen Beispiel 9.14 fort (siehe auch Beispiel 4.22). Setzen wir $X_n := D_1 + \dots + D_n$ für $n \in \mathbb{N}_0$, so ist X ein Martingal.

Die Spielstrategie $H_n := 2^{n-1} \mathbb{1}_{\{D_1=D_2=\dots=D_{n-1}=-1\}}$ für $n \in \mathbb{N}$ und $H_0 = 1$ ist vorhersagbar und lokal beschränkt. Sei $S_n = \sum_{i=1}^n H_i D_i = (H \cdot X)_n$ der Zugewinn nach n Runden. Dann ist S nach dem vorangehenden Satz ein Martingal. Speziell erhalten wir das bereits in Beispiel 4.22 gezeigte Ergebnis, dass $\mathbf{E}[S_n] = 0$ ist für jedes $n \in \mathbb{N}$. Dass dies, wie dort gezeigt, in zumindest vordergründigem Kontrast zu der Aussage $S_n \xrightarrow{n \rightarrow \infty} 1$ f.s. steht, wird uns später noch einmal beschäftigen (siehe Beispiel 11.6).

Für den Moment sei angemerkt, dass das Martingal $S' = (1 - S_n)_{n \in \mathbb{N}_0}$ wie in Beispiel 9.31 die Struktur eines Produkts unabhängiger Zufallsvariablen mit Erwartungswert 1 hat. Es gilt nämlich $S'_n = \prod_{i=1}^n (1 - D_i)$. \diamond

9.4 Diskreter Martingaldarstellungssatz und CRR Modell

Wir haben nun gesehen, dass wir mittels des stochastischen Integrals aus einem Martingal X durch eine Spielstrategie H ein neues Martingal $H \cdot X$ herstellen können. Welche Martingale Y (mit $Y_0 = 0$) sind nun durch eine geeignete Spielstrategie $H = H(Y)$ aus X zu gewinnen? Womöglich alle? Dies ist sicher nicht der Fall, wie das folgende Beispiel zeigt. Allerdings sind alle Martingale darstellbar, wenn für die Zuwächse $X_{n+1} - X_n$ immer nur zwei Werte in Frage kommen (gegeben X_1, \dots, X_n). Wir geben für diesen Fall einen Darstellungssatz an und diskutieren in der Folge den fairen Preis der europäischen Kaufoption (europäischer *Call*) in dem Aktienkursmodell von Cox-Ross-Rubinstein. Wir wollen dabei einen naiven Standpunkt einnehmen und einen in vielerlei Hinsicht idealisierten Markt voraussetzen (keine Handelskosten, gebrochene Anzahlen handelbar, und so fort). Für eine umfassendere Lektüre zum Thema Finanzmathematik eignen sich etwa die Lehrbücher [47], [86], [109], [62], [12] oder [53].

Beispiel 9.41. Wir betrachten ein ganz einfaches Martingal $X = (X_n)_{n=0,1}$ mit nur zwei Zeitpunkten. Es sei $X_0 = 0$ fast sicher und $\mathbf{P}[X_1 = -1] = \mathbf{P}[X_1 = 0] = \mathbf{P}[X_1 = 1] = \frac{1}{3}$. Sei $Y_0 = 0$ sowie $Y_1 = 2$, falls $X_1 = 1$ und $Y_1 = -1$ sonst. Dann ist Y offenbar ein $\sigma(X)$ -Martingal. Allerdings können wir keine Zahl H_1 angeben, sodass $H_1 X_1 = Y_1$ wäre. \diamond

Sei $T \in \mathbb{N}$ ein fester Zeitpunkt. Ist $(Y_n)_{n=0,1,\dots,T}$ ein \mathbb{F} -Martingal, dann ist $Y_n = \mathbf{E}[Y_T | \mathcal{F}_n]$ für jedes $n \leq T$. Durch die Angabe von Y_T ist ein \mathbb{F} -Martingal Y also eindeutig festgelegt (und umgekehrt). Da $(H \cdot X)$ ein Martingal ist, falls X ein Martingal ist, reduziert sich das Darstellungsproblem für Martingale auf das Problem, eine integrierbare Zufallsvariable $V := Y_T$ darzustellen als $v_0 + (H \cdot X)_T$, wobei $v_0 = \mathbf{E}[Y_T]$ ist, falls X ein Martingal ist.

Wir haben eben schon gesehen, dass dies im Allgemeinen nicht möglich ist, wenn die Differenzen $X_{n+1} - X_n$ drei (oder mehr) unterschiedliche Werte annehmen können. Wir betrachten nun also den Fall, wo nur zwei Werte möglich sind. Hier

muss in jedem Schritt ein lineares Gleichungssystem mit zwei Gleichungen und zwei Unbekannten gelöst werden. Im Falle von drei möglichen Werten von $X_{n+1} - X_n$ wären dies drei Gleichungen für zwei Unbekannte, und im Allgemeinen ist dieses Gleichungssystem dann nicht lösbar.

Definition 9.42 (Binäres Modell). Ein stochastischer Prozess X_0, \dots, X_T heißt **binäres Modell**, oder **binär aufspaltend**, falls es Zufallsvariablen D_1, \dots, D_T mit Werten in $\{-1, +1\}$ gibt und Funktionen $f_n : \mathbb{R}^{n-1} \times \{-1, +1\} \rightarrow \mathbb{R}$ für $n = 1, \dots, T$ sowie $x_0 \in \mathbb{R}$, sodass $X_0 = x_0$ und

$$X_n = f_n(X_1, \dots, X_{n-1}, D_n) \quad \text{für jedes } n = 1, \dots, T.$$

Mit $\mathbb{F} = \sigma(X)$ bezeichnen wir dann die von X erzeugte Filtration.

Man beachte, dass X_n nur von X_1, \dots, X_{n-1} und D_n abhängt und nicht von der vollen Information der Werte D_1, \dots, D_n . Man mache sich klar, dass im letzteren Fall eine mehr als binäre Aufspaltung der Werte in einem Zeitschritt möglich wäre.

Satz 9.43 (Darstellungssatz). Sei X ein binäres Modell und V_T eine \mathcal{F}_T -messbare Zufallsvariable. Dann existiert ein beschränkter, vorhersagbarer Prozess H und ein $v_0 \in \mathbb{R}$ mit $V_T = v_0 + (H \cdot X)_T$.

Man beachte, dass \mathbb{F} die von X erzeugte Filtration ist, nicht die im Allgemeinen größere, von D_1, \dots, D_T erzeugte. Für diese ist die Aussage des Satzes im Allgemeinen nicht zutreffend, denn mit der Anlagestrategie H können wir nur an der Wertentwicklung von X partizipieren, nicht jedoch Wetten auf die Werte des zu Grunde liegenden Zufallsmechanismus direkt, also die D_i , abschließen. In dem Extremfall etwa, wo alle f_n konstant sind, können wir nur konstante Zufallsvariablen V_T als stochastisches Integral darstellen.

Beweis. Wir zeigen, dass es \mathcal{F}_{T-1} -messbare Zufallsvariablen V_{T-1} und H_T gibt, sodass $V_T = V_{T-1} + H_T(X_T - X_{T-1})$. Dies liefert per Rückwärtsinduktion die gewünschte Aussage.

Da V_T messbar ist bezüglich \mathcal{F}_T , existiert nach dem Faktorisierungslemma (Korollar 1.97) eine Funktion $g_T : \mathbb{R}^T \rightarrow \mathbb{R}$ mit $V_T = g_T(X_1, \dots, X_T)$. Wir setzen

$$X_T^\pm = f_T(X_1, \dots, X_{T-1}, \pm 1) \quad \text{und} \quad V_T^\pm = g_T(X_0, \dots, X_{T-1}, X_T^\pm).$$

Jede dieser vier Zufallsvariablen ist offenbar \mathcal{F}_{T-1} -messbar. Wir suchen nun also V_{T-1} und H_T , die das folgende lineare Gleichungssystem lösen

$$\begin{aligned} V_{T-1} + H_T(X_T^- - X_{T-1}) &= V_T^-, \\ V_{T-1} + H_T(X_T^+ - X_{T-1}) &= V_T^+. \end{aligned} \tag{9.3}$$

Per Konstruktion ist $X_T^+ - X_T^- \neq 0$, falls $V_T^+ - V_T^- \neq 0$. Also können wir (9.3) auflösen und erhalten

$$H_T := \begin{cases} \frac{V_T^+ - V_T^-}{X_T^+ - X_T^-}, & \text{falls } X_T^+ \neq X_T^-, \\ 0, & \text{sonst,} \end{cases}$$

und $V_{T-1} = V_T^+ - H_T(X_T^+ - X_{T-1}) = V_T^- - H_T(X_T^- - X_{T-1})$. \square

Wir wollen nun X als Kurs einer Aktie auffassen und V_T als Auszahlungsfunktion eines Finanzderivats auf X , eines so genannten **Claims**. Beispielsweise kann V_T eine **europäische Kaufoption (Call)** mit Fälligkeitszeitpunkt (*maturity*) T und Ausübungspreis (*strike price*) $K \geq 0$ sein. In diesem Fall wäre $V_T = (X_T - K)^+$. Ökonomisch ausgedrückt gibt diese Option dem Käufer das Recht (aber nicht die Pflicht) vom Herausgeber der Option zum Zeitpunkt T die Eingehung eines Kaufvertrages über eine Aktie zum Preis K einzufordern. Von diesem Recht macht man sinnvollerweise nur dann Gebrauch, wenn $X_T \geq K$ ist. In diesem Fall kann man die erworbene Aktie zum Preis X_T wieder an der Börse verkaufen und hat einen Gewinn von V_T gemacht.

An den Börsen werden nun aber nicht nur Aktien, sondern auch Derivate auf Aktien gehandelt. Welches ist also der faire Preis $\pi(V_T)$ für den eine Börsenhändlerin den Claim V_T anbieten kann? Gibt es eine Strategie H und ein v_0 , sodass $V_T = v_0 + (H \cdot X)_T$ ist, dann kann die Händlerin gegen Bezahlung von v_0 risikolos V_T nachbilden, indem sie H als Handelsstrategie an der Börse benutzt. Wir nennen den Claim V_T dann **replizierbar** und die Strategie H einen **Hedge**. Ein Markt, in dem jeder Claim replizierbar ist, heißt vollständig. In diesem Sinne ist das Binärmodell also ein vollständiger Markt.

Gäbe es nun eine zweite Strategie H' und ein zweites v'_0 mit $v'_0 + (H' \cdot X)_T = V_T$, so wäre insbesondere $v_0 - v'_0 = ((H' - H) \cdot X)_T$. Wäre $v_0 > v'_0$, so könnte die Händlerin risikolos durch Verfolgen der Strategie $H' - H$ einen Gewinn $v_0 - v'_0$ machen, im Falle $v_0 < v'_0$ hingegen durch die Strategie $H - H'$ einen Gewinn $v'_0 - v_0$. Wenn man solche **Arbitragemöglichkeiten** ausschließt, also einen so genannten **arbitragefreien Markt** annimmt, so ist der faire Preis $\pi(V_T)$ eindeutig festgelegt, sobald man *eine* Strategie H und *ein* v_0 mit $V_T = v_0 + (H \cdot X)_T$ hat.

Wir haben bislang noch nirgends angenommen, dass X ein Martingal sei. Ist nun aber X ein Martingal, so ist $(H \cdot X)$ ein Martingal mit $(H \cdot X)_0 = 0$, also offenbar $\mathbf{E}[(H \cdot X)_T] = 0$. Damit ist dann

$$\pi(V_T) = v_0 = \mathbf{E}[V_T]. \quad (9.4)$$

Da in diesem Fall v_0 nicht mehr von der Handelsstrategie abhängt, also eindeutig ist, ist der Markt automatisch arbitragefrei. In endlichen Märkten ist Arbitragefreiheit äquivalent zur Existenz eines äquivalenten Martingals (siehe unten), und in diesem Fall ist die Vollständigkeit äquivalent zur Eindeutigkeit des Martingals (Satz von

Harrison-Pliska (1981), [74]). In größeren Marktmodellen gilt Äquivalenz nur für einen etwas flexibleren Arbitrage-Begriff (siehe [35]).

Ist nun X kein Martingal, so können wir in manchen Fällen X durch einen anderen Prozess X' ersetzen, der ein Martingal ist, und so, dass die Verteilungen \mathbf{P}_X und $\mathbf{P}_{X'}$ die selben Nullmengen haben, also äquivalent sind. Ein Hedge, der V_T bezüglich X nachbildet, bildet V_T auch bezüglich X' nach. Insbesondere ändert sich der faire Preis $\pi(V_T)$ nicht, wenn wir zum Martingal X' übergehen. Wir können $\pi(V_T)$ also berechnen, indem wir (9.4) auf das äquivalente Martingal anwenden.

Das Maß $\mathbf{P}_{X'}$ wird auch **äquivalentes Martingalmaß** genannt. Während es hier für uns nur von rechnerischem Interesse ist, hat es eine ökonomische Interpretation als Maß für die Preisbildung, falls alle Händler sich risikoneutral verhalten, also den Wert einer künftigen Auszahlung allein nach deren Erwartungswert bemessen (was typischerweise nicht der Fall ist; die meisten Anleger sind risikoavers, lassen sich also Unsicherheiten durch einen Aufschlag bezahlen).

Nun wollen wir aber ein Modell im Detail betrachten.

Definition 9.44. Seien $T \in \mathbb{N}$, $a \in (-1, 0)$ und $b > 0$ sowie $p \in (0, 1)$. Ferner seien D_1, \dots, D_T u.i.v. mit $\mathbf{P}[D_1 = 1] = 1 - \mathbf{P}[D_1 = -1] = p$. Wir definieren $X_0 = x_0 > 0$ und für $n = 1, \dots, T$

$$X_n = \begin{cases} (1 + b) X_{n-1}, & \text{falls } D_n = +1, \\ (1 + a) X_{n-1}, & \text{falls } D_n = -1. \end{cases}$$

X heißt **mehrstufiges Binomialmodell** oder **Cox-Ross-Rubinstein'sches Modell** (ohne Verzinsung).

Nach dem bisher Gezeigten ist das CRR Modell vollständig. Ferner können wir durch die Wahl $p = p^* := \frac{a}{a-b}$ den Prozess X zu einem Martingal machen. Mithin ist das Modell auch arbitragefrei (für jedes $p \in (0, 1)$). Wir wollen nun den Preis des europäischen Calls $V_T := (X_T - K)^+$ explizit ausrechnen. Hierzu können wir wieder $p = p^*$ annehmen. Wir erhalten dann mit $A := \min\{i \in \mathbb{N}_0 : (1 + b)^i(1 + a)^{T-i}x_0 > K\}$,

$$\begin{aligned} \pi(V_T) &= \mathbf{E}_{p^*}[V_T] = \sum_{i=0}^T b_{T,p^*}(\{i\}) [(1 + b)^i(1 + a)^{T-i}x_0 - K]^+ \\ &= x_0 \sum_{i=A}^T \binom{T}{i} (p^*)^i (1 - p^*)^{T-i} [(1 + b)^i(1 + a)^{T-i}] - K \sum_{i=A}^T b_{T,p^*}(\{i\}). \end{aligned}$$

Setzen wir $p' = (1 + b)p^*$, dann ist $p' \in (0, 1)$ und $1 - p' = (1 - p^*)(1 + a)$. Wir erhalten so die Cox-Ross-Rubinstein'sche Formel

$$\pi(V_T) = x_0 b_{T,p'}(\{A, \dots, T\}) - K b_{T,p^*}(\{A, \dots, T\}). \quad (9.5)$$

Dies ist das diskrete Analogon zur berühmten Black–Scholes Formel für die Optionsbewertung in gewissen zeitkontinuierlichen Märkten.

10 Optional Sampling Sätze

Wir haben im vorigen Kapitel gesehen, dass aus Martingalen wieder Martingale werden, wenn man gewisse Spielstrategien anwendet. Wir wollen in diesem Kapitel ähnliche Stabilitätseigenschaften für zufällig gestoppte Martingale zeigen. Um die Aussagen auch für Submartingale und Supermartingale zu bekommen, geben wir im ersten Abschnitt einen Zerlegungssatz für adaptierte Prozesse an. Im zweiten Abschnitt kommen dann die Optional Sampling und Optional Stopping Sätze.

10.1 Doob-Zerlegung und quadratische Variation

Sei $X = (X_n)_{n \in \mathbb{N}_0}$ ein adaptierter Prozess mit $\mathbf{E}[|X_n|] < \infty$ für jedes $n \in \mathbb{N}_0$. Wir wollen X zerlegen in eine Summe aus einem Martingal und einem vorhersagbaren Prozess. Dazu definieren wir für $n \in \mathbb{N}_0$

$$M_n := X_0 + \sum_{k=1}^n (X_k - \mathbf{E}[X_k | \mathcal{F}_{k-1}]) \quad (10.1)$$

und

$$A_n := \sum_{k=1}^n (\mathbf{E}[X_k | \mathcal{F}_{k-1}] - X_{k-1}).$$

Offenbar ist $X_n = M_n + A_n$. Per Konstruktion ist A vorhersagbar mit $A_0 = 0$, und M ist ein Martingal, denn

$$\mathbf{E}[M_n - M_{n-1} | \mathcal{F}_{n-1}] = \mathbf{E}[X_n - \mathbf{E}[X_n | \mathcal{F}_{n-1}] | \mathcal{F}_{n-1}] = 0.$$

Satz 10.1 (Doob-Zerlegung). Sei $X = (X_n)_{n \in \mathbb{N}_0}$ ein adaptierter, integrierbarer Prozess. Dann existiert eine eindeutige Zerlegung $X = M + A$, wobei A vorhersagbar ist mit $A_0 = 0$ und M ein Martingal. Diese Darstellung von X heißt **Doob-Zerlegung**. X ist genau dann ein Submartingal, wenn A monoton wachsend ist.

Beweis. Nur die Eindeutigkeit ist zu zeigen. Seien also $X = M + A = M' + A'$ zwei Zerlegungen mit den genannten Eigenschaften. Dann ist $M - M' = A' - A$ ein

vorhersagbares Martingal, also ist (siehe Übung 9.2.2) $M_n - M'_n = M_0 - M'_0 = 0$ f.s. für jedes $n \in \mathbb{N}_0$. \square

Beispiel 10.2. Sei $I = \mathbb{N}_0$ oder $I = \{0, \dots, N\}$. Sei $(X_n)_{n \in I}$ ein quadratisch integrierbares \mathbb{F} -Martingal (das heißt $\mathbf{E}[X_n^2] < \infty$ für jedes $n \in I$). Nach Satz 9.35 ist $Y := (X_n^2)_{n \in I}$ ein Submartingal. Sei $Y = M + A$ die Doob-Zerlegung von Y . Es ist dann $(X_n^2 - A_n)_{n \in I}$ ein Martingal. Ferner ist $\mathbf{E}[X_{i-1}X_i | \mathcal{F}_{i-1}] = X_{i-1}\mathbf{E}[X_i | \mathcal{F}_{i-1}] = X_{i-1}^2$, also (wie in (10.1))

$$\begin{aligned} A_n &= \sum_{i=1}^n \left(\mathbf{E}[X_i^2 | \mathcal{F}_{i-1}] - X_{i-1}^2 \right) \\ &= \sum_{i=1}^n \left(\mathbf{E}[(X_i - X_{i-1})^2 | \mathcal{F}_{i-1}] - 2X_{i-1}^2 + 2\mathbf{E}[X_{i-1}X_i | \mathcal{F}_{i-1}] \right) \\ &= \sum_{i=1}^n \mathbf{E}[(X_i - X_{i-1})^2 | \mathcal{F}_{i-1}]. \end{aligned} \quad \diamond$$

Definition 10.3. Sei $(X_n)_{n \in I}$ ein quadratisch integrierbares \mathbb{F} -Martingal. Der eindeutig bestimmte vorhersagbare Prozess A , mit dem $(X_n^2 - A_n)_{n \in I}$ ein Martingal wird, heißt **quadratischer Variationsprozess** von X und wird in Formeln mit $(\langle X \rangle_n)_{n \in I} := A$ bezeichnet.

Aus dem vorangehenden Beispiel ergibt sich sofort der folgende Satz.

Satz 10.4. Sei X wie in Definition 10.3. Dann ist für $n \in \mathbb{N}_0$

$$\langle X \rangle_n = \sum_{i=1}^n \mathbf{E}[(X_i - X_{i-1})^2 | \mathcal{F}_{i-1}] \quad (10.2)$$

und

$$\mathbf{E}[\langle X \rangle_n] = \mathbf{Var}[X_n - X_0]. \quad (10.3)$$

Bemerkung 10.5. Sind Y und A wie in Beispiel 10.2, dann ist A monoton wachsend, weil $(X_n^2)_{n \in I}$ ein Submartingal ist (vergleiche Satz 10.1). Deshalb wird A manchmal auch der **wachsende Prozess** von Y genannt. \diamond

Beispiel 10.6. Seien Y_1, Y_2, \dots unabhängige, quadratisch integrierbare, zentrierte Zufallsvariablen. Dann wird durch $X_n := Y_1 + \dots + Y_n$ ein quadratisch integrierbares Martingal definiert mit $\langle X \rangle_n = \sum_{i=1}^n \mathbf{E}[Y_i^2]$, denn es ist $A_n = \sum_{i=1}^n \mathbf{E}[Y_i^2 | Y_1, \dots, Y_{i-1}] = \sum_{i=1}^n \mathbf{E}[Y_i^2]$ (wie in Beispiel 10.2).

Man beachte, dass es für diese einfache Darstellung von $\langle X \rangle$ nicht ausreicht, dass die Y_1, Y_2, \dots unkorreliert sind. \diamond

Beispiel 10.7. Seien Y_1, Y_2, \dots unabhängige, quadratisch integrierbare Zufallsvariablen mit $\mathbf{E}[Y_n] = 1$ für $n \in \mathbb{N}$. Setze $X_n := \prod_{i=1}^n Y_i$ für $n \in \mathbb{N}_0$. Dann ist $X = (X_n)_{n \in \mathbb{N}_0}$ ein quadratisch integrierbares Martingal (warum?) bezüglich $\mathbb{F} = \sigma(X)$ und

$$\mathbf{E}[(X_n - X_{n-1})^2 | \mathcal{F}_{n-1}] = \mathbf{E}[(Y_n - 1)^2 X_{n-1}^2 | \mathcal{F}_{n-1}] = \mathbf{Var}[Y_n] X_{n-1}^2.$$

Also ist $\langle X \rangle_n = \sum_{i=1}^n \mathbf{Var}[Y_i] X_{i-1}^2$. Wir sehen, dass der quadratische Variationsprozess also durchaus ein echt zufälliger Prozess sein kann. \diamond

Beispiel 10.8. Sei $(X_n)_{n \in \mathbb{N}_0}$ die eindimensionale symmetrische einfache Irrfahrt:

$$X_n = \sum_{i=1}^n R_i \quad \text{für jedes } n \in \mathbb{N}_0,$$

wobei R_1, R_2, R_3, \dots u.i.v. sind mit $\mathbf{P}[R_i = 1] = 1 - \mathbf{P}[R_i = -1] = \frac{1}{2}$.

Offenbar ist X ein Martingal, also $|X|$ ein Submartingal. Sei $|X| = M + A$ die Doob-Zerlegung von $|X|$. Dann ist

$$A_n = \sum_{i=1}^n (\mathbf{E}[|X_i| | \mathcal{F}_{i-1}] - |X_{i-1}|).$$

Nun ist

$$|X_i| = \begin{cases} |X_{i-1}| + R_i, & \text{falls } X_{i-1} > 0, \\ |X_{i-1}| - R_i, & \text{falls } X_{i-1} < 0, \\ 1, & \text{falls } X_{i-1} = 0. \end{cases}$$

Also gilt

$$\mathbf{E}[|X_i| | \mathcal{F}_{i-1}] = \begin{cases} |X_{i-1}|, & \text{falls } |X_{i-1}| \neq 0, \\ 1, & \text{falls } |X_{i-1}| = 0. \end{cases}$$

Mithin ist

$$A_n = \#\{i \leq n-1 : |X_i| = 0\}$$

die **Lokalzeit** von X in 0. Es folgt (wegen $\mathbf{P}[X_{2j} = 0] = \binom{2j}{j} 4^{-j}$ und $\mathbf{P}[X_{2j+1} = 0] = 0$):

$$\begin{aligned} \mathbf{E}[|X_n|] &= \mathbf{E}[\#\{i \leq n-1 : X_i = 0\}] \\ &= \sum_{i=0}^{n-1} \mathbf{P}[X_i = 0] = \sum_{j=0}^{\lfloor (n-1)/2 \rfloor} \binom{2j}{j} 4^{-j}. \end{aligned} \quad \diamond$$

Beispiel 10.9. Wir wollen das vorangehende Beispiel jetzt noch etwas verallgemeinern. Offenbar brauchten wir (außer in der letzten Formel) nicht, dass X eine Irrfahrt ist, sondern lediglich, dass die Differenzen $(\Delta X)_n := X_n - X_{n-1}$ nur die Werte

-1 und $+1$ annehmen können. Sei also jetzt X ein Martingal mit $|X_n - X_{n-1}| = 1$ fast sicher für jedes $n \in \mathbb{N}$ und mit $X_0 = x_0 \in \mathbb{Z}$ fast sicher. Sei $f : \mathbb{Z} \rightarrow \mathbb{R}$ eine beliebige Abbildung. Dann ist $Y := (f(X_n))_{n \in \mathbb{N}_0}$ ein integrierbarer, adaptierter Prozess (weil $|f(X_n)| \leq \max_{x \in \{x_0-n, \dots, x_0+n\}} |f(x)|$). Um die Doob-Zerlegung von Y zu bestimmen, definieren wir die erste und zweite (diskrete) Ableitung von f

$$f'(x) := \frac{f(x+1) - f(x-1)}{2}$$

und

$$f''(x) := f(x-1) + f(x+1) - 2f(x).$$

Wir setzen noch $F'_n := f'(X_{n-1})$ und $F''_n := f''(X_{n-1})$. Durch Unterscheidung der Fälle $X_n = X_{n-1} - 1$ und $X_n = X_{n-1} + 1$ sehen wir, dass für jedes $n \in \mathbb{N}$

$$\begin{aligned} f(X_n) - f(X_{n-1}) &= \frac{f(X_{n-1}+1) - f(X_{n-1}-1)}{2}(X_n - X_{n-1}) \\ &\quad + \frac{1}{2}f(X_{n-1}-1) + \frac{1}{2}f(X_{n-1}+1) - f(X_{n-1}) \\ &= f'(X_{n-1})(X_n - X_{n-1}) + \frac{1}{2}f''(X_{n-1}) \\ &= F'_n \cdot (X_n - X_{n-1}) + \frac{1}{2}F''_n. \end{aligned}$$

Insgesamt erhalten wir also die **diskrete Itô-Formel**

$$\begin{aligned} f(X_n) &= f(x_0) + \sum_{i=1}^n f'(X_{i-1})(X_i - X_{i-1}) + \sum_{i=1}^n \frac{1}{2}f''(X_{i-1}) \\ &= f(x_0) + (F' \cdot X)_n + \sum_{i=1}^n \frac{1}{2}F''_i. \end{aligned} \tag{10.4}$$

Hierbei ist $F' \cdot X$ das diskrete stochastische Integral (siehe Definition 9.37). Nun ist $M := f(x_0) + F' \cdot X$ ein Martingal nach Satz 9.39, weil F' vorhersagbar ist (und $|F'_n| \leq \max_{x \in \{x_0-n, \dots, x_0+n\}} |F'(x)|$), und $A := (\sum_{i=1}^n \frac{1}{2}F''_i)_{n \in \mathbb{N}_0}$ ist vorhersagbar. Also ist $f(X) := (f(X_n))_{n \in \mathbb{N}_0} = M + A$ die Doob-Zerlegung von $f(X)$. Speziell ist natürlich $f(X)$ ein Submartingal, wenn $f''(x) \geq 0$ für alle $x \in \mathbb{Z}$, wenn also f konvex ist. Dies wussten wir zwar schon aus Satz 9.35, allerdings haben wir hier auch noch quantifiziert, wie weit $f(X)$ von einem Martingal abweicht.

In den Spezialfällen $f(x) = x^2$ und $f(x) = |x|$ ist $f''(x) = 2$ beziehungsweise $f''(x) = 2 \cdot \mathbb{1}_{\{0\}}(x)$, und wir erhalten aus (10.4) die Aussagen von Satz 10.4 und Beispiel 10.8.

Später werden wir eine zu (10.4) vergleichbare Formel auch für stochastische Prozesse in stetiger Zeit herleiten (siehe Kapitel 25.3). \diamond

10.2 Optional Sampling und Optional Stopping

Lemma 10.10. Sei $I \subset \mathbb{R}$ höchstens abzählbar, $(X_t)_{t \in I}$ ein Martingal, $T \in I$ und τ eine Stoppzeit mit $\tau \leq T$. Dann gilt $X_\tau = \mathbf{E}[X_T | \mathcal{F}_\tau]$ und speziell $\mathbf{E}[X_\tau] = \mathbf{E}[X_0]$.

Beweis. Es reicht zu zeigen, dass $\mathbf{E}[X_T \mathbf{1}_A] = \mathbf{E}[X_\tau \mathbf{1}_A]$ für jedes $A \in \mathcal{F}_\tau$ gilt. Nach der Definition von \mathcal{F}_τ ist $\{\tau = t\} \cap A \in \mathcal{F}_t$ für jedes $t \in I$, also

$$\begin{aligned}\mathbf{E}[X_\tau \mathbf{1}_A] &= \sum_{t \leq T} \mathbf{E}[X_t \mathbf{1}_{\{\tau=t\} \cap A}] = \sum_{t \leq T} \mathbf{E}[\mathbf{E}[X_T | \mathcal{F}_t] \mathbf{1}_{\{\tau=t\} \cap A}] \\ &= \sum_{t \leq T} \mathbf{E}[X_T \mathbf{1}_A \mathbf{1}_{\{\tau=t\}}] = \mathbf{E}[X_T \mathbf{1}_A].\end{aligned}$$

□

Satz 10.11 (Optional Sampling Theorem). Sei $X = (X_n)_{n \in \mathbb{N}_0}$ ein Supermartingal, und seien $\sigma \leq \tau$ Stoppzeiten.

(i) Gibt es ein $T \in \mathbb{N}$ mit $\tau \leq T$, dann ist

$$X_\sigma \geq \mathbf{E}[X_\tau | \mathcal{F}_\sigma]$$

und speziell $\mathbf{E}[X_\sigma] \geq \mathbf{E}[X_\tau]$. Ist X ein Martingal, so gilt jeweils Gleichheit.

(ii) Ist X nichtnegativ und $\tau < \infty$ f.s., so gelten $\mathbf{E}[X_\tau] \leq \mathbf{E}[X_0] < \infty$, $\mathbf{E}[X_\sigma] \leq \mathbf{E}[X_0] < \infty$ und $X_\sigma \geq \mathbf{E}[X_\tau | \mathcal{F}_\sigma]$.

(iii) Ist allgemeiner X lediglich adaptiert und integrierbar, so ist X genau dann ein Martingal, wenn $\mathbf{E}[X_\tau] = \mathbf{E}[X_0]$ für jede beschränkte Stoppzeit τ gilt.

Beweis. (i) Sei $X = M + A$ die Doob-Zerlegung von X , also A vorhersagbar und monoton fallend, $A_0 = 0$, und M ein Martingal. Dann ist nach Lemma 10.10, angewandt auf M ,

$$\begin{aligned}X_\sigma &= A_\sigma + M_\sigma = \mathbf{E}[A_\sigma + M_T | \mathcal{F}_\sigma] \\ &\geq \mathbf{E}[A_\tau + M_T | \mathcal{F}_\sigma] = \mathbf{E}[A_\tau + \mathbf{E}[M_T | \mathcal{F}_\tau] | \mathcal{F}_\sigma] \\ &= \mathbf{E}[A_\tau + M_\tau | \mathcal{F}_\sigma] = \mathbf{E}[X_\tau | \mathcal{F}_\sigma].\end{aligned}$$

Wir haben dabei $\mathcal{F}_\tau \supset \mathcal{F}_\sigma$, die Turmeigenschaft und die Monotonie der bedingten Erwartung (Satz 8.14) ausgenutzt.

(ii) Es gilt $X_{\tau \wedge n} \xrightarrow{n \rightarrow \infty} X_\tau$ fast sicher. Nach (i) gilt $\mathbf{E}[X_{\tau \wedge n}] \leq \mathbf{E}[X_0]$ für jedes $n \in \mathbb{N}$. Nach dem Lemma von Fatou ist also

$$\mathbf{E}[X_\tau] \leq \liminf_{n \rightarrow \infty} \mathbf{E}[X_{\tau \wedge n}] \leq \mathbf{E}[X_0] < \infty.$$

Analog zeigt man $\mathbf{E}[X_\sigma] \leq \mathbf{E}[X_0]$.

Seien nun $m, n \in \mathbb{N}$ mit $m \geq n$. Teil (i) mit den beschränkten Stoppzeiten $\tau \wedge m \geq \sigma \wedge n$ liefert

$$X_{\sigma \wedge n} \geq \mathbf{E}[X_{\tau \wedge m} | \mathcal{F}_{\sigma \wedge n}].$$

Für $A \in \mathcal{F}_\sigma$ ist $\{\sigma < n\} \cap A \in \mathcal{F}_{\sigma \wedge n}$, also ist

$$\mathbf{E}[X_\sigma \mathbb{1}_{\{\sigma < n\} \cap A}] = \mathbf{E}[X_{\sigma \wedge n} \mathbb{1}_{\{\sigma < n\} \cap A}] \geq \mathbf{E}[X_{\tau \wedge m} \mathbb{1}_{\{\sigma < n\} \cap A}].$$

Das Lemma von Fatou liefert

$$\mathbf{E}[X_\tau \mathbb{1}_{\{\sigma < n\} \cap A}] \leq \liminf_{m \rightarrow \infty} \mathbf{E}[X_{\tau \wedge m} \mathbb{1}_{\{\sigma < n\} \cap A}] \leq \mathbf{E}[X_\sigma \mathbb{1}_{\{\sigma < n\} \cap A}].$$

Monotone Konvergenz (für $n \rightarrow \infty$) ergibt daher $\mathbf{E}[X_\tau \mathbb{1}_A] \leq \mathbf{E}[X_\sigma \mathbb{1}_A]$.

(iii) Ist X ein Martingal, so folgt die Aussage aus Lemma 10.10. Gelte nun $\mathbf{E}[X_\tau] = \mathbf{E}[X_0]$ für jede beschränkte Stoppzeit τ . Sei $t > s$ und $A \in \mathcal{F}_s$. Es reicht zu zeigen, dass $\mathbf{E}[X_t \mathbb{1}_A] = \mathbf{E}[X_s \mathbb{1}_A]$. Setze $\tau = s \mathbb{1}_A + t \mathbb{1}_{A^c}$. Dann ist τ eine beschränkte Stoppzeit. Nach Voraussetzung gilt aber

$$\mathbf{E}[X_t \mathbb{1}_A] = \mathbf{E}[X_t] - \mathbf{E}[X_t \mathbb{1}_{A^c}] = \mathbf{E}[X_0] - \mathbf{E}[X_\tau] + \mathbf{E}[X_s \mathbb{1}_A] = \mathbf{E}[X_s \mathbb{1}_A]. \quad \square$$

Korollar 10.12. Sei X ein Martingal (beziehungsweise Submartingal), und sei $(\tau_N)_{N \in \mathbb{N}}$ eine monoton wachsende Folge beschränkter Stoppzeiten (also $\tau_N \leq T_N$, $N \in \mathbb{N}$ für gewisse $T_N \in \mathbb{N}$). Dann ist $(X_{\tau_N})_{N \in \mathbb{N}}$ ein Martingal (beziehungsweise Submartingal) bezüglich der Filtration $(\mathcal{F}_{\tau_N})_{N \in \mathbb{N}}$.

Definition 10.13. Sei $I \subset \mathbb{R}$ höchstens abzählbar und $(X_t)_{t \in I}$ adaptiert sowie τ eine Stoppzeit. Wir definieren den **gestoppten Prozess** X^τ durch

$$X_t^\tau = X_{\tau \wedge t} \quad \text{für jedes } t \in I.$$

Ferner sei \mathbb{F}^τ die Filtration $\mathbb{F}^\tau = (\mathcal{F}_t^\tau)_{t \in I} = (\mathcal{F}_{\tau \wedge t})_{t \in I}$.

Bemerkung 10.14. X^τ ist an \mathbb{F} und an \mathbb{F}^τ adaptiert. \diamond

Satz 10.15 (Optional Stopping). Sei $(X_n)_{n \in \mathbb{N}_0}$ ein (Sub-, Super-)Martingal bezüglich \mathbb{F} und τ eine Stoppzeit. Dann ist X^τ ein (Sub-, Super-)Martingal sowohl bezüglich \mathbb{F} als auch bezüglich \mathbb{F}^τ .

Beweis. Wir führen den Beweis nur für den Fall, dass X ein Submartingal ist. Die anderen Fälle ergeben sich, weil in ihnen $(-X)$ ein Submartingal ist.

Sei also X ein Submartingal. Offenbar ist für jedes $n \in \mathbb{N}_0$

$$\mathbf{E}[|X_n^\tau|] \leq \mathbf{E}\left[\max\{|X_m| : m \leq n\}\right] \leq \mathbf{E}[|X_0|] + \dots + \mathbf{E}[|X_n|] < \infty,$$

also ist X^τ integrierbar.

Wegen $\{\tau > n - 1\} \in \mathcal{F}_{n-1}$ ist

$$\begin{aligned}\mathbf{E}[X_n^\tau - X_{n-1}^\tau \mid \mathcal{F}_{n-1}] &= \mathbf{E}[X_{\tau \wedge n} - X_{\tau \wedge (n-1)} \mid \mathcal{F}_{n-1}] \\ &= \mathbf{E}[(X_n - X_{n-1}) \mathbb{1}_{\{\tau > n-1\}} \mid \mathcal{F}_{n-1}] \\ &= \mathbb{1}_{\{\tau > n-1\}} \mathbf{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] \\ &\geq 0, \text{ da } X \text{ ein } \mathbb{F}\text{-Submartingal ist.}\end{aligned}$$

Also ist X^τ ein \mathbb{F} -Submartingal. Da X^τ an \mathbb{F}^τ adaptiert und \mathbb{F}^τ die kleinere Filtration ist, ist X^τ auch ein \mathbb{F}^τ -Submartingal (siehe Bemerkung 9.29). \square

Beispiel 10.16. Sei X die symmetrische einfache Irrfahrt aus Beispiel 10.8. Seien $a, b \in \mathbb{Z}$, $a < 0$, $b > 0$ sowie

$$\tau_a = \inf\{t \geq 0 : X_t = a\}, \quad \tau_b = \inf\{t \geq 0 : X_t = b\} \text{ und } \tau_{a,b} = \tau_a \wedge \tau_b.$$

$\tau_{a,b}$ ist eine Stoppzeit nach Lemma 9.18. Sei $A = \{\tau_{a,b} = \tau_a\}$ das Ereignis, dass X in a ist bevor es in b ist. Wir wollen $\mathbf{P}[A]$ bestimmen. Nach Übung 2.3.1 ist fast sicher $\limsup_{n \rightarrow \infty} X_n = \infty$ und $\liminf_{n \rightarrow \infty} X_n = -\infty$. Also ist fast sicher $\tau_a < \infty$ und $\tau_b < \infty$. Nach dem Optional Stopping Theorem ist $X^{\tau_{a,b}}$ ein Martingal. Wegen $\tau_{a,b} \wedge n \xrightarrow{n \rightarrow \infty} \tau_{a,b}$ fast sicher, gilt $X_n^{\tau_{a,b}} \xrightarrow{n \rightarrow \infty} X_{\tau_{a,b}}$ fast sicher. Da $|X_n^{\tau_{a,b}}|$ durch $b - a$ beschränkt ist, gilt $X_n^{\tau_{a,b}} \xrightarrow{n \rightarrow \infty} X_{\tau_{a,b}}$ auch in L^1 . Also ist

$$\begin{aligned}0 &= \lim_{n \rightarrow \infty} \mathbf{E}[X_n^{\tau_{a,b}}] = \mathbf{E}[X_{\tau_{a,b}}] = a \cdot \mathbf{P}[\tau_{a,b} = \tau_a] + b \cdot \mathbf{P}[\tau_{a,b} = \tau_b] \\ &= b + (a - b) \mathbf{P}[\tau_{a,b} = \tau_a].\end{aligned}$$

Es folgt $\mathbf{P}[\tau_{a,b} = \tau_a] = \frac{b}{b - a}$. \diamond

Beispiel 10.17. Schließlich wollen wir unsere Maschinerie benutzen, um $\mathbf{E}[\tau_{a,b}]$ und $\mathbf{E}[\tau_a]$ zu berechnen. Der quadratische Variationsprozess $\langle X \rangle$ (vergleiche Definition 10.3) ist gegeben durch

$$\langle X \rangle_n = \sum_{i=1}^n \mathbf{E}[(X_i - X_{i-1})^2 \mid \mathcal{F}_{i-1}] = n,$$

also ist $(X_n^2 - n)_{n \in \mathbb{N}_0}$ ein Martingal. Nach dem Optional Stopping Theorem ist

$$0 = \mathbf{E}[X_{\tau_{a,b} \wedge n}^2 - (\tau_{a,b} \wedge n)] \quad \text{für jedes } n \in \mathbb{N}_0.$$

Monotone Konvergenz liefert

$$\mathbf{E}[\tau_{a,b}] = \mathbf{E}[X_{\tau_{a,b}}^2] = a^2 \mathbf{P}[\tau_{a,b} = \tau_a] + b^2 \mathbf{P}[\tau_{a,b} = \tau_b] = |a| \cdot b.$$

Um $\mathbf{E}[\tau_a]$ zu berechnen, bemerken wir, dass $\tau_{a,b} \uparrow \tau_a$ fast sicher gilt, falls $b \rightarrow \infty$. Der Satz von der monotonen Konvergenz liefert also $\mathbf{E}[\tau_a] = \lim_{b \rightarrow \infty} \mathbf{E}[\tau_{a,b}] = \infty$. \diamond

Bemerkung 10.18. Offenbar ist $X_{\tau_b} = b > 0$, also $X_0 < \mathbf{E}[X_{\tau_b} | \mathcal{F}_0] = b$. Die Aussage des Optional Sampling Theorems gilt also im Allgemeinen nicht, falls die Stopzeit unbeschränkt ist. \diamond

Beispiel 10.19 (Gambler's Ruin Problem). Wir betrachten ein Spiel zwischen zwei Personen A und B . In jeder Runde wird eine Münze geworfen. Je nach Ergebnis erhält A von B eine Geldeinheit oder B von A . Gespielt wird so lange, bis einer der beiden Spieler ruiniert ist. Wir nehmen der Einfachheit halber an, dass A zum Spielbeginn $k_A \in \mathbb{N}$ Geldeinheiten hat, B hingegen $k_B = N - k_A$, wobei $N \in \mathbb{N}$, $N \geq k_A$. Gesucht ist die Ruinwahrscheinlichkeit von B . In Beispiel 10.16 haben wir für den Fall einer fairen Münze bereits ausgerechnet, dass die Ruinwahrscheinlichkeit k_A/N ist. Nun wollen wir den Fall einer unfairen Münze betrachten.

Seien also Y_1, Y_2, \dots unabhängig und $\mathbf{P}[Y_i = 1] = 1 - \mathbf{P}[Y_i = -1] = p$ für alle $i \in \mathbb{N}$ und für gewisses $p \in (0, 1) \setminus \{\frac{1}{2}\}$. Mit $X_n := k_B + \sum_{i=1}^n Y_i$ bezeichnen wir den Kontostand von B nach n Runden, wobei wir formal annehmen, dass die Spiele weiter gehen, auch wenn ein Spieler bereits ruiniert ist. Wir definieren noch wie oben τ_0, τ_N und $\tau_{0,N}$ als die ersten Eintreffzeiten von X in $\{0\}, \{N\}$ beziehungsweise $\{0, N\}$. Die Ruinwahrscheinlichkeit von B ist nun $p_B^N := \mathbf{P}[\tau_{0,N} = \tau_0]$. Da X kein Martingal ist (außer im Falle $p = \frac{1}{2}$, den wir hier ausschließen wollen), behelfen wir uns mit einem Trick: Wir definieren einen neuen Prozess Z durch $Z_n := r^{X_n} = r^{k_B} \prod_{i=1}^n r^{Y_i}$, wobei wir $r > 0$ noch geeignet wählen müssen, so dass Z ein Martingal wird. Nach Beispiel 9.31 ist dies genau dann der Fall, wenn $\mathbf{E}[r^{Y_1}] = pr + (1-p)r^{-1} = 1$ ist, also wenn $r = 1$ oder $r = \frac{1-p}{p}$ ist. Offenbar ist die Wahl $r = 1$ nutzlos, also nehmen wir $r = \frac{1-p}{p}$ an. Wir erhalten so $\tau_0 = \inf\{n \in \mathbb{N}_0 : Z_n = 1\}$ und $\tau_N = \inf\{n \in \mathbb{N}_0 : Z_n = r^N\}$.

(Man beachte, dass wir hier nicht wie oben argumentieren können, um zu zeigen, dass $\tau_0 < \infty$ und $\tau_N < \infty$ fast sicher gilt. In der Tat ist für $p \neq \frac{1}{2}$ auch stets nur genau eine der beiden Aussagen richtig. Allerdings erhält man, beispielsweise durch das starke Gesetz der großen Zahl, dass $\liminf_{n \rightarrow \infty} X_n = \infty$ (und damit $\tau_N < \infty$) fast sicher, falls $p > \frac{1}{2}$. Analog ist $\tau_0 < \infty$ fast sicher, falls $p < \frac{1}{2}$.) Wie in Beispiel 10.16 liefert der Optional Stopping Satz $r^{k_B} = Z_0 = \mathbf{E}[Z_{\tau_{0,N}}] = p_B^N + (1 - p_B^N)r^N$, also ist die Ruinwahrscheinlichkeit von B

$$p_B^N = \frac{r^{k_B} - r^N}{1 - r^N}. \quad (10.5)$$

Ist das Spiel vorteilhaft für B , also $p > \frac{1}{2}$, so ist $r < 1$, und wir erhalten im Limes $N \rightarrow \infty$ (bei konstantem k_B)

$$p_B^\infty := \lim_{N \rightarrow \infty} p_B^N = r^{k_B}. \quad (10.6)$$

\diamond

Übung 10.2.1. Sei X ein quadratisch integrierbares Martingal mit quadratischem Variationsprozess $\langle X \rangle$. Sei τ eine endliche Stoppzeit. Man zeige:

- (i) Ist $\mathbf{E}[\langle X \rangle_\tau] < \infty$, so gilt

$$\mathbf{E}[(X_\tau - X_0)^2] = \mathbf{E}[\langle X \rangle_\tau] \quad \text{und} \quad \mathbf{E}[X_\tau] = \mathbf{E}[X_0]. \quad (10.7)$$

- (ii) Ist $\mathbf{E}[\langle X \rangle_\tau] = \infty$, so braucht keine der Gleichungen in (10.7) zu gelten. ♣

Übung 10.2.2. Wir betrachten hier eine allgemeinere Situation als im vorangehenden Beispiel, indem wir nur noch annehmen, dass Y_1, Y_2, \dots u.i.v. integrierbare Zufallsvariablen sind, die nicht fast sicher konstant sind (und $X_n = Y_1 + \dots + Y_n$). Weiter nehmen wir an, dass es ein $\delta > 0$ gibt mit $\mathbf{E}[\exp(\theta Y_1)] < \infty$ für jedes $\theta \in (-\delta, \delta)$. Wir definieren eine Abbildung $\psi : (-\delta, \delta) \rightarrow \mathbb{R}$ durch $\theta \mapsto \log(\mathbf{E}[\exp(\theta Y_1)])$ und setzen $Z_n^\theta := \exp(\theta X_n - n\psi(\theta))$ für $n \in \mathbb{N}_0$. Man zeige:

- (i) Für jedes $\theta \in (-\delta, \delta)$ ist Z^θ ist ein Martingal.
- (ii) ψ ist strikt konvex.
- (iii) $\mathbf{E}[\sqrt{Z_n^\theta}] \xrightarrow{n \rightarrow \infty} 0$ für $\theta \neq 0$.
- (iv) $Z_n^\theta \xrightarrow{n \rightarrow \infty} 0$ fast sicher.

Interpretieren wir Y_n als die Differenz zwischen Prämieneinnahmen und Schadensauszahlungen einer Versicherungsgesellschaft zur Zeit n , so gibt $k_0 + X_n$ den Kontostand der Versicherung zur Zeit n wieder, wenn das Anfangskapital $k_0 > 0$ beträgt. Wir interessieren uns für die Ruinwahrscheinlichkeit

$$p(k_0) = \mathbf{P}[\inf\{X_n + k_0 : n \in \mathbb{N}_0\} < 0]$$

in Abhängigkeit vom Startkapital.

Man kann davon ausgehen, dass die Prämien so kalkuliert sind, dass $\mathbf{E}[Y_1] > 0$. Man zeige: Falls die Gleichung $\psi(\theta) = 0$ eine Lösung $\theta^* \neq 0$ hat, so ist $\theta^* < 0$. Man zeige, dass in diesem Fall die **Cramér-Lundberg'sche Ungleichung** gilt:

$$p(k_0) \leq \exp(\theta^* k_0). \quad (10.8)$$

In dem Fall, wo Y_i nur die Werte -1 und 1 annimmt und $k_0 \in \mathbb{N}$ ist, gilt Gleichheit, und wir erhalten Gleichung (10.6) mit $r = \exp(\theta^*)$. ♣

10.3 Gleichgradige Integrierbarkeit und Optional Sampling

Wir wollen jetzt das Optional Sampling Theorem auf unbeschränkte Stoppzeiten ausweiten. Dies geht, falls das zugrunde liegende Martingal gleichgradig integrierbar ist (vergleiche Definition 6.16).

Lemma 10.20. Sei X ein gleichgradig integrierbares Martingal. Dann ist die Familie $(X_\tau : \tau \text{ ist endliche Stoppzeit})$ gleichgradig integrierbar.

Beweis. Nach Satz 6.19 gibt es eine monoton wachsende, konvexe Funktion $f : [0, \infty) \rightarrow [0, \infty)$ mit $\liminf_{x \rightarrow \infty} f(x)/x = \infty$ und $L := \sup_{n \in \mathbb{N}_0} \mathbf{E}[f(|X_n|)] < \infty$. Ist $\tau < \infty$ eine endliche Stoppzeit, dann ist nach dem Optional Sampling Theorem für beschränkte Stoppzeiten (Satz 10.11 mit $\tau = n$ und $\sigma = \tau \wedge n$) $\mathbf{E}[X_n | \mathcal{F}_{\tau \wedge n}] = X_{\tau \wedge n}$. Wegen $\{\tau \leq n\} \in \mathcal{F}_{\tau \wedge n}$ liefert die Jensen'sche Ungleichung

$$\begin{aligned} \mathbf{E}[f(|X_\tau|) \mathbf{1}_{\{\tau \leq n\}}] &= \mathbf{E}[f(|X_{\tau \wedge n}|) \mathbf{1}_{\{\tau \leq n\}}] \\ &\leq \mathbf{E}[\mathbf{E}[f(|X_n|) | \mathcal{F}_{\tau \wedge n}] \mathbf{1}_{\{\tau \leq n\}}] \\ &= \mathbf{E}[f(|X_n|) \mathbf{1}_{\{\tau \leq n\}}] \leq L. \end{aligned}$$

Also ist $\mathbf{E}[f(|X_\tau|)] \leq L$. Nach Satz 6.19 ist $(X_\tau, \tau \text{ ist endliche Stoppzeit})$ gleichgradig integrierbar. \square

Satz 10.21 (Optional Sampling und gleichgradige Integrierbarkeit).

Ist $(X_n, n \in \mathbb{N}_0)$ ein gleichgradig integrierbares Martingal (beziehungsweise Supermartingal), und sind $\sigma \leq \tau$ endliche Stoppzeiten, dann gilt $\mathbf{E}[|X_\tau|] < \infty$ und $X_\sigma = \mathbf{E}[X_\tau | \mathcal{F}_\sigma]$ (beziehungsweise $X_\sigma \geq \mathbf{E}[X_\tau | \mathcal{F}_\sigma]$).

Beweis. Sei zunächst X ein Martingal. Für $F \in \mathcal{F}_\sigma$ ist $\{\sigma \leq n\} \cap F \in \mathcal{F}_{\sigma \wedge n}$, also nach dem Optional Sampling Theorem (Satz 10.11)

$$\mathbf{E}[X_{\sigma \wedge n} \mathbf{1}_{\{\sigma \leq n\} \cap F}] = \mathbf{E}[X_{\sigma \wedge n} \mathbf{1}_{\{\sigma \leq n\} \cap F}].$$

Nach Lemma 10.20 ist $(X_{\sigma \wedge n}, n \in \mathbb{N}_0)$ und damit $(X_{\sigma \wedge n} \mathbf{1}_{\{\sigma \leq n\} \cap F}, n \in \mathbb{N}_0)$ gleichgradig integrierbar. Analog gilt dies für X_τ . Nach Satz 6.25 gilt daher

$$\mathbf{E}[X_\tau \mathbf{1}_F] = \lim_{n \rightarrow \infty} \mathbf{E}[X_{\sigma \wedge n} \mathbf{1}_{\{\sigma \leq n\} \cap F}] = \lim_{n \rightarrow \infty} \mathbf{E}[X_{\sigma \wedge n} \mathbf{1}_{\{\sigma \leq n\} \cap F}] = \mathbf{E}[X_\sigma \mathbf{1}_F].$$

Es folgt $\mathbf{E}[X_\tau | \mathcal{F}_\sigma] = X_\sigma$.

Sei nun X ein Supermartingal. Dann hat X die Doob-Zerlegung $X = M + A$, wobei M ein Martingal ist und $A \leq 0$ vorhersagbar und fallend. Wegen

$$\mathbf{E}[|A_n|] = \mathbf{E}[-A_n] \leq \mathbf{E}[|X_n - X_0|] \leq \mathbf{E}[|X_0|] + \sup_{m \in \mathbb{N}_0} \mathbf{E}[|X_m|] < \infty,$$

gilt $A_n \downarrow A_\infty$ für ein $A_\infty \leq 0$ mit $\mathbf{E}[-A_\infty] < \infty$ (nach dem Satz über monotone Konvergenz). Also ist A damit auch $M = X - A$ gleichgradig integrierbar (Satz 6.18(ii)). Es folgt

$$\mathbf{E}[|X_\tau|] \leq \mathbf{E}[-A_\tau] + \mathbf{E}[|M_\tau|] \leq \mathbf{E}[-A_\infty] + \mathbf{E}[|M_\tau|] < \infty.$$

Ferner ist

$$\begin{aligned}
 \mathbf{E}[X_\tau \mid \mathcal{F}_\sigma] &= \mathbf{E}[M_\tau \mid \mathcal{F}_\sigma] + \mathbf{E}[A_\tau \mid \mathcal{F}_\sigma] \\
 &= M_\sigma + A_\sigma + \mathbf{E}[(A_\tau - A_\sigma) \mid \mathcal{F}_\sigma] \\
 &\leq M_\sigma + A_\sigma = X_\sigma. \quad \square
 \end{aligned}$$

Korollar 10.22. *Ist X ein gleichgradig integrierbares Martingal (beziehungsweise Supermartingal), und sind $\tau_1 \leq \tau_2 \leq \dots$ endliche Stopzzeiten, so ist $(X_{\tau_n})_{n \in \mathbb{N}}$ ein Martingal (beziehungsweise Supermartingal).*

11 Martingalkonvergenzsätze und Anwendungen

Wir haben Martingale $X = (X_n)_{n \in \mathbb{N}_0}$ als faire Spiele kennen gelernt und festgestellt, dass sie unter gewissen Transformationen (Optionales Stoppen, diskretes stochastisches Integral) wieder zu Martingalen werden. In diesem Kapitel werden wir sehen, dass unter schwachen Bedingungen (Nichtnegativität oder gleichgradige Integrierbarkeit) Martingale fast sicher konvergieren. Zudem impliziert die Martingalstruktur die L^p -Konvergenz schon unter formal schwächeren Annahmen als denen, die wir in Kapitel 7 gesehen haben. Die grundlegenden Ideen dieses Kapitels liegen in der Doob'schen Ungleichung (Satz 11.2) und in der Aufkreuzungsungleichung (Lemma 11.3).

11.1 Die Doob'sche Ungleichung

Wir haben mit der Kolmogorov'schen Ungleichung (Satz 5.28) eine Ungleichung kennen gelernt, die analog zur Chebyshev'schen Ungleichung die Wahrscheinlichkeit für große Werte des Maximums eines quadratisch integrierbaren Prozesses mit unabhängigen, zentrierten Zuwächsen nach oben abschätzt. An dieser Stelle wollen wir die Ungleichung in mehrere Richtungen verbessern. Einerseits wollen wir Martingale betrachten. Andererseits wollen wir mit weniger als zweiten Momenten auskommen, beziehungsweise bei Anwesenheit höherer Momente die Ungleichung verschärfen.

Sei $I \subset \mathbb{N}_0$ und $X = (X_n)_{n \in I}$ ein stochastischer Prozess. Wir schreiben für $n \in \mathbb{N}$

$$X_n^* = \sup\{X_k : k \leq n\} \quad \text{und} \quad |X|_n^* = \sup\{|X_k| : k \leq n\}.$$

Lemma 11.1. *Ist X ein Submartingal, dann gilt für jedes $\lambda > 0$*

$$\lambda \mathbf{P}[X_n^* \geq \lambda] \leq \mathbf{E}[X_n \mathbf{1}_{\{X_n^* \geq \lambda\}}] \leq \mathbf{E}[|X_n| \mathbf{1}_{\{X_n^* \geq \lambda\}}].$$

Beweis. Die zweite Ungleichung ist trivial. Für die erste betrachte

$$\tau := \inf \{k \in I : X_k \geq \lambda\} \wedge n.$$

Nach Satz 10.11 (Optional Sampling Theorem) ist

$$\begin{aligned}\mathbf{E}[X_n] &\geq \mathbf{E}[X_\tau] = \mathbf{E}[X_\tau \mathbb{1}_{\{X_n^* \geq \lambda\}}] + \mathbf{E}[X_\tau \mathbb{1}_{\{X_n^* < \lambda\}}] \\ &\geq \lambda \mathbf{P}[X_n^* \geq \lambda] + \mathbf{E}[X_n \mathbb{1}_{\{X_n^* < \lambda\}}].\end{aligned}$$

(Merke: $\tau = n$, falls $X_n^* < \lambda$.) Jetzt subtrahiere $\mathbf{E}[X_n \mathbb{1}_{\{X_n^* < \lambda\}}]$. \square

Satz 11.2 (Doob'sche L^p -Ungleichung). Sei X ein Martingal oder ein positives Submartingal.

- (i) Für jedes $p \geq 1$ und $\lambda > 0$ gilt $\lambda^p \mathbf{P}[|X|_n^* \geq \lambda] \leq \mathbf{E}[|X_n|^p]$.
- (ii) Für jedes $p > 1$ gilt

$$\mathbf{E}[|X_n|^p] \leq \mathbf{E}[(|X|_n^*)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbf{E}[|X_n|^p].$$

Beweis. Der Beweis folgt [154].

(i) Nach Satz 9.35 ist $(|X_n|^p)_{n \in I}$ ein Submartingal, und die Behauptung folgt aus Lemma 11.1 (mit λ^p und $|X|^p$ statt λ und X).

(ii) Die erste Ungleichung ist trivial. Die zweite Ungleichung ist trivial, falls $\mathbf{E}[|X_n|^p] = \infty$. Sei also $\mathbf{E}[|X_n|^p] < \infty$. Man beachte, dass nach Lemma 11.1 gilt

$$\lambda \mathbf{P}[|X|_n^* \geq \lambda] \leq \mathbf{E}[|X_n| \mathbb{1}_{\{|X|_n^* \geq \lambda\}}].$$

Also ist für jedes $K > 0$

$$\begin{aligned}\mathbf{E}[(|X|_n^* \wedge K)^p] &= \mathbf{E}\left[\int_0^{|X|_n^* \wedge K} p \lambda^{p-1} d\lambda\right] \\ &= \mathbf{E}\left[\int_0^K p \lambda^{p-1} \mathbb{1}_{\{|X|_n^* \geq \lambda\}} d\lambda\right] \\ &= \int_0^K p \lambda^{p-1} \mathbf{P}[|X|_n^* \geq \lambda] d\lambda \\ &\leq \int_0^K p \lambda^{p-2} \mathbf{E}[|X_n| \mathbb{1}_{\{|X|_n^* \geq \lambda\}}] d\lambda \\ &= p \mathbf{E}\left[|X_n| \int_0^{|X|_n^* \wedge K} \lambda^{p-2} d\lambda\right] = \frac{p}{p-1} \mathbf{E}[|X_n| \cdot (|X|_n^* \wedge K)^{p-1}].\end{aligned}$$

Die Hölder'sche Ungleichung liefert nun

$$\mathbf{E}[(|X|_n^* \wedge K)^p] \leq \frac{p}{p-1} \mathbf{E}[(|X|_n^* \wedge K)^p]^{(p-1)/p} \cdot \mathbf{E}[|X_n|^p]^{1/p}.$$

Indem wir beide Seiten zur p -ten Potenz erheben und durch $\mathbf{E}[(|X|_n^* \wedge K)^p]^{p-1}$ teilen (hier wird das Abschneiden bei K gebraucht, damit wir sicher durch eine endliche Zahl teilen), erhalten wir

$$\mathbf{E}[(|X|_n^* \wedge K)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbf{E}[|X_n|^p].$$

Jetzt lassen wir K nach ∞ gehen. \square

Übung 11.1.1. Sei $(X_n)_{n \in \mathbb{N}_0}$ ein Submartingal oder Supermartingal. Man zeige mit Hilfe von Satz 11.2 und der Doob-Zerlegung, dass für jedes $n \in \mathbb{N}$ und $\lambda > 0$

$$\lambda \mathbf{P}[|X|_n^* \geq \lambda] \leq 12 \mathbf{E}[|X_0|] + 9 \mathbf{E}[|X_n|].$$



11.2 Martingalkonvergenzsätze

In diesem Abschnitt zeigen wir die gängigen Martingalkonvergenzsätze und geben ein paar kleinere Beispiele an. Wir beginnen mit dem Herzstück der Martingalkonvergenzsätze, der Aufkreuzungsungleichung.

Sei $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}_0}$ eine Filtration und $\mathcal{F}_\infty = \sigma(\bigcup_{n \in \mathbb{N}_0} \mathcal{F}_n)$. Sei $(X_n)_{n \in \mathbb{N}_0}$ reellwertig und an \mathbb{F} adaptiert. Seien $a, b \in \mathbb{R}$ mit $a < b$. Stellen wir uns X als Aktienkurs vor, so wäre es eine sinnvolle Handelsstrategie, immer dann die Aktie zu kaufen, wenn ihr Kurs unter a gefallen ist, und zu verkaufen, sobald der Kurs wieder über b gestiegen ist – jedenfalls dann, wenn wir sicher wissen, dass der Kurs immer wieder über b steigt. Jedes Mal, wenn der Kurs der Aktie eine solche *Aufkreuzung* von a nach b vollführt, erhalten wir einen Gewinn von mindestens $b - a$. Indem wir den maximal möglichen Gewinn nach oben abschätzen, erhalten wir eine obere Abschätzung für die Zahl der Aufkreuzungen. Ist diese aber endlich für je zwei Werte $a < b$, so muss der Prozess fast sicher konvergieren.

Wir kommen zu den technischen Details. Wir definieren Stoppzeiten $\sigma_0 \equiv 0$ und

$$\begin{aligned} \tau_k &:= \inf\{n \geq \sigma_{k-1} : X_n \leq a\} \quad \text{für } k \in \mathbb{N}, \\ \sigma_k &:= \inf\{n \geq \tau_k : X_n \geq b\} \quad \text{für } k \in \mathbb{N}. \end{aligned}$$

Man beachte, dass $\tau_k = \infty$ gilt, falls $\sigma_{k-1} = \infty$ ist, und $\sigma_k = \infty$, falls $\tau_k = \infty$. Wir sagen, dass X zwischen τ_k und σ_k die k -te **Aufkreuzung** über $[a, b]$ hat, falls $\sigma_k < \infty$. Für $n \in \mathbb{N}$ definieren wir

$$U_n^{a,b} := \sup\{k \in \mathbb{N}_0 : \sigma_k \leq n\}$$

als Anzahl der Aufkreuzungen bis zur Zeit n . (Man beachte, dass $U_n^{a,b}$ messbar bezüglich \mathcal{F}_∞ ist, denn für jedes $k \in \mathbb{N}$ ist $\{U_n^{a,b} \geq k\} = \{\sigma_k \leq n\}$.)

Lemma 11.3 (Aufkreuzungsungleichung). Es sei $(X_n)_{n \in \mathbb{N}_0}$ ein Submartingal. Dann ist

$$\mathbf{E}[U_n^{a,b}] \leq \frac{\mathbf{E}[(X_n - a)^+] - \mathbf{E}[(X_0 - a)^+]}{b - a}.$$

Beweis. Wir erinnern an das diskrete stochastische Integral (Definition 9.37) $H \cdot X$ und beschreiben formal die oben angedeutete Handelsstrategie H für $m \in \mathbb{N}_0$ durch

$$H_m := \begin{cases} 1, & \text{falls } m \in \{\tau_k + 1, \dots, \sigma_k\} \text{ für ein } k \in \mathbb{N}, \\ 0, & \text{sonst.} \end{cases}$$

H ist nichtnegativ und vorhersagbar, denn für $m \in \mathbb{N}$ ist

$$\{H_m = 1\} = \bigcup_{k=1}^{\infty} (\{\tau_k \leq m - 1\} \cap \{\sigma_k > m - 1\}),$$

und jedes der Ereignisse liegt in \mathcal{F}_{m-1} . Setze $Y = \max(X, a)$. Ist $k \in \mathbb{N}$ und $\sigma_k < \infty$, so ist offenbar $Y_{\sigma_i} - Y_{\tau_i} = Y_{\sigma_i} - a \geq b - a$ für jedes $i \leq k$, also ist

$$(H \cdot Y)_{\sigma_k} = \sum_{i=1}^k \sum_{j=\tau_i+1}^{\sigma_i} (Y_j - Y_{j-1}) = \sum_{i=1}^k (Y_{\sigma_i} - Y_{\tau_i}) \geq k(b - a).$$

Für $j \in \{\sigma_k, \dots, \tau_{k+1}\}$ ist $(H \cdot Y)_j = (H \cdot Y)_{\sigma_k}$, und für $j \in \{\tau_k + 1, \dots, \sigma_k\}$ ist $(H \cdot Y)_j \geq (H \cdot Y)_{\tau_k} = (H \cdot Y)_{\sigma_{k-1}}$. Für $n \in \mathbb{N}$ ist daher $(H \cdot Y)_n \geq (b - a)U_n^{a,b}$.

Nach Korollar 9.34 ist Y ein Submartingal, und damit (nach Satz 9.39) auch $H \cdot Y$ und $(1 - H) \cdot Y$. Nun ist $Y_n - Y_0 = (1 \cdot Y)_n = (H \cdot Y)_n + ((1 - H) \cdot Y)_n$, also

$$\mathbf{E}[Y_n - Y_0] \geq \mathbf{E}[(H \cdot Y)_n] \geq (b - a)\mathbf{E}[U_n^{a,b}]. \quad \square$$

Satz 11.4 (Martingalkonvergenzsatz).

Sei $(X_n)_{n \in \mathbb{N}_0}$ ein Submartingal mit $\sup\{\mathbf{E}[X_n^+]: n \geq 0\} < \infty$. Dann existiert eine \mathcal{F}_∞ -messbare Zufallsvariable X_∞ mit $\mathbf{E}[|X_\infty|] < \infty$ und $X_n \xrightarrow{n \rightarrow \infty} X_\infty$ fast sicher.

Beweis. Für $a < b$ ist (wegen $\mathbf{E}[(X_n - a)^+] \leq |a| + \mathbf{E}[X_n^+]$) nach Lemma 11.3

$$\mathbf{E}[U_n^{a,b}] \leq \frac{|a| + \mathbf{E}[X_n^+]}{b - a}.$$

Offensichtlich existiert der monotone Limes $U^{a,b} := \lim_{n \rightarrow \infty} U_n^{a,b}$. Es gilt nach Voraussetzung $\mathbf{E}[U^{a,b}] = \lim_{n \rightarrow \infty} \mathbf{E}[U_n^{a,b}] < \infty$. Speziell ist $\mathbf{P}[U^{a,b} < \infty] = 1$. Definiere die \mathcal{F}_∞ -messbaren Ereignisse

$$C^{a,b} = \left\{ \liminf_{n \rightarrow \infty} X_n < a \right\} \cap \left\{ \limsup_{n \rightarrow \infty} X_n > b \right\} \subset \{U^{a,b} = \infty\}$$

und

$$C = \bigcup_{\substack{a,b \in \mathbb{Q} \\ a < b}} C^{a,b}.$$

Dann gilt $\mathbf{P}[C^{a,b}] = 0$ und damit auch $\mathbf{P}[C] = 0$. Per Konstruktion ist $(X_n)_{n \in \mathbb{N}}$ jedoch auf C^c konvergent. Also existiert der fast sichere Limes $X_\infty = \lim_{n \rightarrow \infty} X_n$. Jedes der X_n ist \mathcal{F}_∞ -messbar, also ist X_∞ auch \mathcal{F}_∞ -messbar.

Nach dem Lemma von Fatou ist

$$\mathbf{E}[X_\infty^+] \leq \sup \{ \mathbf{E}[X_n^+] : n \geq 0 \} < \infty.$$

Andererseits ist (weil X Submartingal ist), wieder mit Fatous Lemma

$$\begin{aligned} \mathbf{E}[X_\infty^-] &\leq \liminf_{n \rightarrow \infty} \mathbf{E}[X_n^-] = \liminf_{n \rightarrow \infty} (\mathbf{E}[X_n^+] - \mathbf{E}[X_n]) \\ &\leq \sup \{ \mathbf{E}[X_n^+] : n \in \mathbb{N}_0 \} - \mathbf{E}[X_0] < \infty. \end{aligned}$$

□

Korollar 11.5. Ist X ein nichtnegatives Supermartingal, so gibt es eine \mathcal{F}_∞ -messbare Zufallsvariable $X_\infty \geq 0$ mit $\mathbf{E}[X_\infty] \leq \mathbf{E}[X_0]$ und $X_n \xrightarrow{n \rightarrow \infty} X_\infty$ f.s.

Beweis. Der vorige Satz mit $(-X)$ liefert X_∞ als fast sicheren Limes. Das Lemma von Fatou liefert

$$\mathbf{E}[X_\infty] \leq \liminf_{n \rightarrow \infty} \mathbf{E}[X_n] \leq \mathbf{E}[X_0].$$

□

Beispiel 11.6. Sei S_n der Kontostand beim Petersburger Spiel nach der n -ten Runde (siehe Beispiel 9.40). Dann ist S ein Martingal und $S_n \leq 1$ fast sicher für jedes n . Also sind die Voraussetzungen von Satz 11.4 erfüllt und $(S_n)_{n \in \mathbb{N}_0}$ konvergiert für $n \rightarrow \infty$ fast sicher gegen einen endlichen (zufälligen) Grenzwert. Da sich der Kontostand noch ändert, so lange Geld eingesetzt wird, also so lange $S_n < 1$ ist, gilt $\lim_{n \rightarrow \infty} S_n = 1$ fast sicher.

Da $\mathbf{E}[S_n] = 0$ ist für jedes $n \in \mathbb{N}_0$, kann die Konvergenz nicht in L^1 gelten. Dies geht einher mit der Beobachtung, dass S nicht gleichgradig integrierbar ist. ◇

Für gleichgradig integrierbare Martingale gilt ein stärkerer Konvergenzsatz.

Satz 11.7 (Konvergenzsatz für gleichgradig integrierbare Martingale).

Ist $(X_n)_{n \in \mathbb{N}_0}$ ein gleichgradig integrierbares \mathbb{F} - (Sub-, Super-) Martingal, dann existiert eine \mathcal{F}_∞ -messbare Zufallsvariable X_∞ mit $X_n \xrightarrow{n \rightarrow \infty} X_\infty$ f.s. und in L^1 . Weiter gilt, dass

- $X_n = \mathbf{E}[X_\infty | \mathcal{F}_n]$ für jedes $n \in \mathbb{N}$, falls X ein Martingal ist,
- $X_n \leq \mathbf{E}[X_\infty | \mathcal{F}_n]$ für jedes $n \in \mathbb{N}$, falls X ein Submartingal ist,
- $X_n \geq \mathbf{E}[X_\infty | \mathcal{F}_n]$ für jedes $n \in \mathbb{N}$, falls X ein Supermartingal ist.

Bemerkung 11.8. Die Aussage von Satz 11.7 lässt sich so ausdrücken: Der Prozess $(X_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}$ ist ein (Sub-, Super-) Martingal bezüglich $(\mathcal{F}_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}$. \diamond

Beweis. Wir führen den Beweis für den Fall, wo X ein Submartingal ist. Gleichgradige Integrierbarkeit impliziert $\sup\{\mathbf{E}[X_n^+] : n \geq 0\} < \infty$. Nach Satz 11.4 existiert der fast sichere Limes X_∞ . Nach Satz 6.25 gilt dann $\mathbf{E}[|X_n - X_\infty|] \xrightarrow{n \rightarrow \infty} 0$. Nach Korollar 8.21 impliziert diese L^1 -Konvergenz schon die L^1 -Konvergenz der bedingten Erwartungen $\mathbf{E}[|\mathbf{E}[X_n | \mathcal{F}_m] - \mathbf{E}[X_\infty | \mathcal{F}_m]|] \xrightarrow{n \rightarrow \infty} 0$. Die Dreiecksungleichung liefert also

$$\begin{aligned} & \left| \mathbf{E}[(\mathbf{E}[X_\infty | \mathcal{F}_m] - X_m)^-] - \mathbf{E}[(\mathbf{E}[X_n | \mathcal{F}_m] - X_m)^-] \right| \\ & \leq \mathbf{E}\left[|\mathbf{E}[X_\infty | \mathcal{F}_m] - \mathbf{E}[X_n | \mathcal{F}_m]|\right] \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Da X ein Submartingal ist, gilt $(\mathbf{E}[X_n | \mathcal{F}_m] - X_m)^- = 0$ für $n \geq m$. Es folgt also $\mathbf{E}[(\mathbf{E}[X_\infty | \mathcal{F}_m] - X_m)^-] = 0$ und damit $\mathbf{E}[X_\infty | \mathcal{F}_m] - X_m \geq 0$ fast sicher. \square

Korollar 11.9. Sei $X \geq 0$ ein Martingal und $X_\infty = \lim_{n \rightarrow \infty} X_n$. Genau dann ist $\mathbf{E}[X_\infty] = \mathbf{E}[X_0]$, wenn X gleichgradig integrierbar ist.

Beweis. Das folgt direkt aus Satz 6.25. \square

Sei $p \in [1, \infty)$. Ein reellwertiger stochastischer Prozess $(X_i)_{i \in I}$ heißt L^p -beschränkt, falls $\sup_{i \in I} \mathbf{E}[|X_i|^p] < \infty$ (Definition 6.20). Im Allgemeinen folgt aus der L^p -Beschränktheit eines Prozesses $(X_i)_{i \in I}$ nicht, dass $(|X_i|^p)_{i \in I}$ gleichgradig integrierbar wäre. Ist X ein Martingal und $p > 1$, so folgt dies jedoch aus der Doob'schen Ungleichung. Insbesondere folgt dann aus fast sicherer Konvergenz schon die Konvergenz in L^p .

Satz 11.10 (L^p -Konvergenzsatz für Martingale).

Sei $p > 1$ und $(X_n)_{n \in \mathbb{N}_0}$ ein L^p -beschränktes Martingal. Dann existiert eine \mathcal{F}_∞ -messbare Zufallsvariable X_∞ mit $\mathbf{E}[|X_\infty|^p] < \infty$ sowie $X_n \xrightarrow{n \rightarrow \infty} X_\infty$ fast sicher und in L^p . Speziell ist $(|X_n|^p)_{n \in \mathbb{N}_0}$ gleichgradig integrierbar.

Beweis. Nach Korollar 6.21 ist X gleichgradig integrierbar. Also existiert der fast sichere Limes X_∞ . Nach der Doob'schen Ungleichung (Satz 11.2) ist für $n \in \mathbb{N}$

$$\mathbf{E}\left[\sup\{|X_k|^p : k \leq n\}\right] \leq \left(\frac{p}{p-1}\right)^p \mathbf{E}[|X_n|^p].$$

Also ist

$$\mathbf{E}\left[\sup\{|X_k|^p : k \in \mathbb{N}_0\}\right] \leq \left(\frac{p}{p-1}\right)^p \sup\{\mathbf{E}[|X_n|^p] : n \in \mathbb{N}_0\} < \infty.$$

Insbesondere ist $(|X_n|^p)_{n \in \mathbb{N}_0}$ also gleichgradig integrierbar.

Majorisierte Konvergenz liefert (wegen $|X_n - X_\infty|^p \leq 2^p \sup\{|X_n|^p : n \in \mathbb{N}_0\}$)

$$\mathbf{E}[|X_\infty|^p] < \infty \quad \text{und} \quad \mathbf{E}[|X_n - X_\infty|^p] \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Für den Fall quadratintegrierbarer Martingale gibt es ein handliches Kriterium für die L^2 -Beschränktheit, das wir hier als Korollar festhalten (siehe Definition 10.3).

Korollar 11.11. *Sei X ein quadratintegrierbares Martingal mit quadratischem Variationsprozess $\langle X \rangle$. Dann sind folgende vier Aussagen äquivalent:*

- (i) $\sup_{n \in \mathbb{N}} \mathbf{E}[X_n^2] < \infty$,
- (ii) $\lim_{n \rightarrow \infty} \mathbf{E}[\langle X \rangle_n] < \infty$,
- (iii) X konvergiert in L^2 ,
- (iv) X konvergiert fast sicher und in L^2 .

Beweis. „(i) \iff (ii)“ Wegen $\mathbf{Var}[X_n - X_0] = \mathbf{E}[\langle X \rangle_n]$ (siehe Satz 10.4) ist X genau dann in L^2 beschränkt, wenn (ii) gilt.

„(iv) \implies (iii) \implies (i)“ Dies ist trivial.

„(i) \implies (iv)“ Dies ist die Aussage von Satz 11.10. \square

Bemerkung 11.12. Die Aussage von Satz 11.10 ist für $p = 1$ im Allgemeinen falsch. Siehe Übung 11.2.1. \diamond

Lemma 11.13. *Sei X ein quadratintegrierbares Martingal mit quadratischem Variationsprozess $\langle X \rangle$, und sei τ eine Stoppzeit. Dann hat der gestoppte Prozess X^τ den quadratischen Variationsprozess $\langle X^\tau \rangle = \langle X \rangle^\tau := (\langle X \rangle_{\tau \wedge n})_{n \in \mathbb{N}_0}$.*

Beweis. Übung! \square

Nehmen wir statt wie in Korollar 11.11 nicht die Beschränktheit der *Erwartungswerte* der quadratischen Variation an, sondern lediglich die *fast sichere* Beschränktheit, so erhalten wir immerhin noch fast sichere Konvergenz von X , im Allgemeinen nicht jedoch L^2 -Konvergenz.

Satz 11.14. *Sei X ein quadratintegrierbares Martingal mit $\sup_{n \in \mathbb{N}} \langle X \rangle_n < \infty$ fast sicher. Dann konvergiert X fast sicher.*

Beweis. Ohne Einschränkung können wir annehmen, dass $X_0 = 0$ ist, sonst betrachten wir das Martingal $(X_n - X_0)_{n \in \mathbb{N}_0}$, das den selben quadratischen Variationsprozess hat. Betrachte für $K > 0$

$$\tau_K := \inf\{n \in \mathbb{N} : \langle X \rangle_{n+1} \geq K\}.$$

Dies ist eine Stoppzeit, da $\langle X \rangle$ vorhersagbar ist. Offenbar ist $\sup_{n \in \mathbb{N}} \langle X \rangle_{\tau_K \wedge n} \leq K$ fast sicher. Nach Korollar 11.11 konvergiert der gestoppte Prozess X^{τ_K} fast sicher (und in L^2) gegen eine Zufallsvariable, die wir $X_\infty^{\tau_K}$ nennen wollen. Nach Voraussetzung gilt $\mathbf{P}[\tau_K = \infty] \rightarrow 1$ für $K \rightarrow \infty$, also konvergiert X fast sicher. \square

Beispiel 11.15. Sei X die symmetrische einfache Irrfahrt auf \mathbb{Z} , das heißt $X_n = \sum_{k=1}^n R_k$, wobei R_1, R_2, \dots u.i.v. sind mit $\mathbf{P}[R_1 = 1] = 1 - \mathbf{P}[R_1 = -1] = \frac{1}{2}$. Dann ist X ein Martingal, jedoch ist $\limsup_{n \rightarrow \infty} X_n = \infty$ und $\liminf_{n \rightarrow \infty} X_n = -\infty$, also X nicht einmal uneigentlich konvergent. Dies geht nach dem Martingalkonvergenzsatz einher damit, dass X nicht gleichgradig integrierbar ist. \diamond

Beispiel 11.16 (Wählermodell). Wir betrachten ein einfaches Modell zum Verhalten von opportunistischen Wählern, die nur einer von zwei Meinungen (zu einem politischen Thema) fähig sind, sagen wir 0 und 1. Wir betrachten eine Menge $\Lambda \subset \mathbb{Z}^d$, die wir als die Menge der Orte auffassen, an denen je ein Individuum sitzt. Zur Einfachheit nehmen wir an, dass $\Lambda = \{0, \dots, L-1\}^d$ für ein $L \in \mathbb{N}$. Mit $x \in \{0, 1\}^\Lambda$ bezeichnen wir einen generischen Zustand des gesamten Wahlvolkes, wobei $x(i) \in \{0, 1\}$ die Meinung des Individuums $i \in \Lambda$ ist. Wir stellen uns nun vor, dass sich die Meinungen in diskreten Zeitschritten ändern. Zu jedem Zeitpunkt n vergisst ein zufällig gewähltes Individuum I_n seine Meinung und übernimmt stattdessen die Meinung eines zufällig gewählten Nachbarn $I_n + N_n$, wobei wir die Addition als komponentenweise modulo L verstehen (manchmal wird dies auch *periodische Randbedingungen* genannt, weil wir Λ als diskreten Torus auffassen). So erhalten wir eine zufällige Folge $(X_n)_{n \in \mathbb{N}_0}$ von Zuständen in $\{0, 1\}^\Lambda$, die die zufällige Entwicklung der Meinungen darstellt.

Um noch einmal formal zu fassen, was wir gerade beschrieben haben: Wir betrachten unabhängige Zufallsvariablen $(I_n)_{n \in \mathbb{N}}$ und $(N_n)_{n \in \mathbb{N}}$. Für jedes $n \in \mathbb{N}$ sei I_n uniform verteilt auf Λ und N_n uniform verteilt auf den $2d$ direkten Nachbarn des Ursprungs $\mathcal{N} := \{i \in \mathbb{Z}^d : \|i\|_2 = 1\}$. Zudem ist $x = X_0 \in \{0, 1\}^\Lambda$ der ursprüngliche Zustand. Die weiteren Zustände definieren wir induktiv durch

$$X_n(i) = \begin{cases} X_{n-1}(i), & \text{falls } I_n \neq i, \\ X_{n-1}(I_n + N_n), & \text{falls } I_n = i. \end{cases}$$

Wir interessieren uns jetzt für das Langzeitverhalten von $(X_n)_{n \in \mathbb{N}_0}$. Wird es auf lange Sicht einen Konsens unter allen Individuen geben, oder können zwei konkurrierende Meinungen koexistieren? Wir betrachten dazu die Gesamtzahl aller Individuen mit Meinung 1 zur Zeit n , nämlich $M_n := \sum_{i \in \Lambda} X_n(i)$. Sei \mathbb{F} die Filtration

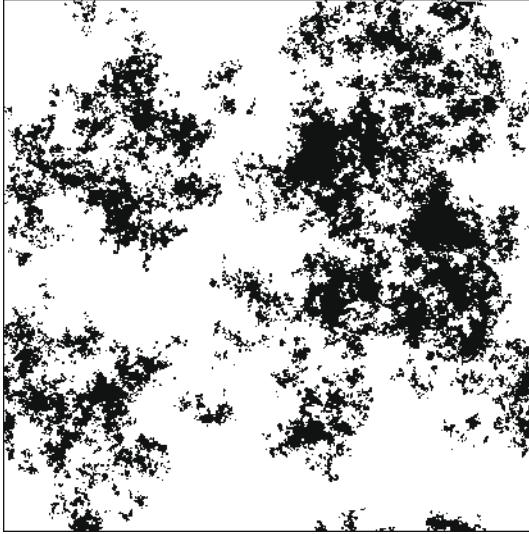


Abb. 11.1. Schnappschuss eines Wählermodells auf einem 800×800 Torus. Die schwarzen Punkte sind die Einen.

$\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}_0}$, wobei $\mathcal{F}_n = \sigma(I_k, N_k : k \leq n)$ ist für jedes $n \in \mathbb{N}_0$. Dann ist M an \mathbb{F} adaptiert und

$$\begin{aligned}\mathbf{E}[M_n | \mathcal{F}_{n-1}] &= M_{n-1} - \mathbf{E}[X_{n-1}(I_n) | \mathcal{F}_{n-1}] + \mathbf{E}[X_{n-1}(I_n + N_n) | \mathcal{F}_{n-1}] \\ &= M_{n-1} - \sum_{i \in \Lambda} \mathbf{P}[I_n = i] X_{n-1}(i) + \sum_{i \in \Lambda} \mathbf{P}[I_n + N_n = i] X_{n-1}(i) \\ &= M_{n-1},\end{aligned}$$

weil $\mathbf{P}[I_n = i] = \mathbf{P}[I_n + N_n = i] = L^{-d}$ für jedes $i \in \Lambda$. Also ist M ein beschränktes \mathbb{F} -Martingal und damit fast sicher und in L^1 konvergent gegen eine Zufallsvariable M_∞ . Da M nur ganzzahlige Werte annimmt, gibt es ein (zufälliges) n_0 , sodass $M_n = M_{n_0}$ für jedes $n \geq n_0$. Damit ist aber auch $X_n = X_{n_0}$ für jedes $n \geq n_0$. Offenbar ist jedoch kein Zustand x mit $x \not\equiv 0$ und $x \not\equiv 1$ stabil, denn hier gilt, falls i und j in Λ benachbart sind und $x(i) \neq x(j)$,

$$\mathbf{P}[X_n \neq X_{n-1} | X_{n-1} = x] \geq \mathbf{P}[I_{n-1} = i, N_{n-1} = j - i] = L^{-d}(2d)^{-1}.$$

Es muss also $M_\infty \in \{0, L^d\}$ gelten. Nun ist aber $\mathbf{E}[M_\infty] = M_0$, also gilt

$$\mathbf{P}[M_\infty = L^d] = \frac{M_0}{L^d} \quad \text{und} \quad \mathbf{P}[M_\infty = 0] = 1 - \frac{M_0}{L^d}.$$

Etwas formaler sehen wir den Sachverhalt, dass nur die beiden extremen Zustände stabil sind, so ein: Wir betrachten den quadratischen Variationsprozess $\langle M \rangle$ von M .

Dann ist

$$\begin{aligned}\langle M \rangle_n &= \sum_{k=1}^n \mathbf{E}[\mathbb{1}_{\{M_k \neq M_{k-1}\}} | \mathcal{F}_{k-1}] \\ &= \sum_{k=1}^n \mathbf{P}[X_{k-1}(I_k) \neq X_{k-1}(I_k + N_k) | \mathcal{F}_{k-1}].\end{aligned}$$

Also ist

$$\begin{aligned}L^{2d} &\geq \mathbf{Var}[M_n] = \mathbf{E}[\langle M \rangle_n] \\ &= \sum_{k=1}^n \mathbf{P}[X_{k-1}(I_k) \neq X_{k-1}(I_k + N_k)] \\ &\geq (2d)^{-1} L^{-d} \sum_{k=1}^n \mathbf{P}[M_{k-1} \notin \{0, L^d\}].\end{aligned}$$

Es folgt, dass $\sum_{k=1}^{\infty} \mathbf{P}[M_{k-1} \notin \{0, L^d\}] \leq 2dL^{3d} < \infty$, also ist nach dem Lemma von Borel-Cantelli $M_{\infty} \in \{0, L^d\}$. \diamond

Beispiel 11.17 (Satz von Radon-Nikodym). Wir wollen mit Hilfe des Martingalkonvergenzsatzes einen alternativen Beweis des Satzes von Radon-Nikodym (Korollar 7.34) angeben.

Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein W-Raum und Q ein weiteres W-Maß auf (Ω, \mathcal{A}) . Wir nehmen zudem an, dass \mathcal{F} abzählbar erzeugt ist, dass es also (höchstens) abzählbar viele Mengen $A_1, A_2, \dots \in \mathcal{F}$ gibt, sodass $\mathcal{F} = \sigma(\{A_1, A_2, \dots\})$. Dies ist beispielsweise dann richtig, wenn \mathcal{F} die Borel'sche σ -Algebra auf einem polnischen Raum ist. Speziell können wir für den Fall $\Omega = \mathbb{R}^d$ offene Kugeln mit rationalen Radien und rationalen Zentren nehmen (vergleiche Bemerkung 1.24).

Wir bilden nun eine Filtration $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$, indem wir $\mathcal{F}_n := \sigma(\{A_1, \dots, A_n\})$ setzen. Offenbar ist $\#\mathcal{F}_n < \infty$ für jedes $n \in \mathbb{N}$. Genauer gilt, dass es eine endliche (eindeutig bestimmte) Teilmenge $Z_n \subset \mathcal{F}_n \setminus \{\emptyset\}$ gibt mit

$$B = \bigcup_{C \in Z_n, C \subset B} C \quad \text{für jedes } B \in \mathcal{F}_n.$$

Z_n ist die Zerlegung von \mathcal{F}_n in „Atome“. Definiere $X = (X_n)_{n \in \mathbb{N}}$ durch

$$X_n := \sum_{C \in Z_n: \mathbf{P}[C] > 0} \frac{Q(C)}{\mathbf{P}[C]} \mathbb{1}_C.$$

Offenbar ist X an \mathbb{F} adaptiert. Sei $B \in \mathcal{F}_n$ und $m \geq n$. Für jedes $C \in Z_m$ gilt entweder $C \cap B = \emptyset$ oder $C \subset B$. Also ist

$$\mathbf{E}[X_m \mathbb{1}_B] = \sum_{C \in Z_m: \mathbf{P}[C] > 0} \frac{Q(C)}{\mathbf{P}[C]} \mathbf{P}[C \cap B] = \sum_{C \in Z_m: C \subset B} Q(C) = Q(B). \quad (11.1)$$

Insbesondere ist X also ein \mathbb{F} -Martingal.

Wir nehmen nun an, dass Q absolutstetig bezüglich P ist. Nach Beispiel 7.39 ist X dann gleichgradig integrierbar. Nach dem Martingalkonvergenzsatz konvergiert X fast sicher und in L^1 gegen eine Zufallsvariable X_∞ . Nach (11.1) ist $\mathbf{E}[X_\infty \mathbf{1}_B] = Q(B)$ für jedes $B \in \bigcup_{n \in \mathbb{N}} \mathcal{F}_n$, also auch für jedes $B \in \mathcal{F}$. Mithin ist X_∞ die Radon-Nikodym-Dichte von Q bezüglich \mathbf{P} .

Man beachte, dass wir für diesen Beweis des Satzes von Radon-Nikodym nirgends die Existenz bedingter Erwartungen vorausgesetzt haben (vielmehr haben wir sie für endliche σ -Algebren explizit konstruiert), also nicht in versteckter Weise auf den Satz selber zurückgegriffen haben.

Man könnte einwenden, dass wir hier nur den Spezialfall von W-Maßen behandeln könnten. Dieser Mangel kann jedoch sehr leicht behoben werden: Sind μ und ν beliebige (jedoch von Null verschiedene) σ -endliche Maße, dann gibt es messbare Funktionen $g, h : \Omega \rightarrow (0, \infty)$ mit $\int g d\mu = 1$ und $\int h d\nu = 1$. Wir setzen nun $\mathbf{P} = g\mu$ und $Q = h\nu$. Offenbar gilt genau dann $Q \ll \mathbf{P}$, wenn $\nu \ll \mu$. In diesem Fall ist $\frac{g}{h} X_\infty$ eine Version der Radon-Nikodym-Ableitung $\frac{d\nu}{d\mu}$.

Auf die Einschränkung, dass \mathcal{F} abzählbar erzeugt werden kann, kann man ebenfalls verzichten. Mit Hilfe der Approximationssätze für Maße kann man zeigen, dass es stets eine abzählbar erzeugte σ -Algebra $\mathcal{G} \subset \mathcal{F}$ gibt, sodass für jedes $A \in \mathcal{F}$ ein $B \in \mathcal{G}$ existiert mit $\mathbf{P}[A \triangle B] = 0$. Hiermit lässt sich der allgemeine Fall beweisen. Wir führen dies hier nicht aus, sondern verweisen auf [180, Kapitel 14.13]. \diamond

Übung 11.2.1. Die Aussage von Satz 11.10 ist für $p = 1$ im Allgemeinen falsch. Man gebe ein Beispiel an für ein nichtnegatives Martingal X mit $\mathbf{E}[X_n] = 1$ für jedes $n \in \mathbb{N}$, aber $X_n \xrightarrow{n \rightarrow \infty} 0$ fast sicher. \clubsuit

Übung 11.2.2. Seien X_1, X_2, \dots unabhängige, quadratisch integrierbare Zufallsvariablen mit $\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbf{Var}[X_n] < \infty$. Man zeige mit Hilfe des Martingalkonvergenzsatzes das starke Gesetz der großen Zahl für $(X_n)_{n \in \mathbb{N}}$. \clubsuit

Übung 11.2.3. Man gebe ein Beispiel an für ein quadratisch integrierbares Martingal, das fast sicher konvergiert, aber nicht in L^2 . \clubsuit

Übung 11.2.4. Man zeige: In Satz 11.14 gilt im Allgemeinen nicht die Umkehrung. Das heißt, es gibt ein quadratintegrierbares Martingal X , das fast sicher konvergiert, für das aber nicht gilt, dass $\lim_{n \rightarrow \infty} \langle X \rangle_n < \infty$ fast sicher. \clubsuit

Übung 11.2.5. Man zeige die folgende Umkehrung von Satz 11.14. Sei $L > 0$, und sei $(X_n)_{n \in \mathbb{N}}$ ein Martingal mit der Eigenschaft

$$|X_{n+1} - X_n| \leq L \quad \text{f.s.} \tag{11.2}$$

Definiere die Ereignisse

$$\begin{aligned} C &:= \{(X_n)_{n \in \mathbb{N}} \text{ konvergiert für } n \rightarrow \infty\}, \\ A^+ &:= \left\{ \limsup_{n \rightarrow \infty} X_n < \infty \right\}, \\ A^- &:= \left\{ \liminf_{n \rightarrow \infty} X_n > -\infty \right\}, \\ F &:= \left\{ \sup_{n \in \mathbb{N}} \langle X \rangle_n < \infty \right\}. \end{aligned}$$

Man zeige:

$$C = A^+ = A^- = F \quad (\text{mod } \mathbf{P}).$$

Dabei heißt Gleichheit (mod \mathbf{P}), dass die jeweiligen Ereignisse bis auf eine \mathbf{P} -Nullmenge übereinstimmen (siehe Definition 1.68(iii)).

Hinweis: Man verwende Stoppzeiten $\sigma_K = \inf\{n \in \mathbb{N} : |X_n| \geq K\}$, $\sigma_K^\pm = \inf\{n \in \mathbb{N} : \pm X_n \geq K\}$ und τ_K wie im Beweis von Satz 11.14. ♣

Übung 11.2.6. Die Bezeichnungen seien wie in Übung 11.2.5, jedoch gelte statt (11.2) lediglich

$$\mathbf{E} \left[\sup_{n \in \mathbb{N}} |X_{n+1} - X_n| \right] < \infty. \quad (11.3)$$

Man zeige

$$C = A^+ = A^- \quad (\text{mod } \mathbf{P}).$$

Hinweis: Man verwende den Martingalkonvergenzsatz (Satz 11.4) für den gestoppten Prozess X^{ϱ_K} für geeignete Stoppzeiten ϱ_K . ♣

Übung 11.2.7 (Bedingte Version des Borel-Cantelli Lemmas). Sei $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$ eine Filtration und $(A_n)_{n \in \mathbb{N}}$ Ereignisse mit $A_n \in \mathcal{F}_n$ für jedes $n \in \mathbb{N}$. Setze $A^* = \limsup_{n \rightarrow \infty} A_n$ und $A_\infty = \left\{ \sum_{n=1}^{\infty} \mathbf{P}[A_n | \mathcal{F}_{n-1}] = \infty \right\}$. Man zeige die bedingte Version des Borel-Cantelli Lemmas: $\mathbf{P}[A_\infty \Delta A^*] = 0$.

Hinweis: Wende Übung 11.2.5 an auf $X_n = \sum_{k=1}^n (\mathbb{1}_{A_k} - \mathbf{P}[A_k | \mathcal{F}_{k-1}])$. ♣

Übung 11.2.8. Sei $p \in [0, 1]$ und $X = (X_n)_{n \in \mathbb{N}_0}$ ein stochastischer Prozess mit Werten in $[0, 1]$. Für jedes $n \in \mathbb{N}_0$ gelte: Gegeben X_0, \dots, X_n ist

$$X_{n+1} = \begin{cases} 1 - p + pX_n & \text{mit Wahrscheinlichkeit } X_n, \\ pX_n & \text{mit Wahrscheinlichkeit } 1 - X_n. \end{cases}$$

Man zeige, dass X ein Martingal ist und fast sicher konvergiert. Man bestimme die Verteilung des fast sicheren Grenzwertes $\lim_{n \rightarrow \infty} X_n$. ♣

Übung 11.2.9. Sei $f \in \mathcal{L}^1(\lambda)$, wobei λ die Einschränkung des Lebesgue-Maßes auf $[0, 1]$ bezeichnet. Sei $I_{n,k} = [k 2^{-n}, (k+1) 2^{-n}]$ für $n \in \mathbb{N}$ und $k = 0, \dots, 2^n - 1$. Definiere $f_n : [0, 1] \rightarrow \mathbb{R}$ durch

$$f_n(x) = 2^n \int_{I_{k,n}} f d\lambda, \quad \text{falls } k \text{ so gewählt ist, dass } x \in I_{k,n}.$$

Zeige: Für λ -fast alle $x \in [0, 1]$ gilt $f_n(x) \xrightarrow{n \rightarrow \infty} f(x)$. ♣

Übung 11.2.10. Sei $(\Omega, \mathcal{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum mit einer Filtration $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$. Sei $\mathcal{F}_\infty := \sigma(\mathcal{F}_n : n \in \mathbb{N})$, und sei \mathcal{M} der Vektorraum der gleichgradig integrierbaren \mathbb{F} -Martingale. Man zeige: die Abbildung $\Phi : \mathcal{L}^1(\mathcal{F}_\infty) \rightarrow \mathcal{M}$, $X_\infty \mapsto (\mathbf{E}[X_\infty | \mathcal{F}_n])_{n \in \mathbb{N}}$ ist ein Vektorraumisomorphismus. ♣

11.3 Beispiel: Verzweigungsprozess

Sei $p = (p_k)_{k \in \mathbb{N}_0}$ ein Wahrscheinlichkeitsvektor auf \mathbb{N}_0 und $(Z_n)_{n \in \mathbb{N}_0}$ der Galton-Watson-Prozess mit einem Urahn und Nachkommenverteilung p (siehe Definition 3.9). Zur Erinnerung geben wir die Konstruktion von Z an. Seien $(X_{n,i})_{n \in \mathbb{N}_0, i \in \mathbb{N}}$ u.i.v. Zufallsvariablen mit $\mathbf{P}[X_{1,1} = k] = p_k$ für $k \in \mathbb{N}_0$.

Setze $Z_0 = 1$ und induktiv

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i} \quad \text{für } n \in \mathbb{N}_0.$$

Wir interpretieren Z_n als Größe einer Population zur Zeit n und $X_{n,i}$ als Anzahl der Nachkommen des i -ten Individuums aus der n -ten Generation.

Seien $m := \mathbf{E}[X_{1,1}] < \infty$ die erwartete Kinderanzahl pro Individuum und $\sigma^2 := \mathbf{Var}[X_{1,1}] \in (0, \infty)$ die Varianz der Kinderzahl. Setze

$$\mathcal{F}_n := \sigma(X_{k,i} : k < n, i \in \mathbb{N}).$$

Dann ist Z an \mathbb{F} adaptiert. Definiere $W_n := m^{-n} Z_n$.

Lemma 11.18. W ist ein Martingal. Speziell ist $\mathbf{E}[Z_n] = m^n$ für jedes $n \in \mathbb{N}$.

Beweis. Wir berechnen die bedingte Erwartung für $n \in \mathbb{N}_0$:

$$\begin{aligned} \mathbf{E}[W_{n+1} | \mathcal{F}_n] &= m^{-(n+1)} \mathbf{E}[Z_{n+1} | \mathcal{F}_n] \\ &= m^{-(n+1)} \mathbf{E}\left[\sum_{i=1}^{Z_n} X_{n,i} | \mathcal{F}_n\right] \\ &= m^{-(n+1)} \sum_{k=1}^{\infty} \mathbf{E}[\mathbb{1}_{\{Z_n=k\}} k \cdot X_{n,i} | \mathcal{F}_n] \\ &= m^{-n} \sum_{k=1}^{\infty} \mathbf{E}[k \cdot \mathbb{1}_{\{Z_n=k\}} | \mathcal{F}_n] \\ &= m^{-n} Z_n = W_n. \end{aligned}$$
□

Satz 11.19. Sei $\text{Var}[X_{1,1}] \in (0, \infty)$. Es existiert der fast sichere Limes $W_\infty = \lim_{n \rightarrow \infty} W_n$, und es gilt

$$m > 1 \iff \mathbf{E}[W_\infty] = 1 \iff \mathbf{E}[W_\infty] > 0.$$

Beweis. W_∞ existiert, weil $W \geq 0$ ein Martingal ist. Ist $m \leq 1$, so folgt, dass $(Z_n)_{n \in \mathbb{N}}$ f.s. gegen ein Z_∞ konvergiert. Wegen $\sigma^2 > 0$ kommt nur $Z_\infty = 0$ in Frage.

Sei nun $m > 1$. Es gilt nach dem Satz von Blackwell-Girshick (Satz 5.10) wegen $\mathbf{E}[Z_{n-1}] = m^{n-1}$ (Lemma 11.18)

$$\begin{aligned} \text{Var}[W_n] &= m^{-2n} (\sigma^2 \mathbf{E}[Z_{n-1}] + m^2 \text{Var}[Z_{n-1}]) \\ &= \sigma^2 m^{-(n+1)} + \text{Var}[W_{n-1}]. \end{aligned}$$

Induktiv folgt $\text{Var}[W_n] = \sigma^2 \sum_{k=2}^{n+1} m^{-k} \leq \frac{\sigma^2 m}{m-1} < \infty$. Also ist W in L^2 beschränkt, und Satz 11.10 liefert, dass $W_n \rightarrow W_\infty$ in L^2 und damit auch in L^1 . Speziell ist $\mathbf{E}[W_\infty] = \mathbf{E}[W_0] = 1$. \square

Unter der Annahme der endlichen Varianz waren die Aussagen von Satz 11.19 nicht schwer zu zeigen. Es gilt aber eine viel stärkere Aussage, die wir hier nur zitieren (siehe [103], beziehungsweise [120] für einen modernen Beweis).

Satz 11.20 (Kesten-Stigum (1966)). Sei $m > 1$. Dann sind äquivalent

- (i) $\mathbf{E}[W_\infty] = 1$,
- (ii) $\mathbf{E}[W_\infty] > 0$,
- (iii) $\mathbf{E}[X_{1,1} \log(X_{1,1})^+] < \infty$.

12 Rückwärtsmartingale und Austauschbarkeit

Bei vielen Datenerhebungen, etwa Telefonumfragen, ist die Reihenfolge, in der die Daten kommen, unerheblich. Mathematisch sprechen wir von austauschbaren Zufallsvariablen, wenn sich die gemeinsame Verteilung unter endlichen Vertauschungen nicht ändert. Der Struktursatz für austauschbare Zufallsvariablen von de Finetti besagt, dass sich eine unendlich große austauschbare Familie von Zufallsvariablen mit Werten im Raum E als Zweistufenexperiment beschreiben lässt: In der ersten Stufe wird eine zufällige Wahrscheinlichkeitsverteilung Ξ auf E ausgewürfelt. In der zweiten Stufe werden die Zufallsvariablen u.i.v. mit Verteilung Ξ realisiert.

Wir definieren zunächst den Begriff der Austauschbarkeit. Danach betrachten wir Rückwärtsmartingale und zeigen den Konvergenzsatz für Rückwärtsmartingale. Dieser ist der Eckstein für den Beweis des Satzes von de Finetti.

12.1 Austauschbare Familien von Zufallsvariablen

Definition 12.1. Sei I eine beliebige Indexmenge und E ein polnischer Raum. Eine Familie $(X_i)_{i \in I}$ von Zufallsvariablen mit Werten in E heißt **austauschbar**, falls für jede endliche Permutation $\varrho : I \rightarrow I$ gilt, dass

$$\mathcal{L} \left[(X_{\varrho(i)})_{i \in I} \right] = \mathcal{L} \left[(X_i)_{i \in I} \right].$$

Als endliche Permutation bezeichnen wir dabei eine Bijektion $\varrho : I \rightarrow I$, die alle bis auf endlich viele Koordinaten unverändert lässt.

Bemerkung 12.2. Offenbar sind äquivalent:

- (i) $(X_i)_{i \in I}$ ist austauschbar.
- (ii) Für $n \in \mathbb{N}$ und paarweise unterschiedliche $i_1, \dots, i_n \in I$ sowie paarweise unterschiedliche $j_1, \dots, j_n \in I$ gilt $\mathcal{L}[(X_{i_1}, \dots, X_{i_n})] = \mathcal{L}[(X_{j_1}, \dots, X_{j_n})]$.

Insbesondere sind austauschbare Zufallsvariablen stets identisch verteilt (dies ist (ii) mit $n = 1$). \diamond

Beispiel 12.3. (i) Ist $(X_i)_{i \in I}$ u.i.v., so ist $(X_i)_{i \in I}$ austauschbar.

(ii) In einer Urne seien N Kugeln, davon M schwarz. Wir ziehen sukzessive ohne Zurücklegen alle Kugeln und setzen

$$X_n := \begin{cases} 1, & \text{falls die } n\text{-te Kugel schwarz ist,} \\ 0, & \text{sonst.} \end{cases}$$

Dann ist $(X_n)_{n=1,\dots,N}$ austauschbar. Dies folgt aus elementarer Kombinatorik, denn für jede Wahl von $x_1, \dots, x_N \in \{0, 1\}$ mit $x_1 + \dots + x_N = M$ ist offenbar

$$\mathbf{P}[X_1 = x_1, \dots, X_N = x_N] = \frac{1}{\binom{N}{M}}.$$

Diese Formel können wir aber auch formal durch eine kleine Rechnung mit bedingten Wahrscheinlichkeiten herleiten, die wir in ähnlicher Form für das Pólya'sche Urnenmodell in Beispiel 12.29 noch einmal durchführen werden. Setze dazu $s_k = x_1 + \dots + x_k$ für $k = 0, \dots, N$ und

$$g_k(x) = \begin{cases} M - s_k, & \text{falls } x = 1, \\ N - M + s_k - k, & \text{falls } x = 0. \end{cases}$$

Dann ist $\mathbf{P}[X_1 = x_1] = g_0(x_1)/N$ und

$$\mathbf{P}[X_{k+1} = x_{k+1} | X_1 = x_1, \dots, X_k = x_k] = \frac{g_k(x_{k+1})}{N - k} \quad \text{für } k = 1, \dots, N - 1.$$

Ferner ist offenbar $g_k(0) = N - M - l$, wobei $l = \#\{i \leq k : x_i = 0\}$. Es folgt

$$\begin{aligned} & \mathbf{P}[X_1 = x_1, \dots, X_N = x_N] \\ &= \mathbf{P}[X_1 = x_1] \prod_{k=1}^{N-1} \mathbf{P}[X_{k+1} = x_{k+1} | X_1 = x_1, \dots, X_k = x_k] \\ &= \frac{1}{N!} \prod_{k=0}^{N-1} g_k(x_{k+1}) = \frac{1}{N!} \prod_{k: x_k=1} g_k(1) \prod_{k: x_k=0} g_k(0) \\ &= \frac{1}{N!} \prod_{l=0}^{M-1} (M - l) \prod_{l=0}^{N-1} (N - M - l) = \frac{M!(N - M)!}{N!}. \end{aligned}$$

(iii) Sei Y eine Zufallsvariable mit Werten in $[0, 1]$, und gegeben Y seien $(X_i)_{i \in I}$ unabhängig und Ber_Y -verteilt. Das heißt, für jedes endliche $J \subset I$ ist

$$\mathbf{P}[X_j = 1 \text{ für jedes } j \in J | Y] = Y^{\#J}.$$

Dann ist $(X_i)_{i \in I}$ austauschbar. ◇

Sei $X = (X_n)_{n \in \mathbb{N}}$ ein stochastischer Prozess mit Werten in einem polnischen Raum E . Sei $S(n)$ die Menge der Permutationen $\varrho : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. Wir fassen ϱ ebenfalls als Abbildung $\mathbb{N} \rightarrow \mathbb{N}$ auf durch $\varrho(k) = k$ für $k > n$. Für $\varrho \in S(n)$ und $x = (x_1, \dots, x_n) \in E^n$ schreiben wir $x^\varrho = (x_{\varrho(1)}, \dots, x_{\varrho(n)})$. Für $x \in E^{\mathbb{N}}$ schreiben wir analog $x^\varrho = (x_{\varrho(1)}, x_{\varrho(2)}, \dots) \in E^{\mathbb{N}}$. Ist E' ein weiterer polnischer Raum, so definieren wir für messbare Abbildungen $f : E^n \rightarrow E'$ und $F : E^{\mathbb{N}} \rightarrow E'$ die Abbildungen f^ϱ und F^ϱ durch $f^\varrho(x) = f(x^\varrho)$ und $F^\varrho(x) = F(x^\varrho)$. Ferner schreiben wir $f(x) = f(x_1, \dots, x_n)$ auch, falls $x \in E^n$.

Definition 12.4. (i) Eine Abbildung $f : E^n \rightarrow E'$ heißt **symmetrisch**, falls $f^\varrho = f$ ist für jedes $\varrho \in S(n)$.

(ii) Eine Abbildung $F : E^{\mathbb{N}} \rightarrow E'$ heißt **n-symmetrisch**, falls $F^\varrho = F$ für jedes $\varrho \in S(n)$. F heißt **symmetrisch**, falls F n-symmetrisch ist für jedes $n \in \mathbb{N}$.

Beispiel 12.5. (i) Für $x \in \mathbb{R}^{\mathbb{N}}$ definieren wir das n -te arithmetische Mittel durch $a_n(x) = \frac{1}{n} \sum_{i=1}^n x_i$. Offenbar ist a_n eine n -symmetrische Abbildung (aber nicht m -symmetrisch für ein $m > n$). Weiter definiert $\bar{a}(x) := \limsup_{n \rightarrow \infty} a_n(x)$ eine symmetrische Abbildung $\mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$.

(ii) Die Abbildung $s : \mathbb{R}^{\mathbb{N}} \rightarrow [0, \infty]$, $x \mapsto \sum_{i=1}^{\infty} |x_i|$ ist symmetrisch. Anders als \bar{a} hängt der Wert von s von jeder einzelnen Koordinate ab, falls er endlich ist.

(iii) Für $x \in E^{\mathbb{N}}$ definieren wir die n -te empirische Verteilung durch $\xi_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ (wobei δ_{x_i} das Dirac-Maß in x_i ist). Offenbar ist ξ_n eine n -symmetrische Abbildung.

(iv) Sei $k \in \mathbb{N}$ und $\varphi : E^k \rightarrow \mathbb{R}$ eine Abbildung. Das n -te symmetrisierte Mittel

$$A_n(\varphi) : E^{\mathbb{N}} \rightarrow \mathbb{R}, \quad x \mapsto \frac{1}{n!} \sum_{\varrho \in S(n)} \varphi(x^\varrho) \quad (12.1)$$

ist eine n -symmetrische Abbildung. \diamond

Definition 12.6. Sei $X = (X_n)_{n \in \mathbb{N}}$ ein stochastischer Prozess mit Werten in E . Für $n \in \mathbb{N}$ sei

$$\mathcal{E}'_n := \sigma(F : F : E^{\mathbb{N}} \rightarrow \mathbb{R} \text{ ist messbar und } n\text{-symmetrisch}),$$

sowie $\mathcal{E}_n := X^{-1}(\mathcal{E}'_n)$ die σ -Algebra der unter allen Permutation $\varrho \in S(n)$ invarianten Ereignisse. Ferner sei

$$\mathcal{E}' := \bigcap_{n=1}^{\infty} \mathcal{E}'_n = \sigma(F : F : E^{\mathbb{N}} \rightarrow \mathbb{R} \text{ ist messbar und symmetrisch}),$$

sowie $\mathcal{E} := \bigcap_{n=1}^{\infty} \mathcal{E}_n = X^{-1}(\mathcal{E}')$ die σ -Algebra der austauschbaren Ereignisse für X , oder kurz die **austauschbare σ -Algebra**.

Bemerkung 12.7. Ist $A \in \sigma(X_n, n \in \mathbb{N})$ ein Ereignis, so gibt es ein messbares $B \subset E^{\mathbb{N}}$ mit $A = \{X \in B\}$. Schreiben wir $A^\varrho = \{X^\varrho \in B\}$ für $\varrho \in S(n)$, so ist $\mathcal{E}_n = \{A : A^\varrho = A \text{ für alle } \varrho \in S(n)\}$. Dies rechtfertigt den Namen „austauschbares Ereignis“. \diamond

Bemerkung 12.8. Schreiben wir $\Xi_n(\omega) := \xi_n(X(\omega)) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}$ für die n -te empirische Verteilung, so ist nach Übung 12.1.1 $\mathcal{E}_n = \sigma(\Xi_n)$. \diamond

Bemerkung 12.9. Bezeichnen wir mit $\mathcal{T} = \bigcap_{n \in \mathbb{N}} \sigma(X_{n+1}, X_{n+2}, \dots)$ die terminale σ -Algebra, so ist $\mathcal{T} \subset \mathcal{E}$, wobei strikte Inklusion gelten kann.

In der Tat: Offenbar ist $\sigma(X_{n+1}, X_{n+2}, \dots) \subset \mathcal{E}_n$ für $n \in \mathbb{N}$, also $\mathcal{T} \subset \mathcal{E}$. Sei nun $E = \{0, 1\}$, und seien X_1, X_2, \dots unabhängig mit $\mathbf{P}[X_n = 1] \in (0, 1)$ für jedes $n \in \mathbb{N}$. Die Zufallsvariable $S := \sum_{n=1}^{\infty} X_n$ ist messbar bezüglich \mathcal{E} , nicht aber bezüglich \mathcal{T} . \diamond

Satz 12.10. Sei $X = (X_n)_{n \in \mathbb{N}}$ austauschbar. Ist $\varphi : E^{\mathbb{N}} \rightarrow \mathbb{R}$ messbar und $\mathbf{E}[|\varphi(X)|] < \infty$, dann gilt für jedes $n \in \mathbb{N}$ und jedes $\varrho \in S(n)$

$$\mathbf{E}[\varphi(X) | \mathcal{E}_n] = \mathbf{E}[\varphi(X^\varrho) | \mathcal{E}_n]. \quad (12.2)$$

Speziell ist

$$\mathbf{E}[\varphi(X) | \mathcal{E}_n] = A_n(\varphi) := \frac{1}{n!} \sum_{\varrho \in S(n)} \varphi(X^\varrho). \quad (12.3)$$

Beweis. Sei $A \in \mathcal{E}_n$. Dann existiert ein $B \in \mathcal{E}'_n$ mit $A = X^{-1}(B)$. Setze $F := \mathbb{1}_B$. Dann ist $F \circ X = \mathbb{1}_A$. Nach der Definition von \mathcal{E}'_n ist $F : E^{\mathbb{N}} \rightarrow \mathbb{R}$ messbar, n -symmetrisch und beschränkt. Daher ist

$$\mathbf{E}[\varphi(X)F(X)] = \mathbf{E}[\varphi(X^\varrho)F(X^\varrho)] = \mathbf{E}[\varphi(X^\varrho)F(X)],$$

wobei wir in der ersten Gleichung die Austauschbarkeit von X benutzt haben, in der zweiten hingegen die Symmetrie von F . Hieraus folgt (12.2). Nun ist aber $A_n(\varphi)$ schon \mathcal{E}_n -messbar, also ist

$$\mathbf{E}[\varphi(X) | \mathcal{E}_n] = \mathbf{E}\left[\frac{1}{n!} \sum_{\varrho \in S(n)} \varphi(X^\varrho) \middle| \mathcal{E}_n\right] = \frac{1}{n!} \sum_{\varrho \in S(n)} \varphi(X^\varrho). \quad \square$$

Heuristik zur Struktur austauschbarer Familien

Wir betrachten eine endliche, austauschbare Familie X_1, \dots, X_N von E -wertigen Zufallsvariablen. Wie sieht für $n \leq N$ die bedingte Verteilung von (X_1, \dots, X_n) gegeben Ξ_N aus? Für jedes messbare $A \subset E$ kommt $\{X_i \in A\}$ für genau $N\Xi_N(A)$

viele $i \in \{1, \dots, N\}$ vor, wobei die Reihenfolge des Auftretens keinen Einfluss auf die Wahrscheinlichkeit hat. Wir sind also in der Situation des Ziehens von gefärbten Kugeln *ohne Zurücklegen*. Genauer gesagt können wir annehmen, dass die paarweise unterschiedlichen $e_1, \dots, e_k \in E$ die Atome von Ξ_N mit Häufigkeiten N_1, \dots, N_k sind, dass also $\Xi_N = \sum_{i=1}^k (N_i/N)\delta_{e_i}$ gilt. Wir haben es also mit Kugeln in k Farben zu tun, wobei von der i -ten Farbe genau N_i Kugeln vorhanden sind. Wir ziehen n dieser Kugeln ohne Zurücklegen, aber mit Beachtung der Reihenfolge. Bis auf die Beachtung der Reihenfolge ist die resultierende Verteilung also die allgemeine hypergeometrische Verteilung (siehe (1.19) auf Seite 46). Es gilt also für paarweise disjunkte, messbare Mengen A_1, \dots, A_k mit $\bigcup_{l=1}^k A_l = E$, für $i_1, \dots, i_n \in \{1, \dots, k\}$, paarweise unterschiedliche $j_1, \dots, j_n \in \{1, \dots, N\}$ und mit der Festlegung $m_l := \#\{r \in \{1, \dots, n\} : i_r = l\}$ für $l \in \{1, \dots, k\}$

$$\mathbf{P}[X_{j_r} \in A_{i_r} \text{ für jedes } r = 1, \dots, n \mid \Xi_N] = \frac{1}{(N)_n} \prod_{l=1}^k (N \Xi_N(A_l))^{m_l}, \quad (12.4)$$

wobei wir $(n)_l := n(n-1) \cdots (n-l+1)$ definieren.

Was passiert nun, wenn wir $N \rightarrow \infty$ gehen lassen? Wir nehmen hier der Einfachheit halber an, dass der Limes $\Xi_\infty(A_l) = \lim_{N \rightarrow \infty} \Xi_N(A_l)$ für jedes $l = 1, \dots, k$ in einem geeigneten Sinne existiert. Dann wird aus (12.4) formal

$$\mathbf{P}[X_{j_r} \in A_{i_r} \text{ für jedes } r = 1, \dots, n \mid \Xi_\infty] = \prod_{l=1}^k \Xi_\infty(A_l)^{m_l}. \quad (12.5)$$

Aus dem Ziehen der Kugeln ohne Zurücklegen wird nun also asymptotisch für große Kugelanzahl das Ziehen mit Zurücklegen. Damit sind die Zufallsvariablen X_1, X_2, \dots unabhängig mit Verteilung Ξ_∞ gegeben Ξ_∞ .

Einen formalen Beweis, der entlang der von dieser Heuristik vorgezeichneten Linie verläuft, bringen wir in Kapitel 13.4.

Um diese Aussage, den so genannten Satz von de Finetti, in Abschnitt 12.3 rigoros zu formulieren und zu beweisen, brauchen wir noch etwas Begriffsbildung (etwa bedingte Unabhängigkeit). Als technisches Hilfsmittel verwenden wir in diesem Kapitel den Konvergenzsatz für Rückwärtsmartingale, den wir in Abschnitt 12.2 formulieren. Als weiterführende Lektüre sei auf [4, 38, 105, 114] verwiesen.

Übung 12.1.1. Sei $n \in \mathbb{N}$. Man zeige, dass sich jede symmetrische Funktion $f : E^n \rightarrow \mathbb{R}$ schreiben lässt als $f(x) = g(\frac{1}{n} \sum_{i=1}^n \delta_{x_i})$, wobei g (abhängig von f) geeignet zu wählen ist. ♣

Übung 12.1.2. Man leite (12.4) formal her. ♣

Übung 12.1.3. Seien X_1, \dots, X_n austauschbare quadratintegrierbare Zufallsvariablen. Man zeige

$$\text{Cov}[X_1, X_2] \geq -\frac{1}{n-1} \text{Var}[X_1]. \quad (12.6)$$

Man gebe für $n \geq 2$ ein (nichttriviales) Beispiel für Gleichheit in (12.6) an. ♣

Übung 12.1.4. Seien X_1, X_2, X_3, \dots austauschbare, quadratintegrierbare Zufallsvariablen. Man zeige, dass $\text{Cov}[X_1, X_2] \geq 0$ gilt. ♣

Übung 12.1.5. Man zeige: Für jedes $n \in \mathbb{N} \setminus \{1\}$ gibt es eine austauschbare Familie von Zufallsvariablen X_1, \dots, X_n , die nicht zu einer unendlichen, austauschbaren Familie X_1, X_2, \dots fortgesetzt werden kann. ♣

12.2 Rückwärtsmartingale

Die Begriffe der Filtration und des Martingals haben nirgends vorausgesetzt, dass die Zeitmenge $I \subset [0, \infty)$ wäre. Wir wollen jetzt den Fall $I = -\mathbb{N}_0$ betrachten.

Definition 12.11 (Rückwärtsmartingal). Sei $\mathbb{F} = (\mathcal{F}_n)_{n \in -\mathbb{N}_0}$ eine Filtration und $X = (X_n)_{n \in -\mathbb{N}_0}$ ein \mathbb{F} -Martingal. Dann nennen wir $X = (X_{-n})_{n \in \mathbb{N}_0}$ ein Rückwärtsmartingal.

Bemerkung 12.12. Ein Rückwärtsmartingal ist stets gleichgradig integrierbar. Dies folgt aus Korollar 8.22 und der Tatsache, dass $X_{-n} = \mathbf{E}[X_0 | \mathcal{F}_{-n}]$ für jedes $n \in \mathbb{N}_0$. ◇

Beispiel 12.13. Seien X_1, X_2, \dots austauschbare, reelle Zufallsvariablen. Für $n \in \mathbb{N}$ setze $\mathcal{F}_{-n} = \mathcal{E}_n$ und

$$Y_{-n} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Die folgende Rechnung zeigt, dass $(Y_{-n})_{n \in \mathbb{N}}$ ein \mathbb{F} -Rückwärtsmartingal ist: Adaptiertheit ist klar. Außerdem ist nach Satz 12.10 (mit $k = n$ und $\varphi(X_1, \dots, X_n) = \frac{1}{n-1}(X_1 + \dots + X_{n-1})$)

$$\mathbf{E}[Y_{-n+1} | \mathcal{F}_{-n}] = \frac{1}{n!} \sum_{\varrho \in S(n)} \frac{1}{n-1} (X_{\varrho(1)} + \dots + X_{\varrho(n-1)}) = Y_{-n}.$$

Betrachten wir statt \mathbb{F} die kleinere Filtration $\mathbb{G} = (\mathcal{G}_n)_{n \in -\mathbb{N}}$, die für $n \in \mathbb{N}$ durch $\mathcal{G}_{-n} = \sigma(Y_{-n}, X_{n+1}, X_{n+2}, \dots) = \sigma(Y_{-n}, Y_{-n-1}, Y_{-n-2}, \dots)$ definiert wird, also die von Y erzeugte Filtration, so ist Y natürlich auch bezüglich \mathbb{G} ein Rückwärtsmartingal (siehe Bemerkung 9.29). ◇

Seien $a < b$ und $n \in \mathbb{N}$. Sei $U_{-n}^{a,b}$ die Anzahl der Aufkreuzungen von X über $[a, b]$ zwischen Zeit $-n$ und 0 sowie $U^{a,b} = \lim_{n \rightarrow \infty} U_{-n}^{a,b}$. Nach der Aufkreuzungsgleichung (Lemma 11.3) ist $\mathbf{E}[U_{-n}^{a,b}] \leq \frac{1}{b-a} \mathbf{E}[(X_0 - a)^+]$, also $\mathbf{P}[U^{a,b} < \infty] = 1$. Wie im Beweis des Martingalkonvergenzsatzes (Satz 11.4) folgt:

Satz 12.14 (Konvergenzsatz für Rückwärtsmartingale). Sei $(X_n)_{n \in \mathbb{N}_0}$ ein Martingal bezüglich $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}_0}$. Dann existiert $X_{-\infty} = \lim_{n \rightarrow \infty} X_{-n}$ fast sicher und in L^1 . Es gilt $X_{-\infty} = \mathbf{E}[X_0 | \mathcal{F}_{-\infty}]$, wobei $\mathcal{F}_{-\infty} = \bigcap_{n=1}^{\infty} \mathcal{F}_{-n}$.

Beispiel 12.15. Seien X_1, X_2, \dots austauschbare integrierbare Zufallsvariablen. Ferner sei $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_m, m \geq n)$ die terminale σ -Algebra der X_1, X_2, \dots und \mathcal{E} die austauschbare σ -Algebra. Dann gilt $\mathbf{E}[X_1 | \mathcal{T}] = \mathbf{E}[X_1 | \mathcal{E}]$ f.s. und

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_1 | \mathcal{E}] \quad \text{f.s. und in } L^1.$$

In der Tat: Setzen wir $Y_{-n} := \frac{1}{n} \sum_{i=1}^n X_i$, so ist (nach Beispiel 12.13) $(Y_{-n})_{n \in \mathbb{N}}$ ein Rückwärtsmartingal bezüglich $(\mathcal{F}_n)_{n \in \mathbb{N}} = (\mathcal{E}_{-n})_{n \in \mathbb{N}}$, und daher gilt

$$Y_{-n} \xrightarrow{n \rightarrow \infty} Y_{-\infty} = \mathbf{E}[X_1 | \mathcal{E}] \quad \text{f.s. und in } L^1.$$

Nun ist nach Beispiel 2.36(ii) $Y_{-\infty}$ schon \mathcal{T} -messbar, also (wegen $\mathcal{T} \subset \mathcal{E}$ und der Turmeigenschaft der bedingten Erwartung) $Y_{-\infty} = \mathbf{E}[X_1 | \mathcal{T}]$. \diamond

Beispiel 12.16 (Starkes Gesetz der großen Zahl). Sind Z_1, Z_2, \dots reell und u.i.v. mit $\mathbf{E}[|Z_1|] < \infty$, dann gilt

$$\frac{1}{n} \sum_{i=1}^n Z_i \xrightarrow{n \rightarrow \infty} \mathbf{E}[Z_1] \quad \text{fast sicher.}$$

Nach dem Kolmogorov'schen 0-1 Gesetz (Satz 2.37) ist die terminale σ -Algebra \mathcal{T} nämlich trivial, also gilt

$$\mathbf{E}[Z_1 | \mathcal{T}] = \mathbf{E}[Z_1] \quad \text{fast sicher.}$$

In Korollar 12.19 werden wir sehen, dass im Falle unabhängiger Zufallsvariablen auch \mathcal{E} schon \mathbf{P} -trivial ist, woraus $\mathbf{E}[Z_1 | \mathcal{E}] = \mathbf{E}[Z_1]$ folgt. \diamond

Wir schließen diesen Abschnitt, indem wir Beispiel 12.15 auf Mittelwerte von Funktionen von $k \in \mathbb{N}$ Variablen verallgemeinern. Diese Schlussfolgerung aus dem Konvergenzsatz für Rückwärtsmartingale wird im folgenden Abschnitt in essenzieller Weise benötigt.

Satz 12.17. Sei $X = (X_n)_{n \in \mathbb{N}}$ eine austauschbare Familie von Zufallsvariablen mit Werten in E , sei $k \in \mathbb{N}$ und $\varphi : E^k \rightarrow \mathbb{R}$ messbar mit $\mathbf{E}[|\varphi(X_1, \dots, X_k)|] < \infty$. Wir schreiben $\varphi(X) = \varphi(X_1, \dots, X_k)$ und setzen $A_n(\varphi) := \frac{1}{n!} \sum_{\varrho \in S(n)} \varphi(X^\varrho)$. Dann gilt

$$\mathbf{E}[\varphi(X) | \mathcal{E}] = \mathbf{E}[\varphi(X) | \mathcal{T}] = \lim_{n \rightarrow \infty} A_n(\varphi) \quad f.s. \text{ und in } L^1. \quad (12.7)$$

Beweis. Nach Satz 12.10 ist $A_n(\varphi) = \mathbf{E}[\varphi(X) | \mathcal{E}_n]$. Also ist $(A_{-n}(\varphi))_{n \geq k}$ ein Rückwärtsmartingal bezüglich $(\mathcal{E}_{-n})_{n \in -\mathbb{N}}$. Nach Satz 12.14 ist also

$$A_n(\varphi) \xrightarrow{n \rightarrow \infty} \mathbf{E}[\varphi(X) | \mathcal{E}] \quad f.s. \text{ und in } L^1. \quad (12.8)$$

Wir können wie für das arithmetische Mittel (Beispiel 12.16) argumentieren, dass $\lim_{n \rightarrow \infty} A_n(\varphi)$ schon \mathcal{T} -messbar ist. In der Tat ist

$$\limsup_{n \rightarrow \infty} \frac{\#\{\varrho \in S(n) : \varrho^{-1}(i) \leq l \text{ für ein } i \in \{1, \dots, k\}\}}{n!} = 0 \quad \text{für jedes } l \in \mathbb{N}.$$

Der Wert von $A_n(\varphi)$ hängt für große n also in zu vernachlässigender Weise von den ersten l Koordinaten ab. Zusammen mit (12.8) folgt (12.7). \square

Korollar 12.18. Sei $X = (X_n)_{n \in \mathbb{N}}$ austauschbar. Dann gibt es für jedes $A \in \mathcal{E}$ ein $B \in \mathcal{T}$ mit $\mathbf{P}[A \Delta B] = 0$.

Man beachte, dass $\mathcal{T} \subset \mathcal{E}$ ist, dass also die Aussage trivialerweise gilt, wenn wir die Rollen von \mathcal{E} und \mathcal{T} vertauschen.

Beweis. Wegen $\mathcal{E} \subset \sigma(X_1, X_2, \dots)$ existiert nach dem Approximationssatz für Maße eine Folge von messbaren Mengen $(A_k)_{k \in \mathbb{N}}$ mit $A_k \in \sigma(X_1, \dots, X_k)$ und $\mathbf{P}[A \Delta A_k] \xrightarrow{k \rightarrow \infty} 0$. Sei $C_k \in E^k$ messbar mit $A_k = \{(X_1, \dots, X_k) \in C_k\}$ für jedes $k \in \mathbb{N}$. Mit $\varphi_k := \mathbb{1}_{C_k}$ folgt aus Satz 12.17

$$\begin{aligned} \mathbb{1}_A &= \mathbf{E}[\mathbb{1}_A | \mathcal{E}] = \mathbf{E}\left[\lim_{k \rightarrow \infty} \varphi_k(X) | \mathcal{E}\right] = \lim_{k \rightarrow \infty} \mathbf{E}[\varphi_k(X) | \mathcal{E}] \\ &= \lim_{k \rightarrow \infty} \mathbf{E}[\varphi_k(X) | \mathcal{T}] =: \psi \quad \text{fast sicher.} \end{aligned}$$

Es gibt also eine \mathcal{T} -messbare Funktion ψ mit $\psi = \mathbb{1}_A$ fast sicher. Wir können nun annehmen, dass $\psi = \mathbb{1}_B$ für ein $B \in \mathcal{T}$. \square

Als weitere Anwendung erhalten wir das 0-1 Gesetz von Hewitt und Savage [78].

Korollar 12.19 (0-1 Gesetz von Hewitt-Savage). Seien X_1, X_2, \dots u.i.v. Zufallsvariablen. Dann ist die austauschbare σ -Algebra \mathbf{P} -trivial, also $\mathbf{P}[A] \in \{0, 1\}$ für jedes $A \in \mathcal{E}$.

Beweis. Nach dem Kolmogorov'schen 0-1 Gesetz (Satz 2.37) ist \mathcal{T} trivial. Die Aussage folgt also ohne weiteres aus Korollar 12.18. \square

12.3 Satz von de Finetti

Wir zeigen in diesem Abschnitt den Struktursatz für (abzählbar) unendliche, austauschbare Familien, den wir heuristisch schon am Ende von Abschnitt 12.1 motiviert hatten. Es soll also gezeigt werden, dass eine unendliche, austauschbare Familie von Zufallsvariablen eine unabhängige, identisch verteilte Familie ist gegeben die austauschbare σ -Algebra \mathcal{E} . Ferner berechnen wir die bedingte Verteilung der einzelnen Zufallsvariablen. Als ersten Schritt geben wir eine Definition der bedingten Unabhängigkeit an.

Definition 12.20 (Bedingte Unabhängigkeit). Seien $(\Omega, \mathcal{F}, \mathbf{P})$ ein W-Raum, $\mathcal{A} \subset \mathcal{F}$ eine Teil- σ -Algebra sowie $(\mathcal{A}_i)_{i \in I}$ eine beliebige Familie von Teil- σ -Algebren von \mathcal{F} . Die Familie $(\mathcal{A}_i)_{i \in I}$ heißt **unabhängig gegeben \mathcal{A}** , falls für jedes endliche $J \subset I$ und jede Wahl von $A_j \in \mathcal{A}_j$ für $j \in J$ gilt

$$\mathbf{P}\left[\bigcap_{j \in J} A_j \mid \mathcal{A}\right] = \prod_{j \in J} \mathbf{P}[A_j \mid \mathcal{A}] \quad \text{fast sicher.} \quad (12.9)$$

Eine Familie $(X_i)_{i \in I}$ von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbf{P})$ heißt **unabhängig (und identisch verteilt) gegeben \mathcal{A}** , falls die erzeugten σ -Algebren $(\sigma(X_i))_{i \in I}$ unabhängig gegeben \mathcal{A} sind (und die bedingten Verteilungen $\mathbf{P}[X_i \in \cdot \mid \mathcal{A}]$ alle gleich sind).

Beispiel 12.21. Jede beliebige Familie $(\mathcal{A}_i)_{i \in I}$ von Teil- σ -Algebren von \mathcal{F} ist unabhängig gegeben \mathcal{F} . In der Tat ist in diesem Fall nämlich (mit $A = \bigcap_{j \in J} A_j$)

$$\mathbf{P}[A \mid \mathcal{F}] = \mathbb{1}_A = \prod_{j \in J} \mathbb{1}_{A_j} = \prod_{j \in J} \mathbf{P}[A_j \mid \mathcal{F}] \quad \text{fast sicher.} \quad \diamond$$

Beispiel 12.22. Ist $(\mathcal{A}_i)_{i \in I}$ eine unabhängige Familie von σ -Algebren, und ist \mathcal{A} trivial, dann ist $(\mathcal{A}_i)_{i \in I}$ unabhängig gegeben \mathcal{A} . \diamond

Beispiel 12.23. Es gibt keine „Monotonie“ bei der bedingten Unabhängigkeit in folgendem Sinne: Sind $\mathcal{F}_1, \mathcal{F}_2$ und \mathcal{F}_3 σ -Algebren mit $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$, und ist $(\mathcal{A}_i)_{i \in I}$ unabhängig sowohl gegeben \mathcal{F}_1 wie auch gegeben \mathcal{F}_3 , so folgt noch nicht die Unabhängigkeit gegeben \mathcal{F}_2 .

Um dies zu illustrieren, nehmen wir an, dass X und Y nichttriviale, unabhängige, reelle Zufallsvariablen sind. Wir wählen $\mathcal{F}_1 = \{\emptyset, \Omega\}$, $\mathcal{F}_2 = \sigma(X + Y)$ und $\mathcal{F}_3 = \sigma(X, Y)$. Dann sind $\sigma(X)$ und $\sigma(Y)$ unabhängig gegeben \mathcal{F}_1 und gegeben \mathcal{F}_3 , nicht jedoch gegeben \mathcal{F}_2 . \diamond

Sei $X = (X_n)_{n \in \mathbb{N}}$ ein stochastischer Prozess auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ mit Werten in einem polnischen Raum E . Sei \mathcal{E} die austauschbare σ -Algebra und \mathcal{T} die terminale σ -Algebra.

Satz 12.24 (de Finetti). Die Familie $X = (X_n)_{n \in \mathbb{N}}$ ist genau dann austauschbar, wenn es eine σ -Algebra $\mathcal{A} \subset \mathcal{F}$ gibt, sodass $(X_n)_{n \in \mathbb{N}}$ u.i.v. gegeben \mathcal{A} ist. In diesem Fall kann $\mathcal{A} = \mathcal{E}$ oder $\mathcal{A} = \mathcal{T}$ gewählt werden.

Beweis. „ \implies “ Sei X austauschbar, und sei $\mathcal{A} = \mathcal{E}$ oder $\mathcal{A} = \mathcal{T}$. Für jedes $n \in \mathbb{N}$ sei $f_n : E \rightarrow \mathbb{R}$ eine messbare und beschränkte Abbildung. Setze

$$\varphi_k(x_1, \dots, x_k) = \prod_{i=1}^k f_i(x_i) \quad \text{für jedes } k \in \mathbb{N}.$$

Dann ist (wobei $A_n(\varphi)$ das symmetrisierte Mittel aus Satz 12.17 ist)

$$\begin{aligned} A_n(\varphi_{k-1})A_n(f_k) &= \frac{1}{n!} \sum_{\varrho \in S(n)} \varphi_{k-1}(X^\varrho) \frac{1}{n} \sum_{i=1}^n f_k(X_i) \\ &= \frac{1}{n!} \sum_{\varrho \in S(n)} \varphi_k(X^\varrho) + R_{n,k} = A_n(\varphi_k) + R_{n,k}, \end{aligned}$$

wobei

$$\begin{aligned} |R_{n,k}| &\leq 2 \|\varphi_{k-1}\|_\infty \cdot \|f_k\|_\infty \cdot \frac{1}{n!} \frac{1}{n} \sum_{\varrho \in S(n)} \sum_{i=1}^n \mathbb{1}_{\{i \in \{\varrho(1), \dots, \varrho(k-1)\}\}} \\ &= 2 \|\varphi_{k-1}\|_\infty \cdot \|f_k\|_\infty \cdot \frac{k-1}{n} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Es folgt zusammen mit Satz 12.17

$$A_n(\varphi_{k-1})A_n(f_k) \xrightarrow{n \rightarrow \infty} \mathbf{E}[\varphi_k(X_1, \dots, X_k) | \mathcal{A}] \text{ f.s. und in } L^1.$$

Andererseits gilt nach Satz 12.17

$$A_n(\varphi_{k-1}) \xrightarrow{n \rightarrow \infty} \mathbf{E}[\varphi_{k-1}(X_1, \dots, X_{k-1}) | \mathcal{A}]$$

und

$$A_n(f_k) \xrightarrow{n \rightarrow \infty} \mathbf{E}[f_k(X_1) | \mathcal{A}],$$

also

$$\mathbf{E}[\varphi_k(X_1, \dots, X_k) | \mathcal{A}] = \mathbf{E}[\varphi_{k-1}(X_1, \dots, X_{k-1}) | \mathcal{A}] \mathbf{E}[f_k(X_1) | \mathcal{A}]$$

und induktiv

$$\mathbf{E}\left[\prod_{i=1}^k f_i(X_i) | \mathcal{A}\right] = \prod_{i=1}^k \mathbf{E}[f_i(X_1) | \mathcal{A}].$$

Mithin ist X u.i.v. gegeben \mathcal{A} .

„ \Leftarrow “ Sei nun X u.i.v. gegeben \mathcal{A} für eine geeignete σ -Algebra $\mathcal{A} \subset \mathcal{F}$. Für jede beschränkte, messbare Funktion $\varphi : E^n \rightarrow \mathbb{R}$ und für jedes $\varrho \in S(n)$ ist dann $\mathbf{E}[\varphi(X) | \mathcal{A}] = \mathbf{E}[\varphi(X^\varrho) | \mathcal{A}]$, also

$$\mathbf{E}[\varphi(X)] = \mathbf{E}[\mathbf{E}[\varphi(X) | \mathcal{A}]] = \mathbf{E}[\mathbf{E}[\varphi(X^\varrho) | \mathcal{A}]] = \mathbf{E}[\varphi(X^\varrho)].$$

Mithin ist X austauschbar. \square

Mit $\mathcal{M}_1(E)$ bezeichnen wir die Menge der W-Maße auf E , ausgestattet mit der Topologie der schwachen Konvergenz (siehe Definition 13.12 und Bemerkung 13.14), das heißt: Eine Folge $(\mu_n)_{n \in \mathbb{N}}$ in $\mathcal{M}_1(E)$ konvergiert genau dann schwach gegen ein $\mu \in \mathcal{M}_1(E)$, wenn $\int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu$ für jede stetige und beschränkte Funktion $f : E \rightarrow \mathbb{R}$. Wir werden die schwache Konvergenz in Kapitel 13 genauer untersuchen. An dieser Stelle wollen wir die Topologie lediglich verwenden, um $\mathcal{M}_1(E)$ zu einem Messraum zu machen, nämlich mit der Borel'schen σ -Algebra $\mathcal{B}(\mathcal{M}_1(E))$. Wir können jetzt Zufallsvariablen mit Werten in $\mathcal{M}_1(E)$ betrachten, so genannte zufällige Maße (vergleiche Kapitel 24.1). Für $x \in E^{\mathbb{N}}$ sei $\xi_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \in \mathcal{M}_1(E)$.

Definition 12.25. Das zufällige Maß $\Xi_n := \xi_n(X) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ heißt **empirische Verteilung** von X_1, \dots, X_n .

Wir betrachten die selben Voraussetzung wie in Satz 12.24.

Satz 12.26 (de Finetti Darstellungssatz). Die Familie $X = (X_n)_{n \in \mathbb{N}}$ ist genau dann austauschbar, wenn es eine σ -Algebra $\mathcal{A} \subset \mathcal{F}$ gibt und eine \mathcal{A} -messbare Zufallsvariable $\Xi_\infty : \Omega \rightarrow \mathcal{M}_1(E)$ mit der Eigenschaft: gegeben Ξ_∞ ist $(X_n)_{n \in \mathbb{N}}$ u.i.v. mit $\mathcal{L}[X_1 | \Xi_\infty] = \Xi_\infty$. In diesem Fall kann $\mathcal{A} = \mathcal{E}$ oder $\mathcal{A} = \mathcal{T}$ gewählt werden.

Beweis. „ \Leftarrow “ Dies ist klar wie im Beweis von Satz 12.24.

„ \Rightarrow “ Sei X austauschbar. Dann existiert nach Satz 12.24 eine σ -Algebra $\mathcal{A} \subset \mathcal{F}$, sodass $(X_n)_{n \in \mathbb{N}}$ u.i.v. gegeben \mathcal{A} ist. Da E polnisch ist, existiert eine reguläre Version der bedingten Verteilung (siehe Satz 8.37) $\Xi_\infty := \mathcal{L}[X_1 | \mathcal{A}]$. Für messbare $A_1, \dots, A_n \subset E$ ist $\mathbf{P}[X_i \in A_i | \mathcal{A}] = \Xi_\infty(A_i)$ für jedes $i = 1, \dots, n$, also

$$\begin{aligned} \mathbf{P}\left[\bigcap_{i=1}^n \{X_i \in A_i\} \mid \Xi_\infty\right] &= \mathbf{E}\left[\mathbf{P}\left[\bigcap_{i=1}^n \{X_i \in A_i\} \mid \mathcal{A}\right] \mid \Xi_\infty\right] \\ &= \mathbf{E}\left[\prod_{i=1}^n \Xi_\infty(A_i) \mid \Xi_\infty\right] = \prod_{i=1}^n \Xi_\infty(A_i). \end{aligned}$$

Mithin ist $\mathcal{L}[X | \Xi_\infty] = \Xi_\infty^{\otimes \mathbb{N}}$. \square

Bemerkung 12.27. (i) In der Situation von Satz 12.26 ist nach dem starken Gesetz der großen Zahl für jede stetige und beschränkte Funktion $f : E \rightarrow \mathbb{R}$

$$\int f d\Xi_n \xrightarrow{n \rightarrow \infty} \int f d\Xi_\infty \quad \text{fast sicher.}$$

Ist E zudem lokalkompakt (zum Beispiel $E = \mathbb{R}^d$), so kann man sogar zeigen, dass

$$\Xi_n \xrightarrow{n \rightarrow \infty} \Xi_\infty \quad \text{fast sicher.}$$

(ii) Für endliche austauschbare Familien gilt die Aussage des Satzes von de Finetti nicht. Eine detaillierte Diskussion findet sich in [38]. \diamond

Beispiel 12.28. Seien $(X_n)_{n \in \mathbb{N}}$ austauschbar und $X_n \in \{0, 1\}$. Dann existiert eine Zufallsvariable $Y : \Omega \rightarrow [0, 1]$, sodass für endliches $J \subset \mathbb{N}$

$$\mathbf{P}[X_j = 1 \text{ für jedes } j \in J | Y] = Y^{\#J}.$$

Mit anderen Worten: Gegeben Y ist $(X_n)_{n \in \mathbb{N}}$ unabhängig und Ber_Y -verteilt. Vergleiche Beispiel 12.3(iii). \diamond

Beispiel 12.29. (Pólya'sches Urnenmodell) (Siehe Beispiel 14.38 sowie [144], [20] und [63].) In einer Urne seien anfangs N Kugeln, davon M schwarz und $N - M$ weiß. In jedem Schritt wird eine Kugel gezogen und zusammen mit einer weiteren Kugel der selben Farbe wieder zurückgelegt. Sei

$$X_n := \begin{cases} 1, & \text{falls die } n\text{-te Kugel schwarz ist,} \\ 0, & \text{sonst,} \end{cases}$$

und $S_n = \sum_{i=1}^n X_i$. Dann ist

$$\mathbf{P}[X_n = 1 | X_1, X_2, \dots, X_{n-1}] = \frac{S_{n-1} + M}{N + n - 1}.$$

Sukzessive erhält man für $x_1, \dots, x_n \in \{0, 1\}$ und $s_k = \sum_{i=1}^k x_i$

$$\begin{aligned} \mathbf{P}[X_i = x_i \text{ für jedes } i = 1, \dots, n] &= \prod_{i \leq n: x_i=1} \frac{M + s_{i-1}}{N + i - 1} \prod_{i \leq n: x_i=0} \frac{N + i - 1 - M - s_{i-1}}{N + i - 1} \\ &= \frac{(N-1)!}{(N-1+n)!} \cdot \frac{(M+s_n-1)!}{(M-1)!} \frac{(N-M-1+(n-s_n))!}{(N-M-1)!}. \end{aligned}$$

Die rechte Seite hängt nur von s_n und nicht von der Reihenfolge der x_1, \dots, x_n ab. Also ist $(X_n)_{n \in \mathbb{N}}$ austauschbar. Sei $Z = \lim_{n \rightarrow \infty} \frac{1}{n} S_n$. Dann ist $(X_n)_{n \in \mathbb{N}}$ unabhängig und identisch Ber_Z -verteilt gegeben Z . Also ist (siehe Beispiel 12.28)

$$\begin{aligned}
\mathbf{E}[Z^n] &= \mathbf{E}[\mathbf{P}[X_1 = \dots = X_n = 1 | Z]] \\
&= \mathbf{P}[S_n = n] \\
&= \frac{(N-1)!}{(M-1)!} \frac{(M+n-1)!}{(N+n-1)!} \quad \text{für jedes } n \in \mathbb{N}.
\end{aligned}$$

Nach Übung 5.1.2 sind dies sind aber gerade die Momente der Beta-Verteilung $\beta_{M,N-M}$ auf $[0, 1]$ mit Parametern $(M, N - M)$ (siehe Beispiel 1.107(ii)). Durch Angabe der Momente ist eine Verteilung auf $[0, 1]$ eindeutig bestimmt (Satz 15.4). Also gilt $Z \sim \beta_{M,N-M}$. \diamond

13 Konvergenz von Maßen

In der Wahrscheinlichkeitstheorie interessiert man sich für Verteilungen, die durch das Zusammenwirken vieler zufälliger Einflüsse zustandekommen. Oftmals lässt sich eine brauchbare Idealisierung erreichen, indem man Grenzwerte solcher Verteilungen anschaut, zum Beispiel, wenn die Anzahl der Einflüsse nach Unendlich geht. Ein Beispiel ist die Konvergenz der Anzahl eingetretener Ereignisse bei vielen seltenen Ereignissen gegen die Poisson-Verteilung (siehe Satz 3.7). Vielfach sind aber auch Skalierungen der ursprünglichen Verteilung notwendig, um das wesentliche Fluktuationsverhalten zu erfassen, etwa im Zentralen Grenzwertsatz. Während diese Sätze mit reellen Zufallsvariablen auskommen, werden wir auch Grenzwertsätze kennenlernen, bei denen die Zufallsvariablen Werte in allgemeineren Räumen annehmen, beispielsweise im Raum aller stetigen Funktionen, wenn wir die zufällige zeitliche Bewegung eines Teilchens modellieren.

In diesem Kapitel wird der Begriff der schwachen Konvergenz von W-Maßen auf allgemeinen (meist polnischen) Räumen eingeführt und untersucht. Hierzu ist eine solide Kenntnis von mengentheoretischer Topologie notwendig. Wir beginnen daher mit einem kurzen Überblick über die verwendeten topologischen Begriffe und Sätze.

Dieses Kapitel soll nur eine knappe Einführung in die für die Wahrscheinlichkeitstheorie wichtigsten Sätze liefern. Als ausführlichere Darstellungen seien [17] und [90] empfohlen.

Beim ersten Lesen mag der Leser dieses eher analytisch geprägte Kapitel vielleicht überspringen. In diesem Fall genügt es fürs Erste, sich mit den Definitionen von schwacher Konvergenz und Straffheit (Definition 13.12 und 13.26) vertraut zu machen, sowie mit den Aussagen des Portemanteau Theorems (Satz 13.16) und des Satzes von Prohorov (Satz 13.29).

13.1 Wiederholung Topologie

Wir geben kurзорisch einige Definitionen und Aussagen der mengentheoretischen Topologie an. Zum Nachlesen eignen sich etwa [97] oder [150].

Im Folgenden sei stets (E, τ) ein topologischer Raum mit der Borel'schen σ -Algebra $\mathcal{E} = \mathcal{B}(E)$ (vergleiche Definition 1.20 und 1.21). Tatsächlich wollen wir auch stets annehmen, dass (E, τ) ein **Hausdorffraum** ist, das heißtt, dass zu je zwei Punkten $x, y \in E$ mit $x \neq y$ offene Mengen U, V existieren mit $x \in U, y \in V$ und $U \cap V = \emptyset$.

Ist $A \subset E$, so bezeichnen wir mit \overline{A} den **Abschluss** von A , mit A° das **Innere** und mit ∂A den **Rand** von A . Eine Menge $A \subset E$ heißtt **dicht**, falls $\overline{A} = E$.

(E, τ) heißtt **metrisierbar**, falls es eine Metrik d auf E gibt, sodass τ durch die offenen Kugeln $B_\varepsilon(x) := \{y \in E : d(x, y) < \varepsilon\}$ erzeugt wird. Eine Metrik d auf E heißtt **vollständig**, falls jede Cauchy-Folge bezüglich d einen Grenzwert in E besitzt. (E, τ) heißtt **vollständig metrisierbar**, falls es eine vollständige Metrik auf E gibt, die τ erzeugt. Ist (E, d) ein metrischer Raum und $A, B \subset E$, so schreiben wir $d(A, B) = \inf\{d(x, y) : x \in A, y \in B\}$ sowie $d(x, B) := d(\{x\}, B)$ für $x \in E$.

Ein metrisierbarer Raum (E, τ) heißtt **separabel**, falls es eine abzählbare, dichte Teilmenge von E gibt. Separabilität in metrisierbaren Räumen ist äquivalent dazu, dass es eine **abzählbare Basis der Topologie** gibt, also eine abzählbare Menge $\mathcal{U} \subset \tau$ mit $A = \bigcup_{U \in \mathcal{U}: U \subset A} U$ für jedes $A \in \tau$. (Man wähle etwa ε -Kugeln um die Punkte aus einer abzählbaren, dichten Teilmenge und lasse ε die positiven rationalen Zahlen durchlaufen.) Ein kompakter, metrischer Raum ist stets separabel (man wähle einfach zu jedem $n \in \mathbb{N}$ eine endliche Überdeckung $\mathcal{U}_n \subset \tau$ mit Kugeln vom Radius $\frac{1}{n}$ und nehme dann $\mathcal{U} := \bigcup_{n \in \mathbb{N}} \mathcal{U}_n$).

Eine Menge $A \subset E$ heißtt **kompakt**, falls zu jeder offenen Überdeckung $\mathcal{U} \subset \tau$ von A (das heißtt, $A \subset \bigcup_{U \in \mathcal{U}} U$) eine endliche Teilüberdeckung existiert, also eine endliche Menge $\mathcal{U}' \subset \mathcal{U}$ mit $A \subset \bigcup_{U \in \mathcal{U}'} U$. Kompakte Mengen sind stets abgeschlossen. Nach dem Satz von Heine-Borel ist eine Teilmenge von \mathbb{R}^d genau dann kompakt, wenn sie beschränkt und abgeschlossen ist. $A \subset E$ heißtt **relativ kompakt**, falls \overline{A} kompakt ist. Hingegen heißtt A **folgenkompakt** (beziehungsweise **relativ folgenkompakt**), falls jede Folge $(x_n)_{n \in \mathbb{N}}$ mit Werten in A eine Teilfolge $(x_{n_k})_{k \in \mathbb{N}}$ hat, die gegen einen Grenzwert $x \in A$ (beziehungsweise $x \in \overline{A}$) konvergiert. In metrisierbaren Räumen fallen die Begriffe *kompakt* und *folgenkompakt* zusammen. Eine Menge $A \subset E$ heißtt **σ -kompakt**, falls A abzählbare Vereinigung von kompakten Mengen ist. E heißtt **lokalkompakt**, falls jeder Punkt $x \in E$ eine offene Umgebung besitzt, deren Abschluss kompakt ist. Ein lokalkompaakter, separabler, metrischer Raum ist offenbar stets σ -kompakt. Ist E lokalkompakt und metrisch und $U \subset E$ offen sowie $K \subset U$ kompakt, so existiert eine kompakte Menge L mit $K \subset L^\circ \subset L \subset U$. (Man bilde etwa zu jedem Punkt $x \in K$ eine offene Kugel $B_{\varepsilon_x}(x)$ vom Radius $\varepsilon_x > 0$, die ganz in U liegt und relativ kompakt ist. Indem man ε_x eventuell nochmal halbiert, kann man annehmen, dass sogar der Abschluss dieser Kugel in U liegt. Da K kompakt ist, gibt es endlich viele $x_1, \dots, x_n \in K$ mit $K \subset V := \bigcup_{i=1}^n B_{\varepsilon_{x_i}}(x_i)$. Nach Konstruktion ist $L = \overline{V} \subset U$ kompakt.)

Einen in der Wahrscheinlichkeitstheorie wichtigen Typ von topologischen Räumen stellen wir in einer separaten Definition vor.

Definition 13.1. Ein topologischer Raum (E, τ) heißt **polnischer Raum**, falls er vollständig metrisierbar und separabel ist.

Polnische Räume sind beispielsweise abzählbare, diskrete Räume (nicht jedoch \mathbb{Q} mit der üblichen Topologie), die euklidischen Räume \mathbb{R}^n , aber auch der Raum $C([0, 1])$ der stetigen Funktionen $[0, 1] \rightarrow \mathbb{R}$, ausgestattet mit der Supremumsnorm $\|\cdot\|_\infty$. Praktisch sind alle Räume, die in der Wahrscheinlichkeitstheorie bedeutsam sind, polnische Räume.

Sei (E, d) ein metrischer Raum. Eine Menge $A \subset E$ heißt **total beschränkt**, falls es zu jedem $\varepsilon > 0$ endlich viele Punkte $x_1, \dots, x_n \in A$ gibt mit $A \subset \bigcup_{i=1}^n B_\varepsilon(x_i)$. Kompakte Mengen sind offenbar total beschränkt. In polnischen Räumen gilt sogar:

Lemma 13.2. Sei (E, τ) polnisch mit vollständiger Metrik d . Eine Teilmenge $A \subset E$ ist genau dann total beschränkt bezüglich d , wenn A relativ kompakt ist.

Beweis. Übung! □

Im Folgenden sei stets (E, τ) ein topologischer Raum mit Borel'scher σ -Algebra $\mathcal{E} = \mathcal{B}(E) := \sigma(\tau)$ und vollständiger Metrik d . Für Maße auf (E, \mathcal{E}) führen wir die folgenden Regularitätsbegriffe ein.

Definition 13.3. Ein σ -endliches Maß μ auf (E, \mathcal{E}) heißt

- (i) **lokal endlich** oder **Borel-Maß**, falls es zu jedem Punkt $x \in E$ eine offene Menge $U \ni x$ gibt mit $\mu(U) < \infty$,
- (ii) **regulär von innen**, falls

$$\mu(A) = \sup \{\mu(K) : K \subset A \text{ ist kompakt}\} \quad \text{für jedes } A \in \mathcal{E}, \quad (13.1)$$

- (iii) **regulär von außen**, falls

$$\mu(A) = \inf \{\mu(U) : U \supset A \text{ ist offen}\} \quad \text{für jedes } A \in \mathcal{E}, \quad (13.2)$$

- (iv) **regulär**, falls μ von innen und von außen regulär ist,

- (v) **Radon-Maß**, falls μ ein von innen reguläres Borel-Maß ist.

Definition 13.4. Wir führen die folgenden Mengen von Maßen auf E ein.

$$\begin{aligned}\mathcal{M}(E) &:= \{\text{Radon-Maße auf } (E, \mathcal{E})\}, \\ \mathcal{M}_f(E) &:= \{\text{endliche Maße auf } (E, \mathcal{E})\}, \\ \mathcal{M}_1(E) &:= \{\mu \in \mathcal{M}_f(E) : \mu(E) = 1\}, \\ \mathcal{M}_{\leq 1}(E) &:= \{\mu \in \mathcal{M}_f(E) : \mu(E) \leq 1\}.\end{aligned}$$

Die Elemente von $\mathcal{M}_{\leq 1}(E)$ nennen wir **Sub-Wahrscheinlichkeitsmaße** auf E .

Ferner vereinbaren wir die folgende Notation für Mengen von stetigen Funktionen

$$\begin{aligned}C(E) &:= \{f : E \rightarrow \mathbb{R} \text{ ist stetig}\}, \\ C_b(E) &:= \{f \in C(E) \text{ ist beschränkt}\}, \\ C_c(E) &:= \{f \in C(E) \text{ hat kompakten Träger}\} \subset C_b(E).\end{aligned}$$

Der Träger einer reellen Funktion f ist dabei $\overline{f^{-1}(\mathbb{R} \setminus \{0\})}$.

Ist nichts anderes vereinbart, so sind die Vektorräume $C(E)$, $C_b(E)$ und $C_c(E)$ mit der Supremumsnorm ausgestattet.

Lemma 13.5. Ist E polnisch und $\mu \in \mathcal{M}_f(E)$, so existiert zu jedem $\varepsilon > 0$ eine kompakte Menge $K \subset E$ mit $\mu(E \setminus K) < \varepsilon$.

Beweis. Sei $\varepsilon > 0$. Zu jedem $n \in \mathbb{N}$ existieren $x_1^n, x_2^n, \dots \in E$ mit $E = \bigcup_{i=1}^{\infty} B_{1/n}(x_i^n)$. Wähle $N_n \in \mathbb{N}$ so, dass $\mu\left(E \setminus \bigcup_{i=1}^{N_n} B_{1/n}(x_i^n)\right) < \frac{\varepsilon}{2^n}$. Setze

$$A := \bigcap_{n=1}^{\infty} \bigcup_{i=1}^{N_n} B_{1/n}(x_i^n).$$

Nach Konstruktion ist A total beschränkt. Da E polnisch ist, ist also \overline{A} kompakt. Außerdem folgt $\mu(E \setminus \overline{A}) \leq \mu(E \setminus A) < \sum_{n=1}^{\infty} \varepsilon 2^{-n} = \varepsilon$. \square

Satz 13.6. Ist E polnisch und $\mu \in \mathcal{M}_f(E)$, so ist μ regulär. Speziell ist dann $\mathcal{M}_f(E) \subset \mathcal{M}(E)$.

Beweis. (Regularität von außen) 1. Schritt. Sei zunächst $B \subset E$ abgeschlossen und $\varepsilon > 0$. Sei d eine vollständige Metrik auf E . Für $\delta > 0$ sei

$$B_\delta := \{x \in E : d(x, B) < \delta\}$$

die offene δ -Umgebung von B . Da B abgeschlossen ist, gilt $\bigcap_{\delta>0} B_\delta = B$. Da μ stetig von oben ist (Satz 1.36), gibt es ein $\delta > 0$ mit $\mu(B_\delta) \leq \mu(B) + \varepsilon$.

2. Schritt. Sei $B \in \mathcal{E}$ und $\varepsilon > 0$. Betrachte das Mengensystem

$$\mathcal{A} := \{V \cap C : V \subset E \text{ offen}, C \subset E \text{ abgeschlossen}\}.$$

Man prüft leicht nach, dass \mathcal{A} ein Semiring ist, und offenbar gilt $\mathcal{E} = \sigma(\mathcal{A})$. Nach dem Approximationssatz für Maße (Satz 1.65) existieren paarweise disjunkte Mengen $A_n = V_n \cap C_n \in \mathcal{A}$, $n \in \mathbb{N}$, mit $B \subset A := \bigcup_{n=1}^{\infty} A_n$ und $\mu(A) \leq \mu(B) + \varepsilon/2$. Wie im 1. Schritt gezeigt, existiert für jedes $n \in \mathbb{N}$ ein offenes $W_n \supset C_n$ mit $\mu(W_n) \leq \mu(C_n) + \varepsilon 2^{-n-1}$. Es ist also $U_n := V_n \cap W_n$ offen, $B \subset U := \bigcup_{n=1}^{\infty} U_n$ und $\mu(U) \leq \mu(A) + \sum_{n=1}^{\infty} \varepsilon 2^{-n-1} \leq \mu(B) + \varepsilon$.

Regularität von innen. Das selbe Argument mit B^c statt B liefert die Existenz einer abgeschlossenen Menge $D \subset B$ mit $\mu(B \setminus D) < \varepsilon/2$. Nach Lemma 13.5 existiert ein Kompaktum K mit $\mu(K^c) < \varepsilon/2$. Setze $C = D \cap K$. Dann ist $C \subset B$ kompakt und $\mu(B \setminus C) < \varepsilon$. Also ist μ auch regulär von innen. \square

Korollar 13.7. Das Lebesgue-Maß λ auf \mathbb{R}^d ist ein reguläres Radon-Maß. Jedoch ist nicht jedes σ -endliche Maß auf \mathbb{R}^d regulär.

Beweis. Offenbar ist \mathbb{R}^d polnisch und λ lokal endlich. Sei $A \in \mathcal{B}(\mathbb{R}^d)$ und $\varepsilon > 0$. Es existiert eine aufsteigende Folge $(K_n)_{n \in \mathbb{N}}$ von kompakten Mengen mit $K_n \uparrow \mathbb{R}^d$. Da jedes K_n beschränkt ist, ist $\lambda(K_n) < \infty$. Es existiert also nach dem vorangehenden Satz zu jedem $n \in \mathbb{N}$ eine offene Menge $U_n \supset A \cap K_n$ mit $\lambda(U_n \setminus A) < \varepsilon/2^n$. Für die offene Menge $U := \bigcup_{n \in \mathbb{N}} U_n$ gilt daher $\lambda(U \setminus A) < \varepsilon$.

Ist $\lambda(A) < \infty$, so existiert ein $n \in \mathbb{N}$ mit $\lambda(A \setminus K_n) < \varepsilon/2$. Nach dem vorangehenden Satz existiert eine kompakte Menge $C \subset A \cap K_n$ mit $\lambda((A \cap K_n) \setminus C) < \varepsilon/2$. Es folgt $\lambda(A \setminus C) < \varepsilon$.

Ist andererseits $\lambda(A) = \infty$, so müssen wir zu jedem $L > 0$ ein Kompaktum $C \subset A$ finden mit $\lambda(C) > L$. Nun gilt aber $\lambda(A \cap K_n) \xrightarrow{n \rightarrow \infty} \infty$, also existiert ein $n \in \mathbb{N}$ mit $\lambda(A \cap K_n) > L + 1$. Nach dem schon Gezeigten existiert ein Kompaktum $C \subset A \cap K_n$ mit $\lambda((A \cap K_n) \setminus C) < 1$, also $\lambda(C) > L$.

Schließlich sei das Maß $\mu = \sum_{q \in \mathbb{Q}} \delta_q$ betrachtet. Dieses Maß ist offenbar σ -endlich, jedoch nicht lokal endlich und auch nicht regulär von außen. \square

Definition 13.8. Seien (E, d_E) und (F, d_F) metrische Räume. Eine Funktion $f : E \rightarrow F$ heißt **Lipschitz-stetig**, falls es eine Zahl $K < \infty$ gibt, die so genannte Lipschitz-Konstante, mit $d_F(f(x), f(y)) \leq K \cdot d_E(x, y)$ für alle $x, y \in E$. Mit $\text{Lip}_K(E; F)$ bezeichnen wir den Raum der Lipschitz-stetigen Funktionen zur Konstanten K , und mit $\text{Lip}(E; F) = \bigcup_{K>0} \text{Lip}_K(E; F)$ den Raum der Lipschitz-stetigen Funktionen auf E .

Wir schreiben kurz $\text{Lip}_K(E) := \text{Lip}_K(E; \mathbb{R})$ und $\text{Lip}(E) := \text{Lip}(E; \mathbb{R})$.

Definition 13.9. Sei $\mathcal{F} \subset \mathcal{M}(E)$ eine Familie von Radon-Maßen. Eine Familie \mathcal{C} messbarer Abbildungen $E \rightarrow \mathbb{R}$ heißt **trennende Familie** für \mathcal{F} , falls für je zwei Maße $\mu, \nu \in \mathcal{F}$ gilt:

$$\left(\int f d\mu = \int f d\nu \quad \text{für jedes } f \in \mathcal{C} \cap \mathcal{L}^1(\mu) \cap \mathcal{L}^1(\nu) \right) \implies \mu = \nu.$$

Lemma 13.10. Sei (E, d) ein metrischer Raum. Zu jeder abgeschlossenen Menge $A \subset E$ und jedem $\varepsilon > 0$ gibt es eine Lipschitz-stetige Abbildung $\rho_{A,\varepsilon} : E \rightarrow [0, 1]$ mit

$$\rho_{A,\varepsilon}(x) = \begin{cases} 1, & \text{falls } x \in A, \\ 0, & \text{falls } d(x, A) \geq \varepsilon. \end{cases}$$

Beweis. Sei $\varphi : \mathbb{R} \rightarrow [0, 1]$, $t \mapsto (t \vee 0) \wedge 1$. Für $x \in E$ setze $\rho_{A,\varepsilon}(x) = 1 - \varphi(\varepsilon^{-1}d(x, A))$. \square

Satz 13.11. Sei (E, d) ein metrischer Raum.

- (i) $\text{Lip}_1(E; [0, 1])$ ist trennend für $\mathcal{M}(E)$.
- (ii) Ist E zudem lokalkompakt, so ist $C_c(E) \cap \text{Lip}_1(E; [0, 1])$ trennend für $\mathcal{M}(E)$.

Beweis. (i) Seien $\mu_1, \mu_2 \in \mathcal{M}(E)$ mit $\int f d\mu_1 = \int f d\mu_2$ für jedes $f \in \text{Lip}_1(E; [0, 1])$. Ist $A \in \mathcal{E}$, so ist $\mu_i(A) = \sup\{\mu_i(K) : K \subset A \text{ ist kompakt}\}$, da das Radon-Maß μ_i von innen regulär ist ($i = 1, 2$). Es reicht also zu zeigen, dass $\mu_1(K) = \mu_2(K)$ für jede kompakte Menge K .

Sei nun $K \subset E$ kompakt. Da μ_1 und μ_2 lokal endlich sind, existiert zu jedem $x \in K$ eine offene Menge $U_x \ni x$ mit $\mu_1(U_x) < \infty$ und $\mu_2(U_x) < \infty$. Da K kompakt ist, können wir endlich viele Punkte $x_1, \dots, x_n \in K$ finden, sodass $K \subset U := \bigcup_{j=1}^n U_{x_j}$. Nach Konstruktion ist $\mu_i(U) < \infty$, also $\mathbb{1}_U \in L^1(\mu_i)$ für $i = 1, 2$. Da U^c abgeschlossen ist, und $U^c \cap K = \emptyset$, ist $\delta := d(U^c, K) > 0$. Für die Abbildung $\rho_{K,\varepsilon}$ aus Lemma 13.10 ist also $\mathbb{1}_K \leq \rho_{K,\varepsilon} \leq \mathbb{1}_U \in L^1(\mu_i)$, falls $\varepsilon \in (0, \delta)$. Wegen $\rho_{K,\varepsilon} \xrightarrow{\varepsilon \rightarrow 0} \mathbb{1}_K$ folgt aus dem Satz über majorisierte Konvergenz (Korollar 6.26) $\mu_i(K) = \lim_{\varepsilon \rightarrow 0} \int \rho_{K,\varepsilon} d\mu_i$. Nun ist aber $\varepsilon \rho_{K,\varepsilon} \in \text{Lip}_1(E; [0, 1])$ für jedes $\varepsilon > 0$, also nach Voraussetzung

$$\int \rho_{K,\varepsilon} d\mu_1 = \varepsilon^{-1} \int (\varepsilon \rho_{K,\varepsilon}) d\mu_1 = \varepsilon^{-1} \int (\varepsilon \rho_{K,\varepsilon}) d\mu_2 = \int \rho_{K,\varepsilon} d\mu_2.$$

Es folgt $\mu_1(K) = \mu_2(K)$, also $\mu_1 = \mu_2$.

(ii) Ist E lokalkompakt, so können wir in (i) die Umgebungen U_x zusätzlich relativ kompakt wählen. Es ist dann U relativ kompakt, also hat $\rho_{K,\varepsilon}$ für $\varepsilon \in (0, \delta)$ einen kompakten Träger, ist also in $C_c(E)$. \square

Übung 13.1.1. (i) Man zeige, dass $C([0, 1])$ eine abzählbare, dichte Teilmenge besitzt.

(ii) Man zeige, dass der Raum $(C_b([0, \infty)), \|\cdot\|_\infty)$ der stetigen, beschränkten Funktionen mit der Supremumsnorm nicht separabel ist.

(iii) Man zeige, dass der Raum $C_c([0, \infty))$ der stetigen Funktionen mit kompaktem Träger, ausgestattet mit der Supremumsnorm, separabel ist. ♣

Übung 13.1.2. Man zeige: Ist μ ein lokal endliches Maß, so ist $\mu(K) < \infty$ für jede kompakte Menge K . ♣

Übung 13.1.3 (Satz von Lusin). Sei Ω ein polnischer Raum, μ ein σ -endliches Maß auf $(\Omega, \mathcal{B}(\Omega))$ und $f : \Omega \rightarrow \mathbb{R}$ eine Abbildung. Man zeige, dass die beiden folgenden Aussagen äquivalent sind:

(i) Es gibt eine Borel-messbare Abbildung $g : \Omega \rightarrow \mathbb{R}$ mit $f = g$ μ -fast überall.

(ii) Zu jedem $\varepsilon > 0$ gibt es eine kompakte Menge K_ε mit $\mu(\Omega \setminus K_\varepsilon) < \varepsilon$, sodass die eingeschränkte Funktion $f|_{K_\varepsilon}$ stetig ist. ♣

Übung 13.1.4. Sei \mathcal{U} eine Familie offener Intervalle in \mathbb{R} so, dass $W := \bigcup_{U \in \mathcal{U}} U$ endliches Lebesgue-Maß $\lambda(W)$ hat. Man zeige: Für jedes $\varepsilon > 0$ gibt es endlich viele, paarweise disjunkte Mengen $U_1, \dots, U_n \in \mathcal{U}$ mit

$$\sum_{i=1}^n \lambda(U_i) > \frac{1-\varepsilon}{3} \lambda(W).$$

Hinweis: Man wähle eine endliche Familie $\mathcal{U}' \subset \mathcal{U}$, sodass $\bigcup_{U \in \mathcal{U}'} U$ das Maß mindestens $(1 - \varepsilon)\lambda(W)$ hat. Hieraus wähle man eine nach absteigender Länge sortierte maximale Folge \mathcal{U}'' *disjunkter* Intervalle aus und zeige, dass jedes $U \in \mathcal{U}'$ in $(x - 3a, x + 3a)$ liegt für ein $(x - a, x + a) \in \mathcal{U}''$. ♣

Übung 13.1.5. Sei $C \subset \mathbb{R}^d$ eine offene, beschränkte und konvexe Menge und $\mathcal{U} \subset \{x + rC : x \in \mathbb{R}^d, r > 0\}$ so gewählt, dass $W := \bigcup_{U \in \mathcal{U}} U$ endliches Lebesgue-Maß $\lambda^d(W)$ hat. Man zeige: Für jedes $\varepsilon > 0$ gibt es endlich viele, paarweise disjunkte Mengen $U_1, \dots, U_n \in \mathcal{U}$ mit

$$\sum_{i=1}^n \lambda^d(U_i) > \frac{1-\varepsilon}{3^d} \lambda(W).$$

Man überlege sich ein Gegenbeispiel, das zeigt, dass man auf die Bedingung der Ähnlichkeit der offenen Mengen aus \mathcal{U} nicht ohne Weiteres verzichten kann. ♣

Übung 13.1.6. Sei μ ein Radon-Maß auf \mathbb{R}^d und $A \in \mathcal{B}(\mathbb{R}^d)$ eine μ -Nullmenge. Man zeige mit Hilfe von Übung 13.1.5, dass für jede beschränkte, konvexe und offene Menge $C \subset \mathbb{R}^d$ mit $0 \in C$ gilt:

$$\lim_{r \downarrow 0} \frac{\mu(x + rC)}{r^d} = 0 \quad \text{für } \lambda^d\text{-fast alle } x \in A.$$

Man folgere: Ist F die Verteilungsfunktion eines Stieltjes-Maßes μ auf \mathbb{R} und $A \in \mathcal{B}(\mathbb{R})$ eine μ -Nullmenge, so gilt $\frac{d}{dx}F(x) = 0$ für λ -fast alle $x \in A$. \clubsuit

Übung 13.1.7 (Hauptsatz der Differential- und Integralrechnung). Sei $f \in \mathcal{L}^1(\mathbb{R}^d)$, $\mu = f \lambda^d$ und $C \subset \mathbb{R}^d$ offen, konvex und beschränkt mit $0 \in C$. Man zeige:

$$\lim_{r \downarrow 0} \frac{\mu(x + rC)}{r^d \lambda^d(C)} = f(x) \quad \text{für } \lambda^d\text{-fast alle } x \in \mathbb{R}^d.$$

Man folgere für den Fall $d = 1$ den Hauptsatz der Differential- und Integralrechnung:

$$\frac{d}{dx} \int_{[0,x]} f d\lambda = f(x) \quad \text{für } \lambda\text{-fast alle } x \in \mathbb{R}.$$

Hinweis: Verwende Übung 13.1.6 mit $\mu_q(dx) = (f(x) - q)^+ \lambda^d(dx)$ für $q \in \mathbb{Q}$ sowie die Ungleichung

$$\frac{\mu(x + rC)}{r^d \lambda^d(C)} \leq q + \frac{\mu_q(x + rC)}{r^d \lambda^d(C)}. \quad \clubsuit$$

Übung 13.1.8. Man zeige, ähnlich wie in Korollar 13.7: Sei E σ -kompakt und polnisch sowie μ ein Maß auf E . Dann ist μ genau dann ein Radon-Maß, wenn $\mu(K) < \infty$ für jedes kompakte $K \subset E$ gilt. \clubsuit

13.2 Schwache und vage Konvergenz

Nachdem wir in Satz 13.11 gesehen haben, dass Integrale stetiger, beschränkter Funktionen, beziehungsweise für lokalkompaktes E sogar stetiger Funktionen mit kompaktem Träger, ein Radon-Maß vollständig bestimmen, liegt es nahe, $C_b(E)$ und $C_c(E)$ auch als Klassen von Testfunktionen für *Konvergenzbegriffe* für Maße heranzuziehen.

Definition 13.12 (Schwache und vage Konvergenz). Sei E ein metrischer Raum.

- (i) Seien $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_f(E)$. Wir sagen, dass $(\mu_n)_{n \in \mathbb{N}}$ **schwach** (weakly) gegen μ konvergiere, in Formeln $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ schwach oder $\mu = \text{w-lim}_{n \rightarrow \infty} \mu_n$, falls

$$\int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu \quad \text{für jedes } f \in C_b(E).$$

- (ii) Es seien $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}(E)$. Wir sagen, dass $(\mu_n)_{n \in \mathbb{N}}$ **vag** (vaguely) gegen μ konvergiert, in Formeln $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ vag oder $\mu = \text{v-lim}_{n \rightarrow \infty} \mu_n$, falls

$$\int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu \quad \text{für jedes } f \in C_c(E).$$

Bemerkung 13.13. Ist E polnisch, so ist nach Satz 13.6 und 13.11 der schwache Limes eindeutig. Das Gleiche gilt für den vagen Limes, falls E lokalkompakt ist. \diamond

Bemerkung 13.14. (i) In der Funktionalanalysis wird unter schwacher Konvergenz etwas anderes verstanden: Ausgehend von einem normierten Vektorraum X (hier der Raum der endlichen signierten Maße mit der Totalvariationsnorm) wird der Raum X' der stetigen linearen Funktionale $X \rightarrow \mathbb{R}$ betrachtet. Die Folge (μ_n) aus X konvergiert *schwach* gegen $\mu \in X$, falls $\Phi(\mu_n) \xrightarrow{n \rightarrow \infty} \Phi(\mu)$ für jedes $\Phi \in X'$. Im hier betrachteten Falle ist dies äquivalent dazu, dass (μ_n) beschränkt ist und $\mu_n(A) \xrightarrow{n \rightarrow \infty} \mu(A)$ für jedes messbare A gilt (siehe [43, Theorem IV.9.5]). Im Vergleich zu Satz 13.16(vi) wird klar, dass es sich um einen stärkeren Konvergenzbegriff handelt.

(ii) Die in Definition 13.12 eingeführte schwache Konvergenz erzeugt auf $\mathcal{M}_f(E)$ die **schwache Topologie** τ_w . Dies ist die grösste Topologie, sodass für jedes $f \in C_b(E)$ die Abbildung $\mathcal{M}_f(E) \rightarrow \mathbb{R}$, $\mu \mapsto \int f d\mu$ stetig ist. In der Funktionalanalysis entspricht τ_w hingegen mehr dem Konzept der Schwach*-Konvergenz. Ausgehend von einem normierten Vektorraum X (hier $X = C_b(E)$ mit der $\|\cdot\|_\infty$ -Norm) wird auf dem Dualraum X' eine Topologie (die Schwach*-Topologie) eingeführt durch $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ genau dann, wenn $\mu_n(x) \xrightarrow{n \rightarrow \infty} \mu(x)$ für jedes $x \in X$. Offenbar ist jedes endliche Maß μ eine stetige Linearform auf $C_b(E)$ durch $f \mapsto \mu(f) := \int f d\mu$, also ist $\mathcal{M}_f(E) \subset C_b(E)'$. Damit ist τ_w die Spurtopologie der Schwach*-Topologie auf $\mathcal{M}_f(E)$.

(iii) Ist E separabel, so kann man zeigen, dass $(\mathcal{M}_f(E), \tau_w)$ metrisierbar ist, zum Beispiel mit der so genannten **Prohorov-Metrik**

$$d_P(\mu, \nu) := \max\{d'_P(\mu, \nu), d'_P(\nu, \mu)\}, \quad (13.3)$$

wobei

$$d'_P(\mu, \nu) := \inf \{ \varepsilon > 0 : \mu(B) \leq \nu(B^\varepsilon) + \varepsilon \text{ für jedes } B \in \mathcal{B}(E) \}, \quad (13.4)$$

und wo $B^\varepsilon = \{x : d(x, B) < \varepsilon\}$ ist, siehe etwa [17, Appendix III, Theorem 5]. (Man kann zeigen, dass $d'_P(\mu, \nu) = d'_P(\nu, \mu)$ ist, falls $\mu, \nu \in \mathcal{M}_1(E)$.) Ist E lokalkompakt und polnisch, so ist $(\mathcal{M}_f(E), \tau_w)$ sogar polnisch (siehe [145, Seite 167]).

(iv) Analog ist die **vage Topologie** τ_v auf $\mathcal{M}(E)$ die grösste Topologie, sodass für jedes $f \in C_c(E)$ die Abbildung $\mathcal{M}(E) \rightarrow \mathbb{R}$, $\mu \mapsto \int f d\mu$ stetig ist. Ist E lokalkompakt, so ist $(\mathcal{M}(E), \tau_v)$ ein Hausdorffraum. Ist E zudem polnisch, so ist $(\mathcal{M}(E), \tau_v)$ ebenfalls polnisch (siehe etwa [90, Sektion 15.7]). \diamond

Während bei der schwachen Konvergenz stets auch Konvergenz der Gesamtmassen gilt (schließlich ist $1 \in C_b(E)$), kann bei der vagen Konvergenz ein Massendefekt im Limes auftreten, jedoch kein Massenzuwachs, wie das folgende Lemma zeigt.

Lemma 13.15. *Sei E ein lokalkompakter, polnischer Raum, und seien $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}(E)$ mit $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ vag. Dann gilt*

$$\mu(E) \leq \liminf_{n \rightarrow \infty} \mu_n(E).$$

Beweis. Sei $(f_N)_{N \in \mathbb{N}}$ eine Folge in $C_c(E; [0, 1])$ mit $f_N \uparrow 1$. Dann gilt

$$\begin{aligned} \mu(E) &= \sup_{N \in \mathbb{N}} \int f_N d\mu = \sup_{N \in \mathbb{N}} \lim_{n \rightarrow \infty} \int f_N d\mu_n \\ &\leq \liminf_{n \rightarrow \infty} \sup_{N \in \mathbb{N}} \int f_N d\mu_n = \liminf_{n \rightarrow \infty} \mu_n(E). \end{aligned} \quad \square$$

Die Folge $(\delta_{1/n})_{n \in \mathbb{N}}$ von W-Maßen auf \mathbb{R} konvergiert offenbar schwach gegen δ_0 , nicht jedoch in der Totalvariationsnorm: Für die abgeschlossene Menge $(-\infty, 0]$ gilt nämlich $\lim_{n \rightarrow \infty} \delta_{1/n}((-\infty, 0]) = 0 < 1 = \delta_0((-\infty, 0])$. Etwas lax gesagt, kann in abgeschlossene Mengen im schwachen Limes Masse an den Rändern einwandern (nicht jedoch auswandern). Die komplementäre Aussage gilt für offene Mengen, denn $\lim_{n \rightarrow \infty} \delta_{1/n}((0, \infty)) = 1 > 0 = \delta_0((0, \infty))$; hier kann also Masse auswandern, nicht jedoch einwandern. Tatsächlich kann man die schwache Konvergenz über diese Eigenschaft charakterisieren. Im folgenden Satz werden wir ein ganzes Bündel solcher Aussagen „auf einen Kleiderbügel (französisch: portemanteau) hängen“.

Für messbares $g : \Omega \rightarrow \mathbb{R}$ sei U_g die Menge der Unstetigkeitsstellen von g . Beachte, dass U_g Borel-messbar ist (nach Übung 1.1.3).

Satz 13.16 (Portemanteau Theorem). Sei E ein metrischer Raum, und seien $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_{\leq 1}(E)$. Dann sind äquivalent:

$$(i) \mu = \text{w-lim}_{n \rightarrow \infty} \mu_n.$$

$$(ii) \int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu \text{ für alle beschränkten, Lipschitz-stetigen } f.$$

$$(iii) \int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu \text{ für alle beschränkten, messbaren } f \text{ mit } \mu(U_f) = 0.$$

(iv) Es gilt $\liminf_{n \rightarrow \infty} \mu_n(E) \geq \mu(E)$ und $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$ für alle abgeschlossenen $F \subset E$.

(v) Es gilt $\limsup_{n \rightarrow \infty} \mu_n(E) \leq \mu(E)$ und $\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu(G)$ für alle offenen $G \subset E$.

$$(vi) \lim_{n \rightarrow \infty} \mu_n(A) = \mu(A) \text{ für alle messbaren } A \text{ mit } \mu(\partial A) = 0.$$

Ist E auch lokalkompakt und polnisch, so sind zudem jeweils äquivalent

$$(vii) \mu = \text{v-lim}_{n \rightarrow \infty} \mu_n \text{ und } \mu(E) = \lim_{n \rightarrow \infty} \mu_n(E).$$

$$(viii) \mu = \text{v-lim}_{n \rightarrow \infty} \mu_n \text{ und } \mu(E) \geq \limsup_{n \rightarrow \infty} \mu_n(E).$$

Beweis. „(iv) \iff (v) \implies (vi)“ Dies ist trivial.

„(iii) \implies (i) \implies (ii)“ Dies ist trivial.

„(ii) \implies (iv)“ Die Konvergenz der Gesamtmassen folgt mit der Testfunktion $1 \in \text{Lip}(E; [0, 1])$. Sei F abgeschlossen und $\rho_{F,\varepsilon}$ wie in Lemma 13.10. Dann ist

$$\limsup_{n \rightarrow \infty} \mu_n(F) \leq \inf_{\varepsilon > 0} \lim_{n \rightarrow \infty} \int \rho_{F,\varepsilon} d\mu_n = \inf_{\varepsilon > 0} \int \rho_{F,\varepsilon} d\mu = \mu(F),$$

weil $\rho_{F,\varepsilon}(x) \xrightarrow{\varepsilon \rightarrow 0} \mathbb{1}_F(x)$ für jedes $x \in E$.

„(viii) \implies (vii)“ Dies ist klar nach Lemma 13.15.

„(i) \implies (vii)“ Wegen $C_c(E) \subset C_b(E)$ und $1 \in C_b(E)$ ist dies klar.

„(vii) \implies (v)“ Sei G offen und $\varepsilon > 0$. Da μ von innen regulär ist (Satz 13.6), gibt es ein Kompaktum $K \subset G$ mit $\mu(G) - \mu(K) < \varepsilon$. Da E lokalkompakt ist, gibt es ein Kompaktum L mit $K \subset L^\circ \subset L \subset G$. Sei $\delta := d(K, L^c) > 0$ und $\rho_{K,\delta}$ wie in Lemma 13.10. Dann ist $\mathbb{1}_K \leq \rho_{K,\delta} \leq \mathbb{1}_L$, also $\rho_{K,\delta} \in C_c(E)$ und daher

$$\liminf_{n \rightarrow \infty} \mu_n(G) \geq \liminf_{n \rightarrow \infty} \int \rho_{K,\delta} d\mu_n = \int \rho_{K,\delta} d\mu \geq \mu(K) \geq \mu(G) - \varepsilon.$$

Indem wir $\varepsilon \rightarrow 0$ gehen lassen, folgt die Aussage von (v).

„(vi) \implies (iii)“ Sei $f : E \rightarrow \mathbb{R}$ beschränkt und messbar mit $\mu(U_f) = 0$. Wir machen die elementare Beobachtung, dass für jedes $D \subset \mathbb{R}$ gilt

$$\partial f^{-1}(D) \subset f^{-1}(\partial D) \cup U_f. \quad (13.5)$$

In der Tat: Falls f in $x \in E$ stetig ist, so existiert zu jedem $\delta > 0$ ein $\varepsilon(\delta) > 0$ mit $f(B_{\varepsilon(\delta)}(x)) \subset B_\delta(f(x))$. Ist $x \in \partial f^{-1}(D)$, so existieren $y \in f^{-1}(D) \cap B_{\varepsilon(\delta)}(x)$ und $z \in f^{-1}(D^c) \cap B_{\varepsilon(\delta)}(x)$. Also ist $f(y) \in B_\delta(f(x)) \cap D \neq \emptyset$ und $f(z) \in B_\delta(f(x)) \cap D^c \neq \emptyset$, also $f(x) \in \partial D$.

Sei $\varepsilon > 0$. Offenbar ist die Menge $A := \{y \in \mathbb{R} : \mu(f^{-1}(\{y\})) > 0\}$ der Atome des endlichen Maßes $\mu \circ f^{-1}$ höchstens abzählbar. Daher gibt es $N \in \mathbb{N}$ und $y_0 \leq -\|f\|_\infty < y_1 < \dots < y_{N-1} < \|f\|_\infty < y_N$ so, dass

$$y_i \in \mathbb{R} \setminus A \quad \text{und} \quad |y_{i+1} - y_i| < \varepsilon \quad \text{für jedes } i.$$

Sei $E_i = f^{-1}([y_{i-1}, y_i])$ für $i = 1, \dots, N$. Dann ist $E = \biguplus_{i=1}^N E_i$ und wegen Gleichung (13.5)

$$\mu(\partial E_i) \leq \mu(f^{-1}(\{y_{i-1}\})) + \mu(f^{-1}(\{y_i\})) + \mu(U_f) = 0.$$

Daher gilt

$$\limsup_{n \rightarrow \infty} \int f d\mu_n \leq \limsup_{n \rightarrow \infty} \sum_{i=1}^N \mu_n(E_i) \cdot y_i = \sum_{i=1}^N \mu(E_i) \cdot y_i \leq \varepsilon + \int f d\mu.$$

Wir lassen $\varepsilon \rightarrow 0$ gehen und erhalten $\limsup_{n \rightarrow \infty} \int f d\mu_n \leq \int f d\mu$. Indem wir $(-f)$ betrachten, folgt die andere Ungleichung $\liminf_{n \rightarrow \infty} \int f d\mu_n \geq \int f d\mu$. \square

Definition 13.17. Seien X, X_1, X_2, \dots Zufallsvariablen mit Werten in E . Wir sagen $(X_n)_{n \in \mathbb{N}}$ **konvergiert in Verteilung** gegen X , in Formeln $X_n \xrightarrow{\mathcal{D}} X$ oder $X_n \xrightarrow{n \rightarrow \infty} X$, wenn die Verteilungen schwach konvergieren, also wenn $\mathbf{P}_X = \text{w-lim}_{n \rightarrow \infty} \mathbf{P}_{X_n}$. Manchmal schreiben wir auch $X_n \xrightarrow{\mathcal{D}} \mathbf{P}_X$ oder $X_n \xrightarrow{n \rightarrow \infty} \mathbf{P}_X$, wenn wir nur die Verteilung \mathbf{P}_X nicht aber die Zufallsvariable X spezifizieren wollen.

Satz 13.18 (Satz von Slutsky). Seien X, X_1, X_2, \dots und Y_1, Y_2, \dots Zufallsvariablen mit Werten in E und $X_n \xrightarrow{\mathcal{D}} X$ sowie $d(X_n, Y_n) \xrightarrow{n \rightarrow \infty} 0$ stochastisch. Dann gilt $Y_n \xrightarrow{\mathcal{D}} X$.

Beweis. Sei $f : E \rightarrow \mathbb{R}$ beschränkt und Lipschitz-stetig mit Konstante K . Dann ist

$$|f(x) - f(y)| \leq K d(x, y) \wedge 2 \|f\|_\infty \quad \text{für alle } x, y \in E.$$

Majorisierte Konvergenz liefert $\limsup_{n \rightarrow \infty} \mathbf{E} [|f(X_n) - f(Y_n)|] = 0$. Also gilt

$$\begin{aligned} \limsup_{n \rightarrow \infty} |\mathbf{E}[f(Y_n)] - \mathbf{E}[f(X)]| \\ \leq \limsup_{n \rightarrow \infty} |\mathbf{E}[f(X)] - \mathbf{E}[f(X_n)]| + \limsup_{n \rightarrow \infty} |\mathbf{E}[f(X_n)] - f(Y_n)| = 0. \quad \square \end{aligned}$$

Korollar 13.19. Gilt $X_n \xrightarrow{n \rightarrow \infty} X$ stochastisch, so gilt auch $X_n \xrightarrow{\mathcal{D}} X$, $n \rightarrow \infty$. Die Umkehrung ist im Allgemeinen falsch.

Beispiel 13.20. Sind X, X_1, X_2, \dots u.i.v. (mit nicht-trivialer Verteilung), so gilt trivialerweise $X_n \xrightarrow{\mathcal{D}} X$, jedoch nicht $X_n \xrightarrow{n \rightarrow \infty} X$ stochastisch. \diamond

Man erinnere sich an die Definition der Verteilungsfunktion eines Wahrscheinlichkeitsmaßes in Definition 1.59.

Definition 13.21. Seien F, F_1, F_2, \dots Verteilungsfunktionen von W-Maßen auf \mathbb{R} . Wir sagen $(F_n)_{n \in \mathbb{N}}$ konvergiere **schwach** gegen F , in Formeln $F_n \xrightarrow{n \rightarrow \infty} F$, $F_n \xrightarrow{\mathcal{D}} F$ oder $F = \text{w-lim}_{n \rightarrow \infty} F_n$, wenn

$$F(x) = \lim_{n \rightarrow \infty} F_n(x) \quad \text{für alle Stetigkeitspunkte } x \text{ von } F. \quad (13.6)$$

Sind F, F_1, F_2, \dots Verteilungsfunktionen von Sub-Wahrscheinlichkeitsmaßen, so setzen wir $F(\infty) := \lim_{x \rightarrow \infty} F(x)$ und fordern für die schwache Konvergenz zusätzlich $F(\infty) \geq \limsup_{n \rightarrow \infty} F_n(\infty)$.

Man beachte, dass aus (13.6) stets $F(\infty) \leq \liminf_{n \rightarrow \infty} F_n(\infty)$ folgt. Gilt nun $F_n \xrightarrow{\mathcal{D}} F$, so ist also $F(\infty) = \lim_{n \rightarrow \infty} F_n(\infty)$.

Beispiel 13.22. Ist F die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes auf \mathbb{R} und $F_n(x) := F(x+n)$ für $x \in \mathbb{R}$, so konvergiert $(F_n)_{n \in \mathbb{N}}$ punktweise gegen 1. Dies ist jedoch keine Verteilungsfunktion, da diese für $x \rightarrow -\infty$ gegen Null konvergieren. Ist andererseits $G_n(x) = F(x-n)$, so konvergiert $(G_n)_{n \in \mathbb{N}}$ punktweise gegen $G \equiv 0$. Nun ist aber $G(\infty) = 0 < \limsup_{n \rightarrow \infty} G_n(\infty) = 1$, also liegt auch in diesem Falle keine schwache Konvergenz vor. In der Tat: es tritt jeweils im Limes ein Massendefekt ein (bei den F_n nach links, bei den G_n nach rechts). Die Definition der schwachen Konvergenz von Verteilungsfunktionen ist aber so angelegt, dass kein Massendefekt im Limes eintreten darf. \diamond

Satz 13.23. Seien $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_{\leq 1}(\mathbb{R})$ mit zugehörigen Verteilungsfunktionen F, F_1, F_2, \dots Dann sind äquivalent

$$(i) \mu = \text{w-lim}_{n \rightarrow \infty} \mu_n,$$

$$(ii) F_n \xrightarrow{\mathcal{D}} F.$$

Beweis. „(i) \Rightarrow (ii)“ Sei F in x stetig. Dann ist $\mu(\partial(-\infty, x]) = \mu(\{x\}) = 0$. Nach Satz 13.16 gilt $F_n(x) = \mu_n((-\infty, x]) \xrightarrow{n \rightarrow \infty} \mu((-\infty, x]) = F(x)$.

„(ii) \Rightarrow (i)“ Sei $f \in \text{Lip}_1(\mathbb{R}; [0, 1])$. Nach Satz 13.16 reicht es zu zeigen, dass

$$\int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu. \quad (13.7)$$

Sei $\varepsilon > 0$. Wähle $N \in \mathbb{N}$ und $N + 1$ Stetigkeitspunkte $y_0 < y_1 < \dots < y_N$ von F so, dass $F(y_0) < \varepsilon$, $F(y_N) > F(\infty) - \varepsilon$ und $y_i - y_{i-1} < \varepsilon$ für jedes i . Dann ist

$$\int f d\mu_n \leq (F_n(y_0) + F_n(\infty) - F_n(y_N)) + \sum_{i=1}^N (f(y_i) + \varepsilon)(F_n(y_i) - F_n(y_{i-1})).$$

Nach Voraussetzung gilt $\lim_{n \rightarrow \infty} F_n(\infty) = F(\infty)$ und $F_n(y_i) \xrightarrow{n \rightarrow \infty} F(y_i)$ für jedes $i = 0, \dots, N$, also

$$\limsup_{n \rightarrow \infty} \int f d\mu_n \leq 3\varepsilon + \sum_{i=1}^N f(y_i)(F(y_i) - F(y_{i-1})) \leq 4\varepsilon + \int f d\mu.$$

Daher gilt

$$\limsup_{n \rightarrow \infty} \int f d\mu_n \leq \int f d\mu.$$

Indem wir f durch $(1 - f)$ ersetzen, folgt (13.7). \square

Korollar 13.24. Seien X, X_1, X_2, \dots reelle Zufallsvariablen mit Verteilungsfunktionen F, F_1, F_2, \dots . Dann sind äquivalent:

- (i) $X_n \xrightarrow{\mathcal{D}} X$,
- (ii) $\mathbf{E}[f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbf{E}[f(X)]$ für jedes $f \in C_b(\mathbb{R})$,
- (iii) $F_n \xrightarrow{\mathcal{D}} F$.

Wie verhält sich die schwache Konvergenz, wenn wir zu Bildmaßen übergehen? Wir brauchen gewisse Stetigkeitseigenschaften, damit die schwache Konvergenz erhalten bleibt. Der folgende einfache Satz ist in den Anwendungen extrem nützlich.

Satz 13.25 (Continuous Mapping Theorem). Seien (E_1, d_1) und (E_2, d_2) metrische Räume und $\varphi : E_1 \rightarrow E_2$ messbar sowie U_φ die Menge der Unstetigkeitsstellen von φ .

- (i) Sind $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_{\leq 1}(E_1)$ mit $\mu(U_\varphi) = 0$ und $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ schwach, so gilt $\mu_n \circ \varphi^{-1} \xrightarrow{n \rightarrow \infty} \mu \circ \varphi^{-1}$ schwach.
- (ii) Sind X, X_1, X_2, \dots E_1 -wertige Zufallsvariablen mit $\mathbf{P}[X \in U_\varphi] = 0$ und $X_n \xrightarrow{\mathcal{D}} X$, so gilt $\varphi(X_n) \xrightarrow{\mathcal{D}} \varphi(X)$.

Beweis. Nach Übung 1.1.3 ist $U_\varphi \subset E_1$ Borel-messbar. Also sind die angegebenen Bedingungen sinnvoll.

(i) Sei $f \in C_b(E_2)$. Dann ist $f \circ \varphi$ beschränkt und messbar, und es ist $U_{f \circ \varphi} \subset U_\varphi$, also $\mu(U_{f \circ \varphi}) = 0$. Nach Satz 13.16 ist

$$\begin{aligned}\lim_{n \rightarrow \infty} \int f d(\mu_n \circ \varphi^{-1}) &= \lim_{n \rightarrow \infty} \int (f \circ \varphi) d\mu_n \\ &= \int (f \circ \varphi) d\mu = \int f d(\mu \circ \varphi^{-1}).\end{aligned}$$

(ii) Dies ist klar, wegen $\mathbf{P}_{\varphi(X)} = \mathbf{P}_X \circ \varphi^{-1}$. \square

Übung 13.2.1. Man zeige: Für d'_P aus (13.4) und $\mu, \nu \in \mathcal{M}_1(E)$ gilt: $d_P(\mu, \nu) = d'_P(\mu, \nu) = d'_P(\nu, \mu)$. \clubsuit

Übung 13.2.2. Man zeige: Die Topologie der schwachen Konvergenz auf $\mathcal{M}_f(E)$ ist größer als die von der Totalvariation (siehe Korollar 7.45) erzeugte Topologie auf $\mathcal{M}_f(E)$. Das heißt, es gilt $\|\mu_n - \mu\|_{TV} \xrightarrow{n \rightarrow \infty} 0$, so gilt $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ schwach. \clubsuit

Übung 13.2.3. Sei $E = \mathbb{R}$ und $\mu_n = \frac{1}{n} \sum_{k=0}^n \delta_{k/n}$ sowie $\mu = \lambda|_{[0,1]}$ das auf $[0, 1]$ eingeschränkte Lebesgue-Maß. Man zeige, dass $\mu = \text{w-lim}_{n \rightarrow \infty} \mu_n$. \clubsuit

Übung 13.2.4. Sei $E = \mathbb{R}$ und λ das Lebesgue-Maß auf \mathbb{R} . Für $n \in \mathbb{N}$ sei $\mu_n = \lambda|_{[-n,n]}$. Man zeige: $\lambda = \text{v-lim}_{n \rightarrow \infty} \mu_n$, jedoch ist $(\mu_n)_{n \in \mathbb{N}}$ nicht schwach konvergent. \clubsuit

Übung 13.2.5. Sei $E = \mathbb{R}$ und $\mu_n = \delta_n$ für $n \in \mathbb{N}$. Man zeige: $\text{v-lim}_{n \rightarrow \infty} \mu_n = 0$, jedoch ist $(\mu_n)_{n \in \mathbb{N}}$ nicht schwach konvergent. \clubsuit

Übung 13.2.6 (Lévy-Abstand). Für zwei Verteilungsfunktionen F und G von Wahrscheinlichkeitsmaßen auf \mathbb{R} wird der Lévy-Abstand definiert als

$$d(F, G) = \inf \{ \varepsilon \geq 0 : G(x - \varepsilon) - \varepsilon \leq F(x) \leq G(x + \varepsilon) + \varepsilon \text{ für alle } x \in \mathbb{R} \}.$$

Zeige:

- (i) d ist eine Metrik auf der Menge der Verteilungsfunktionen.
- (ii) Es gilt $F_n \xrightarrow{n \rightarrow \infty} F$ genau dann, wenn $d(F_n, F) \xrightarrow{n \rightarrow \infty} 0$.
- (iii) Zu jedem $P \in \mathcal{M}_1(\mathbb{R})$ gibt es eine Folge $(P_n)_{n \in \mathbb{N}}$ in $\mathcal{M}_1(\mathbb{R})$, sodass jedes P_n endlichen Träger hat, und sodass $P_n \xrightarrow{n \rightarrow \infty} P$. \clubsuit

Übung 13.2.7. Wir können die Begriffe *schwache Konvergenz* und *vage Konvergenz* auf Ladungsverteilungen ausdehnen, also auf Differenzen $\varphi := \mu^+ - \mu^-$ von Maßen aus $\mathcal{M}_f(E)$ beziehungsweise $\mathcal{M}(E)$, indem wir den Wortlaut von Definition 13.12 auf diese Klassen anwenden. Man zeige, dass man hier die schwache Konvergenz im Allgemeinen nicht metrisieren kann.

Anleitung: Man betrachte $E = [0, 1]$.

- (i) Für $n \in \mathbb{N}$ definiere $\varphi_n = \delta_{1/n} - \delta_{2/n}$. Zeige: Für jedes $C > 0$ konvergiert $(C\varphi_n)_{n \in \mathbb{N}}$ schwach gegen das Nullmaß.
- (ii) Man nehme an, dass es eine Metrik gäbe, die die schwache Konvergenz erzeugt. Man zeige: Dann gäbe es eine Folge $(C_n)_{n \in \mathbb{N}}$ mit $C_n \uparrow \infty$ und $0 = \text{w-lim}_{n \rightarrow \infty} (C_n \varphi_n)$.
- (iii) Wähle ein $f \in C([0, 1])$ mit $f(2^{-n}) = (-1)^n C_n^{-1/2}$ für jedes $n \in \mathbb{N}$ und zeige: $\left(\int f d(C_n \varphi_n) \right)_{n \in \mathbb{N}}$ konvergiert nicht gegen Null.
- (iv) Man führe diese Konstruktion zum Widerspruch mit der Metrisierbarkeitsannahme. ♣

Übung 13.2.8. Man zeige, dass durch (13.3) eine Metrik auf $\mathcal{M}_1(E)$ definiert wird, und dass diese die Topologie der schwachen Konvergenz erzeugt. ♣

Übung 13.2.9. Man zeige die Implikation „(vi) \implies (iv)“ aus Satz 13.16 direkt. ♣

Übung 13.2.10. Seien X, X_1, X_2, \dots und Y_1, Y_2, \dots reelle Zufallsvariablen. Es gelte $\mathbf{P}_{Y_n} = \mathcal{N}_{0,1/n}$ für jedes $n \in \mathbb{N}$. Man zeige: Es gilt genau dann $X_n \xrightarrow{\mathcal{D}} X$, wenn $X_n + Y_n \xrightarrow{\mathcal{D}} X$. ♣

Übung 13.2.11. Für jedes $n \in \mathbb{N}$ sei X_n eine geometrisch verteilte Zufallsvariable mit Parameter $p_n \in (0, 1)$. Wie muss die Folge $(p_n)_{n \in \mathbb{N}}$ gewählt sein, damit $\mathbf{P}_{X_n/n}$ schwach gegen die Exponentialverteilung mit Parameter $\alpha > 0$ konvergiert? ♣

Übung 13.2.12. Seien X, X_1, X_2, \dots reelle Zufallsvariablen mit $X_n \xrightarrow{n \rightarrow \infty} X$. Zeige:

- (i) $\mathbf{E}[|X|] \leq \liminf_{n \rightarrow \infty} \mathbf{E}[|X_n|]$.
- (ii) Ist $p > 0$ und $\sup_{n \in \mathbb{N}} \mathbf{E}[|X_n|^p] < \infty$ für ein $r > p$, so gilt $\mathbf{E}[|X|^p] = \lim_{n \rightarrow \infty} \mathbf{E}[|X_n|^p]$. ♣

Übung 13.2.13. Seien F, F_1, F_2, \dots Verteilungsfunktionen von W-Maßen auf \mathbb{R} , und gelte $F_n \xrightarrow{n \rightarrow \infty} F$. Sei $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$, $u \in (0, 1)$, die linksstetige Inverse von F (vergleiche Beweis von Satz 1.104). Zeige:

$F_n^{-1}(u) \xrightarrow{n \rightarrow \infty} F^{-1}(u)$ in jedem Stetigkeitspunkt u von F^{-1} .

Folgere: $F^{-1}(u) \xrightarrow{n \rightarrow \infty} F^{-1}(u)$ für Lebesgue fast alle $u \in (0, 1)$. 

Übung 13.2.14. Seien μ, μ_1, μ_2, \dots W-Maße auf \mathbb{R} mit $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ schwach. Zeige: Es gibt einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$ und reelle Zufallsvariablen X, X_1, X_2, \dots auf $(\Omega, \mathcal{A}, \mathbf{P})$ mit Verteilungen $\mathbf{P}_X = \mu, \mathbf{P}_{X_n} = \mu_n, n \in \mathbb{N}$, so dass

$$X_n \xrightarrow{n \rightarrow \infty} X \quad \mathbf{P}\text{-f.s.}$$

Hinweis: Verwende Übung 13.2.13. 

Übung 13.2.15. Sei (E, d) ein metrischer Raum, und seien μ, μ_1, μ_2, \dots Wahrscheinlichkeitsmaße auf E . Eine messbare Abbildung $f : E \rightarrow \mathbb{R}$ heißt gleichgradig integrierbar bezüglich $(\mu_n)_{n \in \mathbb{N}}$, falls

$$\inf_{a > 0} \sup_{n \in \mathbb{N}} \int_{\{|f| > a\}} |f| d\mu_n = 0.$$

Sei f stetig und gleichgradig integrierbar bezüglich $(\mu_n)_{n \in \mathbb{N}}$, und gelte $\mu_n \xrightarrow{n \rightarrow \infty} \mu$ schwach. Zeige, dass $\int |f| d\mu < \infty$ gilt und

$$\int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu.$$

Hinweis: Wende Übung 13.2.14 auf die Bildmaße $\mu_n \circ f^{-1}$ an. 

13.3 Der Satz von Prohorov

Sei E stets ein polnischer Raum mit Borel'scher σ -Algebra \mathcal{E} . Eine grundlegende Frage ist, wann eine Folge $(\mu_n)_{n \in \mathbb{N}}$ von Maßen auf (E, \mathcal{E}) einen schwachen Grenzwert besitzt, oder wenigstens einen schwachen Häufungspunkt. Eine offensichtlich notwendige Bedingung ist, dass $(\mu_n(E))_{n \in \mathbb{N}}$ beschränkt ist, deshalb werden wir ohne Beschränkung der Allgemeinheit nur Folgen in $\mathcal{M}_{\leq 1}(E)$ betrachten. Dies ist allerdings nicht hinreichend, denn beispielsweise konvergiert die Folge $(\delta_n)_{n \in \mathbb{N}}$ von W-Maßen auf \mathbb{R} nicht schwach. Wir müssen also zusätzlich noch sicherstellen, dass keine Masse „ins Unendliche auswandert“. Dies liefert gerade die Bedingung der Straffheit.

Wir beginnen diesen Abschnitt, indem wir zunächst als Hauptsatz den Satz von Prohorov [145] vorstellen. Wir geben den Beweis erst in dem Spezialfall $E = \mathbb{R}$ an und kommen dann zu Anwendungen, bevor wir den Satz am Ende des Abschnitts in voller Allgemeinheit beweisen.

Definition 13.26 (Straffheit). Eine Familie $\mathcal{F} \subset \mathcal{M}_f(E)$ heißt **straß**, falls zu jedem $\varepsilon > 0$ eine kompakte Menge $K \subset E$ existiert mit

$$\sup \{\mu(E \setminus K) : \mu \in \mathcal{F}\} < \varepsilon.$$

Bemerkung 13.27. Ist E polnisch, so ist nach Lemma 13.5 jede einelementige Familie $\{\mu\} \subset \mathcal{M}_f(E)$ straff und damit auch jede endliche Familie. \diamond

Beispiele 13.28. (i) Ist E kompakt, so sind $\mathcal{M}_1(E)$ und $\mathcal{M}_{\leq 1}(E)$ straff.

(ii) Ist $(X_i)_{i \in I}$ eine beliebige Familie von reellen Zufallsvariablen mit

$$C := \sup \{\mathbf{E}[|X_i|] : i \in I\} < \infty,$$

so ist $\{\mathbf{P}_{X_i} : i \in I\}$ straff. In der Tat: Für $\varepsilon > 0$ und $K = [-C/\varepsilon, C/\varepsilon]$ ist nach der Markov'schen Ungleichung $\mathbf{P}_{X_i}(\mathbb{R} \setminus K) = \mathbf{P}[|X_i| > C/\varepsilon] \leq \varepsilon$.

(iii) Die Familie $(\delta_n)_{n \in \mathbb{N}}$ von W-Maßen auf \mathbb{R} ist nicht straff.

(iv) Die Familie $(\mathcal{U}_{[-n,n]})_{n \in \mathbb{N}}$ von uniformen Verteilungen auf den Intervallen $[-n, n]$, aufgefasst als W-Maße auf \mathbb{R} , ist nicht straff. \diamond

Satz 13.29 (Satz von Prohorov (1956)). Sei (E, d) ein metrischer Raum und $\mathcal{F} \subset \mathcal{M}_{\leq 1}(E)$.

(i) Es gilt

$$\mathcal{F} \text{ ist straff} \implies \mathcal{F} \text{ ist schwach relativ folgenkompakt.}$$

(ii) Ist E zudem polnisch, so gilt auch die Umkehrung

$$\mathcal{F} \text{ ist straff} \iff \mathcal{F} \text{ ist schwach relativ folgenkompakt.}$$

Korollar 13.30. Sei E ein kompakter, metrischer Raum. Dann sind die Mengen $\mathcal{M}_{\leq 1}(E)$ und $\mathcal{M}_1(E)$ schwachfolgenkompakt.

Korollar 13.31. Ist E ein lokalkompakter, separabler, metrischer Raum, so ist $\mathcal{M}_{\leq 1}(E)$ vag folgenkompakt.

Beweis. Sei $(\mu_n)_{n \in \mathbb{N}}$ eine Folge in $\mathcal{M}_{\leq 1}(E)$. Wir müssen zeigen, dass es eine vag konvergente Teilfolge gibt.

Seien x_1, x_2, \dots dicht in E . Da E lokalkompakt ist, existiert zu jedem $n \in \mathbb{N}$ eine offene Umgebung $U_n \ni x_n$, deren Abschluss \overline{U}_n kompakt ist. Setzen wir

$V_n := \bigcup_{i=1}^n V_i$, so ist auch \overline{V}_n kompakt für jedes $n \in \mathbb{N}$. Für jede kompakte Menge $K \subset E$ ist daher $N(K) := \min\{m : K \subset V_m\}$ endlich. Definiere induktiv $W_1 := V_1$ und $W_{n+1} := W_{N(\overline{V}_n)}$, $n \in \mathbb{N}$. Es ist dann W_n offen, \overline{W}_n kompakt und $\overline{W}_n \subset W_{n+1}$ für $n \in \mathbb{N}$. Ferner gilt $W_n \uparrow E$.

Indem wir auf die Maße $(\mu_k \mathbf{1}_{\overline{W}_n})_{k \in \mathbb{N}}$ den Satz von Prohorov (hier Korollar 13.30) anwenden, erhalten wir für jedes $n \in \mathbb{N}$ eine Folge $(k_l^n)_{l \in \mathbb{N}}$ und ein Maß $\tilde{\mu}_n := \text{w-lim}_{l \rightarrow \infty} \mu_{k_l^n} \mathbf{1}_{\overline{W}_n}$ mit Träger in \overline{W}_n . Wir können annehmen, dass die Folgen $(k_l^n)_{l \in \mathbb{N}}$ sukzessive so gewählt wurden, dass (k_l^{n+1}) eine Teilfolge von (k_l^n) ist.

Beachte, dass im Allgemeinen nur $\tilde{\mu}_n(\overline{W}_n) \leq \tilde{\mu}_{n+1}(\overline{W}_n)$ gilt, nicht aber Gleichheit.

Für $f \in C_c(E)$ existiert nun ein $n_0 \in \mathbb{N}$, so dass der Träger von f in W_{n_0} enthalten ist. Es gilt also für $m \geq n \geq n_0$

$$\begin{aligned} \int f d\tilde{\mu}_n &= \lim_{l \rightarrow \infty} \int f \mathbf{1}_{\overline{W}_n} d\mu_{k_l^n} \\ &= \lim_{l \rightarrow \infty} \int f \mathbf{1}_{\overline{W}_n} d\mu_{k_l^m} \\ &= \lim_{l \rightarrow \infty} \int f \mathbf{1}_{\overline{W}_m} d\mu_{k_l^m} = \int f d\tilde{\mu}_m \end{aligned}$$

und damit

$$\int f d\tilde{\mu}_n = \lim_{m \rightarrow \infty} \int f d\mu_{k_m^n}.$$

Für jede messbare relativ kompakte Menge $A \subset E$ gilt daher

$$\tilde{\mu}_m(A) = \tilde{\mu}_{N(\overline{A})} \quad \text{für jedes } m \geq N(\overline{A}).$$

Definiere nun für messbares $A \subset E$

$$\mu(A) := \sup_{n \in \mathbb{N}} \sup_{m > n} \tilde{\mu}_m(A \cap W_n) = \sup_{n \in \mathbb{N}} \tilde{\mu}_{n+1}(A \cap W_n).$$

Man prüft leicht nach, dass μ ein von unten stetiger Inhalt ist, also ein Maß (Satz 1.36). Nach Konstruktion gilt für jedes $f \in C_c(E)$

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f d\mu_{k_n^n}.$$

Also ist $\mu = \text{v-lim}_{n \rightarrow \infty} \mu_{k_n^n}$. □

Bemerkung 13.32. Die Implikation in Satz 13.29(ii) ist die weitaus einfachere, wenn auch weniger nützliche. Hier wird benötigt, dass E polnisch ist, denn eine eingleitige Familie ist offenbar immer schwach kompakt, jedoch nur unter Zusatzannahmen straff – beispielsweise eben, wenn E polnisch ist (Lemma 13.5). ◇

Beweis (von Satz 13.29(ii)). Wir gehen zunächst ähnlich vor wie im Beweis von Lemma 13.5. Sei $\{x_1, x_2, \dots\} \subset E$ dicht. Für $n \in \mathbb{N}$ setze $A_{n,N} := \bigcup_{i=1}^N B_{1/n}(x_i)$. Dann gilt $A_{n,N} \uparrow E$ für $N \rightarrow \infty$ für jedes $n \in \mathbb{N}$. Sei

$$\delta := \sup_{n \in \mathbb{N}} \inf_{N \in \mathbb{N}} \sup_{\mu \in \mathcal{F}} \mu(A_{n,N}^c).$$

Dann gibt es ein $n \in \mathbb{N}$, sodass für jedes $N \in \mathbb{N}$ ein $\mu_N \in \mathcal{F}$ existiert mit $\mu_N(A_{n,N}^c) \geq \delta/2$. Da \mathcal{F} schwach relativ folgenkompakt ist, besitzt $(\mu_N)_{N \in \mathbb{N}}$ eine schwach konvergente Teilfolge $(\mu_{N_k})_{k \in \mathbb{N}}$ mit einem schwachen Limes $\mu \in \mathcal{M}_{\leq 1}(E)$. Nach dem Portemanteau Theorem (Satz 13.16(iv)) gilt für jedes $N \in \mathbb{N}$

$$\mu(A_{n,N}^c) \geq \liminf_{k \rightarrow \infty} \mu_{N_k}(A_{n,N}^c) \geq \liminf_{k \rightarrow \infty} \mu_{N_k}(A_{n,N_k}^c) \geq \delta/2.$$

Andererseits gilt $A_{n,N}^c \downarrow \emptyset$ für $N \rightarrow \infty$, also $\mu(A_{n,N}^c) \xrightarrow{N \rightarrow \infty} 0$. Mithin ist $\delta = 0$.

Sei nun $\varepsilon > 0$ beliebig. Nach dem eben Gezeigten können wir zu jedem $n \in \mathbb{N}$ ein $N'_n \in \mathbb{N}$ wählen, sodass $\mu(A_{n,N'_n}^c) < \varepsilon/2^n$ für jedes $\mu \in \mathcal{F}$. Die Menge $A := \bigcap_{n=1}^{\infty} A_{n,N'_n}$ ist nach Konstruktion total beschränkt, also relativ kompakt. Ferner ist für jedes $\mu \in \mathcal{F}$

$$\mu((\overline{A})^c) \leq \mu(A^c) \leq \sum_{n=1}^{\infty} \mu(A_{n,N'_n}^c) \leq \varepsilon.$$

Also ist \mathcal{F} straff. □

Die andere Implikation im Satz von Prohorov ist schwieriger zu zeigen, vor allem im Fall allgemeiner (metrischer) Räume. Wir geben daher zunächst nur für $E = \mathbb{R}$ einen Beweis an und kommen dann zu Anwendungen, bevor wir die schwierigere Implikation im allgemeinen Fall beweisen.

Das Problem besteht darin, sich einen Kandidaten für einen schwachen Limespunkt zu verschaffen. Für Verteilungen auf \mathbb{R} ist das Problem äquivalent dazu, sich einen schwachen Limespunkt einer Folge von Verteilungsfunktionen zu verschaffen. Hier hilft uns der Helly'sche Auswahlssatz, der auf einem Diagonalfolgenargument beruht, das wir in ähnlicher Form im Beweis des Satzes von Prohorov in allgemeiner Form wieder gebrauchen werden.

Sei

$$V = \{F : \mathbb{R} \rightarrow \mathbb{R} \text{ ist rechtsstetig, monoton wachsend und beschränkt}\}$$

die Menge der Verteilungsfunktionen von endlichen Maßen auf \mathbb{R} .

Satz 13.33 (Auswahlssatz von Helly). Sei $(F_n)_{n \in \mathbb{N}}$ eine gleichmäßig beschränkte Folge in V . Dann existiert ein $F \in V$ und eine Teilfolge $(F_{n_k})_{k \in \mathbb{N}}$ mit

$$F_{n_k}(x) \xrightarrow{k \rightarrow \infty} F(x) \quad \text{in allen Stetigkeitsstellen von } F.$$

Beweis. Wir benutzen ein Diagonalfolgenargument. Wir wählen eine Abzählung $\mathbb{Q} = \{q_1, q_2, q_3, \dots\}$ der rationalen Zahlen. Nach dem Satz von Bolzano-Weierstraß besitzt $(F_n(q_1))_{n \in \mathbb{N}}$ eine konvergente Teilfolge $(F_{n_k^1}(q_1))_{k \in \mathbb{N}}$. Ebenso finden wir eine Teilfolge $(n_k^2)_{k \in \mathbb{N}}$ von $(n_k^1)_{k \in \mathbb{N}}$, sodass $(F_{n_k^2}(q_2))_{k \in \mathbb{N}}$ konvergiert. Iterativ erhalten wir Teilfolgen $(n_k^l) \supset (n_k^{l-1}) \supset (n_k^{l-2}) \supset \dots$, sodass $(F_{n_k^l}(q_l))_{k \in \mathbb{N}}$ für jedes $l \in \mathbb{N}$ konvergiert. Setze jetzt $n_k := n_k^k$. Dann konvergiert $(F_{n_k}(q))_{k \in \mathbb{N}}$ für jedes $q \in \mathbb{Q}$. Setze $\tilde{F}(q) = \lim_{k \rightarrow \infty} F_{n_k}(q)$ und

$$F(x) = \inf \{\tilde{F}(q) : q \in \mathbb{Q} \text{ mit } q > x\}.$$

Da \tilde{F} monoton wachsend ist, ist F rechtsstetig und monoton wachsend.

Ist F stetig in x , so existieren zu $\varepsilon > 0$ Zahlen $q^-, q^+ \in \mathbb{Q}$, $q^- < x < q^+$ mit $\tilde{F}(q^-) \geq F(x) - \varepsilon$ und $\tilde{F}(q^+) \leq F(x) + \varepsilon$. Nach Konstruktion ist dann

$$\limsup_{k \rightarrow \infty} F_{n_k}(x) \leq \lim_{k \rightarrow \infty} F_{n_k}(q^+) = \tilde{F}(q^+) \leq F(x) + \varepsilon.$$

Also ist $\limsup_{k \rightarrow \infty} F_{n_k}(x) \leq F(x)$ und analog mit q^- auch $\liminf_{k \rightarrow \infty} F_{n_k}(x) \geq F(x)$. \square

Beweis (von Satz 13.29(i) für den Fall $E = \mathbb{R}$).

Sei \mathcal{F} straff und $(\mu_n)_{n \in \mathbb{N}}$ eine Folge in \mathcal{F} mit Verteilungsfunktionen $F_n : x \mapsto \mu_n((-\infty, x])$. Nach dem Satz von Helly gibt es eine monotone, rechtsstetige Funktion $F : \mathbb{R} \rightarrow [0, 1]$ und eine Teilfolge $(F_{n_k})_{k \in \mathbb{N}}$ von $(F_n)_{n \in \mathbb{N}}$ mit $F_{n_k}(x) \xrightarrow{k \rightarrow \infty} F(x)$ in allen Stetigkeitspunkten x von F . Nach Satz 13.23 reicht es zu zeigen, dass $F(\infty) \geq \limsup_{k \rightarrow \infty} F_{n_k}(\infty)$.

Da \mathcal{F} straff ist, existiert zu $\varepsilon > 0$ ein $K < \infty$ mit $F_n(\infty) - F_n(x) < \varepsilon$ für jedes $n \in \mathbb{N}$ und $x > K$. Ist $x > K$ eine Stetigkeitsstelle von F , dann gilt $\limsup_{k \rightarrow \infty} F_{n_k}(\infty) \leq \limsup_{k \rightarrow \infty} F_{n_k}(x) + \varepsilon = F(x) + \varepsilon \leq F(\infty) + \varepsilon$. \square

Wir kommen zu einer ersten Anwendung des Satzes von Prohorov. Die ganze Stärke des folgenden Satzes wird erst deutlich, wenn wir geeignete trennende Funktionenklassen zur Verfügung haben. Diese werden wir in Kapitel 15 genauer untersuchen.

Satz 13.34. Sei E polnisch, und seien $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_{\leq 1}(E)$. Dann sind äquivalent:

$$(i) \mu = \text{w-lim}_{n \rightarrow \infty} \mu_n.$$

$$(ii) (\mu_n)_{n \in \mathbb{N}} \text{ ist straff, und es gibt eine trennende Familie } \mathcal{C} \subset C_b(E) \text{ mit}$$

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f d\mu_n \quad \text{für jedes } f \in \mathcal{C}. \quad (13.8)$$

Beweis. „(i) \implies (ii)“ Nach der einfachen Implikation im Satz von Prohorov (Satz 13.29(ii)) folgt aus der schwachen Konvergenz die Straffheit.

„(ii) \implies (i)“ Sei $(\mu_n)_{n \in \mathbb{N}}$ straff und $\mathcal{C} \subset C_b(E)$ trennend mit (13.8). Wir nehmen an, $(\mu_n)_{n \in \mathbb{N}}$ konvergiere nicht schwach gegen μ . Dann existieren $\varepsilon > 0$, $f \in C_b(E)$ und $(n_k)_{k \in \mathbb{N}}$ mit $n_k \uparrow \infty$ und

$$\left| \int f d\mu_{n_k} - \int f d\mu \right| > \varepsilon \quad \text{für alle } k \in \mathbb{N}. \quad (13.9)$$

Nach dem Satz von Prohorov (Satz 13.29) existiert ein $\nu \in \mathcal{M}_{\leq 1}(E)$ und eine Teilfolge $(n'_k)_{k \in \mathbb{N}}$ von $(n_k)_{k \in \mathbb{N}}$ mit $\mu_{n'_k} \rightarrow \nu$ schwach. Wegen (13.9) ist $|\int f d\mu - \int f d\nu| \geq \varepsilon$, also $\mu \neq \nu$. Andererseits ist

$$\int h d\mu = \lim_{k \rightarrow \infty} \int h d\mu_{n'_k} = \int h d\nu \quad \text{für jedes } h \in \mathcal{C},$$

also $\mu = \nu$. Damit ist die Annahme zum Widerspruch geführt, und es gilt (i). \square

Wir wollen den Zusammenhang zwischen schwacher und vager Konvergenz näher beleuchten.

Satz 13.35. Sei E ein lokalkompakter, polnischer Raum, und seien $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_f(E)$. Dann sind äquivalent:

- (i) $\mu = \text{w-lim}_{n \rightarrow \infty} \mu_n$,
- (ii) $\mu = \text{v-lim}_{n \rightarrow \infty} \mu_n$ und $\mu(E) = \lim_{n \rightarrow \infty} \mu_n(E)$,
- (iii) $\mu = \text{v-lim}_{n \rightarrow \infty} \mu_n$ und $\mu(E) \geq \limsup_{n \rightarrow \infty} \mu_n(E)$,
- (iv) $\mu = \text{v-lim}_{n \rightarrow \infty} \mu_n$ und $\{\mu_n, n \in \mathbb{N}\}$ ist straff.

Beweis. „(i) \iff (ii) \iff (iii)“ Dies folgt aus dem Portemanteau Theorem.

„(ii) \implies (iv)“ Es reicht zu zeigen, dass für jedes $\varepsilon > 0$ ein Kompaktum $K \subset E$ existiert mit $\limsup_{n \rightarrow \infty} \mu_n(E \setminus K) \leq \varepsilon$. Da μ regulär ist (Satz 13.6), existiert eine kompakte Menge $L \subset E$ mit $\mu(E \setminus L) < \varepsilon$. Da E lokalkompakt ist, existiert ein Kompaktum $K \subset E$ mit $K^\circ \supset L$ sowie ein $\rho_{L,K} \in C_c(E)$ mit $\mathbb{1}_L \leq \rho_{L,K}(x) \leq \mathbb{1}_K$. Es folgt

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mu_n(E \setminus K) &\leq \limsup_{n \rightarrow \infty} \left(\mu_n(E) - \int \rho_{L,K} d\mu_n \right) \\ &= \mu(E) - \int \rho_{L,K} d\mu \leq \mu(E \setminus L) < \varepsilon. \end{aligned}$$

„(iv) \implies (i)“ Sei $L \subset E$ kompakt mit $\mu_n(E \setminus L) \leq 1$ für jedes $n \in \mathbb{N}$. Sei $\rho \in C_c(E)$ mit $\rho \geq \mathbb{1}_L$. Dann ist

$$\sup_{n \in \mathbb{N}} \mu_n(E) \leq 1 + \sup_{n \in \mathbb{N}} \mu_n(L) \leq 1 + \sup_{n \in \mathbb{N}} \int \rho d\mu_n < \infty,$$

weil nach Voraussetzung $\int \rho d\mu_n$ konvergiert. Es ist also auch

$$C := \max(\mu(E), \sup\{\mu_n(E) : n \in \mathbb{N}\}) < \infty,$$

und wir können zu μ/C und μ_n/C übergehen und ohne Einschränkung annehmen, dass alle Maße in $\mathcal{M}_{\leq 1}(E)$ liegen. Da $C_c(E)$ trennend ist für $\mathcal{M}_{\leq 1}(E)$ (siehe Satz 13.11), folgt (i) aus Satz 13.34. \square

Beweis des Satzes von Prohorov, Teil (i), allgemeiner Fall. Es gibt prinzipiell zwei Möglichkeiten, den Satz im allgemeinen Fall zu beweisen. Die eine Möglichkeit besteht darin, den Satz zunächst für Maße auf \mathbb{R}^d zu zeigen (das haben wir für $d = 1$ bereits getan, siehe auch Übung 13.3.4 für $d \geq 2$). In einem zweiten Schritt wird die Aussage auf Folgenräume $\mathbb{R}^{\mathbb{N}}$ angehoben. Schließlich wird im dritten Schritt eine Einbettung von E in $\mathbb{R}^{\mathbb{N}}$ konstruiert. Diesen Weg findet man beispielsweise in [15] oder [91].

Wir folgen hier der anderen Route, wie sie etwa in [16] (beziehungsweise [17]) oder [49] dargestellt wird. Der Hauptpunkt des Beweises besteht darin, einen Kandidaten für einen schwachen Häufungspunkt der Familie \mathcal{F} zu finden. Wir werden diesen zunächst als Inhalt auf einem abzählbaren Mengensystem konstruieren und dann ein äußeres Maß daraus ableiten. Schließlich zeigen wir, dass die abgeschlossenen Mengen messbar bezüglich dieses äußeren Maßes sind. Die Argumentation verläuft also in Teilen recht ähnlich wie beim Beweis des Satzes von Carathéodory.

Sei (E, d) ein metrischer Raum und $\mathcal{F} \subset \mathcal{M}_{\leq 1}(E)$ straff. Dann existiert eine aufsteigende Folge $K_1 \subset K_2 \subset K_3 \subset \dots$ von kompakten Mengen in E , sodass $\mu(K_n^c) < \frac{1}{n}$ gilt für jedes $\mu \in \mathcal{F}$ und jedes $n \in \mathbb{N}$. Setze $E' := \bigcup_{n=1}^{\infty} K_n$. Dann ist E' ein σ -kompakter, metrischer Raum, also insbesondere separabel. Da nach Konstruktion $\mu(E \setminus E') = 0$ für jedes $\mu \in \mathcal{F}$ gilt, können wir jedes μ als Maß auf E' auffassen. Ohne Beschränkung der Allgemeinheit können wir also annehmen, dass E σ -kompakt und damit separabel ist. Es existiert also eine abzählbare Basis \mathcal{U} der Topologie $\tau|_E$ auf E , das heißt eine abzählbare Menge \mathcal{U} von offenen Mengen, sodass $A = \bigcup_{U \in \mathcal{U}, U \subset A} U$ für jedes offene $A \subset E$ ist. Setze $\mathcal{C}' := \{\overline{U} \cap K_n : U \in \mathcal{U}, n \in \mathbb{N}\}$ und

$$\mathcal{C} := \left\{ \bigcup_{n=1}^N C_n : N \in \mathbb{N} \text{ und } C_1, \dots, C_N \in \mathcal{C}' \right\}.$$

Offenbar ist \mathcal{C} eine abzählbare Menge kompakter Mengen in E , und \mathcal{C} ist vereinigungsstabil. Jedes K_n hat eine endliche Überdeckung mit Mengen aus \mathcal{U} , also ist $K_n \in \mathcal{C}$.

Sei nun $(\mu_n)_{n \in \mathbb{N}}$ eine Folge in \mathcal{F} . Mit Hilfe des Diagonalfolgenarguments (siehe Beweis des Auswahlssatzes von Helly, Satz 13.33) können wir eine Teilfolge $(\mu_{n_k})_{k \in \mathbb{N}}$ finden, für die der Grenzwert

$$\alpha(C) := \lim_{k \rightarrow \infty} \mu_{n_k}(C) \quad (13.10)$$

für jedes $C \in \mathcal{C}$ existiert. Angenommen es gibt ein Maß μ auf der Borel'schen σ -Algebra \mathcal{E} von E , sodass

$$\mu(A) = \sup \{ \alpha(C) : C \in \mathcal{C} \text{ mit } C \subset A \} \quad \text{für } A \subset E \text{ offen.} \quad (13.11)$$

Dann ist

$$\begin{aligned} \mu(E) &\geq \sup_{n \in \mathbb{N}} \alpha(K_n) = \sup_{n \in \mathbb{N}} \lim_{k \rightarrow \infty} \mu_{n_k}(K_n) \\ &\geq \sup_{n \in \mathbb{N}} \limsup_{k \rightarrow \infty} \left(\mu_{n_k}(E) - \frac{1}{n} \right) \\ &= \limsup_{k \rightarrow \infty} \mu_{n_k}(E). \end{aligned}$$

Für offenes A und für $C \in \mathcal{C}$ mit $C \subset A$ ist ferner

$$\alpha(C) = \lim_{k \rightarrow \infty} \mu_{n_k}(C) \leq \liminf_{k \rightarrow \infty} \mu_{n_k}(A),$$

also $\mu(A) \leq \liminf_{k \rightarrow \infty} \mu_{n_k}(A)$. Nach dem Portemanteau Theorem (Satz 13.16) ist $\mu = \lim_{k \rightarrow \infty} \mu_{n_k}$, und damit ist \mathcal{F} als schwach relativ folgenkompakt erkannt. Es bleibt zu zeigen, dass es ein Maß μ auf (E, \mathcal{E}) gibt, das (13.11) erfüllt.

Die Mengenfunktion α auf \mathcal{C} ist offenbar monoton, additiv und subadditiv:

$$\begin{aligned} \alpha(C_1) &\leq \alpha(C_2), && \text{falls } C_1 \subset C_2, \\ \alpha(C_1 \cup C_2) &= \alpha(C_1) + \alpha(C_2), && \text{falls } C_1 \cap C_2 = \emptyset, \\ \alpha(C_1 \cup C_2) &\leq \alpha(C_1) + \alpha(C_2). \end{aligned} \quad (13.12)$$

Wir definieren

$$\beta(A) := \sup \{ \alpha(C) : C \in \mathcal{C} \text{ mit } C \subset A \} \quad \text{für } A \subset E \text{ offen}$$

und

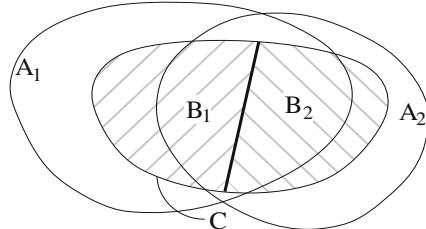
$$\mu^*(G) := \inf \{ \beta(A) : A \supset G \text{ ist offen} \} \quad \text{für } G \in 2^E.$$

Offenbar ist $\beta(A) = \mu^*(A)$ für jedes offene A . Es reicht zu zeigen (Schritte 1-3 unten), dass μ^* ein äußeres Maß ist (siehe Definition 1.46), und (4. Schritt) dass die σ -Algebra der μ^* -messbaren Mengen (siehe Definition 1.48 und Lemma 1.52) die abgeschlossenen Mengen und damit ganz \mathcal{E} enthält. Nach Lemma 1.52 ist dann nämlich μ^* ein Maß auf der σ -Algebra der μ^* -messbaren Mengen, und die Einschränkung $\mu := \mu^*|_{\mathcal{E}}$ erfüllt $\mu(A) = \mu^*(A) = \beta(A)$ für alle offenen A , also gilt Gleichung (13.11).

Offenbar ist $\mu^*(\emptyset) = 0$, und μ^* ist monoton. Um zu zeigen, dass μ^* ein äußeres Maß ist, müssen wir nur noch die σ -Subadditivität nachweisen.

1. Schritt (endliche Subadditivität von β). Seien $A_1, A_2 \subset E$ offen und $C \in \mathcal{C}$ mit $C \subset A_1 \cup A_2$. Sei $n \in N$ mit $C \subset K_n$. Wir definieren zwei Mengen

$$\begin{aligned} B_1 &:= \{x \in C : d(x, A_1^c) \geq d(x, A_2^c)\}, \\ B_2 &:= \{x \in C : d(x, A_1^c) \leq d(x, A_2^c)\}. \end{aligned}$$



Offenbar ist $B_1 \subset A_1$ und $B_2 \subset A_2$. Da $x \mapsto d(x, A_i^c)$ stetig ist für $i = 1, 2$, sind B_1 und B_2 als abgeschlossene Teilmengen von C kompakt. Also ist $d(B_1, A_1^c) > 0$. Es existiert also eine offene Menge D_1 mit $B_1 \subset D_1 \subset \overline{D}_1 \subset A_1$. (Man wähle etwa D_1 als Vereinigung einer endlichen Überdeckung von B_1 mit Kugeln vom Radius $d(B_1, A_1^c)/2$. Diese Kugeln liegen nebst ihren Abschlüssen in A_1 .) Sei $\mathcal{U}_{D_1} := \{U \in \mathcal{U} : U \subset D_1\}$. Dann ist $B_1 \subset D_1 = \bigcup_{U \in \mathcal{U}_{D_1}} U$. Wähle nun eine endliche Teilüberdeckung $\{U_1, \dots, U_N\} \subset \mathcal{U}_{D_1}$ von B_1 und setze $C_1 := \bigcup_{i=1}^N \overline{U}_i \cap K_n$. Dann ist $B_1 \subset C_1 \subset A_1$ und $C_1 \in \mathcal{C}$. Wähle analog ein $C_2 \in \mathcal{C}$ mit $B_2 \subset C_2 \subset A_2$. Es folgt

$$\alpha(C) \leq \alpha(C_1 \cup C_2) \leq \alpha(C_1) + \alpha(C_2) \leq \beta(A_1) + \beta(A_2).$$

Also gilt auch

$$\beta(A_1 \cup A_2) = \sup \{\alpha(C) : C \in \mathcal{C} \text{ mit } C \subset A_1 \cup A_2\} \leq \beta(A_1) + \beta(A_2).$$

2. Schritt (σ -Subadditivität von β). Seien A_1, A_2, \dots offene Mengen und $C \in \mathcal{C}$ mit $C \subset \bigcup_{i=1}^{\infty} A_i$. Da C kompakt ist, existiert ein $n \in \mathbb{N}$ mit $C \subset \bigcup_{i=1}^n A_i$. Die schon gezeigte endliche Subadditivität von β impliziert

$$\alpha(C) \leq \beta\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^{\infty} \beta(A_i).$$

Indem wir das Supremum über solche C bilden, folgt

$$\beta\left(\bigcup_{i=1}^{\infty} A_i\right) = \sup \left\{ \alpha(C) : C \in \mathcal{C} \text{ mit } C \subset \bigcup_{i=1}^{\infty} A_i \right\} \leq \sum_{i=1}^{\infty} \beta(A_i).$$

3. Schritt (σ -Subadditivität von μ^*). Seien $G_1, G_2, \dots \in 2^E$. Sei $\varepsilon > 0$. Wähle für jedes $n \in \mathbb{N}$ eine offene Menge $A_n \supset G_n$ mit $\beta(A_n) < \mu^*(G_n) + \varepsilon/2^n$. Dann gilt aufgrund der σ -Subadditivität von β

$$\mu^*\left(\bigcup_{n=1}^{\infty} G_n\right) \leq \beta\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \beta(A_n) \leq \varepsilon + \sum_{n=1}^{\infty} \mu^*(G_n).$$

Da $\varepsilon > 0$ beliebig war, folgt $\mu^*(\bigcup_{n=1}^{\infty} G_n) \leq \sum_{n=1}^{\infty} \mu^*(G_n)$. Mithin ist μ^* ein äußeres Maß.

4. Schritt (Abgeschlossene Mengen sind μ^* -messbar). Nach Lemma 1.49 ist eine Menge $B \subset E$ genau dann μ^* -messbar, wenn

$$\mu^*(B \cap G) + \mu^*(B^c \cap G) \leq \mu^*(G) \quad \text{für alle } G \in 2^E.$$

Indem wir das Infimum über alle offenen Mengen $A \supset G$ bilden, reicht es zu zeigen, dass für jedes abgeschlossene B und jedes offene $A \subset E$ gilt, dass

$$\mu^*(B \cap A) + \mu^*(B^c \cap A) \leq \beta(A). \quad (13.13)$$

Sei $\varepsilon > 0$. Wähle $C_1 \in \mathcal{C}$ mit $C_1 \subset A \cap B^c$ und $\alpha(C_1) > \beta(A \cap B^c) - \varepsilon$. Wähle ferner $C_2 \in \mathcal{C}$ mit $C_2 \subset A \cap C_1^c$ und $\alpha(C_2) > \beta(A \cap C_1^c) - \varepsilon$. Wegen $C_1 \cap C_2 = \emptyset$ und $C_1 \cup C_2 \subset A$ folgt

$$\begin{aligned} \beta(A) &\geq \alpha(C_1 \cup C_2) = \alpha(C_1) + \alpha(C_2) \geq \beta(A \cap B^c) + \beta(A \cap C_1^c) - 2\varepsilon \\ &\geq \mu^*(A \cap B^c) + \mu^*(A \cap B) - 2\varepsilon. \end{aligned}$$

Indem wir $\varepsilon \rightarrow 0$ gehen lassen, folgt (13.13). Damit ist der Beweis des Satzes von Prohorov vollständig. \square

Übung 13.3.1. Man zeige: Eine Familie $\mathcal{F} \subset \mathcal{M}_f(\mathbb{R})$ ist genau dann straff, wenn es eine messbare Abbildung $f : \mathbb{R} \rightarrow [0, \infty)$ gibt mit $f(x) \rightarrow \infty$ für $|x| \rightarrow \infty$ und $\sup_{\mu \in \mathcal{F}} \int f d\mu < \infty$. \clubsuit

Übung 13.3.2. Sei $L \subset \mathbb{R} \times (0, \infty)$ sowie $\mathcal{F} = \{\mathcal{N}_{\mu, \sigma^2} : (\mu, \sigma^2) \in L\}$ eine Familie von Normalverteilungen mit Parametern in L . Man zeige: \mathcal{F} ist genau dann straff, wenn L beschränkt ist. \clubsuit

Übung 13.3.3. Ist P ein W-Maß auf $[0, \infty)$ mit $m_P := \int x P(dx) \in (0, \infty)$, so definieren wir die **größenverzerrte Verteilung** \widehat{P} auf $[0, \infty)$ durch

$$\widehat{P}(A) = m_P^{-1} \int_A x P(dx). \quad (13.14)$$

Sei nun $(X_i)_{i \in I}$ eine Familie von Zufallsvariablen auf $[0, \infty)$ mit $\mathbf{E}[X_i] = 1$. Man zeige: $(\widetilde{\mathbf{P}_{X_i}})_{i \in I}$ ist genau dann straff, wenn $(X_i)_{i \in I}$ gleichgradig integrierbar ist. \clubsuit

Übung 13.3.4 (Satz von Helly im \mathbb{R}^d). Seien $x = (x^1, \dots, x^d) \in \mathbb{R}^d$ und $y = (y^1, \dots, y^d) \in \mathbb{R}^d$. Wir schreiben $x \leq y$, falls $x^i \leq y^i$ für jedes $i = 1, \dots, d$. Eine Abbildung $F : \mathbb{R}^d \rightarrow \mathbb{R}$ heißt monoton wachsend, falls $F(x) \leq F(y)$ für $x \leq y$. F heißt rechtsstetig, falls für jedes $x \in \mathbb{R}^d$ und jede Folge $(x_n)_{n \in \mathbb{N}}$ in \mathbb{R}^d mit $x_1 \geq x_2 \geq x_3 \geq \dots$ und $x = \lim_{n \rightarrow \infty} x_n$ gilt, dass $F(x) = \lim_{n \rightarrow \infty} F(x_n)$. Mit V_d bezeichnen wir die Menge der monoton wachsenden, rechtsstetigen, beschränkten Funktionen auf \mathbb{R}^d .

- (i) Man zeige, dass der Satz von Helly mit V_d statt V gilt.
- (ii) Man folgere die Gültigkeit des Satzes von Prohorov für $E = \mathbb{R}^d$. ♣

13.4 Anwendung: Satz von de Finetti – anders angeschaut

(Nach einer Idee von Götz Kersting.) Sei E ein polnischer Raum und X_1, X_2, \dots eine austauschbare Folge von Zufallsvariablen mit Werten in E . Wir wollen hier, alternativ zu dem Rückwärtsmartingalargument aus Kapitel 12.3, einen Beweis für den Satz von de Finetti (Satz 12.26) angeben, der besagt, dass es ein zufälliges W-Maß Ξ auf E gibt, sodass, gegeben Ξ , die Zufallsvariablen X_1, X_2, \dots unabhängig und Ξ -verteilt sind. Für $x = (x_1, x_2, \dots) \in E^{\mathbb{N}}$ seien $\xi_n(x) := \frac{1}{n} \sum_{l=1}^n \delta_{x_l}$ die empirische Verteilung von x_1, \dots, x_n ,

$$\mu_{n,k}(x) := \xi_n(x)^{\otimes k} = n^{-k} \sum_{i_1, \dots, i_k=1}^n \delta_{(x_{i_1}, \dots, x_{i_k})}$$

die Verteilung auf E^k des k -fachen unabhängigen Ziehens (mit Beachtung der Reihenfolge) aus (x_1, \dots, x_n) mit Zurücklegen und

$$\nu_{n,k}(x) := \frac{(n-k)!}{n!} \sum_{\substack{i_1, \dots, i_k=1 \\ \#\{i_1, \dots, i_k\}=k}}^n \delta_{(x_{i_1}, \dots, x_{i_k})}$$

die Verteilung auf E^k des k -fachen unabhängigen Ziehens (mit Beachtung der Reihenfolge) aus (x_1, \dots, x_n) ohne Zurücklegen. Für jedes $x \in E^{\mathbb{N}}$ gilt

$$\|\mu_{n,k}(x) - \nu_{n,k}(x)\|_{TV} \leq R_{n,k} := \frac{k(k-1)}{n}.$$

In der Tat ist die Wahrscheinlichkeit $p_{n,k}$, beim k -maligen Ziehen (mit Zurücklegen) aus n unterscheidbaren Kugeln keine zwei gleichen Kugeln zu ziehen,

$$p_{n,k} = \prod_{l=1}^{k-1} (1 - l/n)$$

und $R_{n,k} \geq 2(1 - p_{n,k})$.

Wir erhalten also die, intuitiv klare, Aussage, dass sich für $n \rightarrow \infty$ die Verteilungen des k -maligen Ziehens mit und ohne Zurücklegen annähern

$$\lim_{n \rightarrow \infty} \sup_{x \in E^{\mathbb{N}}} \|\mu_{n,k}(x) - \nu_{n,k}(x)\|_{TV} = 0.$$

Seien nun $f_1, \dots, f_k \in C_b(E)$ und $F(x_1, \dots, x_k) := f_1(x_1) \cdots f_k(x_k)$. Dann gilt wegen der Austauschbarkeit der Folge X_1, X_2, \dots für jede Wahl von paarweise unterschiedlichen Zahlen $1 \leq i_1, \dots, i_k \leq n$

$$\mathbf{E}[F(X_1, \dots, X_k)] = \mathbf{E}[F(X_{i_1}, \dots, X_{i_k})].$$

Indem wir über alle solchen Wahlen von i_1, \dots, i_k mitteln, erhalten wir

$$\mathbf{E}[f_1(X_1) \cdots f_k(X_k)] = \mathbf{E}[F(X_1, \dots, X_k)] = \mathbf{E}\left[\int F d\nu_{n,k}(X)\right].$$

Also ist

$$\begin{aligned} & \left| \mathbf{E}[f_1(X_1) \cdots f_k(X_k)] - \mathbf{E}\left[\int f_1 d\xi_n(X) \cdots \int f_k d\xi_n(X)\right] \right| \\ &= \left| \mathbf{E}\left[\int F d\nu_{n,k}(X)\right] - \mathbf{E}\left[\int F d\mu_{n,k}(X)\right] \right| \\ &\leq \|F\|_{\infty} R_{n,k} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Wir machen uns jetzt das folgende Kriterium für die Straffheit von Teilmengen von $\mathcal{M}_1(\mathcal{M}_1(E))$ zu Nutze.

Übung 13.4.1. Man zeige: Eine Teilmenge $\mathcal{K} \subset \mathcal{M}_1(\mathcal{M}_1(E))$ ist genau dann straff, wenn für jedes $\varepsilon > 0$ eine kompakte Menge $K \subset E$ existiert mit der Eigenschaft

$$\tilde{\mu}(\{\mu \in \mathcal{M}_1(E) : \mu(K^c) > \varepsilon\}) < \varepsilon \quad \text{für jedes } \tilde{\mu} \in \mathcal{K}.$$



Da E polnisch ist, ist \mathbf{P}_{X_1} straff. Zu $\varepsilon > 0$ existiert also ein Kompaktum $K \subset E$ mit $\mathbf{P}[X_1 \in K^c] < \varepsilon^2$. Es folgt, dass

$$\mathbf{P}[\xi_n(X)(K^c) > \varepsilon] \leq \varepsilon^{-1} \mathbf{E}[\xi_n(X)(K^c)] = \varepsilon^{-1} \mathbf{P}[X_1 \in K^c] \leq \varepsilon.$$

Also ist die Familie $(\mathbf{P}_{\xi_n(X)})_{n \in \mathbb{N}}$ straff. Sei Ξ_{∞} eine Zufallsvariable (mit Werten in $\mathcal{M}_1(E)$), sodass $\mathbf{P}_{\Xi_{\infty}} = \lim_{l \rightarrow \infty} \mathbf{P}_{\xi_{n_l}(X)}$ für eine geeignete Teilfolge $(n_l)_{l \in \mathbb{N}}$.

Die Abbildung $\xi \mapsto \int F d\xi = \int f_1 d\xi \cdots \int f_k d\xi$ ist beschränkt und (als Produkt stetiger Abbildungen) stetig bezüglich der Topologie der schwachen Konvergenz auf $\mathcal{M}_1(E)$, also aus $C_b(\mathcal{M}_1(E))$. Daher gilt

$$\begin{aligned} \mathbf{E}\left[\int F d\Xi_{\infty}^{\otimes k}\right] &= \lim_{l \rightarrow \infty} \mathbf{E}\left[\int f_1 d\xi_{n_l}(X) \cdots \int f_k d\xi_{n_l}(X)\right] \\ &= \mathbf{E}[f_1(X_1) \cdots f_k(X_k)]. \end{aligned}$$

Nun hängt der Grenzwert aber nicht mehr von der gewählten Teilfolge ab und ist damit eindeutig. Es folgt, noch einmal komplett ausgeschrieben, dass

$$\mathbf{E}[f_1(X_1) \cdots f_k(X_k)] = \mathbf{E}\left[\int f_1 d\Xi_\infty \cdots \int f_k d\Xi_\infty\right].$$

Durch diese Integrale ist aber die Verteilung von (X_1, \dots, X_k) vollständig bestimmt, und es folgt, dass $\mathbf{P}_{(X_1, \dots, X_k)} = \mathbf{P}_{\Xi_\infty^{\otimes k}}$, oder als Zufallsvariablen ausgedrückt: $(X_1, \dots, X_k) \stackrel{\mathcal{D}}{=} (Y_1, \dots, Y_k)$, wo, gegeben Ξ_∞ , die Zufallsvariablen Y_1, \dots, Y_k unabhängig mit Verteilung Ξ_∞ .

Übung 13.4.2. Man zeige: Eine Familie $(X_n)_{n \in \mathbb{N}}$ von Zufallsvariablen ist genau dann austauschbar, wenn für jede Wahl von natürlichen Zahlen mit $1 \leq n_1 < n_2 < n_3 \dots$ gilt

$$(X_1, X_2, \dots) \stackrel{\mathcal{D}}{=} (X_{n_1}, X_{n_2}, \dots).$$

Obacht: Die eine Richtung ist schwierig zu zeigen. 

14 W-Maße auf Produkträumen

Als Motivation betrachten wir das folgende Beispiel. Sei X eine uniform auf $[0, 1]$ verteilte Zufallsvariable. Sobald wir den Wert von X kennen, wollen wir n mal eine Münze werfen, die Erfolgswahrscheinlichkeit X hat. Die Ergebnisse seien Y_1, \dots, Y_n .

Wie konstruieren wir einen geeigneten Wahrscheinlichkeitsraum, auf dem alle Zufallsvariablen definiert sind?

Eine Möglichkeit: Wir wissen schon, wie wir $n + 1$ unabhängige Zufallsvariablen Z_0, \dots, Z_n konstruieren, die uniform auf $[0, 1]$ verteilt sind (siehe etwa Korollar 2.23). Setze nun $X = Z_0$ und

$$Y_k = \begin{cases} 1, & \text{falls } Z_k < X, \\ 0, & \text{falls } Z_k \geq X. \end{cases}$$

Intuitiv entspricht das unserer Vorstellung davon, dass die Y_1, \dots, Y_n unabhängig sind, sobald wir den Wert von X kennen, und mit Wahrscheinlichkeit X einen Erfolg zeigen.

In diesem Beispiel haben wir uns von Hand ein **Zweistufenexperiment** gebastelt: In der ersten Stufe haben wir den Wert von X festgelegt, in der zweiten Stufe wird, abhängig vom Wert von X , der Wert von $Y = (Y_1, \dots, Y_n)$ bestimmt. Offenbar haben wir bei der Konstruktion Nutzen aus der speziellen Struktur des Problems gezogen. Wir suchen nun nach einem allgemeinen Rahmen, in dem sich mehrstufige Experimente beschreiben und allgemein konstruieren lassen. Anders als in Kapitel 2 brauchen die Zufallsvariablen nicht mehr unabhängig zu sein. Außerdem wollen wir *systematisch* auch unendlich große Familien von Zufallsvariablen mit gegebenen Verteilungen konstruieren.

Im ersten Abschnitt beginnen wir mit Produkten von Messräumen. Danach kommen wir zu endlichen Produkten von Maßräumen sowie Produktmaßen mit Übergangskernen, um schließlich unendliche Produkte von Wahrscheinlichkeitsräumen zu betrachten. Das Hauptergebnis ist der Erweiterungssatz von Kolmogorov.

14.1 Produkträume

Definition 14.1 (Produktraum). Sei $(\Omega_i, i \in I)$ eine beliebige Familie von Mengen. Mit $\Omega = \bigtimes_{i \in I} \Omega_i$ bezeichnen wir die Menge der Abbildungen $\omega : I \rightarrow \bigcup_{i \in I} \Omega_i$ mit der Eigenschaft, dass $\omega(i) \in \Omega_i$ für jedes $i \in I$ gilt. Ω heißt das **Produkt** der $(\Omega_i, i \in I)$, oder kurz **Produktraum**. Sind speziell alle Ω_i gleich, etwa $\Omega_i = \Omega_0$, so schreiben wir $\Omega = \bigtimes_{i \in I} \Omega_i = \Omega_0^I$.

Beispiele 14.2. (i) Ist $\Omega_1 = \{1, \dots, 6\}$ und $\Omega_2 = \{1, 2, 3\}$, so ist

$$\Omega_1 \times \Omega_2 = \{\omega = (\omega_1, \omega_2) : \omega_1 \in \{1, \dots, 6\}, \omega_2 \in \{1, 2, 3\}\}.$$

- (ii) Ist $\Omega_0 = \mathbb{R}$ und $I = \{1, 2, 3\}$, so ist $\mathbb{R}^{\{1, 2, 3\}}$ isomorph zum üblichen \mathbb{R}^3 .
- (iii) Ist $\Omega_0 = \mathbb{R}$ und $I = \mathbb{N}$, so ist $\mathbb{R}^{\mathbb{N}}$ der Raum der Folgen $(\omega(n), n \in \mathbb{N})$ in \mathbb{R} .
- (iv) Ist $I = \mathbb{R}$ und $\Omega_0 = \mathbb{R}$, so ist $\mathbb{R}^{\mathbb{R}}$ die Menge der Abbildungen $\mathbb{R} \rightarrow \mathbb{R}$. ◇

Definition 14.3 (Koordinatenabbildung). Ist $i \in I$, so bezeichnet $X_i : \Omega \rightarrow \Omega_i$, $\omega \mapsto \omega(i)$ die *i-te Koordinatenabbildung*. Allgemeiner nennen wir für $J \subset J' \subset I$ die eingeschränkte Abbildung

$$X_J^{J'} : \bigtimes_{j \in J'} \Omega_j \longrightarrow \bigtimes_{j \in J} \Omega_j, \quad \omega' \mapsto \omega'|_J \quad (14.1)$$

die kanonische Projektion. Speziell schreiben wir $X_J := X_J^I$.

Definition 14.4 (Produkt- σ -Algebra). Seien $(\Omega_i, \mathcal{A}_i)$, $i \in I$, Messräume. Die **Produkt- σ -Algebra**

$$\mathcal{A} = \bigotimes_{i \in I} \mathcal{A}_i$$

ist die kleinste σ -Algebra auf Ω , sodass für jedes $i \in I$ die Abbildung X_i messbar bezüglich $\mathcal{A} - \mathcal{A}_i$ ist:

$$\mathcal{A} = \sigma(X_i, i \in I) := \sigma(X_i^{-1}(\mathcal{A}_i), i \in I).$$

Ist $(\Omega_i, \mathcal{A}_i) = (\Omega_0, \mathcal{A}_0)$ für jedes $i \in I$, so schreiben wir auch $\mathcal{A} = \mathcal{A}_0^{\otimes I}$.

Für $J \subset I$ schreiben wir $\Omega_J = \bigtimes_{j \in J} \Omega_j$ und $\mathcal{A}_J = \bigotimes_{j \in J} \mathcal{A}_j$.

Bemerkung 14.5. Die Begriffsbildung der Produkt- σ -Algebra ist analog zu der der **Produkttopologie**: Sind $((\Omega_i, \tau_i), i \in I)$ topologische Räume, so ist die Produkttopologie τ auf $\Omega = \bigtimes_{i \in I} \Omega_i$ die grösste Topologie, bezüglich der alle Koordinatenabbildungen $X_i : \Omega \longrightarrow \Omega_i$ stetig sind. ◇

Definition 14.6. Sei $I \neq \emptyset$ eine beliebige Indexmenge, (E, \mathcal{E}) ein Messraum, $(\Omega, \mathcal{A}) = (E^I, \mathcal{E}^{\otimes I})$ und $X_t : \Omega \rightarrow E$ die Koordinatenabbildung für jedes $t \in I$. Dann nennen wir die Familie $(X_t)_{t \in I}$ den **kanonischen Prozess** auf (Ω, \mathcal{A}) .

Lemma 14.7. Sei $\emptyset \neq J \subset I$. Dann ist X_J^I messbar bezüglich $\mathcal{A}_I - \mathcal{A}_J$.

Beweis. Für jedes $j \in J$ ist $X_j = X_j^J \circ X_J^I$ messbar bezüglich $\mathcal{A} - \mathcal{A}_j$. Nach Korollar 1.82 ist daher X_J^I messbar. \square

Satz 14.8. Sei I höchstens abzählbar, und für jedes $i \in I$ sei (Ω_i, τ_i) polnisch mit Borel'scher σ -Algebra $\mathcal{B}_i = \sigma(\tau_i)$. Es sei τ die Produkttopologie auf $\Omega = \prod_{i \in I} \Omega_i$ und $\mathcal{B} = \sigma(\tau)$.

Dann ist (Ω, τ) polnisch und $\mathcal{B} = \bigotimes_{i \in I} \mathcal{B}_i$. Speziell ist $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}$ für $d \in \mathbb{N}$.

Beweis. Ohne Einschränkung sei $I = \mathbb{N}$. Für $i \in \mathbb{N}$ sei d_i eine vollständige Metrik, die τ_i erzeugt. Man prüft leicht nach, dass dann

$$d(\omega, \omega') := \sum_{i=1}^{\infty} 2^{-i} \frac{d_i(\omega(i), \omega'(i))}{1 + d_i(\omega(i), \omega'(i))} \quad (14.2)$$

eine vollständige Metrik auf Ω ist, die τ erzeugt.

Für jedes $i \in \mathbb{N}$ sei nun $D_i \subset \Omega_i$ eine abzählbare, dichte Teilmenge und $y_i \in D_i$ ein beliebiger fester Punkt. Die Menge

$$D = \left\{ x \in \bigtimes_{i \in \mathbb{N}} D_i : x_i \neq y_i \text{ nur endlich oft} \right\}$$

ist, wie man leicht prüft, eine abzählbare, dichte Teilmenge von Ω . Also ist Ω separabel und damit polnisch.

Sei nun $\beta_i = \{B_\varepsilon(x_i) : x_i \in D_i, \varepsilon \in \mathbb{Q}^+\}$ für jedes $i \in I$ eine abzählbare Basis der Topologie von Ω_i aus ε -Kugeln. Setze

$$\beta := \bigcup_{N=1}^{\infty} \left\{ \bigcap_{i=1}^N X_i^{-1}(B_i) : B_1 \in \beta_1, \dots, B_N \in \beta_N \right\}.$$

Dann ist β eine abzählbare Basis der Topologie τ , also ist jede offene Menge $A \subset \Omega$ (abzählbare) Vereinigung von Mengen in $\beta \subset \bigotimes_{i \in \mathbb{N}} \mathcal{B}_i$. Mithin ist $\tau \subset \bigotimes_{i \in \mathbb{N}} \mathcal{B}_i$ und damit $\mathcal{B} \subset \bigotimes_{i \in \mathbb{N}} \mathcal{B}_i$.

Andererseits ist jedes X_i stetig, also messbar bezüglich $\mathcal{B} - \mathcal{B}_i$ und damit $\mathcal{B} \supset \bigotimes_{i \in \mathbb{N}} \mathcal{B}_i$. \square

Definition 14.9 (Zylindermengen). Für jedes $i \in I$ sei $\mathcal{E}_i \subset \mathcal{A}_i$ ein Teilsystem der messbaren Mengen.

Für jedes $A \in \mathcal{A}_J$ heißt $X_J^{-1}(A) \subset \Omega$ **Zylindermenge** mit Basis J . Die Menge dieser Zylindermengen wird mit \mathcal{Z}_J bezeichnet. Ist speziell $A = \bigtimes_{j \in J} A_j$ für gewisse $A_j \in \mathcal{A}_j$, so heißt $X_J^{-1}(A)$ **Rechteckzylinder** mit Basis J . Die Menge dieser Rechteckzylinder wird mit \mathcal{Z}_J^R bezeichnet, die Menge aller Rechteckzylinder, für die zusätzlich $A_j \in \mathcal{E}_j$ für jedes $j \in J$ gilt, mit $\mathcal{Z}_J^{\mathcal{E}, R}$.

Wir schreiben

$$\mathcal{Z} = \bigcup_{J \subset I \text{ endlich}} \mathcal{Z}_J, \quad (14.3)$$

und definieren analog \mathcal{Z}^R und $\mathcal{Z}^{\mathcal{E}, R}$. Ferner definieren wir

$$\mathcal{Z}_*^R = \bigcup_{N=1}^{\infty} \left\{ \bigcup_{n=1}^N A_n : A_1, \dots, A_N \in \mathcal{Z}^R \right\}$$

und analog $\mathcal{Z}_*^{\mathcal{E}, R}$.

Bemerkung 14.10. Jedes \mathcal{Z}_J ist eine σ -Algebra, und \mathcal{Z} und \mathcal{Z}_*^R sind Algebren. Außerdem gilt $\bigotimes_{i \in I} \mathcal{A}_i = \sigma(\mathcal{Z})$. \diamond

Lemma 14.11. Ist jedes \mathcal{E}_i schnittstabil, so ist $\mathcal{Z}^{\mathcal{E}, R}$ schnittstabil.

Beweis. Übung! \square

Satz 14.12. Für jedes $i \in I$ sei $\mathcal{E}_i \subset \mathcal{A}_i$ ein Erzeuger von \mathcal{A}_i .

(i) Für jedes endliche $J \subset I$ gilt $\bigotimes_{j \in J} \mathcal{A}_j = \sigma\left(\bigtimes_{j \in J} E_j : E_j \in \mathcal{E}_j\right)$.

(ii) Es gilt $\bigotimes_{i \in I} \mathcal{A}_i = \sigma(\mathcal{Z}^R) = \sigma(\mathcal{Z}^{\mathcal{E}, R})$.

(iii) Sei μ ein σ -endliches Maß auf \mathcal{A} , und sei jedes \mathcal{E}_i zudem schnittstabil. Ferner gebe es eine Folge $(E_n)_{n \in \mathbb{N}}$ in $\mathcal{Z}^{\mathcal{E}, R}$ mit $E_n \uparrow \Omega$ und $\mu(E_n) < \infty$ für jedes $n \in \mathbb{N}$ (speziell ist diese Bedingung natürlich erfüllt, wenn μ endlich ist und $\Omega_i \in \mathcal{E}_i$ für jedes $i \in I$). Dann ist μ durch Angabe von $\mu(A)$ für jedes $A \in \mathcal{Z}^{\mathcal{E}, R}$ eindeutig festgelegt.

Beweis. (i) Sei $\mathcal{A}'_J = \sigma\left(\bigtimes_{j \in J} E_j : E_j \in \mathcal{E}_j \text{ für jedes } j \in J\right)$. Es ist

$$\bigtimes_{j \in J} E_j = \bigcap_{j \in J} (X_j^J)^{-1}(E_j) \in \mathcal{A}_J,$$

also $\mathcal{A}'_J \subset \mathcal{A}_J$. Umgekehrt gilt $(X_j^J)^{-1}(E_j) \in \mathcal{A}'_J$ für $j \in J$ und $E_j \in \mathcal{E}_j$. Da \mathcal{E}_i ein Erzeuger von \mathcal{A}_i ist, ist $(X_j^J)^{-1}(A_j) \in \mathcal{A}'_J$ für jedes $A_j \in \mathcal{A}_j$, also gilt $\mathcal{A}_J \subset \mathcal{A}'_J$.

(ii) Offenbar ist $\mathcal{Z}^{\mathcal{E},R} \subset \mathcal{Z}^R \subset \mathcal{A}$, also auch $\sigma(\mathcal{Z}^{\mathcal{E},R}) \subset \sigma(\mathcal{Z}^R) \subset \mathcal{A}$. Nach Satz 1.81 gilt $\sigma(\mathcal{Z}_{\{i\}}^{\mathcal{E},R}) = \sigma(X_i)$ für jedes $i \in I$, also $\sigma(X_i) \subset \sigma(\mathcal{Z}^{\mathcal{E},R})$ und damit $\mathcal{A}_I \subset \sigma(\mathcal{Z}^{\mathcal{E},R})$.

(iii) Nach (ii) und Lemma 14.11 ist $\mathcal{Z}^{\mathcal{E},R}$ ein schnittstabiler Erzeuger von \mathcal{A} . Die Aussage folgt daher aus Lemma 1.42. \square

Übung 14.1.1. Man zeige:

$$\bigotimes_{i \in I} \mathcal{A}_i = \bigcup_{J \subset I \text{ abzählbar}} \mathcal{Z}_J. \quad (14.4)$$

Hinweis: Man zeige, dass die rechte Seite eine σ -Algebra ist. \clubsuit

14.2 Endliche Produkte und Übergangskerne

Wir betrachten jetzt die Situation endlich vieler σ -endlicher Maßräume $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, \dots, n$, wobei $n \in \mathbb{N}$.

Lemma 14.13. Sei $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ und $f : \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}$ eine $\mathcal{A}_1 \otimes \mathcal{A}_2$ -messbare Abbildung. Dann gilt für jedes $\tilde{\omega}_1 \in \Omega_1$ und $\tilde{\omega}_2 \in \Omega_2$

$$\begin{aligned} A_{\tilde{\omega}_1} &:= \{\omega_2 \in \Omega_2 : (\tilde{\omega}_1, \omega_2) \in A\} \in \mathcal{A}_2, \\ A_{\tilde{\omega}_2} &:= \{\omega_1 \in \Omega_1 : (\omega_1, \tilde{\omega}_2) \in A\} \in \mathcal{A}_1, \\ f_{\tilde{\omega}_1} : \Omega_2 &\rightarrow \overline{\mathbb{R}}, \quad \omega_2 \mapsto f(\tilde{\omega}_1, \omega_2) \quad \text{ist } \mathcal{A}_2\text{-messbar,} \\ f_{\tilde{\omega}_2} : \Omega_1 &\rightarrow \overline{\mathbb{R}}, \quad \omega_1 \mapsto f(\omega_1, \tilde{\omega}_2) \quad \text{ist } \mathcal{A}_1\text{-messbar.} \end{aligned}$$

Beweis. Für $\tilde{\omega}_1$ definiere die Einbettung $i : \Omega_2 \rightarrow \Omega_1 \times \Omega_2$ durch $i(\omega_2) = (\tilde{\omega}_1, \omega_2)$. Da $X_1 \circ i$ konstant gleich $\tilde{\omega}_1$ ist (also \mathcal{A}_1 -messbar), und $X_2 \circ i = \text{id}_{\Omega_2}$ (also \mathcal{A}_2 -messbar), ist nach Korollar 1.82 die Abbildung i messbar bezüglich $\mathcal{A}_2 - (\mathcal{A}_1 \otimes \mathcal{A}_2)$. Mithin ist $A_{\tilde{\omega}_1} = i^{-1}(A) \in \mathcal{A}_2$ und $f_{\tilde{\omega}_1} = f \circ i$ messbar bezüglich \mathcal{A}_2 . \square

Der folgende Satz verallgemeinert Satz 1.61.

Satz 14.14 (Endliche Produktmaß). Es existiert genau ein σ -endliches Maß μ auf $\mathcal{A} := \bigotimes_{i=1}^n \mathcal{A}_i$ mit

$$\mu(A_1 \times \cdots \times A_n) = \prod_{i=1}^n \mu_i(A_i) \quad \text{für } A_i \in \mathcal{A}_i, i = 1, \dots, n. \quad (14.5)$$

Wir nennen $\bigotimes_{i=1}^n \mu_i := \mu_1 \otimes \cdots \otimes \mu_n := \mu$ das **Produktmaß** der μ_i .

Sind alle Räume gleich $(\Omega_0, \mathcal{A}_0, \mu_0)$, so schreiben wir $\mu_0^{\otimes n} := \bigotimes_{i=1}^n \mu_0$.

Beweis. Sei $\tilde{\mu}$ auf \mathcal{Z}^R wie μ in (14.5) festgesetzt. Offenbar ist $\tilde{\mu}(\emptyset) = 0$, und man überlegt sich leicht, dass $\tilde{\mu}$ σ -endlich ist. Seien $A^1, A^2, \dots \in \mathcal{Z}^R$ paarweise disjunkt und $A \in \mathcal{Z}^R$ mit $A \subset \bigcup_{k=1}^{\infty} A^k$. Dann ist nach dem Satz über monotone Konvergenz

$$\begin{aligned} \tilde{\mu}(A) &= \int \mu_1(d\omega_1) \cdots \int \mu_n(d\omega_n) \mathbb{1}_A((\omega_1, \dots, \omega_n)) \\ &\leq \int \mu_1(d\omega_1) \cdots \int \mu_n(d\omega_n) \sum_{k=1}^{\infty} \mathbb{1}_{A^k}((\omega_1, \dots, \omega_n)) = \sum_{k=1}^{\infty} \tilde{\mu}(A^k). \end{aligned}$$

Ist speziell $A = A^1 \uplus A^2$, so erhält man analog $\tilde{\mu}(A) = \tilde{\mu}(A^1) + \tilde{\mu}(A^2)$. Mithin ist $\tilde{\mu}$ eine σ -endliche, additive, σ -subadditive Mengenfunktion auf dem Semiring \mathcal{Z}^R mit $\tilde{\mu}(\emptyset) = 0$. Nach dem Fortsetzungssatz (Satz 1.53) kann $\tilde{\mu}$ in eindeutiger Weise zu einem σ -endlichen Maß fortgesetzt werden. \square

Beispiel 14.15. Für $i = 1, \dots, n$ sei $(\Omega_i, \mathcal{A}_i, \mathbf{P}_i)$ ein Wahrscheinlichkeitsraum. Auf dem Raum $(\Omega, \mathcal{A}, \mathbf{P}) := (\bigtimes_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i, \bigotimes_{i=1}^n \mathbf{P}_i)$ sind die Koordinatenabbildungen $X_i : \Omega \rightarrow \Omega_i$ unabhängig mit Verteilung $\mathbf{P}_{X_i} = \mathbf{P}_i$. \diamond

Satz 14.16 (Fubini). Seien $(\Omega_i, \mathcal{A}_i, \mu_i)$ σ -endliche Maßräume, $i = 1, 2$, sowie $f : \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}$ messbar bezüglich $\mathcal{A}_1 \otimes \mathcal{A}_2$. Ist $f \geq 0$ oder $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$, dann gelten

$$\begin{aligned}\omega_1 &\mapsto \int f(\omega_1, \omega_2) \mu_2(d\omega_2) \text{ ist } \mathcal{A}_1\text{-messbar,} \\ \omega_2 &\mapsto \int f(\omega_1, \omega_2) \mu_1(d\omega_1) \text{ ist } \mathcal{A}_2\text{-messbar,}\end{aligned}\tag{14.6}$$

und

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f d(\mu_1 \otimes \mu_2) &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1) \right) \mu_2(d\omega_2).\end{aligned}\tag{14.7}$$

Beweis. Der Beweis folgt dem üblichen Schema der schrittweisen Approximation ausgehend von einfachen Funktionen.

Sei zunächst $f = \mathbb{1}_A$ für $A = A_1 \times A_2$ mit $A_1 \in \mathcal{A}_1$ und $A_2 \in \mathcal{A}_2$. Dann gelten (14.6) und (14.7) trivialerweise. Durch endliche Summenbildung gilt dies nun auch für $A \in \mathcal{Z}_*^R$ (Algebra der endlichen Vereinigungen von Rechtecken).

Sei nun $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Nach dem Approximationssatz (Satz 1.65) gibt es eine Folge von Mengen $(A^n)_{n \in \mathbb{N}}$ in \mathcal{Z}_*^R , die A dem Maße $\mu_1 \otimes \mu_2$ nach approximieren. Da Limiten messbarer Funktionen wieder messbar sind, und nach Konstruktion die Integrale konvergieren, gelten (14.6) und (14.7) jetzt auch für $f = \mathbb{1}_A$ und $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Durch endliche Summenbildung gelten nun (14.6) und (14.7) auch für den Fall, wo f eine Elementarfunktion ist.

Wir betrachten jetzt $f \geq 0$. Dann existiert nach Satz 1.96 eine Folge von Elementarfunktionen $(f_n)_{n \in \mathbb{N}}$ mit $f_n \uparrow f$. Nach dem Satz von der monotonen Konvergenz (Satz 4.20) gelten (14.6) und (14.7) nun auch für dieses f .

Ist $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$, dann ist $f = f^+ - f^-$ mit integrierbaren Funktionen $f^+, f^- \geq 0$, für die (14.6) und (14.7) gelten, also auch für f . \square

In Definition 2.32 hatten wir die Faltung zweier W-Maße μ und ν als die Verteilung der Summe zweier unabhängiger, wie μ und ν verteilter Zufallsvariablen definiert. Als eine einfache Anwendung des Satzes von Fubini wollen wir eine neue Definition geben (die natürlich mit der alten konsistent ist), die alle endlichen Maße auf dem \mathbb{R}^n umfasst. Haben diese Maße zusätzlich Dichten bezüglich des Lebesgue-Maßes, so können wir eine explizite Formel zur Berechnung der Faltung angeben.

Seien also X und Y \mathbb{R}^n -wertige Zufallsvariablen mit Dichten f_X und f_Y . Das heißt, $f_X, f_Y : \mathbb{R}^n \rightarrow [0, \infty]$ sind messbar und integrierbar bezüglich des n -dimensionalen Lebesgue-Maßes λ^n , und es gilt für jedes $x \in \mathbb{R}^n$

$$\mathbf{P}[X \leq x] = \int_{(-\infty, x]} f_X(t) \lambda^n(dt) \quad \text{und} \quad \mathbf{P}[Y \leq x] = \int_{(-\infty, x]} f_Y(t) \lambda^n(dt).$$

Hierbei ist $(-\infty, x] = \{y \in \mathbb{R}^n : y_i \leq x_i \text{ für } i = 1, \dots, n\}$ (vergleiche (1.5)).

Definition 14.17. Sei $n \in \mathbb{N}$. Für zwei Lebesgue-integrierbare Abbildungen $f, g : \mathbb{R}^n \rightarrow [0, \infty]$ definieren wir die **Faltung** $f * g : \mathbb{R}^n \rightarrow [0, \infty]$ durch

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y) g(x - y) \lambda^n(dy).$$

Für zwei endliche Maße $\mu, \nu \in \mathcal{M}_f(\mathbb{R}^n)$ definieren wir die Faltung $\mu * \nu \in \mathcal{M}_f(\mathbb{R}^n)$ durch

$$(\mu * \nu)((-\infty, x]) = \int \int \mathbb{1}_{A_x}(u, v) \mu(du) \nu(dv),$$

wobei $A_x := \{(u, v) \in \mathbb{R}^n \times \mathbb{R}^n : u + v \leq x\}$ ist.

Lemma 14.18. Die Abbildung $f * g$ ist messbar, und es gelten $f * g = g * f$ und

$$\int_{\mathbb{R}^n} (f * g) d\lambda^n = \left(\int_{\mathbb{R}^n} f d\lambda^n \right) \left(\int_{\mathbb{R}^n} g d\lambda^n \right).$$

Ebenso gelten $\mu * \nu = \nu * \mu$ und $(\mu * \nu)(\mathbb{R}^n) = \mu(\mathbb{R}^n) \nu(\mathbb{R}^n)$.

Beweis. Die Aussagen folgen direkt aus dem Satz von Fubini. □

Satz 14.19 (Faltung von n -dimensionalen Maßen).

- (i) Sind X und Y unabhängige \mathbb{R}^n -wertige Zufallsvariablen mit Dichten f_X und f_Y , so hat $X + Y$ die Dichte $f_X * f_Y$.
- (ii) Sind $\mu = f \lambda^n$ und $\nu = g \lambda^n$ endliche Maße mit Dichten f und g bezüglich des Lebesgue-Maßes, so gilt $\mu * \nu = (f * g) \lambda^n$.

Beweis. (i) Sei $x \in \mathbb{R}^n$ und $A := \{(u, v) \in \mathbb{R}^n \times \mathbb{R}^n : u + v \leq x\}$. Dann gilt nach mehrfacher Anwendung des Satzes von Fubini (sowie der Translationsinviananz von λ^n)

$$\begin{aligned}
\mathbf{P}[X + Y \leq x] &= \mathbf{P}[(X, Y) \in A] \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbb{1}_A(u, v) f_X(u) f_Y(v) (\lambda^n)^{\otimes 2}(d(u, v)) \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \mathbb{1}_A(u, v) f_X(u) \lambda^n(du) \right) f_Y(v) \lambda^n(dv) \\
&= \int_{\mathbb{R}^n} \left(\int_{(-\infty, x-v]} f_X(u) \lambda^n(du) \right) f_Y(v) \lambda^n(dv) \\
&= \int_{\mathbb{R}^n} \left(\int_{(-\infty, x]} f_X(u-v) \lambda^n(du) \right) f_Y(v) \lambda^n(dv) \\
&= \int_{(-\infty, x]} \left(\int_{\mathbb{R}^n} f_X(u-v) f_Y(v) \lambda^n(dv) \right) \lambda^n(du) \\
&= \int_{(-\infty, x]} (f_X * f_Y)(u) d\lambda^n.
\end{aligned}$$

(ii) Ersetze in (i) $\mu = \mathbf{P}_X$ und $\nu = \mathbf{P}_Y$. Die Aussage folgt unmittelbar. \square

Wir kommen zu einer Begriffsbildung, die diejenige der Produktmaße verallgemeinert und in Richtung unseres Eingangsbeispiels steuert.

Wir erinnern an den Begriff des Übergangskerns aus Definition 8.25.

Lemma 14.20. Sei κ ein endlicher Übergangskern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$, und sei $f : \Omega_1 \times \Omega_2 \rightarrow [0, \infty]$ messbar bezüglich $\mathcal{A}_1 \otimes \mathcal{A}_2 - \mathcal{B}([0, \infty])$. Dann ist die Abbildung

$$\begin{aligned}
I_f : \Omega_1 &\rightarrow [0, \infty] \\
\omega_1 &\mapsto \int f(\omega_1, \omega_2) \kappa(\omega_1, d\omega_2)
\end{aligned}$$

wohldefiniert und \mathcal{A}_1 -messbar.

Beweis. Nach Lemma 14.13 ist für jedes $\omega_1 \in \Omega_1$ die Abbildung f_{ω_1} messbar bezüglich \mathcal{A}_2 , also ist $I_f(\omega_1) = \int f_{\omega_1}(\omega_2) \kappa(\omega_1, d\omega_2)$ wohldefiniert. Wir müssen also nur noch die Messbarkeit von I_f zeigen.

Ist $g = \mathbb{1}_{A_1 \times A_2}$ für $A_1 \in \mathcal{A}_1$ und $A_2 \in \mathcal{A}_2$, so ist $I_g(\omega_1) = \mathbb{1}_{A_1}(\omega_1) \kappa(\omega_1, A_2)$ offenbar messbar. Sei nun $\mathcal{D} = \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : I_{\mathbb{1}_A} \text{ ist } \mathcal{A}_1\text{-messbar}\}$. Wir zeigen, dass \mathcal{D} ein Dynkin-System ist:

- (i) Offenbar ist $\Omega_1 \times \Omega_2 \in \mathcal{D}$.
- (ii) Sind $A, B \in \mathcal{D}$ mit $A \subset B$, so ist $I_{\mathbb{1}_{B \setminus A}} = I_{\mathbb{1}_B} - I_{\mathbb{1}_A}$ messbar, wobei wir die Endlichkeit von κ ausgenutzt haben, also ist $B \setminus A \in \mathcal{D}$.

- (iii) Sind $A_1, A_2, \dots \in \mathcal{D}$ paarweise disjunkt und $A := \bigcup_{n=1}^{\infty} A_n$, so ist $I_{\mathbb{1}_A} = \sum_{n=1}^{\infty} I_{\mathbb{1}_{A_n}}$ messbar, also $A \in \mathcal{D}$.

Nun ist \mathcal{D} also ein Dynkin-System, das den schnittstabilen Erzeuger aller Rechteckmengen in $\mathcal{A}_1 \otimes \mathcal{A}_2$ enthält, also ist (nach Satz 1.19) $\mathcal{D} = \mathcal{A}_1 \otimes \mathcal{A}_2$. Mithin ist $I_{\mathbb{1}_A}$ messbar für jedes $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Es folgt, dass I_g messbar ist für jede Elementarfunktion. Sei nun $(f_n)_{n \in \mathbb{N}}$ eine Folge von Elementarfunktionen mit $f_n \uparrow f$. Für jedes feste $\omega_1 \in \Omega_1$ gilt nach dem Satz von der monotonen Konvergenz $I_f(\omega_1) = \lim_{n \rightarrow \infty} I_{f_n}(\omega_1)$, und I_f ist als Limes messbarer Funktionen messbar. \square

Bemerkung 14.21. Wir schreiben im Folgenden oft $\int \kappa(\omega_1, d\omega_2) f(\omega_1, \omega_2)$ statt $\int f(\omega_1, \omega_2) \kappa(\omega_1, d\omega_2)$, denn bei Mehrfachintegralen erlaubt diese Notation es, den Integrator näher an das betreffende Integralzeichen heran zu rücken. \diamond

Satz 14.22. Seien $(\Omega_i, \mathcal{A}_i)$, $i = 0, 1, 2$, Messräume und κ_1 ein endlicher Übergangskern von $(\Omega_0, \mathcal{A}_0)$ nach $(\Omega_1, \mathcal{A}_1)$ sowie κ_2 ein endlicher Übergangskern von $(\Omega_0 \times \Omega_1, \mathcal{A}_0 \otimes \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$. Dann ist die Abbildung

$$\kappa_1 \otimes \kappa_2 : \Omega_0 \times (\mathcal{A}_1 \otimes \mathcal{A}_2) \rightarrow [0, \infty)$$

$$(\omega_0, A) \mapsto \int_{\Omega_1} \kappa_1(\omega_0, d\omega_1) \int_{\Omega_2} \kappa_2((\omega_0, \omega_1), d\omega_2) \mathbb{1}_A((\omega_1, \omega_2))$$

wohldefiniert und ist ein σ -endlicher (aber nicht notwendig endlicher) Übergangskern von $(\Omega_0, \mathcal{A}_0)$ nach $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$. Sind κ_1 und κ_2 (sub-)stochastisch, so ist $\kappa_1 \otimes \kappa_2$ (sub-)stochastisch. Wir nennen $\kappa_1 \otimes \kappa_2$ das **Produkt** von κ_1 und κ_2 .

Ist κ_2 ein Kern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$, so definieren wir das Produkt $\kappa_1 \otimes \kappa_2$ analog, indem wir κ_2 einfach formal als Kern von $(\Omega_0 \times \Omega_1, \mathcal{A}_0 \otimes \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$ auffassen, der nicht von der Ω_0 -Koordinate abhängt.

Beweis. Sei $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Die Abbildung

$$g_A : (\omega_0, \omega_1) \mapsto \int \kappa_2((\omega_0, \omega_1), d\omega_2) \mathbb{1}_A(\omega_1, \omega_2)$$

ist nach Lemma 14.20 wohldefiniert und messbar bezüglich $\mathcal{A}_0 \otimes \mathcal{A}_1$. Daher ist, wiederum nach Lemma 14.20, die Abbildung

$$\omega_0 \mapsto \kappa_1 \otimes \kappa_2(\omega_0, A) = \int \kappa_1(\omega_0, d\omega_1) g_A(\omega_0, \omega_1)$$

wohldefiniert und \mathcal{A}_0 -messbar. Für festes ω_0 ist nach dem Satz über monotone Konvergenz die Abbildung $A \mapsto \kappa_1 \otimes \kappa_2(\omega_0, A)$ σ -additiv, also ein Maß.

Für $\omega_0 \in \Omega_0$ und $n \in \mathbb{N}$ sei $A_{\omega_0, n} := \{\omega_1 \in \Omega_1 : \kappa_2((\omega_0, \omega_1), \Omega_2) < n\}$. Da κ_2 endlich ist, gilt $\bigcup_{n \geq 1} A_{\omega_0, n} = \Omega_1$ für jedes $\omega_0 \in \Omega_0$, und es gilt $\kappa_1 \otimes \kappa_2(\omega_0, A_n \times \Omega_2) \leq n \cdot \kappa_1(\omega_0, A_n) < \infty$. Also ist $\kappa_1 \otimes \kappa_2(\omega_0, \cdot)$ σ -endlich und damit ein Übergangskern.

Der Zusatz ist trivial. \square

Korollar 14.23 (Produkte mit Kernen). Sei $(\Omega_1, \mathcal{A}_1, \mu)$ ein endlicher Maßraum, $(\Omega_2, \mathcal{A}_2)$ ein Messraum und κ ein endlicher Übergangskern von Ω_1 nach Ω_2 . Dann existiert ein eindeutig bestimmtes, σ -endliches Maß $\mu \otimes \kappa$ auf $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ mit

$$\mu \otimes \kappa(A_1 \times A_2) = \int_{A_1} \kappa(\omega_1, A_2) \mu(d\omega_1) \quad \text{für alle } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

Ist κ stochastisch und μ ein W-Maß, so ist $\mu \otimes \kappa$ ein W-Maß.

Beweis. Wende Satz 14.22 an mit $\kappa_2 = \kappa$ und $\kappa_1(\omega_0, \cdot) = \mu$. \square

Korollar 14.24. Seien $n \in \mathbb{N}$ und $(\Omega_i, \mathcal{A}_i)$, $i = 0, \dots, n$, Messräume. Für $i = 1, \dots, n$ sei κ_i ein substochastischer Kern von $\left(\bigtimes_{k=0}^{i-1} \Omega_k, \bigotimes_{k=0}^{i-1} \mathcal{A}_k \right)$ nach $(\Omega_i, \mathcal{A}_i)$ oder von $(\Omega_{i-1}, \mathcal{A}_{i-1})$ nach $(\Omega_i, \mathcal{A}_i)$. Dann definiert die Rekursion $\kappa_1 \otimes \dots \otimes \kappa_i := (\kappa_1 \otimes \dots \otimes \kappa_{i-1}) \otimes \kappa_i$ für jedes $i = 1, \dots, n$ einen substochastischen Kern $\bigotimes_{k=1}^i \kappa_k := \kappa_1 \otimes \dots \otimes \kappa_i$ von $(\Omega_0, \mathcal{A}_0)$ nach $\left(\bigtimes_{k=1}^i \Omega_k, \bigotimes_{k=1}^i \mathcal{A}_k \right)$. Sind alle κ_i stochastisch, so ist jedes $\bigotimes_{k=1}^i \kappa_k$ stochastisch.

Ist μ ein endliches Maß auf $(\Omega_0, \mathcal{A}_0)$, so ist $\mu_i := \mu \otimes \bigotimes_{k=1}^i \kappa_k$ ein endliches Maß auf $\left(\bigtimes_{k=0}^i \Omega_k, \bigotimes_{k=0}^i \mathcal{A}_k \right)$. Ist μ ein W-Maß und jedes κ_i stochastisch, so ist μ_i ein Wahrscheinlichkeitsmaß.

Beweis. Die Aussagen folgen per Induktion aus Satz 14.22. \square

Definition 14.25 (Verkettung von Kernen). Seien $(\Omega_i, \mathcal{A}_i)$ Messräume, $i = 0, 1, 2$, und κ_i ein substochastischer Kern von $(\Omega_{i-1}, \mathcal{A}_{i-1})$ nach $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$. Wir definieren die Verkettung von κ_1 und κ_2 durch

$$\begin{aligned} \kappa_1 \cdot \kappa_2 : \Omega_0 \times \mathcal{A}_2 &\rightarrow [0, \infty) \\ (\omega_0, A_2) &\mapsto \int_{\Omega_1} \kappa_1(\omega_0, d\omega_1) \kappa_2(\omega_1, A_2). \end{aligned}$$

Satz 14.26. Bezeichnen wir mit $\pi_2 : \Omega_1 \times \Omega_2 \rightarrow \Omega_2$ die Projektion auf die zweite Koordinate, so ist

$$(\kappa_1 \cdot \kappa_2)(\omega_0, A_2) = (\kappa_1 \otimes \kappa_2)(\omega_0, \pi_2^{-1}(A_2)) \quad \text{für jedes } A_2 \in \mathcal{A}_2.$$

Speziell ist die Verkettung $\kappa_1 \cdot \kappa_2$ ein (sub-)stochastischer Kern von $(\Omega_0, \mathcal{A}_0)$ nach $(\Omega_2, \mathcal{A}_2)$.

Beweis. Klar. □

Lemma 14.27 (Kerne und Faltung). Seien μ und ν W-Maße auf \mathbb{R}^d und die Kerne $\kappa_i : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $i = 1, 2$, definiert durch $\kappa_1(x, dy) = \mu(dy)$ sowie $\kappa_2(y, dz) = (\delta_y * \nu)(dz)$. Dann ist $\kappa_1 \cdot \kappa_2 = \mu * \nu$.

Beweis. Das ist trivial. □

Satz 14.28 (Kerne und Faltung). Seien X_1, X_2, \dots unabhängige \mathbb{R}^d -wertige Zufallsvariablen mit Verteilungen $\mu_i := \mathbf{P}_{X_i}$, $i = 1, \dots, n$. Setze $S_k := X_1 + \dots + X_k$ für $k = 1, \dots, n$ und definiere stochastische Kerne von \mathbb{R}^d nach \mathbb{R}^d durch $\kappa_k(x, \cdot) = \delta_x * \mu_k$ für $k = 1, \dots, n$. Dann gilt

$$\left(\bigotimes_{k=1}^n \kappa_k \right) (0, \cdot) = \mathbf{P}_{(S_1, \dots, S_n)}. \quad (14.8)$$

Beweis. Für $k = 1, \dots, n$ definiere die messbare Bijektion $\varphi_k : (\mathbb{R}^d)^k \rightarrow (\mathbb{R}^d)^k$ durch

$$\varphi_k(x_1, \dots, x_k) = (x_1, x_1 + x_2, \dots, x_1 + \dots + x_k).$$

Offenbar ist $\mathcal{B}((\mathbb{R}^d)^n) = \sigma(\varphi_n(A_1 \times \dots \times A_n) : A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}^d))$. Es reicht also (14.8) für Mengen von diesem Typ nachzuweisen, also zu zeigen, dass

$$\left(\bigotimes_{k=1}^n \kappa_k \right) (0, \varphi_n(A_1 \times \dots \times A_n)) = \mathbf{P}_{(S_1, \dots, S_n)}(\varphi_n(A_1 \times \dots \times A_n)) = \prod_{k=1}^n \mu_k(A_k).$$

Für $n = 1$ ist die Aussage klar. Per Definition ist $\kappa_n(y_{n-1}, y_{n-1} + A_n) = \mu_n(A_n)$. Induktiv folgt

$$\begin{aligned} & \left(\bigotimes_{k=1}^n \kappa_k \right) (0, \varphi_n(A_1 \times \dots \times A_n)) \\ &= \int_{\varphi_{n-1}(A_1 \times \dots \times A_{n-1})} \left(\bigotimes_{k=1}^{n-1} \kappa_k \right) (0, d(y_1, \dots, y_{n-1})) \kappa_n(y_{n-1}, y_{n-1} + A_n) \\ &= \left(\prod_{k=1}^{n-1} \mu_k(A_k) \right) \mu_n(A_n). \end{aligned} \quad \square$$

Satz 14.29 (Fubini für Übergangskerne). Seien $(\Omega_i, \mathcal{A}_i)$ Messräume, $i = 1, 2$, μ ein endliches Maß auf $(\Omega_1, \mathcal{A}_1)$, κ ein endlicher Übergangskern von Ω_1 nach Ω_2 sowie $f : \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}$ messbar bezüglich $\mathcal{A}_1 \otimes \mathcal{A}_2$. Ist $f \geq 0$ oder $f \in \mathcal{L}^1(\mu \otimes \kappa)$, dann gilt

$$\int_{\Omega_1 \times \Omega_2} f d(\mu \otimes \kappa) = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \kappa(\omega_1, d\omega_2) \right) \mu(d\omega_1). \quad (14.9)$$

Beweis. Für $f = \mathbb{1}_{A_1 \times A_2}$ mit $A_1 \in \mathcal{A}_1$ und $A_2 \in \mathcal{A}_2$ gilt die Aussage per Definition. Für allgemeine f verwenden wir die Approximationsargumente wie in Satz 14.16. \square

Beispiel 14.30. Wir kommen auf das Beispiel vom Anfang des Kapitels zurück. Sei $n \in \mathbb{N}$ und $(\Omega_2, \mathcal{A}_2) = (\{0, 1\}^n, (2^{\{0,1\}})^{\otimes n})$ der Raum des n -fachen Münzwurfs. Für jedes $p \in [0, 1]$ definieren wir

$$P_p = (\text{Ber}_p)^{\otimes n} = ((1-p)\delta_0 + p\delta_1)^{\otimes n}.$$

P_p ist dasjenige W-Maß auf $(\Omega_2, \mathcal{A}_2)$, mit dem die Koordinatenabbildungen Y_i unabhängige Bernoulli-Zufallsvariablen mit Erfolgswahrscheinlichkeit p sind.

Ferner sei $\Omega_1 = [0, 1]$ und $\mathcal{A}_1 = \mathcal{B}([0, 1])$ die Borel'sche σ -Algebra auf Ω_1 sowie $\mu = \mathcal{U}_{[0,1]}$ die Gleichverteilung auf $[0, 1]$. Die identische Abbildung $X : \Omega_1 \rightarrow [0, 1]$, ist dann eine uniform auf $[0, 1]$ verteilte Zufallsvariable auf $(\Omega_1, \mathcal{A}_1, \mu)$.

Schließlich betrachten wir den stochastischen Kern von Ω_1 nach Ω_2

$$\kappa(\omega_1, \cdot) = P_{\omega_1}.$$

Setzen wir $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ und $\mathbf{P} = \mu \otimes \kappa$, dann beschreiben X und Y_1, \dots, Y_n genau die Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbf{P})$ aus dem Beispiel am Anfang des Kapitels. \diamond

Bemerkung 14.31. Das Verfahren lässt sich natürlich für n -stufige Experimente erweitern. Sei $(\Omega_i, \mathcal{A}_i)$ der Messraum des i -ten Experiments, $i = 0, \dots, n-1$. Sei P_0 ein W-Maß auf $(\Omega_0, \mathcal{A}_0)$. Für $i = 1, \dots, n-1$ sei die Verteilung auf $(\Omega_i, \mathcal{A}_i)$ abhängig von $(\omega_1, \dots, \omega_{i-1})$ und gegeben durch einen stochastischen Kern κ_i von $\Omega_0 \times \dots \times \Omega_{i-1}$ nach Ω_i . Das gesamte n -stufige Experiment wird dann durch die Koordinatenabbildungen auf dem Wahrscheinlichkeitsraum $\left(\bigtimes_{i=0}^{n-1} \Omega_i, \bigotimes_{i=0}^{n-1} \mathcal{A}_i, P_0 \otimes \bigotimes_{i=1}^{n-1} \kappa_i \right)$ beschrieben. \diamond

Übung 14.2.1. Man zeige die Faltungsformeln:

- (i) Normalverteilung: $\mathcal{N}_{\mu_1, \sigma_1^2} * \mathcal{N}_{\mu_2, \sigma_2^2} = \mathcal{N}_{\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2}$ für alle $\mu_1, \mu_2 \in \mathbb{R}$ und $\sigma_1^2, \sigma_2^2 > 0$.
- (ii) Gamma-Verteilung: $\Gamma_{\theta, r} * \Gamma_{\theta, s} = \Gamma_{\theta, r+s}$ für alle $\theta, r, s > 0$.
- (iii) Cauchy-Verteilung: $\text{Cau}_r * \text{Cau}_s = \text{Cau}_{r+s}$ für alle $r, s > 0$. 

Übung 14.2.2 (Hilbert-Schmidt Operator). Seien $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, 2$, σ -endliche Maßräume und $a : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ messbar mit

$$\int \mu_1(dt_1) \int \mu_2(dt_2) a(t_1, t_2)^2 < \infty.$$

Für $f \in \mathcal{L}^2(\mu_1)$ definiere

$$(Af)(t_2) = \int a(t_1, t_2) f(t_1) \mu_1(dt_1).$$

Zeige: A ist ein stetiger linearer Operator von $\mathcal{L}^2(\mu_1)$ nach $\mathcal{L}^2(\mu_2)$. ♣

Übung 14.2.3 (Partielle Integration). Seien F_μ und F_ν die Verteilungsfunktionen der lokal endlichen Maße μ und ν auf \mathbb{R} . Für $x \in \mathbb{R}$ definieren wir den linksseitigen Grenzwert $F(x-) = \sup_{y < x} F(y)$ und die Sprunghöhe $\Delta F(x) = F(x) - F(x-)$. Man zeige: Für $a < b$ gilt

$$\begin{aligned} \int_{(a,b]} F_\mu d\nu &= F_\mu(b)F_\nu(b) - F_\mu(a)F_\nu(a) - \int_{(a,b]} F_\nu(x-) \mu(dx) \\ &= F_\mu(b)F_\nu(b) - F_\mu(a)F_\nu(a) - \int_{(a,b]} F_\nu d\mu + \sum_{a < x \leq b} \Delta F_\mu(x) \Delta F_\nu(x). \end{aligned} \quad \clubsuit$$

14.3 Satz von Ionescu-Tulcea und Projektive Familien

Im vorigen Abschnitt hatten wir gesehen, wie wir n -stufige Experimente auf einem W-Raum realisieren können. In diesem Abschnitt wollen wir zunächst zeigen, wie wir abzählbar viele sukzessive Experimente auf einem W-Raum realisieren können (Satz von Ionescu-Tulcea). Danach werden wir auch Maße auf überabzählbaren Produkträumen konstruieren (Erweiterungssatz von Kolmogorov).

Seien $(\Omega_i, \mathcal{A}_i)$, $i \in \mathbb{N}_0$, Messräume und P_0 ein W-Maß auf $(\Omega_0, \mathcal{A}_0)$. Setze $\Omega^i := \bigtimes_{k=0}^i \Omega_k$ und $\mathcal{A}^i = \bigotimes_{k=0}^i \mathcal{A}_k$ sowie

$$\Omega := \bigtimes_{k=0}^{\infty} \Omega_k \quad \text{und} \quad \mathcal{A} = \bigotimes_{k=0}^{\infty} \mathcal{A}_k.$$

Es sei κ_i ein stochastischer Kern von $(\Omega^{i-1}, \mathcal{A}^{i-1})$ nach $(\Omega_i, \mathcal{A}_i)$ für jedes $i \in \mathbb{N}$. Wir hatten in Korollar 14.24 rekursiv W-Maße $P_i = P_0 \otimes \bigotimes_{k=1}^i \kappa_k$ auf $(\Omega^i, \mathcal{A}^i)$ definiert. Per Konstruktion gilt

$$P_i(A \times \Omega_{k+1} \times \cdots \times \Omega_i) = P_j(A \times \Omega_{k+1} \times \cdots \times \Omega_j) \quad (14.10)$$

für $i, j \geq k$ und $A \in \mathcal{A}^k$. Wir wollen nun ein Wahrscheinlichkeitsmaß P auf (Ω, \mathcal{A}) definieren, sodass für $k \in \mathbb{N}_0$ und $A \in \mathcal{A}^k$ gilt:

$$P\left(A \times \bigtimes_{i=k+1}^{\infty} \Omega_i\right) = P_k(A). \quad (14.11)$$

Satz 14.32 (Ionescu-Tulcea). *Es gibt ein eindeutig bestimmtes W-Maß auf (Ω, \mathcal{A}) , sodass (14.11) gilt.*

Beweis. Die Eindeutigkeit ist klar, weil die endlichdimensionalen Rechteckzylinder einen schnittstabilen Erzeuger von \mathcal{A} bilden. Es bleibt die Existenz zu zeigen.

Wir definieren eine Mengenfunktion P auf den Zylindermengen \mathcal{Z} durch (14.11). Offenbar ist P additiv, also ein Inhalt. Ist P nun aber \emptyset -stetig, so ist P nach Satz 1.36 ein Prämaß und lässt sich nach dem Satz von Carathéodory (Satz 1.41) eindeutig zu einem Maß auf \mathcal{A} fortsetzen.

Sei also $A_0 \supset A_1 \supset A_2 \supset \dots$ eine Folge in \mathcal{Z} mit $\alpha := \inf_{n \in \mathbb{N}_0} P(A_n) > 0$. Es reicht zu zeigen, dass $\bigcap_{n=0}^{\infty} A_n \neq \emptyset$. Ohne Einschränkung können wir annehmen, dass $A_n = A'_n \times \times_{k=n+1}^{\infty} \Omega_k$ für gewisses $A'_n \in \mathcal{A}^n$. Für $n \geq m$ setze

$$h_{m,n}(\omega_0, \dots, \omega_m) := \left(\delta_{(\omega_0, \dots, \omega_m)} \otimes \bigotimes_{k=m+1}^n \kappa_k \right) (A'_n)$$

und $h_m := \inf_{n \geq m} h_{m,n}$. Wir zeigen induktiv, dass es $\varrho_i \in \Omega_i$, $i \in \mathbb{N}_0$, gibt mit

$$h_m(\varrho_0, \dots, \varrho_m) \geq \alpha. \quad (14.12)$$

Wegen $A'_{n+1} \subset A'_n \times \Omega_{n+1}$ gilt

$$\begin{aligned} h_{m,n+1}(\omega_0, \dots, \omega_m) &= \left(\delta_{(\omega_0, \dots, \omega_m)} \otimes \bigotimes_{k=m+1}^{n+1} \kappa_k \right) (A'_{n+1}) \\ &\leq \left(\delta_{(\omega_0, \dots, \omega_m)} \otimes \bigotimes_{k=m+1}^{n+1} \kappa_k \right) (A'_n \times \Omega_{n+1}) \\ &= \left(\delta_{(\omega_0, \dots, \omega_m)} \otimes \bigotimes_{k=m+1}^n \kappa_k \right) (A'_n) = h_{m,n}(\omega_0, \dots, \omega_m). \end{aligned}$$

Also gilt $h_{m,n} \downarrow h_m$ für $n \rightarrow \infty$ und nach dem Satz von der majorisierten Konvergenz

$$\int h_m dP_m = \inf_{n \geq m} \int h_{m,n} dP_m = \inf_{n \in \mathbb{N}_0} P_n(A'_n) = \alpha.$$

Daher gilt (14.12) für $m = 0$. Gelte nun (14.12) für ein $m \in \mathbb{N}_0$. Dann ist

$$\begin{aligned} &\int h_{m+1}(\varrho_0, \dots, \varrho_m, \omega_{m+1}) \kappa_{m+1}((\varrho_0, \dots, \varrho_m), d\omega_{m+1}) \\ &= \inf_{n \geq m+1} \int h_{m+1,n}(\varrho_0, \dots, \varrho_m, \omega_{m+1}) \kappa_{m+1}((\varrho_0, \dots, \varrho_m), d\omega_{m+1}) \\ &= \inf_{n \geq m+1} h_{m,n}(\varrho_0, \dots, \varrho_m) = h_m(\varrho_0, \dots, \varrho_m) \geq \alpha. \end{aligned}$$

Es folgt (14.12) für $m + 1$.

Sei $\varrho := (\varrho_0, \varrho_1, \dots) \in \Omega$. Nach Konstruktion ist $\alpha \leq h_{m,m}(\varrho_0, \dots, \varrho_m) = \mathbb{1}_{A'_m}(\varrho_0, \dots, \varrho_m)$, also $\varrho \in A_m$ für jedes $m \in \mathbb{N}_0$ und damit $\bigcap_{i=0}^{\infty} A_i \neq \emptyset$. \square

Korollar 14.33 (Produktmaß). Für jedes $n \in \mathbb{N}_0$ sei P_n ein W-Maß auf $(\Omega_n, \mathcal{A}_n)$. Dann existiert ein eindeutig bestimmtes W-Maß P auf (Ω, \mathcal{A}) mit

$$P \left(A_0 \times \cdots \times A_n \times \bigotimes_{i=n+1}^{\infty} \Omega_i \right) = \prod_{k=0}^n P_k(A_k)$$

für $A_i \in \mathcal{A}_i$, $i = 0, \dots, n$ und $n \in \mathbb{N}_0$.

Wir nennen $\bigotimes_{i=0}^{\infty} P_i := P$ das Produkt der Maße P_0, P_1, \dots . Unter P sind die Koordinatenabbildungen $(X_i)_{i \in \mathbb{N}_0}$ unabhängig.

Beweis. Wende den Satz von Ionescu-Tulcea mit $\kappa_i((\omega_0, \dots, \omega_{i-1}), \cdot) = P_i$ an. \square

Wir wollen nun eine dem Satz von Ionescu-Tulcea vergleichbare Aussage treffen, dabei jedoch auf die Annahme verzichten, dass die Maße P_k auf \mathcal{A}^k *a priori* durch Kerne definiert werden. Bevor wir den Satz formulieren, wollen wir die Konsistenzbedingung (14.10) verallgemeinern. (Erinnerung: für $L \subset J \subset I$ bezeichnet $X_L^J : \Omega_J \rightarrow \Omega_L$ die Projektion.)

Definition 14.34. Eine Familie $(P_J, J \subset I$ endlich) von W-Maßen auf $(\Omega_J, \mathcal{A}_J)$ heißt **projektive Familie**, falls für alle endlichen $L \subset J \subset I$ gilt, dass

$$P_L = P_J \circ (X_L^J)^{-1}.$$

Ist P ein W-Maß auf (Ω, \mathcal{A}) , wobei

$$\Omega = \bigotimes_{i \in I} \Omega_i \quad \text{und} \quad \mathcal{A} = \bigotimes_{i \in I} \mathcal{A}_i,$$

so ist wegen $X_L = X_L^J \circ X_J$, die Familie $(P_J := P \circ X_J^{-1}, J \subset I$ endlich) projektiv. Projektivität ist also eine notwendige Bedingung für die Existenz des Maßes P auf dem Produktraum. Sind alle beteiligten Messräume Borel'sche Räume (siehe Definition 8.35) – also beispielsweise $\mathbb{R}^d, \mathbb{Z}^d, C([0, 1])$ oder allgemeiner polnische Räume –, so ist diese Bedingung auch ausreichend. Wir formulieren diese Aussage zunächst für abzählbare Indexmengen.

Satz 14.35. Sei I höchstens abzählbar, und seien $(\Omega_i, \mathcal{A}_i)$ Borel'sche Räume, $i \in I$. Sei $(P_J, J \subset I$ endlich) eine projektive Familie von W-Maßen. Dann gibt es ein eindeutig bestimmtes W-Maß P auf (Ω, \mathcal{A}) mit $P_J = P \circ X_J^{-1}$ für jedes endliche $J \subset I$.

Beweis. Ohne Einschränkung sei $I = \mathbb{N}_0$. Sei $P_n := P_{\{0, \dots, n\}}$, $\Omega^n := \Omega_{\{0, \dots, n\}}$ und $\mathcal{A}^n := \mathcal{A}_{\{0, \dots, n\}}$. Man prüft leicht nach, dass endliche Produkte von Borel'schen Räumen wieder Borel'sche Räume sind, also ist $(\Omega^n, \mathcal{A}^n)$ Borel'sch für jedes $n \in \mathbb{N}_0$.

Sei $\mathcal{F} := \{A \times \Omega_{n+1} : A \in \mathcal{A}^n\}$, $Y : \Omega^{n+1} \rightarrow \Omega_{n+1}$, $(\omega_0, \dots, \omega_{n+1}) \mapsto \omega_{n+1}$ und $Z : \Omega^{n+1} \rightarrow \Omega^n$, $(\omega_0, \dots, \omega_{n+1}) \mapsto (\omega_0, \dots, \omega_n)$. Nach Satz 8.37 (mit $\Omega = \Omega^{n+1}$, $\mathcal{A} = \mathcal{A}^{n+1}$ und $E = \Omega_{n+1}$) existiert ein stochastischer Kern κ'_{n+1} von $(\Omega^{n+1}, \mathcal{F})$ nach $(\Omega_{n+1}, \mathcal{A}_{n+1})$, sodass κ'_{n+1} eine reguläre Version bedingten Verteilung von Y gegeben \mathcal{F} (unter P_{n+1}) ist. Das heißt, für $A \in \mathcal{A}^n$ und $B \in \mathcal{A}_{n+1}$ gilt (siehe (8.10))

$$\begin{aligned} P_{n+1}(A \times B) &= \int \mathbb{1}_B(Y) \mathbb{1}_{A \times \Omega_{n+1}} dP_{n+1} \\ &= \int \kappa'_{n+1}(\cdot, B) \mathbb{1}_{A \times \Omega_{n+1}} dP_{n+1}. \end{aligned}$$

Da $\kappa'_{n+1}(\cdot, B)$ messbar bezüglich \mathcal{F} ist, gibt es einen stochastischen Kern κ_{n+1} von $(\Omega^n, \mathcal{A}^n)$ nach $(\Omega_{n+1}, \mathcal{A}_{n+1})$, sodass

$$\kappa_{n+1}((\omega_0, \dots, \omega_n), \cdot) = \kappa'_{n+1}((\omega_0, \dots, \omega_{n+1}), \cdot) \quad \text{für alle } \omega_0, \dots, \omega_{n+1}.$$

Also gilt

$$\kappa'_{n+1}(\cdot, B) = \kappa_{n+1}(Z(\cdot), B) \quad \text{und} \quad \mathbb{1}_{A \times \Omega_{n+1}} = \mathbb{1}_A(Z).$$

Wir erhalten so

$$\begin{aligned} P_{n+1}(A \times B) &= \int \kappa_{n+1}(Z, B) \mathbb{1}_A(Z) dP_{n+1} \\ &= \int \kappa_{n+1}(\cdot, B) \mathbb{1}_A d(P_{n+1} \circ Z^{-1}) \\ &= \int_A \kappa_{n+1}(\cdot, B) dP_n, \end{aligned}$$

wobei wir im letzten Schritt die Projektivität von $(P_n)_{n \in \mathbb{N}_0}$ ausgenutzt haben. Nach Korollar 14.23 gilt also $P_{n+1} = P_n \otimes \kappa_{n+1}$. Rekursiv erhalten wir $P_n = P_0 \otimes \bigotimes_{k=1}^n \kappa_k$ für jedes $n \in \mathbb{N}$. Also liefert Satz 14.32 die Behauptung. \square

Als letzten Schritt in unserer Konstruktion wollen wir in Satz 14.35 die abzählbare Indexmenge durch eine beliebige Indexmenge ersetzen.

Satz 14.36 (Kolmogorov'scher Erweiterungssatz). Sei I eine beliebige Indexmenge, und seien $(\Omega_i, \mathcal{A}_i)$ Borel'sche Räume, $i \in I$. Sei $(P_J, J \subset I \text{ endlich})$ eine projektive Familie von W-Maßen. Dann gibt es ein eindeutig bestimmtes W-Maß P auf (Ω, \mathcal{A}) mit $P_J = P \circ X_J^{-1}$ für jedes endliche $J \subset I$. Wir nennen P den **projektiven Limes** und bezeichnen ihn in Formeln mit $P =: \varprojlim_{J \uparrow I} P_J$.

Beweis. Für abzählbares $J \subset I$ gibt es nach Satz 14.35 ein eindeutig bestimmtes W-Maß P_J auf $(\Omega_J, \mathcal{A}_J)$ mit $P_J \circ (X_K^J)^{-1} = P_K$ für endliches $K \subset J$. Durch $\tilde{P}_J(X_J^{-1}(A_J)) := P_J(A_J)$, $A_J \in \mathcal{A}_J$, wird hieraus ein Maß auf $(\Omega, \sigma(X_J))$.

Seien $J, J' \subset I$ höchstens abzählbar und $A \in \sigma(X_J) \cap \sigma(X_{J'}) \cap \mathcal{Z}$ ein $\sigma(X_J) \cap \sigma(X_{J'})$ -messbarer Zylinder mit endlicher Basis. Dann existiert ein endliches $K \subset J \cap J'$ und $A_K \in \mathcal{A}_K$ mit $A = X_K^{-1}(A_K)$. Also ist $\tilde{P}_J(A) = P_K(A_K) = \tilde{P}_{J'}(A)$. Nach Satz 14.12 ist dann aber auch $\tilde{P}_J(A) = P_K(A_K) = \tilde{P}_{J'}(A)$ für alle $A \in \sigma(X_J) \cap \sigma(X_{J'})$. Nun gibt es (nach Übung 14.1.1) für jedes $A \in \mathcal{A}$ ein abzählbares $J \subset I$ mit $A \in \sigma(X_J)$, also können wir auf eindeutige Weise (und unabhängig von der Wahl von J) eine Mengenfunktion P auf \mathcal{A} definieren durch $P(A) = \tilde{P}_J(A)$. Es bleibt zu zeigen, dass P ein W-Maß ist. Offenbar ist $P(\Omega) = 1$. Sind $A_1, A_2, \dots \in \mathcal{A}$ paarweise disjunkt und $A := \bigcup_{n=1}^{\infty} A_n$, so gibt es abzählbare $J_n \subset I$ mit $A_n \in \sigma(X_{J_n})$ für $n \in \mathbb{N}$. Setze $J = \bigcup_{n \in \mathbb{N}} J_n$. Dann ist jedes A_n in $\sigma(X_J)$ und damit auch $A \in \sigma(X_J)$, also

$$P(A) = \tilde{P}_J(A) = \sum_{n=1}^{\infty} \tilde{P}_J(A_n) = \sum_{n=1}^{\infty} P(A_n).$$

Damit ist P als W-Maß erkannt. \square

Beispiel 14.37. Sei $((\Omega_i, \tau_i), i \in I)$ eine beliebige Familie von polnischen Räumen (Erinnerung: nach Satz 8.36 sind polnische Räume auch Borel'sche Räume), $\mathcal{A}_i = \sigma(\tau_i)$ und P_i ein beliebiges W-Maß auf $(\Omega_i, \mathcal{A}_i)$. Für endliches $J \subset I$ sei $P_J = \bigotimes_{j \in J} P_j$ das Produktmaß der P_j , $j \in J$. Offenbar ist die Familie $(P_J, J \subset I$ endlich) projektiv. Wir nennen

$$P = \bigotimes_{i \in I} P_i := \varprojlim_{J \uparrow I} P_J$$

das **Produktmaß** auf (Ω, \mathcal{A}) . Unter P sind alle Projektionen X_j unabhängig. \diamond

Beispiel 14.38 (Pólya'sches Urnenmodell). (Vergleiche Beispiel 12.29.) Zunächst befinden sich k rote und $n - k$ blaue Kugeln in einer Urne. Es wird in jedem Zeitschritt eine Kugel gezogen und zusammen mit einer *weiteren* Kugel der selben Farbe zurückgelegt. Zur Zeit $i \in \mathbb{N}_0$ befinden sich also $n + i$ Kugeln in der Urne, wobei die Anzahl X_i der roten Kugeln zufällig ist.

Formal definieren wir das Modell so: Sei $n \in \mathbb{N}$ und $k \in \{0, \dots, n\}$. Sei $I = \mathbb{N}_0$, $\Omega_i = \{0, \dots, n+i\}$, $i \in \mathbb{N}$. Setze $P_0[\{k\}] = 1$ und definiere die stochastischen Kerne κ_i von Ω_i nach Ω_{i+1} durch

$$\kappa_i(x_i, \{x_{i+1}\}) = \begin{cases} \frac{x_i}{n+i}, & \text{falls } x_{i+1} = x_i + 1, \\ 1 - \frac{x_i}{n+i}, & \text{falls } x_{i+1} = x_i, \\ 0, & \text{sonst.} \end{cases}$$

Setze nun $P_{i+1} = P_i \otimes \kappa_i$. Unter dem Maß $\mathbf{P} = \varprojlim_{i \rightarrow \infty} P_i$ beschreiben die Projektionen $(X_i, i \in \mathbb{N}_0)$ gerade das Pólya'sche Urnenmodell. \diamond

14.4 Markov'sche Halbgruppen

Definition 14.39. Sei E ein polnischer Raum. Sei $I \subset \mathbb{R}$ eine nichtleere Indexmenge und $(\kappa_{s,t} : s, t \in I, s < t)$ eine Familie von stochastischen Kernen von E nach E . Wir sagen, dass die Familie **konsistent** ist, falls für je drei Zahlen $r, s, t \in I$ mit $r < s < t$ gilt, dass $\kappa_{r,s} \cdot \kappa_{s,t} = \kappa_{r,t}$.

Definition 14.40. Sei E ein polnischer Raum. Sei $I \subset [0, \infty)$ eine (additive) Halbgruppe (zum Beispiel $I = \mathbb{N}_0$ oder $I = [0, \infty)$). Eine Familie $(\kappa_t : t \in I)$ von stochastischen Kernen heißt eine **Halbgruppe von stochastischen Kernen**, oder **Markov'sche Halbgruppe**, falls

$$\kappa_0(\omega, \cdot) = \delta_\omega \quad \text{für alle } \omega \in \Omega \quad (14.13)$$

gilt und die **Chapman-Kolmogorov'sche Gleichung** erfüllt ist:

$$\kappa_s \cdot \kappa_t = \kappa_{s+t} \quad \text{für alle } s, t \in I. \quad (14.14)$$

Tatsächlich ist $(\{\kappa_t : t \in I\}, \cdot)$ eine Halbgruppe im algebraischen Sinne, und die Abbildung $t \rightarrow \kappa_t$ ist ein Halbgruppenhomomorphismus. Insbesondere kommutieren die Kerne in dem Sinne, dass $\kappa_s \cdot \kappa_t = \kappa_t \cdot \kappa_s$ für alle $s, t \in I$.

Lemma 14.41. Ist $(\kappa_t : t \in I)$ eine Markov'sche Halbgruppe, so ist die für $t > s$ durch $\tilde{\kappa}_{s,t} := \kappa_{t-s}$ definierte Familie von Kernen konsistent.

Beweis. Das ist trivial. □

Satz 14.42 (Kern durch konsistente Familie von Kernen). Sei $I \subset [0, \infty)$ mit $0 \in I$ und $(\kappa_{s,t} : s, t \in I, s < t)$ eine konsistente Familie von stochastischen Kernen auf dem polnischen Raum E . Dann existiert ein Kern κ von $(E, \mathcal{B}(E))$ nach $(E^I, \mathcal{B}(E)^{\otimes I})$, sodass für jedes $x \in E$ und je endlich viele Zahlen $0 = j_0 < j_1 < j_2 < \dots < j_n$ aus I sowie $J := \{j_0, \dots, j_n\}$ gilt

$$\kappa(x, \cdot) \circ X_J^{-1} = \delta_x \otimes \bigotimes_{k=0}^{n-1} \kappa_{j_k, j_{k+1}}. \quad (14.15)$$

Beweis. Wir zeigen zunächst, dass für festes $x \in E$ durch (14.15) ein W-Maß $\kappa(x, \cdot)$ definiert wird. Hierfür reicht es nach dem Erweiterungssatz zu zeigen, dass die durch $P_J := \delta_x \otimes \bigotimes_{k=0}^{n-1} \kappa_{j_k, j_{k+1}}$ definierte Familie $(P_J : J \subset I \text{ endlich}, 0 \in J)$ von W-Maßen konsistent ist. (Denn für endliches $J \subset I$ mit $0 \notin J$ können wir P_J als die Projektion von $P_{J \cup \{0\}}$ auf E^J definieren und erhalten so eine projektive Familie $(P_J : J \subset I \text{ endlich})$.) Sei also $0 \in L \subset J \subset I$ mit $J \subset I$ endlich. Dann

ist zu zeigen, dass $P_J \circ (X_L^J)^{-1} = P_L$ ist. Wir können annehmen, dass $L = J \setminus \{j_l\}$ ist für ein $l = 1, \dots, n$. Den allgemeinen Fall erhält man hieraus induktiv.

Sei zunächst $l = n$.

Seien $A_{j_0}, \dots, A_{j_{n-1}} \in \mathcal{B}(E)$ und $A := \times_{j \in L} A_j$. Dann ist

$$\begin{aligned} P_J \circ (X_L^J)^{-1}(A) &= P_J(A \times E) = P_L \otimes \kappa_{j_{n-1}, j_n}(A \times E) \\ &= \int_A P_L(d(\omega_0, \dots, \omega_{n-1})) \kappa_{j_{n-1}, j_n}(\omega_{n-1}, E) = P_L(A). \end{aligned}$$

Sei nun $l \in \{1, \dots, n-1\}$ und $A_j \in \mathcal{B}(E)$ für jedes $j \in L$ sowie $A_{j_l} := E$. Setze $A := \times_{j \in L} A_j$ und, zur Abkürzung, $A' = \times_{k=0}^{l-1} A_{j_k}$ und $P' = \delta_x \otimes \bigotimes_{k=0}^{l-2} \kappa_{j_k, j_{k+1}}$ sowie für $i = 0, \dots, n-1$

$$f_i(\omega_i) = \left(\bigotimes_{k=i}^{n-1} \kappa_{j_k, j_{k+1}} \right) (\omega_i, A_{j_{i+1}} \times \dots \times A_{j_n}).$$

Nach Voraussetzung (und dem Satz von Fubini) ist

$$\begin{aligned} f_{l-1}(\omega_{l-1}) &= \int_E \kappa_{j_{l-1}, j_l}(\omega_{l-1}, d\omega_l) \int_{A_{j_{l+1}}} \kappa_{j_l, j_{l+1}}(\omega_l, d\omega_{l+1}) f_{l+1}(\omega_{l+1}) \\ &= \int_{A_{j_{l+1}}} \kappa_{j_{l-1}, j_{l+1}}(\omega_{l-1}, d\omega_{l+1}) f_{l+1}(\omega_{l+1}). \end{aligned}$$

Es folgt

$$\begin{aligned} P_J \circ (X_L^J)^{-1}(A) &= \int_{A'} P'(d(\omega_0, \dots, \omega_{l-1})) f_{l-1}(\omega_{l-1}) \\ &= \int_{A'} P'(d(\omega_0, \dots, \omega_{l-1})) \int_{A_{j_{l+1}}} (\kappa_{j_{l-1}, j_{l+1}})(\omega_{l-1}, d\omega_{l+1}) f_{l+1}(\omega_{l+1}) \\ &= P_L(A). \end{aligned}$$

Wir müssen nun noch zeigen, dass κ ein stochastischer Kern ist, also dass $x \mapsto \kappa(x, A)$ messbar ist bezüglich $\mathcal{B}(E) - \mathcal{B}(E)^{\otimes I}$. Nach Bemerkung 8.26 reicht es aus, dies für Rechteckmengen mit endlicher Basis $A \in \mathcal{Z}^R$ zu prüfen, denn \mathcal{Z}^R ist ein schnittstabilier Erzeuger von $\mathcal{B}(E)^{\otimes I}$. Seien also $0 = t_0 < t_1 < \dots < t_n$ und $B_0, \dots, B_n \in \mathcal{B}(E)$ sowie $A = \bigcap_{i=0}^n X_{t_i}^{-1}(B_i)$. Nach Korollar 14.24 ist aber die folgende Abbildung messbar

$$x \mapsto \mathbf{P}_x[A] = \left(\delta_x \otimes \bigotimes_{i=0}^{n-1} \kappa_{t_i, t_{i+1}} \right) \left(\bigotimes_{i=0}^n B_i \right). \quad \square$$

Korollar 14.43 (Maß durch konsistente Familie von Kernen). Unter den Bedingungen von Satz 14.42 existiert zu jedem W-Maß μ auf E genau ein W-Maß \mathbf{P}_μ auf

$(E^I, \mathcal{B}(E)^{\otimes I})$ mit der Eigenschaft: Für je endlich viele Zahlen $0 = j_0 < j_1 < j_2 < \dots < j_n$ aus I und $J := \{j_0, \dots, j_n\}$ ist $\mathbf{P}_\mu \circ X_J^{-1} = \mu \otimes \bigotimes_{k=0}^{n-1} \kappa_{j_k, j_{k+1}}$.

Beweis. Man setze $\mathbf{P}_\mu = \int \mu(dx) \kappa(x, \cdot)$. □

Als einfache Schlussfolgerung aus Lemma 14.41 und Satz 14.42 ergibt sich die folgende Aussage, die wir separat formulieren, weil wir sie später noch an zentraler Stelle benötigen werden.

Korollar 14.44 (Maße durch Markov'sche Halbgruppen). Sei $(\kappa_t : t \in I)$ eine Markov'sche Halbgruppe auf dem polnischen Raum E . Dann existiert genau ein stochastischer Kern κ von $(E, \mathcal{B}(E))$ nach $(E^I, \mathcal{B}(E)^{\otimes I})$ mit der Eigenschaft: Für jedes $x \in E$ und je endlich viele Zahlen $0 = t_0 < t_1 < t_2 < \dots < t_n$ aus I sowie $J := \{t_0, \dots, t_n\}$ ist

$$\kappa(x, \cdot) \circ X_J^{-1} = \delta_x \otimes \bigotimes_{k=0}^{n-1} \kappa_{t_{k+1}-t_k}. \quad (14.16)$$

Für jedes W-Maß μ auf E existiert genau ein W-Maß \mathbf{P}_μ auf $(E^I, \mathcal{B}(E)^{\otimes I})$ mit der Eigenschaft: Für je endlich viele Zahlen $0 = t_0 < t_1 < t_2 < \dots < t_n$ aus I und $J := \{t_0, \dots, t_n\}$ ist $\mathbf{P}_\mu \circ X_J^{-1} = \mu \otimes \bigotimes_{k=0}^{n-1} \kappa_{t_{k+1}-t_k}$. Wir schreiben $\mathbf{P}_x = \mathbf{P}_{\delta_x} = \kappa(x, \cdot)$ für $x \in E$.

Beispiel 14.45 (Unabhängige, normalverteilte Zuwächse). Seien $I = [0, \infty)$ und $\Omega_i = \mathbb{R}$, $i \in [0, \infty)$, mit der Borel'schen σ -Algebra $\mathcal{B} = \mathcal{B}(\mathbb{R})$, $\Omega = \mathbb{R}^{[0, \infty)}$, $\mathcal{A} = \mathcal{B}^{\otimes [0, \infty)}$ und X_t die kanonische Projektion für $t \in [0, \infty)$. Im Sinne von Definition 14.6 ist $X = (X_t)_{t \geq 0}$ also der kanonische Prozess auf (Ω, \mathcal{A}) .

Wir wollen ein W-Maß \mathbf{P} auf (Ω, \mathcal{A}) konstruieren, sodass der stochastische Prozess X unabhängige, stationäre, normalverteilte Zuwächse hat (siehe Definition 9.7). Das heißt, es sollen gelten:

$$\text{für } 0 =: t_0 < t_1 < \dots < t_n \text{ ist } (X_{t_i} - X_{t_{i-1}})_{i=1, \dots, n} \text{ unabhängig,} \quad (14.17)$$

$$\text{für } t > s \text{ ist } \mathbf{P}_{X_t - X_s} = \mathcal{N}_{0, t-s}. \quad (14.18)$$

Dazu definieren wir die stochastischen Kerne $\kappa_t(x, dy) := \delta_x * \mathcal{N}_{0,t}(dy)$ für $t \in [0, \infty)$, wobei $\mathcal{N}_{0,0} = \delta_0$. Nach Lemma 14.27 gilt die Chapman-Kolmogorov'sche Gleichung, denn (vergleiche Übung 14.2.1(i))

$$\kappa_s \cdot \kappa_t(x, dy) = \delta_x * (\mathcal{N}_{0,s} * \mathcal{N}_{0,t})(dy) = \delta_x * \mathcal{N}_{0,s+t}(dy) = \kappa_{s+t}(x, dy).$$

Sei $P_0 = \delta_0$ und \mathbf{P} das nach Korollar 14.44 zu P_0 und $(\kappa_t : t \geq 0)$ gehörige W-Maß auf Ω . Nach Satz 14.28 gelten (14.17) und (14.18).

Wir haben mit $(X_t)_{t \geq 0}$ fast schon die so genannte **Brown'sche Bewegung** konstruiert. Zusätzlich zu den gezeigten Eigenschaften hat die Brown'sche Bewegung die Eigenschaft, dass die *Pfade*, also die Abbildungen $t \mapsto X_t$, stetig sind. In der Tat ist mit der bisherigen Konstruktion sogar die Messbarkeit der Pfade noch ungeklärt. Hier muss noch gearbeitet werden, und wir kommen darauf in Kapitel 21 zurück. \diamond

Die Konstruktion im vorangehenden Beispiel hängt nicht von den Details der Normalverteilung ab, sondern nur von der Eigenschaft $\mathcal{N}_{0,s+t} = \mathcal{N}_{0,s} * \mathcal{N}_{0,t}$. Wir können also in (14.18) die Normalverteilung durch eine beliebige parametrisierte Familie von Verteilungen $(\nu_t, t \geq 0)$ mit der Eigenschaft $\nu_{t+s} = \nu_t * \nu_s$ ersetzen. Dies gilt speziell für die Familie der Gammaverteilungen $\nu_t = \Gamma_{\theta,t}$ (für festes $\theta > 0$), die Poissonverteilung $\nu_t = \text{Poi}_t$, die negative Binomialverteilung $\nu_t = b_{t,p}^-$ (für festes $p \in (0, 1]$), die Cauchy-Verteilung $\nu_t = \text{Cau}_t$ und andere (vergleiche Satz 15.12 und Korollar 15.13). Wir halten dieses Ergebnis in einem Satz fest.

Definition 14.46 (Faltungshalbgruppe). Sei $I \subset [0, \infty)$ eine Halbgruppe. Eine Familie $\nu = (\nu_t : t \in I)$ von Wahrscheinlichkeitsverteilungen auf \mathbb{R}^d heißt **Faltungshalbgruppe**, falls $\nu_{s+t} = \nu_s * \nu_t$ gilt für alle $s, t \in I$.

Ist $I = [0, \infty)$ und gilt zudem $\nu_t \xrightarrow{t \rightarrow 0} \delta_0$, so heißt die Faltungshalbgruppe **stetig** (im Sinne der schwachen Konvergenz).

Ist $d = 1$ und $\nu_t((-\infty, 0)) = 0$ für jedes $t \in I$, so heißt ν eine **nichtnegative Faltungshalbgruppe**.

Für den folgenden Satz vergleiche Definition 9.7.

Satz 14.47. Zu jeder Faltungshalbgruppe $(\nu_t : t \in I)$ und jedem $x \in \mathbb{R}^d$ existiert ein W-Maß \mathbf{P}_x auf dem Produktraum $(\Omega, \mathcal{A}) = ((\mathbb{R}^d)^I, \mathcal{B}(\mathbb{R}^d)^{\otimes I})$, sodass der kanonische Prozess $(X_t)_{t \in I}$ ein stochastischer Prozess mit $\mathbf{P}_x[X_0 = x] = 1$ und stationären unabhängigen Zuwächsen ist mit $\mathbf{P}_x \circ (X_t - X_s)^{-1} = \nu_{t-s}$ für $t > s$. Umgekehrt definiert jeder stochastische Prozess $(X_t)_{t \in I}$ (auf einem beliebigen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$) mit stationären unabhängigen Zuwächsen eine Faltungshalbgruppe durch $\nu_t = \mathbf{P} \circ (X_t - X_0)^{-1}$ für jedes $t \in I$.

Übung 14.4.1. Sei $(\nu_t : t \geq 0)$ eine stetige Faltungshalbgruppe. Man zeige: Für jedes $t > 0$ gilt $\nu_t = \lim_{s \rightarrow t} \nu_s$. \clubsuit

Übung 14.4.2. Sei $(\nu_t : t \geq 0)$ eine Faltungshalbgruppe. Man zeige: $\nu_{t/n} \xrightarrow{n \rightarrow \infty} \delta_0$. \clubsuit

Übung 14.4.3. Man zeige: Eine nichtnegative Faltungshalbgruppe ist stetig. \clubsuit

Übung 14.4.4. Man zeige: Eine stetige, reelle Faltungshalbgruppe $(\nu_t : t \geq 0)$ mit $\nu_t((-\infty, 0)) = 0$ für ein $t > 0$ ist nichtnegativ. \clubsuit

15 Charakteristische Funktion und Zentraler Grenzwertsatz

Hauptziel dieses Kapitels ist der Zentrale Grenzwertsatz für Summen unabhängiger Zufallsvariablen (Satz 15.37) und für unabhängige Schemata (Satz von Lindeberg-Feller, Satz 15.43), wobei wir für den letzteren nur die eine Richtung beweisen (Satz von Lindeberg).

Das Hilfsmittel der Wahl für die Behandlung von Zentralen Grenzwertsätzen sind charakteristische Funktionen, also Fouriertransformierte von W-Maßen. Wir beginnen mit einer sehr allgemeinen Betrachtung über Klassen von Testfunktionen, die schwache Konvergenz charakterisieren können, und betrachten dann Fouriertransformierte im Detail. Der nachfolgende Abschnitt beweist mit Hilfe von charakteristischen Funktionen den Zentralen Grenzwertsatz für reelle Zufallsvariablen. Im fünften Abschnitt zeigen wir den mehrdimensionalen Zentralen Grenzwertsatz.

15.1 Trennende Funktionenklassen

Sei (E, d) ein metrischer Raum mit Borel'scher σ -Algebra $\mathcal{E} = \mathcal{B}(E)$.

Mit $\mathbb{C} = \{u + iv : u, v \in \mathbb{R}\}$ bezeichnen wir den Körper der komplexen Zahlen. Mit $\operatorname{Re}(u + iv) = u$ und $\operatorname{Im}(u + iv) = v$ bezeichnen wir den Realteil und den Imaginärteil von $z = u + iv \in \mathbb{C}$, mit $\bar{z} = u - iv$ die zu z komplex konjugierte Zahl und mit $|z| = \sqrt{u^2 + v^2}$ den Betrag von z . Von prominenter Bedeutung wird für uns die komplexe Exponentialfunktion $\exp : \mathbb{C} \rightarrow \mathbb{C}$ sein, die wir durch $\exp(z) = \exp(u)(\cos(v) + i \sin(v))$ oder durch die Potenzreihe $\exp(z) = \sum_{n=0}^{\infty} z^n / n!$ definieren können. Bekanntlich gilt $\exp(z_1 + z_2) = \exp(z_1) \cdot \exp(z_2)$. Man beachte, dass aus $\operatorname{Re}(z) = (z + \bar{z})/2$ und $\operatorname{Im}(z) = (z - \bar{z})/2i$ folgt, dass

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2} \quad \text{und} \quad \sin(x) = \frac{e^{ix} - e^{-ix}}{2i} \quad \text{für jedes } x \in \mathbb{R}.$$

Eine Abbildung $f : E \rightarrow \mathbb{C}$ ist genau dann messbar, wenn $\operatorname{Re}(f)$ und $\operatorname{Im}(f)$ messbar sind (siehe Satz 1.90 mit $\mathbb{C} \cong \mathbb{R}^2$). Insbesondere ist jede stetige Funktion $E \rightarrow \mathbb{C}$ messbar. Ist $\mu \in \mathcal{M}(E)$, so definieren wir

$$\int f \, d\mu := \int \operatorname{Re}(f) \, d\mu + i \int \operatorname{Im}(f) \, d\mu,$$

wenn beide Integrale existieren und endlich sind. Mit $C_b(E; \mathbb{C})$ bezeichnen wir den Banachraum der stetigen, beschränkten, komplexwertigen Funktionen auf E , ausgestattet mit der Supremumsnorm $\|f\|_\infty = \sup\{|f(x)| : x \in E\}$. Wir nennen $\mathcal{C} \subset C_b(E; \mathbb{C})$ trennend für $\mathcal{M}_f(E)$, falls es für je zwei Maße $\mu, \nu \in \mathcal{M}_f(E)$ mit $\mu \neq \nu$ ein $f \in \mathcal{C}$ gibt mit $\int f d\mu \neq \int f d\nu$. Satz 13.34 gilt für $\mathcal{C} \subset C_b(E; \mathbb{C})$ sinngemäß.

Definition 15.1. Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Eine Teilmenge $\mathcal{C} \subset C_b(E; \mathbb{K})$ heißt **Algebra**, falls

- (i) $1 \in \mathcal{C}$,
- (ii) für $f, g \in \mathcal{C}$ sind $f \cdot g \in \mathcal{C}$ und $f + g \in \mathcal{C}$,
- (iii) für $f \in \mathcal{C}$ und $\alpha \in \mathbb{K}$ ist $(\alpha f) \in \mathcal{C}$.

\mathcal{C} heißt **Punkte trennend**, falls es zu je zwei Punkten $x, y \in E$ mit $x \neq y$ ein $f \in \mathcal{C}$ gibt mit $f(x) \neq f(y)$.

Satz 15.2 (Stone-Weierstraß). Sei E ein kompakter Hausdorffraum. Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Sei $\mathcal{C} \subset C_b(E; \mathbb{K})$ eine Punkte trennende Algebra. Ist $\mathbb{K} = \mathbb{C}$, so sei \mathcal{C} zusätzlich abgeschlossen bezüglich komplexer Konjugation (das heißt, mit f ist stets auch die komplexe konjugierte Funktion \bar{f} in \mathcal{C}).

Dann liegt \mathcal{C} dicht in $C_b(E; \mathbb{K})$ bezüglich der Supremumsnorm.

Beweis. Wir folgen der Darstellung in Dieudonné ([39, Kapitel VII.3]). Sei zunächst der Fall $\mathbb{K} = \mathbb{R}$ betrachtet. Wir gehen in mehreren Schritten vor.

1. Schritt. Nach dem Weierstraß'schen Approximationssatz (Beispiel 5.15) gibt es eine Folge $(p_n)_{n \in \mathbb{N}}$ von Polynomen, die die Abbildung $[0, 1] \rightarrow [0, 1]$, $t \mapsto \sqrt{t}$ gleichmäßig approximiert. Ist $f \in \mathcal{C}$, so ist also $|f| = \|f\|_\infty \lim_{n \rightarrow \infty} p_n(f^2 / \|f\|_\infty^2)$ im Abschluss $\overline{\mathcal{C}}$ von \mathcal{C} in $C_b(E; \mathbb{R})$.

2. Schritt. Indem wir den 1. Schritt auf die Algebra $\overline{\mathcal{C}}$ anwenden, folgt, dass mit $f, g \in \overline{\mathcal{C}}$ auch $f \vee g = \frac{1}{2}(f + g + |f - g|)$ und $f \wedge g = \frac{1}{2}(f + g - |f - g|)$ in $\overline{\mathcal{C}}$ liegen.

3. Schritt. Für jedes $f \in C_b(E; \mathbb{R})$, jedes $x \in E$ und jedes $\varepsilon > 0$ existiert ein $g_x \in \overline{\mathcal{C}}$ mit $g_x(x) = f(x)$ und $g_x(y) \leq f(y) + \varepsilon$ für jedes $y \in E$. Da \mathcal{C} Punkte trennt, existiert für jedes $z \in E \setminus \{x\}$ ein $H_z \in \mathcal{C}$ mit $H_z(z) \neq H(x) = 0$. Für diese z definieren wir $h_z \in \mathcal{C}$ durch $h_z(y) = f(x) + \frac{f(z) - f(x)}{H_z(z)} H_z(y)$ für jedes $y \in E$. Zudem setzen wir $h_x := f$. Dann ist $h_z(x) = f(x)$ und $h_z(z) = f(z)$ für jedes $z \in E$. Da f und h_z stetig sind, existiert zu jedem $z \in E$ eine offene Umgebung $U_z \ni z$ mit $h(y) \leq f(y) + \varepsilon$ für jedes $y \in U_z$. Wir bilden eine endliche Überdeckung U_{z_1}, \dots, U_{z_n} von E mit solchen Umgebungen und setzen $g_x = \min(h_{z_1}, \dots, h_{z_n})$. Nach Schritt 2 ist $g_x \in \overline{\mathcal{C}}$.

4. Schritt. Sei $f \in C_b(E; \mathbb{R})$, $\varepsilon > 0$ und g_x wie im 3. Schritt für jedes $x \in E$. Da f und g_x stetig sind, existiert zu jedem $x \in E$ eine offene Umgebung $V_x \ni x$ mit $g_x(y) \geq f(y) - \varepsilon$ für jedes $y \in V_x$. Wir bilden eine endliche Überdeckung V_{x_1}, \dots, V_{x_n} von E und definieren $g := \max(g_{x_1}, \dots, g_{x_n})$. Dann ist $g \in \overline{\mathcal{C}}$ nach Schritt 2 und $\|g - f\|_\infty < \varepsilon$ nach Konstruktion. Da $\varepsilon > 0$ beliebig war, gilt also $\overline{\mathcal{C}} = C_b(E; \mathbb{R})$.

5. Schritt. Sei nun $\mathbb{K} = \mathbb{C}$ betrachtet. Nach Voraussetzung sind mit f auch der Realteil $\operatorname{Re}(f) = (f + \bar{f})/2$ und der Imaginärteil $\operatorname{Im}(f) = (f - \bar{f})/2i$ in \mathcal{C} . Speziell ist $\mathcal{C}_0 := \{\operatorname{Re}(f) : f \in \mathcal{C}\} \subset \mathcal{C}$ eine reelle Algebra, die nach Voraussetzung Punkte trennt und die konstanten Funktionen enthält. Also ist \mathcal{C}_0 dicht in $C_b(E; \mathbb{R})$. Wegen $\mathcal{C} = \mathcal{C}_0 + i\mathcal{C}_0$ folgt, dass \mathcal{C} dicht in $C_b(E; \mathbb{C})$ ist. \square

Korollar 15.3. Sei E ein kompakter, metrischer Raum. Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Sei $\mathcal{C} \subset C_b(E; \mathbb{K})$ eine Punkte trennende Familie, die stabil ist unter Multiplikation und 1 enthält. Ist $\mathbb{K} = \mathbb{C}$, so sei \mathcal{C} zusätzlich abgeschlossen bezüglich komplexer Konjugation.

Dann ist \mathcal{C} eine trennende Familie für $\mathcal{M}_f(E)$.

Beweis. Seien $\mu_1, \mu_2 \in \mathcal{M}_f(E)$ mit $\int g \, d\mu_1 = \int g \, d\mu_2$ für jedes $g \in \mathcal{C}$. Sei \mathcal{C}' die Algebra der endlichen Linearkombinationen von Elementen aus \mathcal{C} . Aufgrund der Linearität des Integrals gilt $\int g \, d\mu_1 = \int g \, d\mu_2$ für jedes $g \in \mathcal{C}'$.

Zu jedem $f \in C_b(E; \mathbb{R})$ und jedem $\varepsilon > 0$ existiert nach dem Satz von Stone-Weierstraß ein $g \in \mathcal{C}'$ mit $\|f - g\|_\infty < \varepsilon$. Nach der Dreiecksungleichung ist

$$\begin{aligned} \left| \int f \, d\mu_1 - \int f \, d\mu_2 \right| &\leq \left| \int f \, d\mu_1 - \int g \, d\mu_1 \right| + \left| \int g \, d\mu_1 - \int g \, d\mu_2 \right| \\ &\quad + \left| \int g \, d\mu_2 - \int f \, d\mu_2 \right| \\ &\leq \varepsilon (\mu_1(E) + \mu_2(E)). \end{aligned}$$

Da $\varepsilon > 0$ beliebig war, gilt Gleichheit und damit $\mu_1 = \mu_2$ (nach Satz 13.11). \square

Als einfache Schlussfolgerungen bekommen wir die folgenden Sätze.

Satz 15.4. Die Verteilung einer beschränkten reellen Zufallsvariablen X ist durch die Angabe aller Momente eindeutig bestimmt.

Beweis. Ohne Einschränkung können wir annehmen, dass X nur Werte in $E := [0, 1]$ annimmt. Für $n \in \mathbb{N}$ definiere die Abbildung $f_n : [0, 1] \rightarrow [0, 1]$ durch $f_n : x \mapsto x^n$. Ferner sei $f_0 \equiv 1$. Die Familie $\mathcal{C} = \{f_n, n \in \mathbb{N}_0\}$ ist Punkte trennend und abgeschlossen unter Multiplikation, also trennend für $\mathcal{M}_f(E)$. \mathbf{P}_X ist also eindeutig festgelegt durch Angabe der Momente $\mathbf{E}[X^n] = \int x^n \, \mathbf{P}_X(dx)$, $n \in \mathbb{N}$. \square

Beispiel 15.5 (nach [79]). Im vorangehenden Satz können wir nicht ohne Weiteres auf die Beschränktheit von X verzichten, selbst wenn alle Momente existieren (es gibt allerdings schwächere Bedingungen, siehe Korollar 15.32). Wir betrachten hierzu $X := \exp(Y)$, wobei $Y \sim \mathcal{N}_{0,1}$. Die Verteilung von X heißt auch **Log-Normalverteilung**. Für jedes $n \in \mathbb{N}$ ist nY verteilt wie die Summe von n^2 unabhängigen, standardnormalverteilten Zufallsvariablen $nY \stackrel{\mathcal{D}}{=} Y_1 + \dots + Y_{n^2}$. Also ist für $n \in \mathbb{N}$

$$\begin{aligned} \mathbf{E}[X^n] &= \mathbf{E}[e^{nY}] = \mathbf{E}[e^{Y_1 + \dots + Y_{n^2}}] = \prod_{i=1}^{n^2} \mathbf{E}[e^{Y_i}] = \mathbf{E}[e^Y]^{n^2} \\ &= \left(\int_{-\infty}^{\infty} (2\pi)^{-1/2} e^y e^{-y^2/2} dy \right)^{n^2} = e^{n^2/2}. \end{aligned} \quad (15.1)$$

Wir wollen nun gleich eine ganze Familie von Verteilungen konstruieren, die die gleichen Momente wie X besitzen. Nach der Transformationsformel für Dichten (Satz 1.101) hat die Verteilung von X die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-1} \exp\left(-\frac{1}{2} \log(x)^2\right) \quad \text{für } x > 0.$$

Für $\alpha \in [-1, 1]$ definieren wir Wahrscheinlichkeitsdichten f_α auf $(0, \infty)$ durch

$$f_\alpha(x) = f(x)(1 + \alpha \sin(2\pi \log(x))).$$

Um zu zeigen, dass f_α eine Dichte ist und die selben Momente wie f besitzt, reicht es zu zeigen, dass für jedes $n \in \mathbb{N}_0$ gilt

$$m(n) := \int_0^\infty x^n f(x) \sin(2\pi \log(x)) dx = 0.$$

Mit der Substitution $y = \log(x) - n$ erhalten wir (wegen $\sin(2\pi(y + n)) = \sin(2\pi y)$)

$$\begin{aligned} m(n) &= \int_{-\infty}^{\infty} e^{yn+n^2} (2\pi)^{-1/2} e^{-(y+n)^2/2} \sin(2\pi(y+n)) dy \\ &= (2\pi)^{-1/2} e^{n^2/2} \int_{-\infty}^{\infty} e^{-y^2/2} \sin(2\pi y) dy = 0, \end{aligned}$$

wobei die letzte Gleichheit folgt, weil der Integrand eine ungerade Funktion ist. ◇

Satz 15.6 (Laplace-Transformation). Ein endliches Maß μ auf $[0, \infty)$ ist eindeutig bestimmt durch Angabe der Laplace-Transformierten

$$\mathcal{L}_\mu(\lambda) := \int e^{-\lambda x} \mu(dx) \quad \text{für } \lambda \geq 0.$$

Beweis. Dem Problem, dass der Raum $[0, \infty)$ nicht kompakt ist, begegnen wir, indem wir zur (Einpunkt-) Kompaktifizierung $E = [0, \infty]$ übergehen. Wir definieren für $\lambda \geq 0$ die stetige Funktion $f_\lambda : [0, \infty] \rightarrow [0, 1]$ durch $f_\lambda(x) = e^{-\lambda x}$, falls $x < \infty$ und $f_\lambda(\infty) = \lim_{x \rightarrow \infty} e^{-\lambda x}$. Dann ist $\mathcal{C} = \{f_\lambda, \lambda \geq 0\}$ Punkte trennend, $f_0 = 1 \in \mathcal{C}$ und $f_\mu \cdot f_\lambda = f_{\mu+\lambda} \in \mathcal{C}$. Nach Korollar 15.3 ist \mathcal{C} trennend für $\mathcal{M}_f([0, \infty])$, also auch für $\mathcal{M}_f([0, \infty))$. \square

Definition 15.7. Für $\mu \in \mathcal{M}_f(\mathbb{R}^d)$ heißt die Abbildung $\varphi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$, die durch

$$\varphi_\mu(t) := \int e^{i\langle t, x \rangle} \mu(dx)$$

definiert wird, die **charakteristische Funktion** von μ .

Satz 15.8 (Charakteristische Funktion). Ein endliches Maß $\mu \in \mathcal{M}_f(\mathbb{R}^d)$ ist durch Angabe der charakteristischen Funktion eindeutig festgelegt.

Beweis. Seien $\mu_1, \mu_2 \in \mathcal{M}_f(\mathbb{R}^d)$ mit $\varphi_{\mu_1}(t) = \varphi_{\mu_2}(t)$ für alle $t \in \mathbb{R}^d$. Nach Satz 13.11(ii) ist $C_c(\mathbb{R}^d)$ trennend für $\mathcal{M}_f(\mathbb{R}^d)$. Es reicht also zu zeigen, dass $\int f d\mu_1 = \int f d\mu_2$ für jedes $f \in C_c(\mathbb{R}^d)$ gilt.

Sei nun $f : \mathbb{R}^d \rightarrow \mathbb{R}$ stetig mit kompaktem Träger und $\varepsilon > 0$. Sei $K > 0$ so groß, dass $f(x) = 0$ für $x \notin (-K/2, K/2)^d$ gilt und $\mu_i(\mathbb{R}^d \setminus (-K, K)^d) < \varepsilon$, $i = 1, 2$. Betrachte den Torus $E := \mathbb{R}^d / (2K\mathbb{Z}^d)$ und definiere $\tilde{f} : E \rightarrow \mathbb{R}$ durch

$$\tilde{f}(x + 2K\mathbb{Z}^d) = f(x) \quad \text{für } x \in [-K, K]^d.$$

Da der Träger von f in $(-K, K)^d$ liegt, ist \tilde{f} stetig. Für $m \in \mathbb{Z}^d$ definiere $g_m : \mathbb{R}^d \rightarrow \mathbb{C}$ durch $x \mapsto \exp(i\langle \pi m/K, x \rangle)$. Sei \mathcal{C} die Algebra der endlichen Linearkombinationen der g_m . Für $g \in \mathcal{C}$ ist $g(x) = g(x + 2Kn)$ für jedes $x \in \mathbb{R}^d$ und $n \in \mathbb{Z}^d$. Daher ist $\tilde{g} : E \rightarrow \mathbb{C}$, $\tilde{g}(x + 2K\mathbb{Z}^d) = g(x)$ wohldefiniert, stetig und beschränkt. Also ist $\tilde{\mathcal{C}} := \{\tilde{g} : g \in \mathcal{C}\} \subset C_b(E; \mathbb{C})$ eine Punkte trennende Algebra, die abgeschlossen unter komplexer Konjugation ist. Da E kompakt ist, existiert nach dem Satz von Stone-Weierstraß also ein $g \in \mathcal{C}$ mit $\|\tilde{g} - \tilde{f}\|_\infty < \varepsilon$. Es folgt

$$\|(f - g)\mathbb{1}_{[-K, K]^d}\|_\infty < \varepsilon$$

und

$$\|(f - g)\mathbb{1}_{\mathbb{R}^d \setminus [-K, K]^d}\|_\infty \leq \|g\|_\infty = \|\tilde{g}\|_\infty \leq \|\tilde{f}\|_\infty + \varepsilon = \|f\|_\infty + \varepsilon.$$

Nach Voraussetzung ist $\int g d\mu_1 = \int g d\mu_2$, also folgt mit der Dreiecksungleichung

$$\begin{aligned} \left| \int f d\mu_1 - \int f d\mu_2 \right| &\leq \int |f - g| d\mu_1 + \int |f - g| d\mu_2 \\ &\leq \varepsilon(2\|f\|_\infty + 2\varepsilon + \mu_1(\mathbb{R}^d) + \mu_2(\mathbb{R}^d)). \end{aligned}$$

Da $\varepsilon > 0$ beliebig war, stimmen die Integrale überein. \square

Korollar 15.9. Ein endliches Maß μ auf \mathbb{Z}^d ist durch die Werte

$$\varphi_\mu(t) = \int e^{i\langle t, x \rangle} \mu(dx), \quad t \in [-\pi, \pi]^d,$$

eindeutig festgelegt.

Beweis. Dies ist klar, weil $\varphi_\mu(t + 2\pi k) = \varphi_\mu(t)$ für jedes $k \in \mathbb{Z}^d$. \square

Während das vorangehende Korollar nur eine abstrakte Eindeutigkeitsaussage liefert, wird uns manchmal eine explizite Inversionsformel von Nutzen sein.

Satz 15.10 (Diskrete Fourier-Inversionsformel). Sei $\mu \in \mathcal{M}_f(\mathbb{Z}^d)$ mit charakteristischer Funktion φ_μ . Dann gilt für jedes $x \in \mathbb{Z}^d$

$$\mu(\{x\}) = (2\pi)^{-d} \int_{[-\pi, \pi]^d} e^{-i\langle t, x \rangle} \varphi_\mu(t) dt.$$

Beweis. Nach dem Satz über majorisierte Konvergenz ist

$$\begin{aligned} \int_{[-\pi, \pi]^d} e^{-i\langle t, x \rangle} \varphi_\mu(t) dt &= \int_{[-\pi, \pi]^d} e^{-i\langle t, x \rangle} \left(\lim_{n \rightarrow \infty} \sum_{|y| \leq n} e^{i\langle t, y \rangle} \mu(\{y\}) \right) dt \\ &= \lim_{n \rightarrow \infty} \int_{[-\pi, \pi]^d} e^{-i\langle t, x \rangle} \sum_{|y| \leq n} e^{i\langle t, y \rangle} \mu(\{y\}) dt \\ &= \sum_{y \in \mathbb{Z}^d} \mu(\{y\}) \int_{[-\pi, \pi]^d} e^{i\langle t, y - x \rangle} dt. \end{aligned}$$

Die Behauptung folgt, weil für $y \in \mathbb{Z}^d$ gilt

$$\int_{[-\pi, \pi]^d} e^{i\langle t, y - x \rangle} dt = \begin{cases} (2\pi)^d, & \text{falls } x = y, \\ 0, & \text{sonst.} \end{cases}$$

Ähnliche Inversionsformeln gelten für Maße μ auf \mathbb{R}^d . Besonders einfach ist der Fall, wo μ eine integrierbare Dichte $f := \frac{d\mu}{d\lambda}$ bezüglich des d -dimensionalen Lebesgue-Maßes λ hat. In diesem Fall gilt die Fourier-Inversionsformel

$$f(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} \varphi_\mu(t) \lambda(dt). \quad (15.2)$$

Es gilt die Plancherel'sche Gleichung: Es ist genau dann $f \in \mathcal{L}^2(\lambda)$, wenn $\varphi_\mu \in \mathcal{L}^2(\lambda)$. In diesem Fall ist $\|f\|_2 = \|\varphi\|_2$.

Da wir diese Aussagen jedoch nicht weiter verwenden werden, verweisen wir lediglich auf die einschlägigen Lehrbücher (etwa [179, Kapitel V.2] oder [59, Theorem XV.3.3 und Gleichung (XV.3.8)]).

Übung 15.1.1. Man zeige, dass im Satz von Stone-Weierstraß auf die Kompaktheit von E nicht verzichtet werden kann. *Hinweis:* Man wähle etwa $E = \mathbb{R}$, nutze aus, dass $C_b(\mathbb{R}) = C_b(\mathbb{R}; \mathbb{R})$ nicht separabel ist und konstruiere eine abzählbare, Punkte trennende Algebra $\mathcal{C} \subset C_b(\mathbb{R})$. 

Übung 15.1.2. Sei $d \in \mathbb{N}$ und μ ein endliches Maß auf $[0, \infty)^d$. Man zeige: μ ist durch Angabe der Laplace-Transformierten $\mathcal{L}_\mu(\lambda) = \int e^{-\langle \lambda, x \rangle} \mu(dx)$, $\lambda \in [0, \infty)^d$ eindeutig bestimmt. 

Übung 15.1.3. Man zeige, dass unter den Voraussetzungen von Satz 15.10 die **Plancherel'sche Gleichung** gilt:

$$\sum_{x \in \mathbb{Z}^d} \mu(\{x\})^2 = (2\pi)^{-d} \int_{[-\pi, \pi]^d} |\varphi_\mu(t)|^2 dt. \quad \clubsuit$$

Übung 15.1.4 (Mellin-Transformierte). Sei X eine nichtnegative reelle Zufallsvariable. Für $s \geq 0$ definieren wir die Mellin-Transformierte von \mathbf{P}_X

$$m_X(s) = \mathbf{E}[X^s]$$

mit Werten in $[0, \infty]$.

Man zeige: Gibt es ein $\varepsilon_0 > 0$ mit $m_X(\varepsilon_0) < \infty$ (beziehungsweise $m_X(-\varepsilon_0) < \infty$), so ist für jedes $\varepsilon > 0$ die Verteilung \mathbf{P}_X eindeutig bestimmt durch die Werte $m_X(s)$ (beziehungsweise $m_X(-s)$), $s \in [0, \varepsilon]$.

Anleitung: Für stetiges $f : [0, \infty) \rightarrow [0, \infty)$ sei

$$\phi_f(z) = \int_0^\infty t^{z-1} f(t) dt,$$

für diejenigen $z \in \mathbb{C}$, für die dies wohldefiniert ist. Aus der Funktionentheorie ist bekannt: Ist $\phi_f(s) < \infty$ für ein $s > 1$, so ist ϕ_f holomorph in $\{z \in \mathbb{C} : \operatorname{Re}(z) \in (1, s)\}$ (und damit durch die Werte $\phi_f(r)$, $r \in (1, 1 + \varepsilon)$ eindeutig festgelegt für jedes $\varepsilon > 0$), und es gilt für jedes $r \in (1, s)$

$$f(t) = \frac{1}{2\pi i} \int_{-\infty}^\infty t^{-(r+i\rho)} \phi_f(r+i\rho) d\rho.$$

- (i) Man folgere die Aussage für X mit stetiger Dichte.
- (ii) Für $\delta > 0$ sei $Y_\delta \sim \mathcal{U}_{[1-\delta, 1]}$ und unabhängig von X . Man zeige, dass XY_δ eine stetige Dichte hat.
- (iii) Man bestimme m_{XY_δ} und zeige, dass $m_{XY_\delta} \rightarrow m_X$ für $\delta \downarrow 0$.
- (iv) Man zeige, dass $XY_\delta \Rightarrow X$ für $\delta \downarrow 0$. 

Übung 15.1.5. Seien X, Y, Z unabhängige nichtnegative reelle Zufallsvariablen, sodass $\mathbf{P}[Z > 0] > 0$, und sodass die Mellin-Transformierte $m_{XZ}(s) < \infty$ ist für ein $s > 0$.

Zeige: Gilt $XZ \stackrel{\mathcal{D}}{=} YZ$, so ist $X \stackrel{\mathcal{D}}{=} Y$. ♣

Übung 15.1.6. Sei μ ein W-Maß auf \mathbb{R} mit integrierbarer charakteristischer Funktion φ_μ , also $\varphi_\mu \in \mathcal{L}^1(\lambda)$, wobei λ das Lebesgue-Maß auf \mathbb{R} ist. Man zeige, dass μ absolutstetig ist und die stetige und beschränkte Dichte $f = \frac{d\mu}{d\lambda}$ gegeben ist durch

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_\mu(t) dt \quad \text{für jedes } x \in \mathbb{R}.$$

Anleitung: Man zeige dies zunächst für die Normalverteilung $\mathcal{N}_{0,\varepsilon}$, $\varepsilon > 0$. Man zeige dann, dass $\mu * \mathcal{N}_{0,\varepsilon}$ absolutstetig ist mit Dichte f_ε , die punktweise gegen f konvergiert. ♣

Übung 15.1.7. Sei (Ω, τ) ein separabler topologischer Raum, der das $T_{3\frac{1}{2}}$ -Trennungsaxiom erfüllt: Zu jeder abgeschlossenen Menge $A \subset \Omega$ und jedem Punkt $x \in \Omega \setminus A$ existiert eine stetige Funktion $f : \Omega \rightarrow [0, 1]$ mit $f(x) = 0$ und $f(y) = 1$ für jedes $y \in A$. (Insbesondere ist jeder metrische Raum ein $T_{3\frac{1}{2}}$ -Raum.)

Man zeige: $\sigma(C_b(\Omega)) = \mathcal{B}(\Omega)$, das heißt, die Borel'sche σ -Algebra wird durch die beschränkten, stetigen Funktionen $\Omega \rightarrow \mathbb{R}$ erzeugt. ♣

15.2 Charakteristische Funktionen: Beispiele

Lemma 15.11. Sei X eine Zufallsvariable mit Werten in \mathbb{R}^d und charakteristischer Funktion $\varphi_X(t) = \mathbf{E}[e^{i\langle t, X \rangle}]$. Dann gelten

- (i) $|\varphi_X(t)| \leq 1$ für jedes $t \in \mathbb{R}^d$ und $\varphi_X(0) = 1$,
- (ii) $\varphi_{aX+b}(t) = \varphi_X(at) e^{i\langle b, t \rangle}$ für jedes $a \in \mathbb{R}$ und $b \in \mathbb{R}^d$,
- (iii) $\mathbf{P}_X = \mathbf{P}_{-X}$ genau dann, wenn φ reellwertig ist,
- (iv) Sind X und Y unabhängig, so ist $\varphi_{X+Y} = \varphi_X \cdot \varphi_Y$.
- (v) Für jedes $t \in \mathbb{R}^d$ gilt für den Realteil $0 \leq 1 - \operatorname{Re}(\varphi_X(2t)) \leq 4(1 - \operatorname{Re}(\varphi_X(t)))$.

Beweis. (i) und (ii) sind trivial.

(iii) $\overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t)$.

(iv) Da $e^{i\langle t, X \rangle}$ und $e^{i\langle t, Y \rangle}$ unabhängige Zufallsvariablen sind, gilt

$$\varphi_{X+Y}(t) = \mathbf{E}[e^{i\langle t, X \rangle} \cdot e^{i\langle t, Y \rangle}] = \mathbf{E}[e^{i\langle t, X \rangle}] \mathbf{E}[e^{i\langle t, Y \rangle}] = \varphi_X(t) \varphi_Y(t).$$

(v) Nach dem Additionstheorem für Winkelfunktionen ist

$$1 - \cos(\langle 2t, X \rangle) = 2(1 - (\cos(\langle t, X \rangle))^2) \leq 4(1 - \cos(\langle t, X \rangle)).$$

Bilde jetzt den Erwartungswert. \square

Satz 15.12 (Charakteristische Funktionen wichtiger Verteilungen). Wir geben für verschiedene Verteilungen P mit Dichte $x \mapsto f(x)$ auf \mathbb{R} oder Gewichten $P(\{k\})$, $k \in \mathbb{N}_0$, die charakteristische Funktion $\varphi(t)$ an:

Verteilung				Char. Fkt.
Name Symbol	Parameter	auf	Dichte / Gewicht	$\varphi(t)$
Normal $\mathcal{N}_{\mu, \sigma^2}$	$\mu \in \mathbb{R}$ $\sigma^2 > 0$	\mathbb{R}	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$e^{i\mu t} \cdot e^{-\sigma^2 t^2/2}$
Gleichvert. $\mathcal{U}_{[0,a]}$	$a > 0$	$[0, a]$	$1/a$	$\frac{e^{iat}-1}{iat}$
Gleichvert. $\mathcal{U}_{[-a,a]}$	$a > 0$	$[-a, a]$	$1/2a$	$\frac{\sin(at)}{at}$
Dreieck Tri_a	$a > 0$	$[-a, a]$	$\frac{1}{a}(1 - x /a)^+$	$2 \frac{1 - \cos(at)}{a^2 t^2}$
N.N.	$a > 0$	\mathbb{R}	$\frac{1}{\pi} \frac{1 - \cos(ax)}{ax^2}$	$(1 - t /a)^+$
Gamma	$\theta > 0$ $r > 0$	$[0, \infty)$	$\frac{\theta^r}{\Gamma(r)} x^{r-1} e^{-\theta x}$	$(1 - it/\theta)^{-r}$
Exponential \exp_θ	$\theta > 0$	$[0, \infty)$	$\theta e^{-\theta x}$	$\frac{\theta}{\theta - it}$
zweiseitig Exponential \exp_θ^2	$\theta > 0$	\mathbb{R}	$\frac{\theta}{2} e^{-\theta x }$	$\frac{1}{1 + (t/a)^2}$
Cauchy Cau_a	$a > 0$	\mathbb{R}	$\frac{1}{a\pi} \frac{1}{1 + (x/a)^2}$	$e^{-a t }$
Binomial $b_{n,p}$	$n \in \mathbb{N}$ $p \in [0, 1]$	$\{0, \dots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$((1-p) + pe^{it})^n$
Negativ Binomial $b_{r,p}^-$	$r > 0$ $p \in (0, 1]$	\mathbb{N}_0	$\binom{-r}{k} (-1)^k p^r (1-p)^k$	$\left(\frac{p}{1 - (1-p)e^{it}}\right)^r$
Poisson Poi_λ	$\lambda > 0$	\mathbb{N}_0	$e^{-\lambda} \frac{\lambda^k}{k!}$	$\exp(\lambda(e^{it} - 1))$

Beweis. **(i) (Normalverteilung)** Nach Lemma 15.11 reicht es, den Fall $\mu = 0$ und $\sigma^2 = 1$ zu betrachten. Mit Hilfe des Differentiationslemmas (Satz 6.28) und durch partielle Integration erhalten wir

$$\frac{d}{dt} \varphi(t) = \int_{-\infty}^{\infty} e^{itx} ix e^{-x^2/2} dx = -t \varphi(t).$$

Diese lineare Differentialgleichung mit Anfangswert $\varphi(0) = 1$ hat die eindeutige Lösung $\varphi(t) = e^{-t^2/2}$.

(ii) (Gleichverteilung) Dies ist unmittelbar.

(iii) (Dreieck) Es gilt $\text{Tri}_a = \mathcal{U}_{[-a/2, a/2]} * \mathcal{U}_{[-a/2, a/2]}$, also ist

$$\varphi_{\text{Tri}_a}(t) = \varphi_{\mathcal{U}_{[-a/2, a/2]}}(t)^2 = 4 \frac{\sin(at/2)^2}{a^2 t^2} = 2 \frac{1 - \cos(at)}{a^2 t^2},$$

wobei wir ausgenutzt haben, dass nach dem Additionstheorem gilt

$$1 - \cos(x) = \sin(x/2)^2 + \cos(x/2)^2 - \cos(x) = 2 \sin(x/2)^2.$$

(iv) (N.N.) Dies lässt sich entweder direkt ausrechnen, oder mit Hilfe der Fourier-Inversionsformel (Gleichung (15.2)) aus (iii) folgern.

(v) (Gammaverteilung) Es reicht wiederum, den Fall $\theta = 1$ zu betrachten. Für $0 \leq b < c \leq \infty$ und $t \in \mathbb{R}$ sei $\gamma_{b,c,t}$ der gerade Weg in \mathbb{C} von $b - ibt$ nach $c - ict$ und $\delta_{b,t}$ der gerade Weg von b nach $b - ibt$ sowie $\epsilon_{c,t}$ der gerade Weg von $c - ict$ nach c . Mit der Substitution $z = (1 - it)x$ ist

$$\varphi(t) = \frac{1}{\Gamma(r)} \int_0^\infty x^{r-1} e^{-x} e^{itx} dx = \frac{(1 - it)^{-r}}{\Gamma(r)} \int_{\gamma_{0,\infty,t}} z^{r-1} e^{-z} dz.$$

Daher reicht es zu zeigen, dass $\int_{\gamma_{0,\infty,t}} z^{r-1} \exp(-z) dz = \Gamma(r)$.

Die Funktion $z \mapsto z^{r-1} \exp(-z)$ ist holomorph in der rechten komplexen Halbebene. Nach dem Residuensatz ist daher für $0 < b < c < \infty$

$$\begin{aligned} \int_b^c x^{r-1} \exp(-x) dx &= \int_{\gamma_{b,c,t}} z^{r-1} \exp(-z) dz \\ &\quad + \int_{\delta_{b,t}} z^{r-1} \exp(-z) dz + \int_{\epsilon_{c,t}} z^{r-1} \exp(-z) dz. \end{aligned}$$

Wegen $\int_0^\infty x^{r-1} \exp(-x) dx =: \Gamma(r)$, reicht es zu zeigen, dass die Integrale entlang $\delta_{b,t}$ und $\epsilon_{c,t}$ verschwinden, wenn $b \rightarrow 0$ und $c \rightarrow \infty$.

Nun ist aber $|z^{r-1} \exp(-z)| \leq (1 + t^2)^{(r-1)/2} b^{r-1} \exp(-b)$ für $z \in \delta_{b,t}$. Da der Weg $\delta_{b,t}$ die Länge $b|t|$ hat, ist also $\left| \int_{\delta_{b,t}} z^{r-1} \exp(-z) dz \right| \leq b^r e^{-b} (1 + t^2)^{r/2} \rightarrow$

0, wenn $b \rightarrow 0$. Analog ist $\left| \int_{\epsilon_{c,t}} z^{r-1} \exp(-z) dz \right| \leq c^r \exp(-c)(1+t^2)^{r/2} \rightarrow 0$, wenn $c \rightarrow \infty$.

(vi) (Exponentialverteilung) Wegen $\exp_\theta = \Gamma_{\theta,1}$ folgt dies aus (v).

(vii) (Zweiseitige Exponentialverteilung) Sind X und Y unabhängige \exp_θ verteilte Zufallsvariablen, so ist $X - Y \sim \exp_\theta^2$ (Nachrechnen!). Also ist

$$\varphi_{\exp_\theta^2}(t) = \varphi_{\exp_\theta}(t) \varphi_{\exp_\theta}(-t) = \frac{1}{1-it/\theta} \frac{1}{1+it/\theta} = \frac{1}{1+(t/\theta)^2}.$$

(viii) (Cauchy Verteilung) Dies lässt sich entweder mit Hilfe des Residuenkalküls direkt ausrechnen, oder mit Hilfe der Fourier-Inversionsformel (Gleichung (15.2)) aus der Aussage für die zweiseitige Exponentialverteilung folgern.

(ix) (Binomialverteilung) Nach dem binomischen Lehrsatz ist

$$\varphi(t) = \sum_{k=0}^n \binom{n}{k} (1-p)^{n-k} (pe^{it})^k = (1-p+pe^{it})^n.$$

(x) (Negative Binomialverteilung) Nach dem verallgemeinerten binomischen Lehrsatz (Lemma 3.5) ist für jedes $x \in \mathbb{C}$ mit $|x| < 1$

$$(1-x)^{-r} = \sum_{k=0}^{\infty} \binom{-r}{k} (-x)^k.$$

Wenn wir $x = (1-p)e^{it}$ setzen, folgt die Behauptung.

(xi) (Poissonverteilung) Es ist $\varphi_{\text{Poi}_\lambda}(t) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{(\lambda e^{it})^n}{n!} = e^{\lambda(e^{it}-1)}$. \square

Korollar 15.13. Es gelten die folgenden Faltungsformeln:

- (i) $\mathcal{N}_{\mu_1, \sigma_1^2} * \mathcal{N}_{\mu_2, \sigma_2^2} = \mathcal{N}_{\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2}$ für $\mu_1, \mu_2 \in \mathbb{R}$ und $\sigma_1^2, \sigma_2^2 > 0$,
- (ii) $\Gamma_{\theta,r} * \Gamma_{\theta,s} = \Gamma_{\theta,r+s}$ für $\theta, r, s > 0$,
- (iii) $\text{Cau}_a * \text{Cau}_b = \text{Cau}_{a+b}$ für $a, b > 0$,
- (iv) $b_{m,p} * b_{n,p} = b_{m+n,p}$ für $m, n \in \mathbb{N}$ und $p \in [0, 1]$,
- (v) $b_{r,p}^- * b_{s,p}^- = b_{r+s,p}^-$ für $r, s > 0$ und $p \in (0, 1]$,
- (vi) $\text{Poi}_\lambda * \text{Poi}_\mu = \text{Poi}_{\lambda+\mu}$ für $\lambda, \mu \geq 0$.

Beweis. Die Aussagen folgen aus dem vorangehenden Satz zusammen mit $\varphi_{\mu*\nu} = \varphi_\mu \varphi_\nu$ (Lemma 15.11). \square

Zwei einfache Verfahren, um charakteristische Funktionen von zusammengesetzten Verteilungen auszurechnen, liefert der folgende Satz:

Satz 15.14. (i) Seien $\mu_1, \mu_2, \dots \in \mathcal{M}_f(\mathbb{R}^d)$ und p_1, p_2, \dots nichtnegative Zahlen mit $\sum_{n=1}^{\infty} p_n \mu_n(\mathbb{R}^d) < \infty$. Dann hat das Maß $\mu := \sum_{n=1}^{\infty} p_n \mu_n \in \mathcal{M}_f(\mathbb{R}^d)$ die charakteristische Funktion

$$\varphi_{\mu} = \sum_{n=1}^{\infty} p_n \varphi_{\mu_n}. \quad (15.3)$$

(ii) Es seien N, X_1, X_2, \dots unabhängige Zufallsvariablen. Die X_1, X_2, \dots seien identisch verteilt auf \mathbb{R}^d mit charakteristischer Funktion φ_X . N habe Werte in \mathbb{N}_0 und die Erzeugendenfunktion f_N . Dann hat $Y := \sum_{n=1}^N X_n$ die charakteristische Funktion $\varphi_Y(t) = f_N(\varphi_X(t))$.

(iii) Ist in (ii) speziell $N \sim \text{Poi}_{\lambda}$, so ist $\varphi_Y(t) = \exp(\lambda(\varphi_X(t) - 1))$.

Beweis. (i) Setzen wir $\nu_n = \sum_{k=1}^n p_k \mu_k$, so gilt $\varphi_{\nu_n} = \sum_{k=1}^n p_k \varphi_{\mu_k}$ wegen der Linearität des Integrals. Nach Voraussetzung ist $\mu = \text{w-lim}_{n \rightarrow \infty} \nu_n$, also auch $\varphi_{\mu}(t) = \lim_{n \rightarrow \infty} \varphi_{\nu_n}(t)$.

(ii) Es ist

$$\begin{aligned} \varphi_Y(t) &= \sum_{n=0}^{\infty} \mathbf{P}[N=n] \mathbf{E}[e^{i\langle t, X_1 + \dots + X_n \rangle}] \\ &= \sum_{n=0}^{\infty} \mathbf{P}[N=n] \varphi_X(t)^n = f_N(\varphi(t)). \end{aligned}$$

(iii) Der Spezialfall folgt, weil hier $f_N(z) = e^{\lambda(z-1)}$ für $z \in \mathbb{C}$ mit $|z| \leq 1$. \square

Beispiel 15.15. Sei $n \in \mathbb{N}$, und seien Punkte $0 = a_0 < a_1 < \dots < a_n$ und $1 = y_0 > y_1 > \dots > y_n = 0$ gegeben. Sei $\varphi : \mathbb{R} \rightarrow [0, \infty)$ diejenige gerade Funktion (also $\varphi(x) = \varphi(-x)$), die $\varphi(a_k) = y_k$ für jedes $k = 0, \dots, n$ erfüllt und zwischen den Punkten a_k linear interpoliert ist sowie ferner $\varphi(x) = 0$ für $|x| > a_n$ erfüllt. Wir wollen zusätzlich annehmen, dass die y_k so gewählt sind, dass φ auf $[0, \infty)$ konvex ist. Das ist äquivalent zu der Bedingung, dass $m_1 \leq m_2 \leq \dots \leq m_n \leq 0$, wo $m_k := \frac{y_k - y_{k-1}}{a_k - a_{k-1}}$ die Steigung im k -ten Intervall ist. Wir wollen zeigen, dass φ die charakteristische Funktion eines W-Maßes $\mu \in \mathcal{M}_1(\mathbb{R})$ ist.

Setze $p_k = a_k(m_{k+1} - m_k)$ für $k = 1, \dots, n$.

Sei $\mu_k \in \mathcal{M}_1(\mathbb{R})$ die Verteilung auf \mathbb{R} mit Dichte $\frac{1}{\pi} \frac{1-\cos(a_k \pi)}{a_k x^2}$. Nach Satz 15.12 hat μ_k die charakteristische Funktion $\varphi_{\mu_k}(t) = \left(1 - \frac{|t|}{a_k}\right)^+$. Die charakteristische Funktion φ_{μ} von $\mu := \sum_{k=1}^n p_k \mu_k$ ist dann

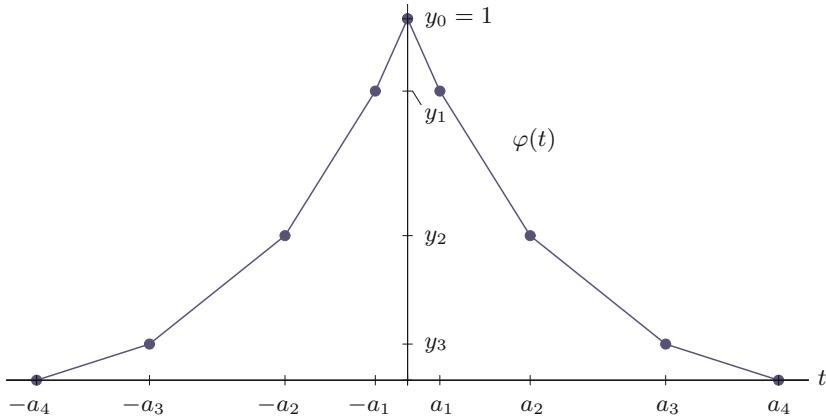


Abb. 15.1. Die charakteristische Funktion φ aus Beispiel 15.15 mit $n = 4$.

$$\varphi_\mu(t) = \sum_{k=1}^n p_k (1 - |t|/a_k)^+.$$

Dies ist eine stetige, symmetrische, reelle Funktion mit $\varphi_\mu(0) = 1$, die auf den Intervallen $[a_{k-1}, a_k]$ jeweils linear ist. Durch partielle Summation erhalten wir (wegen $m_{n+1} = 0$) für jedes $k = 1, \dots, n$

$$\begin{aligned} \varphi_\mu(a_l) &= \sum_{k=1}^n a_k (m_{k+1} - m_k) \left(1 - \frac{a_l}{a_k}\right)^+ = \sum_{k=l}^n (a_k - a_l)(m_{k+1} - m_k) \\ &= \left[(a_n - a_l)m_{n+1} - (a_l - a_l)m_l \right] - \sum_{k=l+1}^n (a_k - a_{k-1})m_k \\ &= - \sum_{k=l+1}^n (y_k - y_{k-1}) = y_l = \varphi(a_l). \end{aligned}$$

Also ist $\varphi_\mu = \varphi$. ◇

Beispiel 15.16. Wir betrachten die Funktion $\varphi : \mathbb{R} \rightarrow [0, 1]$, die periodisch mit Periode 2π ist, und die für $t \in [-\pi, \pi]$ definiert ist durch $\varphi(t) = 1 - 2|t|/\pi$. Durch die diskrete Fourier-Inversionsformel (Satz 15.10) erhalten wir, dass φ die charakteristische Funktion des W-Maßes $\mu \in \mathcal{M}_1(\mathbb{Z})$ mit $\mu(\{x\}) = (2\pi)^{-1} \int_{-\pi}^{\pi} \cos(tx) \varphi(t) dt$ ist, wenn wir zeigen können, dass alle diese Zahlen $\mu(\{x\})$ nichtnegativ sind. Offenbar ist $\mu(\{0\}) = 0$. Für $x \in \mathbb{Z} \setminus \{0\}$ berechnen wir das Integral mit Hilfe partieller Integration

$$\begin{aligned}
\int_{-\pi}^{\pi} \cos(tx) \varphi(t) dt &= 2 \int_0^{\pi} \cos(tx) (1 - 2t/\pi) dt \\
&= \frac{4}{x} \left(1 - \frac{2}{\pi}\right) \sin(\pi x) - \frac{4}{x} \sin(0) + \frac{4}{\pi x} \int_0^{\pi} \sin(tx) dt \\
&= \frac{4}{\pi x^2} (1 - \cos(\pi x)).
\end{aligned}$$

Insgesamt erhalten wir

$$\mu(\{x\}) = \begin{cases} \frac{4}{\pi^2 x^2}, & \text{falls } x \text{ ungerade ist,} \\ 0, & \text{sonst.} \end{cases}$$

Wegen $\mu(\mathbb{Z}) = \varphi(0) = 1$ ist μ tatsächlich ein W-Maß. \diamond

Beispiel 15.17. Wir betrachten die Funktion $\psi : \mathbb{R} \rightarrow [0, 1]$, die periodisch mit Periode π ist, und die für $t \in [-\pi/2, \pi/2)$ definiert ist durch $\psi(t) = 1 - 2|t|/\pi$. Ist φ die charakteristische Funktion zum Maß μ aus dem vorangehenden Beispiel, so ist offenbar $\psi(t) = |\varphi(t)|$. Andererseits ist $\psi(t) = \frac{1}{2} + \frac{1}{2}\varphi(2t)$. Nach Satz 15.14 und Lemma 15.11(ii) ist daher ψ die charakteristische Funktion des Maßes ν mit $\nu(A) = \frac{1}{2}\delta_0(A) + \frac{1}{2}\mu(A/2)$ für $A \subset \mathbb{R}$. Also ist

$$\nu(\{x\}) = \begin{cases} \frac{1}{2}, & \text{falls } x = 0, \\ \frac{8}{\pi^2 x^2}, & \text{falls } \frac{x}{2} \in \mathbb{Z} \text{ ungerade ist,} \\ 0, & \text{sonst.} \end{cases} \quad \diamond$$

Beispiel 15.18. Sei $\varphi(t) = (1 - 2|t|/\pi)^+$ die charakteristische Funktion der Verteilung „N.N.“ aus Satz 15.12 (mit $a = \pi/2$) und ψ die charakteristische Funktion aus dem vorangehenden Beispiel. Man beachte, dass $\varphi(t) = \psi(t)$ für $|t| \leq \pi/2$ und $\varphi(t) = 0$ für $|t| > \pi/2$, also $\varphi^2 = \varphi \cdot \psi$. Seien nun X, Y, Z unabhängige, reelle Zufallsvariablen mit charakteristischen Funktionen $\varphi_X = \varphi_Y = \varphi$ und $\varphi_Z = \psi$. Dann ist $\varphi_X \varphi_Y = \varphi_X \varphi_Z$, also $X + Y \stackrel{\mathcal{D}}{=} X + Z$, jedoch stimmen die Verteilungen von Y und Z nicht überein. \diamond

Übung 15.2.1. Sei φ die charakteristische Funktion der d -dimensionalen Zufallsvariablen X . Man zeige: Ist $\varphi(t) = 1$ für ein $t \neq 0$, so ist $\mathbf{P}[X \in H_t] = 1$, wo

$$\begin{aligned}
H_t &= \{x \in \mathbb{R}^d : \langle x, t \rangle \in 2\pi\mathbb{Z}\} \\
&= \{y + z \cdot (2\pi t / \|t\|_2^2) : z \in \mathbb{Z}, y \in \mathbb{R}^d \text{ mit } \langle y, t \rangle = 0\}.
\end{aligned}$$

Man folgere, dass $\varphi(t+s) = \varphi(s)$ ist für jedes $s \in \mathbb{R}^d$. \clubsuit

Übung 15.2.2. Man zeige: Es gibt reelle Zufallsvariablen X, X' und Y, Y' mit $X \stackrel{\mathcal{D}}{=} X'$ und $Y \stackrel{\mathcal{D}}{=} Y'$, sodass X' und Y' unabhängig sind und $X + Y \stackrel{\mathcal{D}}{=} X' + Y'$ gilt, jedoch X und Y nicht unabhängig sind. \clubsuit

Übung 15.2.3. Sei X eine reelle Zufallsvariable mit charakteristischer Funktion φ . X heißt **gitterverteilt**, wenn es $a, d \in \mathbb{R}$ gibt, sodass $\mathbf{P}[X \in a + d\mathbb{Z}] = 1$. Zeige: X ist genau dann gitterverteilt, wenn es ein $u \neq 0$ gibt mit $|\varphi(u)| = 1$. ♣

Übung 15.2.4. Sei X eine reelle Zufallsvariable mit charakteristischer Funktion φ . Es gebe eine Folge $(t_n)_{n \in \mathbb{N}}$ reeller Zahlen mit $|t_n| \downarrow 0$ und $|\varphi(t_n)| = 1$ für jedes n . Man zeige: Es gibt ein $b \in \mathbb{R}$ mit $X = b$ fast sicher. Gilt sogar $\varphi(t_n) = 1$ für alle n , so ist $X = 0$ fast sicher. ♣

15.3 Der Lévy'sche Stetigkeitssatz

Die Hauptaussage dieses Abschnitts ist der Stetigkeitssatz von Lévy (Satz 15.23), der, grob gesprochen, besagt, dass eine Folge von charakteristischen Funktionen genau dann punktweise gegen eine stetige Funktion konvergiert, wenn der Grenzwert wieder eine charakteristische Funktion ist und die zugehörigen Wahrscheinlichkeitsmaße schwach konvergieren. Wir bereiten den Beweis des Satzes mit ein paar analytischen Aussagen vor.

Lemma 15.19. Sei $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ mit charakteristischer Funktion φ . Dann gilt

$$|\varphi(t) - \varphi(s)|^2 \leq 2(1 - \operatorname{Re}(\varphi(t-s))) \quad \text{für alle } s, t \in \mathbb{R}^d.$$

Beweis. Nach der Cauchy-Schwarz'schen Ungleichung gilt

$$\begin{aligned} |\varphi(t) - \varphi(s)|^2 &= \left| \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} - e^{i\langle s, x \rangle} \mu(dx) \right|^2 \\ &= \left| \int_{\mathbb{R}^d} (e^{i\langle t-s, x \rangle} - 1) e^{i\langle s, x \rangle} \mu(dx) \right|^2 \\ &\leq \int_{\mathbb{R}^d} |e^{i\langle t-s, x \rangle} - 1|^2 \mu(dx) \cdot \int_{\mathbb{R}^d} |e^{i\langle s, x \rangle}|^2 \mu(dx) \\ &= \int_{\mathbb{R}^d} (e^{i\langle t-s, x \rangle} - 1)(e^{-i\langle t-s, x \rangle} - 1) \mu(dx) \\ &= 2(1 - \operatorname{Re}(\varphi(t-s))). \end{aligned}$$
□

Definition 15.20. Sei (E, d) ein metrischer Raum. Eine Familie $(f_i, i \in I)$ von Abbildungen $E \rightarrow \mathbb{R}$ heißt **gleichgradig gleichmäßig stetig**, falls für jedes $\varepsilon > 0$ ein $\delta > 0$ existiert, sodass $|f_i(t) - f_i(s)| < \varepsilon$ für alle $i \in I$ und alle $s, t \in E$ mit $d(s, t) < \delta$.

Satz 15.21. Ist $\mathcal{F} \subset \mathcal{M}_1(\mathbb{R}^d)$ eine straffe Familie, so ist $\{\varphi_\mu : \mu \in \mathcal{F}\}$ gleichgradig gleichmäßig stetig. Speziell ist jede charakteristische Funktion gleichmäßig stetig.

Beweis. Zu zeigen ist, dass für jedes $\varepsilon > 0$ ein $\delta > 0$ existiert, sodass für jedes $t \in \mathbb{R}^d$, jedes $s \in \mathbb{R}^d$ mit $|t - s| < \delta$ und jedes $\mu \in \mathcal{F}$ gilt: $|\varphi_\mu(t) - \varphi_\mu(s)| < \varepsilon$.

Da \mathcal{F} straff ist, existiert ein $N \in \mathbb{N}$ mit $\mu([-N, N]^d) > 1 - \varepsilon^2/6$ für jedes $\mu \in \mathcal{F}$. Weiterhin existiert ein $\delta > 0$, sodass für $x \in [-N, N]^d$ und $u \in \mathbb{R}^d$ mit $|u| < \delta$ gilt, dass $|1 - e^{i\langle u, x \rangle}| < \varepsilon^2/6$. Wir erhalten so für jedes $\mu \in \mathcal{F}$

$$\begin{aligned} 1 - \operatorname{Re}(\varphi_\mu(u)) &\leq \int_{\mathbb{R}^d} |1 - e^{i\langle u, x \rangle}| \mu(dx) \\ &\leq \frac{\varepsilon^2}{3} + \int_{[-N, N]^d} |1 - e^{i\langle u, x \rangle}| \mu(dx) \\ &\leq \frac{\varepsilon^2}{3} + \frac{\varepsilon^2}{6} = \frac{\varepsilon^2}{2} \end{aligned}$$

und damit für $|t - s| < \delta$ nach Lemma 15.19 $|\varphi_\mu(t) - \varphi_\mu(s)| \leq \varepsilon$. \square

Lemma 15.22. Sei (E, d) ein metrischer Raum und f, f_1, f_2, \dots Abbildungen $E \rightarrow \mathbb{R}$ mit $f_n \xrightarrow{n \rightarrow \infty} f$ punktweise. Ist $(f_n)_{n \in \mathbb{N}}$ gleichgradig gleichmäßig stetig, so ist f gleichmäßig stetig und $(f_n)_{n \in \mathbb{N}}$ konvergiert gegen f gleichmäßig auf kompakten Mengen: Für jede kompakte Menge $K \subset E$ gilt $\sup_{s \in K} |f_n(s) - f(s)| \xrightarrow{n \rightarrow \infty} 0$.

Beweis. Sei $\varepsilon > 0$ vorgegeben und $\delta > 0$ so gewählt, dass $|f_n(t) - f_n(s)| < \varepsilon$ gilt für alle $n \in \mathbb{N}$ und alle $s, t \in E$ mit $d(s, t) < \delta$. Für diese s, t gilt dann

$$|f(s) - f(t)| = \lim_{n \rightarrow \infty} |f_n(s) - f_n(t)| \leq \varepsilon.$$

Also ist f gleichmäßig stetig.

Sei nun $K \subset E$ kompakt. Da kompakte Mengen total beschränkt sind, existiert ein $N \in \mathbb{N}$ sowie $t_1, \dots, t_N \in K$ mit $K \subset \bigcup_{i=1}^N B_\delta(t_i)$. Wähle $n_0 \in \mathbb{N}$ so groß, dass $|f_n(t_i) - f(t_i)| \leq \varepsilon$ ist für jedes $i = 1, \dots, N$ und jedes $n \geq n_0$.

Sei nun $s \in K$ und $n \geq n_0$. Wähle ein t_i mit $d(s, t_i) < \delta$. Dann gilt

$$|f_n(s) - f(s)| \leq |f_n(s) - f_n(t_i)| + |f_n(t_i) - f(t_i)| + |f(t_i) - f(s)| \leq 3\varepsilon.$$

Da $\varepsilon > 0$ beliebig war, gilt $f_n \xrightarrow{n \rightarrow \infty} f$ gleichmäßig auf K . \square

Wir nennen eine Abbildung $f : \mathbb{R}^d \rightarrow \mathbb{R}$ **partiell stetig** im Punkte $x = (x_1, \dots, x_d)$, falls für jedes $i = 1, \dots, d$ die Abbildung $y_i \mapsto f(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_d)$ in $y_i = x_i$ stetig ist.

Satz 15.23 (Lévy'scher Stetigkeitssatz).

Seien $P, P_1, P_2, \dots \in \mathcal{M}_1(\mathbb{R}^d)$ mit charakteristischen Funktionen $\varphi, \varphi_1, \varphi_2, \dots$

(i) Ist $P = \text{w-lim}_{n \rightarrow \infty} P_n$, so gilt $\varphi_n \xrightarrow{n \rightarrow \infty} \varphi$ gleichmäßig auf kompakten Mengen.

(ii) Gilt $\varphi_n \xrightarrow{n \rightarrow \infty} f$ punktweise für eine in 0 partiell stetige Funktion $f : \mathbb{R}^d \rightarrow \mathbb{C}$, so existiert ein W-Maß Q mit $\varphi_Q = f$ und $Q = \text{w-lim}_{n \rightarrow \infty} P_n$.

Beweis. (i) Es gilt $\varphi_n \xrightarrow{n \rightarrow \infty} \varphi$ punktweise. Da die Familie $(P_n)_{n \in \mathbb{N}}$ straff ist, ist $(\varphi_n)_{n \in \mathbb{N}}$ nach Satz 15.21 gleichgradig gleichmäßig stetig. Nach Lemma 15.22 folgt hieraus die gleichmäßige Konvergenz auf kompakten Mengen.

(ii) Nach Satz 13.34 reicht es zu zeigen, dass die Folge $(P_n)_{n \in \mathbb{N}}$ straff ist. Zu diesem Zweck reicht es zu zeigen, dass für jedes $k = 1, \dots, n$ die durch $P_n^k = P_n \circ \pi_k^{-1}$ (wo $\pi_k : \mathbb{R}^d \rightarrow \mathbb{R}$ die Projektion auf die k -te Koordinate ist) definierte Folge $(P_n^k)_{n \in \mathbb{N}}$ der k -ten Marginalverteilungen straff ist. Bezeichnet e_k den k -ten Einheitsvektor in \mathbb{R}^d , so ist die charakteristische Funktion von P_n^k gegeben durch $\varphi_{P_n^k}(t) = \varphi_n(t e_k)$. Nach Voraussetzung konvergiert nun aber $\varphi_{P_n^k}$ für $n \rightarrow \infty$ punktweise gegen eine in 0 stetige Funktion f_k . Wir haben damit das Problem auf den eindimensionalen Fall zurückgeführt und können daher im Folgenden annehmen, dass $d = 1$ gilt.

Da $\varphi_n(0) = 1$ ist für jedes $n \in \mathbb{N}$, haben wir $f(0) = 1$. Wir definieren die Abbildung $h : \mathbb{R} \rightarrow [0, \infty)$ durch $h(x) = 1 - \sin(x)/x$ für $x \neq 0$ und $h(0) = 0$. Offenbar ist h auf ganz \mathbb{R} stetig differenzierbar. Wir erhalten leicht, dass $\alpha := \inf\{h(x) : |x| \geq 1\} = 1 - \sin(1) > 0$. Jetzt berechnen wir (unter Benutzung der Markov'schen Ungleichung und des Satzes von Fubini) für $K > 0$

$$\begin{aligned} P_n([-K, K]^c) &\leq \alpha^{-1} \int_{[-K, K]^c} h(x/K) P_n(dx) \\ &\leq \alpha^{-1} \int_{\mathbb{R}} h(x/K) P_n(dx) \\ &= \alpha^{-1} \int_{\mathbb{R}} \left(\int_0^1 (1 - \cos(tx/K)) dt \right) P_n(dx) \\ &= \alpha^{-1} \int_0^1 \left(\int_{\mathbb{R}} (1 - \cos(tx/K)) P_n(dx) \right) dt \\ &= \alpha^{-1} \int_0^1 (1 - \operatorname{Re}(\varphi_n(t/K))) dt. \end{aligned}$$

Wir erhalten nun (mit dem Satz von der majorisierten Konvergenz)

$$\begin{aligned} \limsup_{n \rightarrow \infty} P_n([-K, K]^c) &\leq \alpha^{-1} \limsup_{n \rightarrow \infty} \int_0^1 (1 - \operatorname{Re}(\varphi_n(t/K))) dt \\ &= \alpha^{-1} \int_0^1 \left(\lim_{n \rightarrow \infty} (1 - \operatorname{Re}(\varphi_n(t/K))) \right) dt \\ &= \alpha^{-1} \int_0^1 (1 - \operatorname{Re}(f(t/K))) dt. \end{aligned}$$

Da f stetig und $f(0) = 1$ ist, konvergiert das letzte Integral gegen 0, wenn $K \rightarrow \infty$. Also ist $(P_n)_{n \in \mathbb{N}}$ straff. \square

Eine einfache Anwendung des Lévy'schen Stetigkeitssatzes auf Beispiel 15.15 liefert den folgenden Satz von Pólya.

Satz 15.24 (Pólya). Sei $f : \mathbb{R} \rightarrow [0, 1]$ stetig und gerade mit $f(0) = 1$. Ferner sei f auf $[0, \infty)$ konvex. Dann ist f die charakteristische Funktion eines W-Maßes.

Beweis. Wir können f auf $[0, \infty)$ durch konvexe Polygonzüge f_n approximieren, indem wir $f_n(k/n) = f(k/n)$ setzen für $k = 0, \dots, n^2$ und f_n zwischen den Stützstellen linear interpolieren und rechts von n konstant fortsetzen. Für $x < 0$ setzen wir $f_n(x) = f_n(-x)$. Nach Beispiel 15.15 ist jedes f_n charakteristische Funktion eines W-Maßes μ_n . Offenbar gilt $f_n \xrightarrow{n \rightarrow \infty} f$ punktweise, also ist f die charakteristische Funktion eines W-Maßes $\mu = \lim_{n \rightarrow \infty} \mu_n$ auf \mathbb{R} . \square

Korollar 15.25. Für jedes $\alpha \in (0, 1]$ und $r > 0$ ist $\varphi_{\alpha,r}(t) = e^{-|r t|^\alpha}$ die charakteristische Funktion eines symmetrischen W-Maßes $\mu_{\alpha,r}$ auf \mathbb{R} .

Bemerkung 15.26. Tatsächlich ist $\varphi_{\alpha,r}$ auch für $\alpha \in (0, 2]$ eine charakteristische Funktion (für $\alpha = 2$ die der Normalverteilung), siehe Kapitel 16.2. Die Verteilungen $\mu_{\alpha,r}$ haben die Eigenschaft **α -stabil** zu sein (siehe Definition 16.20): Sind X_1, X_2, \dots, X_n unabhängig und $\mu_{\alpha,a}$ -verteilt, so ist $\varphi_{X_1+\dots+X_n}(t) = \varphi_X(t)^n = \varphi_X(n^{1/\alpha} t)$, also $X_1 + \dots + X_n \stackrel{\mathcal{D}}{=} n^{1/\alpha} X_1$. \diamond

Wir haben mit dem Satz von Stone-Weierstraß gesehen, dass charakteristische Funktionen Verteilungen eindeutig bestimmen. Der Satz von Pólya bietet eine hinreichende Bedingung dafür, dass eine symmetrische reelle Funktion eine charakteristische Funktion ist. Dass diese Bedingung nicht notwendig ist, sieht man schon daran, dass die charakteristische Funktion der Normalverteilung sie nicht erfüllt. Wir geben nun, gewissermaßen zur Allgemeinbildung und ohne Beweis, den Satz von Bochner an, der eine notwendige und hinreichende Bedingung dafür formuliert, dass eine Funktion $\varphi : \mathbb{R}^d \rightarrow \mathbb{C}$ die charakteristische Funktion eines W-Maßes ist.

Definition 15.27. Eine Funktion $f : \mathbb{R}^d \rightarrow \mathbb{C}$ heißt **positiv semidefinit**, falls für jedes $n \in \mathbb{N}$ und alle $t_1, \dots, t_n \in \mathbb{R}^d$ sowie $y_1, \dots, y_n \in \mathbb{C}$ gilt

$$\sum_{k,l=1}^n y_k \bar{y}_l f(t_k - t_l) \geq 0,$$

mit anderen Worten, falls die Matrix $(f(t_k - t_l))_{k,l=1,\dots,n}$ positiv semidefinit ist.

Lemma 15.28. Ist $\mu \in \mathcal{M}_f(\mathbb{R}^d)$ mit charakteristischer Funktion φ , so ist φ positiv semidefinit.

Beweis. Es gilt

$$\begin{aligned}
 \sum_{k,l=1}^n y_k \bar{y}_l \varphi(t_k - t_l) &= \sum_{k,l=1}^n y_k \bar{y}_l \int e^{ix(t_k - t_l)} \mu(dx) \\
 &= \int \sum_{k,l=1}^n y_k e^{ixt_k} \overline{y_l e^{ixt_l}} \mu(dx) \\
 &= \int \left| \sum_{k=1}^n y_k e^{ixt_k} \right|^2 \mu(dx) \geq 0. \quad \square
 \end{aligned}$$

Der folgende Satz geht im Falle $d = 1$ auf Bochner (1932) zurück.

Satz 15.29 (Bochner). Eine stetige Funktion $f : \mathbb{R}^d \rightarrow \mathbb{C}$ ist genau dann die charakteristische Funktion einer Wahrscheinlichkeitsverteilung auf \mathbb{R}^d , wenn f positiv semidefinit ist und $f(0) = 1$ gilt.

Die Aussage gilt ebenfalls, wenn wir \mathbb{R}^d durch eine lokalkompakte, abelsche Gruppe ersetzen.

Beweis. Für den Fall $d = 1$ siehe [22, §20, Satz 23] oder [59, Kapitel XIX.2, Seite 622]. Für den ganz allgemeinen Fall siehe etwa [77, Seite 293, Theorem 33.3]. \square

Übung 15.3.1. (Vergleiche [55] und [4].) Man zeige: Es gibt zwei austauschbare Folgen $X = (X_n)_{n \in \mathbb{N}}$ und $Y = (Y_n)_{n \in \mathbb{N}}$ reeller Zufallsvariablen mit $\mathbf{P}_X \neq \mathbf{P}_Y$, jedoch mit

$$\sum_{k=1}^n X_k \stackrel{\mathcal{D}}{=} \sum_{k=1}^n Y_k \quad \text{für jedes } n \in \mathbb{N}. \quad (15.4)$$

Anleitung:

- (i) Definiere die charakteristischen Funktionen (siehe Satz 15.12) $\varphi_1(t) = \frac{1}{1+t^2}$ und $\varphi_2(t) = (1-t/2)^+$. Zeige mit dem Satz von Pólya, dass

$$\psi_1(t) := \begin{cases} \varphi_1(t), & \text{falls } |t| \leq 1, \\ \varphi_2(t), & \text{falls } |t| > 1, \end{cases}$$

und

$$\psi_2(t) := \begin{cases} \varphi_2(t), & \text{falls } |t| \leq 1, \\ \varphi_1(t), & \text{falls } |t| > 1, \end{cases}$$

charakteristische Funktionen von Wahrscheinlichkeitsverteilungen auf \mathbb{R} sind.

- (ii) Definiere unabhängige Zufallsvariablen $X_{n,i}, Y_{n,i}, n \in \mathbb{N}, i = 1, 2$, und $\Theta_n, n \in \mathbb{N}$ mit: $X_{n,i}$ hat charakteristische Funktion φ_i , $Y_{n,i}$ hat charakteristische Funktion ψ_i und $\mathbf{P}[\Theta_n = 1] = \mathbf{P}[\Theta_n = -1] = \frac{1}{2}$. Setze $X_n = X_{n,\Theta_n}$ und $Y_n = Y_{n,\Theta_n}$. Zeige, dass (15.4) gilt.

- (iii) Bestimme $\mathbf{E}[e^{it_1 X_1 + it_2 X_2}]$ und $\mathbf{E}[e^{it_1 Y_1 + it_2 Y_2}]$ für $t_1 = \frac{1}{2}$ und $t_2 = 2$ und folgere, dass $(X_1, X_2) \not\stackrel{\mathcal{D}}{=} (Y_1, Y_2)$ und damit $\mathbf{P}_X \neq \mathbf{P}_Y$. ♣

Übung 15.3.2. Man zeige (wie im Beweis des Lévy'schen Stetigkeitssatzes): Für jedes $\delta > 0$ und $\varepsilon > 0$ gibt es ein $C = C(\varepsilon, \delta) < \infty$, so dass für jedes $\mu \in \mathcal{M}_1(\mathbb{R})$ mit charakteristischer Funktion φ gilt:

$$\mu([-\delta, \delta]^c) \leq C \int_0^\varepsilon (1 - \operatorname{Re}(\varphi(t))) dt.$$

Für $\varepsilon\delta \leq 3$ kann $C(\delta, \varepsilon) = 12/\varepsilon^3\delta^2$ gewählt werden. ♣

Übung 15.3.3. Sei $(\mu_n)_{n \in \mathbb{N}}$ eine Folge von W-Maßen auf \mathbb{R} mit charakteristischen Funktionen $(\varphi_n)_{n \in \mathbb{N}}$. Es gelte $\varphi_n(t) \xrightarrow{n \rightarrow \infty} 1$ für t in einer Umgebung von 0. Zeige mit Hilfe von Übung 15.3.2, dass $\mu_n \xrightarrow{n \rightarrow \infty} \delta_0$. ♣

15.4 Charakteristische Funktion und Momente

Wir wollen den Zusammenhang zwischen den Ableitungen der charakteristischen Funktion φ_X einer reellen Zufallsvariablen X und den Momenten von X untersuchen. Wir beginnen mit einem elementaren Lemma.

Lemma 15.30. Für $t \in \mathbb{R}$ und $n \in \mathbb{N}$ gilt $\left| e^{it} - 1 - \frac{it}{1!} - \dots - \frac{(it)^{n-1}}{(n-1)!} \right| \leq \frac{|t|^n}{n!}$.

Beweis. Dies folgt direkt aus der Taylorformel, da die n -te Ableitung von e^{it} dem Betrage nach 1 ist. □

Satz 15.31 (Momente und Differenzierbarkeit). Sei X eine reelle Zufallsvariable mit charakteristischer Funktion φ .

(i) Ist $\mathbf{E}[|X|^n] < \infty$, so ist φ n -mal stetig differenzierbar mit Ableitungen

$$\varphi^{(k)}(t) = \mathbf{E}[(iX)^k e^{itX}] \quad \text{für jedes } k = 0, \dots, n.$$

(ii) Ist speziell $\mathbf{E}[X^2] < \infty$, so ist

$$\varphi(t) = 1 + it\mathbf{E}[X] - \frac{1}{2}t^2\mathbf{E}[X^2] + \varepsilon(t)t^2$$

mit $\varepsilon(t) \rightarrow 0$ für $t \rightarrow 0$.

(iii) Sei $h \in \mathbb{R}$. Gilt $\lim_{n \rightarrow \infty} \frac{|h|^n \mathbf{E}[|X|^n]}{n!} = 0$, so ist für jedes $t \in \mathbb{R}$

$$\varphi(t+h) = \sum_{k=0}^{\infty} \frac{(ih)^k}{k!} \mathbf{E}[e^{itX} X^k].$$

Speziell gilt dies, falls $\mathbf{E}[e^{|hX|}] < \infty$.

Beweis. (i) Für $t \in \mathbb{R}$, $h \in \mathbb{R} \setminus \{0\}$ und $k \in \{1, \dots, n\}$ sei

$$Y_k(t, h, x) = k! h^{-k} e^{itx} \left(e^{ihx} - \sum_{l=0}^{k-1} \frac{(ihx)^l}{l!} \right).$$

Dann ist

$$\mathbf{E}[Y_k(t, h, X)] = k! h^{-k} \left(\varphi(t+h) - \varphi(t) - \sum_{l=1}^{k-1} \mathbf{E}[e^{itX} (iX)^l] \frac{h^l}{l!} \right).$$

Existiert nun der Limes $\varphi_k(t) := \lim_{h \rightarrow 0} \mathbf{E}[Y_k(t, h, X)]$, so ist φ k -mal differenzierbar in t mit $\varphi^{(k)}(t) = \varphi_k(t)$.

Es gilt aber (nach Lemma 15.30 mit $n = k+1$) $Y_k(t, h, x) \xrightarrow{h \rightarrow 0} (ix)^k e^{itx}$ für jedes $x \in \mathbb{R}$ und (nach Lemma 15.30 mit $n = k$) $|Y_k(t, h, x)| \leq |x|^k$. Da nach Voraussetzung $\mathbf{E}[|X|^k] < \infty$ gilt, folgt mit dem Satz über majorisierte Konvergenz, dass $\mathbf{E}[Y_k(t, h, X)] \xrightarrow{h \rightarrow 0} \mathbf{E}[(iX)^k e^{itX}] = \varphi^{(k)}(t)$. Eine einfache Anwendung des Stetigkeitslemmas (Satz 6.27) liefert die Stetigkeit von $\varphi^{(k)}$.

(ii) Dies folgt direkt aus (i).

(iii) Nach Voraussetzung gilt

$$\begin{aligned} \left| \varphi(t+h) - \sum_{k=0}^{n-1} \frac{(ih)^k}{k!} \mathbf{E}[e^{itX} X^k] \right| &= \frac{h^n}{n!} |\mathbf{E}[Y_n(t, h, X)]| \\ &\leq \frac{h^n \mathbf{E}[|X|^n]}{n!} \xrightarrow{n \rightarrow \infty} 0. \end{aligned} \quad \square$$

Korollar 15.32 (Momentenproblem). Sei X eine reelle Zufallsvariable mit

$$\alpha := \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbf{E}[|X|^n]^{1/n} < \infty.$$

Dann ist die charakteristische Funktion φ von X analytisch, und die Verteilung von X ist durch die Angabe der Momente $\mathbf{E}[X^n]$, $n \in \mathbb{N}$, eindeutig bestimmt. Speziell gilt dies, falls $\mathbf{E}[e^{t|X|}] < \infty$ ist für ein $t > 0$.

Beweis. Nach der Stirling'schen Formel ist $\lim_{n \rightarrow \infty} \frac{1}{n!} n^n e^{-n} \sqrt{2\pi n} = 1$. Für $|h| < 1/(3\alpha)$ gilt daher

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbf{E}[|X|^n] \cdot |h|^n / n! &= \limsup_{n \rightarrow \infty} \sqrt{2\pi n} \left(\mathbf{E}[|X|^n]^{1/n} \cdot |h| \cdot e/n \right)^n \\ &\leq \limsup_{n \rightarrow \infty} \sqrt{2\pi n} (e/3)^n = 0. \end{aligned}$$

Die charakteristische Funktion ist also um jeden Punkt $t \in \mathbb{R}$ in eine Potenzreihe entwickelbar mit Konvergenzradius mindestens $1/(3\alpha)$, ist insbesondere also analytisch. Damit ist sie festgelegt durch die Koeffizienten der Potenzreihe um $t = 0$, also durch die Momente von X . \square

Beispiele 15.33. (i) Sei $X \sim \mathcal{N}_{\mu, \sigma^2}$. Dann ist für jedes $t \in \mathbb{R}$

$$\begin{aligned} \mathbf{E}[e^{tX}] &= (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} e^{tx} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= e^{\mu t + t^2\sigma^2/2} (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} e^{-(x-\mu-t\sigma^2)^2/2\sigma^2} dx \\ &= e^{\mu t + t^2\sigma^2/2} < \infty. \end{aligned}$$

Also ist die Verteilung von X durch Angabe aller Momente komplett bestimmt. Die charakteristische Funktion $\varphi(t) = e^{i\mu t} e^{-\sigma^2 t^2/2}$, die wir durch die obige Rechnung mit it statt t erhalten, ist in der Tat analytisch.

(ii) Sei X exponentialverteilt mit Parameter $\theta > 0$. Dann ist für $t \in (0, \theta)$

$$\mathbf{E}[e^{tX}] = \theta \int_0^{\infty} e^{tx} e^{-\theta x} dx = \frac{\theta}{\theta - t} < \infty.$$

Also ist die Verteilung von X durch Angabe aller Momente bestimmt. Die selbe Rechnung mit it statt t liefert $\varphi(t) = \theta/(\theta - it)$, und diese Funktion ist in der Tat

analytisch. Der Umstand, dass φ im Komplexen eine Singularität bei $t = -i\theta$ hat, impliziert, dass die Potenzreihe von φ um 0 den Konvergenzradius θ hat. Insbesondere folgt hieraus, dass nicht alle exponentiellen Momente existieren können. Dies wird reflektiert durch die obige Rechnung, die zeigt, dass für $t \geq \theta$ keine exponentiellen Momente existieren.

(iii) Sei X log-normalverteilt (siehe Beispiel 15.5). Dann ist $\mathbf{E}[X^n] = e^{n^2/2}$. Speziell ist in diesem Fall $\alpha = \infty$. Tatsächlich hatten wir in Beispiel 15.5 gesehen, dass die Momente in diesem Fall nicht die Verteilung von X bestimmen.

(iv) Hat X Werte in \mathbb{N}_0 und gilt $\beta := \limsup_{n \rightarrow \infty} \mathbf{E}[X^n]^{1/n} < 1$, so gilt nach den Hadamard-Kriterium $\psi_X(z) := \sum_{k=1}^{\infty} \mathbf{P}[X = k] z^k < \infty$ für $|z| < 1/\beta$. Speziell ist die Erzeugendenfunktion von X durch die Ableitungen $\psi_X^{(n)}(1)$, $n \in \mathbb{N}$, und damit durch die Momente von X eindeutig festgelegt. Vergleiche Satz 3.2(iii). \diamond

Satz 15.34. Sei X eine reelle Zufallsvariable und φ die charakteristische Funktion von X . Sei $n \in \mathbb{N}$, und φ sei $2n$ -mal differenzierbar in 0 mit Ableitung $\varphi^{(2n)}(0)$. Dann gilt $\mathbf{E}[X^{2n}] = (-1)^n \varphi^{(2n)}(0) < \infty$.

Beweis. Wir führen den Beweis per Induktion nach $n \in \mathbb{N}_0$. Für $n = 0$ ist die Aussage trivialerweise richtig. Sei nun $n \in \mathbb{N}$, und φ sei $2n$ -mal differenzierbar in 0. Wir setzen $u(t) = \operatorname{Re}(\varphi(t))$. Dann ist u ebenfalls $2n$ -mal differenzierbar in 0 und $u^{(2k-1)}(0) = 0$ für $k = 1, \dots, n$, weil u gerade ist. Da $\varphi^{(2n)}(0)$ existiert, ist $\varphi^{(2n-1)}$ stetig in 0 und $\varphi^{(2n-1)}(t)$ existiert für $t \in (-\varepsilon, \varepsilon)$ für gewisses $\varepsilon > 0$. Ferner existiert dann $\varphi^{(k)}$ in $(-\varepsilon, \varepsilon)$ und ist dort stetig für jedes $k = 0, \dots, 2n-2$. Nach der Taylorformel gilt also für jedes $t \in (-\varepsilon, \varepsilon)$

$$\left| u(t) - \sum_{k=0}^{n-1} u^{(2k)}(0) \frac{t^{2k}}{(2k)!} \right| \leq \frac{|t|^{2n-1}}{(2n-1)!} \sup_{\theta \in (0,1]} |u^{(2n-1)}(\theta t)|. \quad (15.5)$$

Wir definieren eine stetige Funktion $f_n : \mathbb{R} \rightarrow [0, \infty)$ durch $f_n(0) = 1$ und für $x \neq 0$

$$f_n(x) = (-1)^n (2n)! x^{-2n} \left[\cos(x) - \sum_{k=0}^{n-1} (-1)^k \frac{x^{2k}}{(2k)!} \right].$$

Nach Induktionsvoraussetzung ist $\mathbf{E}[X^{2k}] = u^{(2k)}(0)$ für jedes $k = 1, \dots, n-1$. Es folgt mit (15.5)

$$\mathbf{E}[f_n(tX) X^{2n}] \leq \frac{2n}{|t|} \sup_{\theta \in (0,1]} |u^{(2n-1)}(\theta t)| \leq g_n(t) := 2n \sup_{\theta \in (0,1]} \frac{|u^{(2n-1)}(\theta t)|}{\theta |t|}.$$

Mit dem Lemma von Fatou folgt

$$\begin{aligned} \mathbf{E}[X^{2n}] &= \mathbf{E}[f_n(0) X^{2n}] \leq \liminf_{t \rightarrow 0} \mathbf{E}[f_n(tX) X^{2n}] \\ &\leq \liminf_{t \rightarrow 0} g_n(t) = 2n |u^{(2n)}(0)| < \infty. \end{aligned}$$

Nach Satz 15.31 folgt hieraus $\mathbf{E}[X^{2n}] = (-1)^n u^{(2n)}(0) = (-1)^n \varphi^{(2n)}(0)$. \square

Bemerkung 15.35. Für ungerade Momente allgemeiner Zufallsvariablen gilt die Aussage des Satzes nicht (siehe etwa Übung 15.4.4 für das erste Moment). In der Tat ist φ in 0 genau dann differenzierbar mit Ableitung $i m$ für ein $m \in \mathbb{R}$, wenn $x \mathbf{P}[|X| > x] \xrightarrow{x \rightarrow \infty} 0$ und $\mathbf{E}[X \mathbf{1}_{\{|X| \leq x\}}] \xrightarrow{x \rightarrow \infty} m$. (Siehe [59, Kapitel XVII.2a, Seite 565].) \diamond

Übung 15.4.1. Es seien X und Y nichtnegative Zufallsvariablen mit

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbf{E}[|X|^n]^{1/n} < \infty, \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbf{E}[|Y|^n]^{1/n} < \infty,$$

und

$$\mathbf{E}[X^m Y^n] = \mathbf{E}[X^m] \mathbf{E}[Y^n] \quad \text{für alle } m, n \in \mathbb{N}_0.$$

Man zeige: X und Y sind unabhängig.

Hinweis: Verwende Korollar 15.32 für die Zufallsvariable Y bezüglich des Wahrscheinlichkeitsmaßes $X^m \mathbf{P}[\cdot] / \mathbf{E}[X^m]$ um zu zeigen, dass

$$\mathbf{E}[X^m \mathbf{1}_A(Y)] / \mathbf{E}[X^m] = \mathbf{P}[Y \in A] \quad \text{für jedes } A \in \mathcal{B}(R) \text{ und } m \in \mathbb{N}_0.$$

Verwende nun Korollar 15.32 für X bezüglich des W-Maßes $\mathbf{P}[\cdot | Y \in A]$. \clubsuit

Übung 15.4.2. Seien $r, s > 0$, und seien $Z \sim \Gamma_{1,r+s}$ und $B \sim \beta_{r,s}$ unabhängig (siehe Beispiel 1.107). Man zeige mit Hilfe von Übung 15.4.1: Die Zufallsvariablen $X := BZ$ und $Y := (1 - B)Z$ sind unabhängig mit $X \sim \Gamma_{1,r}$ und $Y \sim \Gamma_{1,s}$. \clubsuit

Übung 15.4.3. Man zeige, dass für $\alpha > 2$ die Funktion $\phi_\alpha(t) = e^{-|t|^\alpha}$ keine charakteristische Funktion ist.

(*Hinweis:* Man nehme das Gegenteil an und zeige, dass dann die zugehörige Verteilung verschwindende Varianz hätte.) \clubsuit

Übung 15.4.4. Seien X_1, X_2, \dots u.i.v. reelle Zufallsvariablen mit charakteristischer Funktion φ . Man zeige:

- (i) Ist φ differenzierbar in 0, so ist $\varphi'(0) = i m$ für ein $m \in \mathbb{R}$.
- (ii) φ ist differenzierbar in 0 mit $\varphi'(0) = i m$ genau dann, wenn $(X_1 + \dots + X_n)/n \xrightarrow{n \rightarrow \infty} m$ stochastisch.
- (iii) Sei φ ist differenzierbar in 0 und $X_1 \geq 0$ fast sicher. Dann gilt $\mathbf{E}[X_1] = -i \varphi'(0) < \infty$. *Hinweis:* Verwende (ii) und das Gesetz der großen Zahl.
- (iv) Die Verteilung von X_1 kann so gewählt werden, dass φ differenzierbar in 0 ist, aber $\mathbf{E}[|X_1|] = \infty$. \clubsuit

Übung 15.4.5. Seien X_1, X_2, \dots reelle Zufallsvariablen. Für $r > 0$ bezeichne $M_r(X_n) = \mathbf{E}[|X_n|^r]$ das absolute r -te Moment. Für $k \in \mathbb{N}$ bezeichne $m_k(X_n) = \mathbf{E}[X_n^k]$ das k -te Moment, falls $M_k(X_n) < \infty$.

- (i) Sei X eine reelle Zufallsvariable und $(X_{n_l})_{l \in \mathbb{N}}$ eine Teilfolge mit

$$\mathbf{P}_{X_{n_l}} \xrightarrow{l \rightarrow \infty} \mathbf{P}_X \text{ schwach.}$$

Es gebe es ein $r > 0$ mit $\sup_{n \in \mathbb{N}} M_r(X_n) < \infty$. Zeige: Für jedes $k \in \mathbb{N} \cap (0, r)$ und jedes $s \in (0, r)$ gilt $M_s(X) < \infty$ sowie

$$M_s(X_{n_l}) \xrightarrow{l \rightarrow \infty} M_s(X) \quad \text{und} \quad m_k(X_{n_l}) \xrightarrow{l \rightarrow \infty} m_k(X).$$

- (ii) Für jedes $k \in \mathbb{N}$ existiere der Grenzwert $m_k := \lim_{n \rightarrow \infty} m_k(X_n)$ und sei endlich (dabei dürfen für jedes k endlich viele der $m_k(X_n)$ undefiniert sein.) Zeige: Es existiert eine reelle Zufallsvariable X mit $m_k(X) = m_k$ für jedes $k \in \mathbb{N}$ sowie eine Teilfolge $(X_{n_l})_{l \in \mathbb{N}}$ mit

$$\mathbf{P}_{X_{n_l}} \xrightarrow{l \rightarrow \infty} \mathbf{P}_X \text{ schwach.}$$

- (iii) Zeige den Satz von Fréchet-Shohat: Ist in (ii) die Verteilung von X durch die Momente $m_k(X) = m_k$, $k \in \mathbb{N}$, eindeutig bestimmt (siehe Korollar 15.32), so gilt

$$\mathbf{P}_{X_n} \xrightarrow{n \rightarrow \infty} \mathbf{P}_X \text{ schwach.} \quad \clubsuit$$

Übung 15.4.6. Seien X_1, X_2, \dots u.i.v. reelle Zufallsvariablen mit $\mathbf{E}[X_1] = 0$ und $\mathbf{E}[|X_1|^k] < \infty$ für jedes $k \in \mathbb{N}$.

- (i) Zeige: Es gibt (von der Verteilung \mathbf{P}_{X_1} abhängige) endliche Zahlen $(d_k)_{k \in \mathbb{N}}$, sodass für jedes $k, n \in \mathbb{N}$ gilt:

$$|\mathbf{E}[(X_1 + \dots + X_n)^{2k-1}]| \leq d_{2k-1} n^{k-1}$$

und

$$\left| \mathbf{E}[(X_1 + \dots + X_n)^{2k}] - \frac{(2k)!}{2^k k!} \mathbf{E}[X_1^2]^k n^k \right| \leq d_{2k} n^{k-1}.$$

Hinweis: Man multipliziere die Klammer aus, sortiere nach den unterschiedlichen gemischten Momenten und bestimme die Anzahl der jeweiligen Summanden. Eine besondere Bedeutung kommt der Anzahl der Summanden vom Typ $\mathbf{E}[X_{l_1}^2 \cdots X_{l_k}^2]$ für unterschiedliche l_1, \dots, l_k zu.

- (ii) Sei $Y \sim \mathcal{N}_{0,1}$. Zeige mit Hilfe von Satz 15.31(i): Für jedes $k \in \mathbb{N}$ gilt

$$\mathbf{E}[Y^{2k-1}] = 0 \quad \text{und} \quad \mathbf{E}[Y^{2k}] = \frac{(2k)!}{2^k k!}.$$

- (iii) Sei $\mathbf{Var}[X_1] > 0$ und $S_n^* = (X_1 + \dots + X_n)/\sqrt{n \mathbf{Var}[X_1]}$. Folgere mit Hilfe von Aufgabe 15.4.5 die Aussage des zentralen Grenzwertsatzes (vergleiche Satz 15.37)

$$\mathbf{P}_{S_n^*} \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0,1} \text{ schwach.} \quad \clubsuit$$

15.5 Der Zentrale Grenzwertsatz

Während wir im starken Gesetz der großen Zahl gesehen haben, dass Summen $S_n = X_1 + \dots + X_n$ u.i.v. integrierbarer Zufallsvariablen Werte in etwa von der Größe $n \cdot \mathbf{E}[X_1]$ annehmen, wollen wir jetzt anschauen, wie groß und von welcher Form die *typischen Abweichungen* von diesem Wert sind – jedenfalls unter der zusätzlichen Annahme, dass $\mathbf{Var}[X_1] \in (0, \infty)$ gilt.

Wir bereiten den Beweis des zentralen Grenzwertsatzes mit einem Lemma vor.

Lemma 15.36. Seien X_1, X_2, \dots u.i.v. reelle Zufallsvariablen mit $\mathbf{E}[X_1] = \mu$ und $\mathbf{Var}[X_1] = \sigma^2 \in (0, \infty)$. Sei

$$S_n^* := \frac{1}{\sqrt{n\sigma^2}} \sum_{k=1}^n (X_k - \mu)$$

die standardisierte n -te Partialsumme. Dann gilt

$$\lim_{n \rightarrow \infty} \varphi_{S_n^*}(t) = e^{-t^2/2} \quad \text{für jedes } t \in \mathbb{R}.$$

Beweis. Sei $\varphi = \varphi_{X_k - \mu}$. Dann ist nach Satz 15.31(ii)

$$\varphi(t) = 1 - \frac{\sigma^2}{2} t^2 + \varepsilon(t) t^2,$$

wobei $\varepsilon(t) \rightarrow 0$, wenn $t \rightarrow 0$. Nach Lemma 15.11(iv) und (ii) ist

$$\varphi_{S_n^*}(t) = \varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right)^n.$$

Nun ist $\left(1 - \frac{t^2}{2n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-t^2/2}$ und

$$\begin{aligned} \left| \left(1 - \frac{t^2}{2n}\right)^n - \varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right)^n \right| &\leq n \left| 1 - \frac{t^2}{2n} - \varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right) \right| \\ &\leq n \frac{t^2}{n\sigma^2} \left| \varepsilon\left(\frac{t}{\sqrt{n\sigma^2}}\right) \right| \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

(Beachte: $|u^n - v^n| \leq |u - v| \cdot n \cdot \max(|u|, |v|)^{n-1}$ für alle $u, v \in \mathbb{C}$.) □

Satz 15.37 (Zentraler Grenzwertsatz). Seien X_1, X_2, \dots u.i.v. reelle Zufallsvariablen mit $\mu := \mathbf{E}[X_1] \in \mathbb{R}$ und $\sigma^2 := \mathbf{Var}[X_1] \in (0, \infty)$. Für $n \in \mathbb{N}$ sei $S_n^* := \frac{1}{\sqrt{\sigma^2 n}} \sum_{i=1}^n (X_i - \mu)$. Dann gilt

$$\mathbf{P}_{S_n^*} \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0,1} \text{ schwach.}$$

Für $-\infty \leq a < b \leq +\infty$ gilt $\lim_{n \rightarrow \infty} \mathbf{P}[S_n^* \in [a, b]] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$.

Beweis. Nach Lemma 15.36 und dem Lévy'schen Stetigkeitssatz (Satz 15.23) konvergiert $\mathbf{P}_{S_n^*}$ gegen die Verteilung mit charakteristischer Funktion $\varphi(t) = e^{-t^2/2}$. Nach Satz 15.12(i) ist dies $\mathcal{N}_{0,1}$. Der Zusatz folgt mit dem Portemanteau Theorem (Satz 13.16), weil $\mathcal{N}_{0,1}$ eine Dichte hat, also $\mathcal{N}_{0,1}(\partial[a, b]) = 0$ gilt. \square

Bemerkung 15.38. Man kann ohne Benutzung des Stetigkeitssatzes auch so argumentieren: Für jedes $K > 0$ und $n \in \mathbb{N}$ ist $\mathbf{P}[|S_n^*| > K] \leq \mathbf{Var}[S_n^*]/K^2 = 1/K^2$, also ist die Folge $\mathbf{P}_{S_n^*}$ straff. Da die charakteristischen Funktionen verteilungsbestimmend sind, folgt die Aussage mit Satz 13.34. \diamond

Wir wollen uns nun von der Annahme von Satz 15.37 lösen, dass die Zufallsvariablen identisch verteilt sind. Tatsächlich können wir sogar Partialsummen bilden, die jeweils ganz unterschiedliche zentrierte Zufallsvariablen aufsummieren. Entscheidend ist, dass die Varianz der normierten Summe 1 ist, und dass jeder einzelne Summanden nur einen kleinen Beitrag liefert.

Definition 15.39. Für jedes $n \in \mathbb{N}$ sei $k_n \in \mathbb{N}$ und seien $X_{n,1}, \dots, X_{n,k_n}$ reelle Zufallsvariablen. Wir nennen $(X_{n,l}) = (X_{n,l}, l = 1, \dots, k_n, n \in \mathbb{N})$ ein **Schema von Zufallsvariablen**. Wir definieren stets $S_n = X_{n,1} + \dots + X_{n,k_n}$ als die Zeilensumme. Das Schema heißt

- **unabhängig**, falls für jedes $n \in \mathbb{N}$ die Familie $(X_{n,l})_{l=1, \dots, k_n}$ unabhängig ist,
- **zentriert**, falls $X_{n,l} \in \mathcal{L}^1(\mathbf{P})$ und $\mathbf{E}[X_{n,l}] = 0$ ist für jedes n und l ,
- **normiert**, falls $X_{n,l} \in \mathcal{L}^2(\mathbf{P})$ und $\sum_{l=1}^{k_n} \mathbf{Var}[X_{n,l}] = 1$ ist für jedes $n \in \mathbb{N}$.

Ein zentriertes Schema heißt **asymptotisch vernachlässigbar**, falls für jedes $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \max_{1 \leq l \leq k_n} \mathbf{P}[|X_{n,l}| > \varepsilon] = 0.$$

Definition 15.40. Ein zentriertes Schema $(X_{n,l})$ mit $X_{n,l} \in \mathcal{L}^2(\mathbf{P})$ für jedes $n \in \mathbb{N}$ und $l = 1, \dots, k_n$ erfüllt die **Lindeberg-Bedingung**, falls für jedes $\varepsilon > 0$ gilt, dass

$$L_n(\varepsilon) := \frac{1}{\mathbf{Var}[S_n]} \sum_{l=1}^{k_n} \mathbf{E}\left[X_{n,l}^2 \mathbb{1}_{\{X_{n,l}^2 > \varepsilon^2 \mathbf{Var}[S_n]\}}\right] \xrightarrow{n \rightarrow \infty} 0. \quad (15.6)$$

Das Schema erfüllt die **Lyapunov-Bedingung**, falls für ein $\delta > 0$ gilt

$$\lim_{n \rightarrow \infty} \frac{1}{\mathbf{Var}[S_n]^{1+(\delta/2)}} \sum_{l=1}^{k_n} \mathbf{E}[|X_{n,l}|^{2+\delta}] = 0. \quad (15.7)$$

Lemma 15.41. Die Lyapunov-Bedingung impliziert die Lindeberg-Bedingung.

Beweis. Für $x \in \mathbb{R}$ ist $x^2 \mathbb{1}_{\{|x|>\varepsilon'\}} \leq (\varepsilon')^{-\delta} |x|^{2+\delta} \mathbb{1}_{\{|x|>\varepsilon'\}} \leq (\varepsilon')^{-\delta} |x|^{2+\delta}$. Mit $\varepsilon' := \varepsilon \sqrt{\text{Var}[S_n]}$ folgt $L_n(\varepsilon) \leq \varepsilon^{-\delta} \frac{1}{\text{Var}[S_n]^{1+(\delta/2)}} \sum_{l=1}^{k_n} \mathbf{E}[|X_{n,l}|^{2+\delta}]$. \square

Beispiel 15.42. Seien $(Y_n)_{n \in \mathbb{N}}$ u.i.v. mit $\mathbf{E}[Y_n] = 0$ und $\text{Var}[Y_n] = 1$. Sei $k_n = n$ und $X_{n,l} = \frac{Y_l}{\sqrt{n}}$. Dann ist $(X_{n,l})$ unabhängig, zentriert und normiert. Es gilt $\mathbf{P}[|X_{n,l}| > \varepsilon] = \mathbf{P}[|Y_1| > \varepsilon\sqrt{n}] \xrightarrow{n \rightarrow \infty} 0$, also ist $(X_{n,l})$ asymptotisch vernachlässigbar. Es gilt $L_n(\varepsilon) = \mathbf{E}[Y_1^2 \mathbb{1}_{\{|Y_1|>\sqrt{\varepsilon n}\}}] \xrightarrow{n \rightarrow \infty} 0$, also erfüllt $(X_{n,l})$ die Lindeberg-Bedingung.

Gilt $Y_1 \in \mathcal{L}^{2+\delta}(\mathbf{P})$ für ein $\delta > 0$, so ist

$$\sum_{l=1}^n \mathbf{E}[|X_{n,l}|^{2+\delta}] = n^{-(\delta/2)} \mathbf{E}[|Y_1|^{2+\delta}] \xrightarrow{n \rightarrow \infty} 0.$$

In diesem Fall erfüllt $(X_{n,l})$ auch die Lyapunov Bedingung. \diamond

Der folgende Satz geht auf Lindeberg (1922) für die Richtung (i) \Rightarrow (ii) und Feller (1935 und 1937) für die Richtung (ii) \Rightarrow (i) zurück. In den Anwendungen interessiert meist nur die Richtung von Lindeberg (i) \Rightarrow (ii), daher beweisen wir nur diesen Teil. Für die Richtung (ii) \Rightarrow (i) siehe etwa [164, Theorem III.4.3].

Satz 15.43 (Zentraler Grenzwertsatz von Lindeberg-Feller). Sei $(X_{n,l})$ ein unabhängiges, zentriertes und normiertes Schema reeller Zufallsvariablen sowie $S_n = X_{n,1} + \dots + X_{n,k_n}$ für jedes $n \in \mathbb{N}$. Dann sind äquivalent:

- (i) Es gilt die Lindeberg-Bedingung.
- (ii) $(X_{n,l})$ ist asymptotisch vernachlässigbar, und es gilt $\mathbf{P}_{S_n} \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0,1}$.

Wir bereiten den Beweis des Satzes von Lindeberg mit ein paar Lemmata vor.

Lemma 15.44. Gilt (i) in Satz 15.43, so ist $(X_{n,l})$ asymptotisch vernachlässigbar.

Beweis. Für $\varepsilon > 0$ ist nach der Chebyshev'schen Ungleichung

$$\sum_{l=1}^{k_n} \mathbf{P}[|X_{n,l}| > \varepsilon] \leq \varepsilon^{-2} \sum_{l=1}^{k_n} \mathbf{E}[X_{n,l}^2 \mathbb{1}_{\{|X_{n,l}|>\varepsilon\}}] = L_n(\varepsilon) \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Seien im Folgenden stets $\varphi_{n,l}$ und φ_n die charakteristischen Funktionen von $X_{n,l}$ und S_n .

Lemma 15.45. Für jedes $n \in \mathbb{N}$ und $t \in \mathbb{R}$ gilt $\sum_{l=1}^{k_n} |1 - \varphi_{n,l}(t)| \leq \frac{t^2}{2}$.

Beweis. Für jedes $x \in \mathbb{R}$ ist $|e^{itx} - 1 - itx| \leq \frac{t^2 x^2}{2}$. Wegen $\mathbf{E}[X_{n,l}] = 0$ ist

$$\begin{aligned} \sum_{l=1}^{k_n} |\varphi_{n,l}(t) - 1| &= \sum_{l=1}^{k_n} |\mathbf{E}[e^{itX_{n,l}} - 1]| \\ &\leq \sum_{l=1}^{k_n} \mathbf{E}[|e^{itX_{n,l}} - itX_{n,l} - 1|] + |\mathbf{E}[itX_{n,l}]| \\ &\leq \sum_{l=1}^{k_n} \frac{t^2}{2} \mathbf{E}[X_{n,l}^2] = \frac{t^2}{2}. \end{aligned} \quad \square$$

Lemma 15.46. Gilt (i) in Satz 15.43, so ist $\lim_{n \rightarrow \infty} \left| \log \varphi_n(t) - \sum_{l=1}^{k_n} \mathbf{E}[e^{itX_{n,l}} - 1] \right| = 0$.

Beweis. Setze $m_n := \max_{l=1,\dots,k_n} |\varphi_{n,l}(t) - 1|$. Beachte, dass für jedes $\varepsilon > 0$ gilt:

$$|e^{itx} - 1| \leq \begin{cases} 2x^2/\varepsilon^2, & \text{falls } |x| > \varepsilon, \\ \varepsilon t, & \text{falls } |x| \leq \varepsilon. \end{cases}$$

Hieraus folgt

$$\begin{aligned} |\varphi_{n,l}(t) - 1| &\leq \mathbf{E}\left[|e^{itX_{n,l}} - 1| \mathbb{1}_{\{|X_{n,l}| \leq \varepsilon\}}\right] + \mathbf{E}\left[|e^{itX_{n,l}} - 1| \mathbb{1}_{\{|X_{n,l}| > \varepsilon\}}\right] \\ &\leq \varepsilon t + 2\varepsilon^{-2} \mathbf{E}\left[X_{n,l}^2 \mathbb{1}_{\{|X_{n,l}| > \varepsilon\}}\right]. \end{aligned}$$

Also ist für jedes $\varepsilon > 0$

$$\limsup_{n \rightarrow \infty} m_n \leq \limsup_{n \rightarrow \infty} (\varepsilon t + 2\varepsilon^{-2} L_n(\varepsilon)) = \varepsilon t,$$

und damit $\lim_{n \rightarrow \infty} m_n = 0$. Nun ist für $x \in \mathbb{C}$ mit $|x| \leq \frac{1}{2}$ stets $|\log(1+x) - x| \leq x^2$.

Ist n groß genug, sodass $m_n < \frac{1}{2}$, dann ist

$$\begin{aligned}
\left| \log \varphi_n(t) - \sum_{l=1}^{k_n} \mathbf{E}[e^{itX_{n,l}} - 1] \right| &= \left| \sum_{l=1}^{k_n} \log(\varphi_{n,l}(t)) - \mathbf{E}[e^{itX_{n,l}} - 1] \right| \\
&\leq \sum_{l=1}^{k_n} (\varphi_{n,l}(t) - 1)^2 \\
&\leq m_n \sum_{l=1}^{k_n} |\varphi_{n,l}(t) - 1| \\
&\leq \frac{1}{2} m_n t^2 \quad (\text{nach Lemma 15.45}) \\
&\longrightarrow 0 \quad \text{für } n \rightarrow \infty. \quad \square
\end{aligned}$$

Der eigentliche Trick besteht in der Einführung der Funktion

$$f_t(x) := \begin{cases} \frac{1+x^2}{x^2} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right), & \text{falls } x \neq 0, \\ -\frac{t^2}{2}, & \text{falls } x = 0, \end{cases} \quad (15.8)$$

sowie der Maße $\mu_n, \nu_n \in \mathcal{M}_f(\mathbb{R})$, $n \in \mathbb{N}$,

$$\nu_n(dx) := \sum_{l=1}^{k_n} x^2 \mathbf{P}_{X_{n,l}}(dx) \quad \text{und} \quad \mu_n(dx) := \sum_{l=1}^{k_n} \frac{x^2}{1+x^2} \mathbf{P}_{X_{n,l}}(dx).$$

Lemma 15.47. Für jedes $t \in \mathbb{R}$ gilt $f_t \in C_b(\mathbb{R})$.

Beweis. Für jedes $|x| \geq 1$ ist $\frac{1+x^2}{x^2} \leq 2$, also gilt

$$|f_t(x)| \leq 2 \left(|e^{itx}| + 1 + \frac{tx}{1+x^2} \right) \leq 4 + 2|t|.$$

Wir müssen zeigen, dass f_t stetig in 0 ist. Die Taylorformel (Lemma 15.30) liefert

$$e^{itx} = 1 + itx - \frac{t^2 x^2}{2} + R(tx)$$

mit $|R(tx)| \leq \frac{1}{6}|tx|^3$. Also ist für festes t

$$\lim_{0 \neq x \rightarrow 0} f_t(x) = \lim_{0 \neq x \rightarrow 0} \frac{1}{x^2} \left(itx \left(1 - \frac{1}{1+x^2} \right) - \frac{t^2 x^2}{2} + R(tx) \right) = -\frac{t^2}{2}. \quad \square$$

Lemma 15.48. Gilt (i) in Satz 15.43, so gilt $\nu_n \xrightarrow{n \rightarrow \infty} \delta_0$ schwach.

Beweis. Für jedes $n \in \mathbb{N}$ ist $\nu_n \in \mathcal{M}_1(\mathbb{R})$, denn

$$\nu_n(\mathbb{R}) = \sum_{l=1}^{k_n} \int x^2 \mathbf{P}_{X_{n,l}}(dx) = \sum_{l=1}^{k_n} \mathbf{Var}[X_{n,l}] = 1.$$

Für $\varepsilon > 0$ gilt aber $\nu_n((-\varepsilon, \varepsilon)^c) = L_n(\varepsilon) \xrightarrow{n \rightarrow \infty} 0$, also $\nu_n \xrightarrow{n \rightarrow \infty} \delta_0$. \square

Lemma 15.49. Gilt (i) in Satz 15.43, so gilt

$$\int f_t(x) \mu_n(dx) + it \int \frac{1}{x} \mu_n(dx) \xrightarrow{n \rightarrow \infty} -\frac{t^2}{2}.$$

Beweis. Wegen $(x \mapsto f_t(x)/(1+x^2)) \in C_b(\mathbb{R})$ ist nach Lemma 15.48

$$\int f_t(x) \mu_n(dx) = \int f_t(x) \frac{1}{1+x^2} \nu_n(dx) \xrightarrow{n \rightarrow \infty} f_t(0) = -\frac{t^2}{2}.$$

Nun ist $(x \mapsto x/(1+x^2)) \in C_b(\mathbb{R})$ und $\mathbf{E}[X_{n,l}] = 0$ für jedes n und l , also

$$\begin{aligned} \int \frac{1}{x} \mu_n(dx) &= \sum_{l=1}^{k_n} \mathbf{E} \left[\frac{X_{n,l}}{1+X_{n,l}^2} \right] = \sum_{l=1}^{k_n} \mathbf{E} \left[\frac{X_{n,l}}{1+X_{n,l}^2} - X_{n,l} \right] \\ &= - \sum_{l=1}^{k_n} \mathbf{E} \left[X_{n,l}^2 \cdot \frac{X_{n,l}}{1+X_{n,l}^2} \right] \\ &= - \int \frac{x}{1+x^2} \nu_n(dx) \xrightarrow{n \rightarrow \infty} 0. \end{aligned} \quad \square$$

Beweis von Satz 15.43.

„(i) \Rightarrow (ii)“ Wir müssen für jedes $t \in \mathbb{R}$ zeigen, dass $\lim_{n \rightarrow \infty} \log \varphi_n(t) = -\frac{t^2}{2}$. Nach Lemma 15.46 ist dies äquivalent zu

$$\lim_{n \rightarrow \infty} \sum_{l=1}^{k_n} (\varphi_{n,l}(t) - 1) = -\frac{t^2}{2}.$$

Nun ist $f_t(x) \frac{x^2}{1+x^2} = e^{itx} - 1 - \frac{itx}{1+x^2}$. Also gilt

$$\begin{aligned} \sum_{l=1}^{k_n} (\varphi_{n,l}(t) - 1) &= \sum_{l=1}^{k_n} \int \left(f_t(x) \frac{x^2}{1+x^2} + \frac{itx}{1+x^2} \right) \mathbf{P}_{X_{n,l}}(dx) \\ &= \int f_t d\mu_n + it \int \frac{1}{x} \mu_n(dx) \\ &\xrightarrow{n \rightarrow \infty} -\frac{t^2}{2} \quad (\text{nach Lemma 15.49}). \end{aligned} \quad \square$$

Als eine Anwendung des Satzes von Lindeberg-Feller bringen wir den so genannten **Dreireihensatz**, der auf Kolmogorov zurückgeht.

Satz 15.50 (Kolmogorov'scher Dreireihensatz). Seien X_1, X_2, \dots unabhängige reelle Zufallsvariablen. Es sei $K > 0$ und $Y_n := X_n \mathbb{1}_{\{|X_n| \leq K\}}$ für jedes $n \in \mathbb{N}$.

Die Reihe $\sum_{n=1}^{\infty} X_n$ konvergiert genau dann fast sicher, wenn die folgenden drei Bedingungen gelten:

- (i) $\sum_{n=1}^{\infty} \mathbf{P}[|X_n| > K] < \infty$,
- (ii) $\sum_{n=1}^{\infty} \mathbf{E}[Y_n]$ konvergiert,
- (iii) $\sum_{n=1}^{\infty} \text{Var}[Y_n] < \infty$.

Beweis. „ \Leftarrow “ Es gelten (i), (ii) und (iii). Nach Übung 7.1.1 konvergiert wegen (iii) die Reihe $\sum_{n=1}^{\infty} (Y_n - \mathbf{E}[Y_n])$ f.s. Wegen (ii) konvergiert also $\sum_{n=1}^{\infty} Y_n$ f.s. Nach dem Lemma von Borel-Cantelli existiert ein $N = N(\omega)$, sodass $|X_n| < K$, also $X_n = Y_n$, für $n \geq N$. Also ist $\sum_{n=1}^{\infty} X_n = \sum_{n=1}^{N-1} X_n + \sum_{n=N}^{\infty} Y_n$ f.s. konvergent.

„ \Rightarrow “ Es konvergiere $\sum_{n=1}^{\infty} X_n$ f.s. Offenbar folgt daraus, dass (i) gilt (andernfalls wäre nach dem Lemma von Borel-Cantelli $|X_n| > K$ unendlich oft, was der Annahme widerspräche).

Wir nehmen an, dass (iii) nicht gilt und führen dies zum Widerspruch. Wir setzen $\sigma_n^2 = \sum_{k=1}^n \text{Var}[Y_k]$ und definieren ein Schema $(X_{n,l}; l = 1, \dots, n, n \in \mathbb{N})$ durch $X_{n,l} = (Y_l - \mathbf{E}[Y_l])/\sigma_n$. Das Schema ist zentriert und normiert. Wegen $\sigma_n^2 \xrightarrow{n \rightarrow \infty} \infty$, gilt für jedes $\varepsilon > 0$ und großes $n \in \mathbb{N}$, dass $2K < \varepsilon \sigma_n$, also $|X_{n,l}| \leq \varepsilon$ für alle $l = 1, \dots, n$. Es folgt $L_n(\varepsilon) \xrightarrow{n \rightarrow \infty} 0$, wobei $L_n(\varepsilon) = \sum_{l=1}^n \mathbf{E}[X_{n,l}^2 \mathbb{1}_{\{|X_{n,l}| \geq \varepsilon\}}]$ die Größe aus der Lindeberg-Bedingung ist (siehe (15.6)).

Nach dem Satz von Lindeberg-Feller gilt also $S_n := X_{n,1} + \dots + X_{n,n} \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0,1}$. Wie im ersten Teil des Beweises gezeigt, folgt aus der fast sicheren Konvergenz von $\sum_{n=1}^{\infty} X_n$ und aus (i).

$$\sum_{n=1}^{\infty} Y_n \quad \text{konvergiert fast sicher.} \quad (15.9)$$

Insbesondere gilt $T_n := (Y_1 + \dots + Y_n)/\sigma_n \xrightarrow{n \rightarrow \infty} 0$. Nach dem Satz von Slutsky gilt also auch $(S_n - T_n) \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0,1}$. Andererseits ist $S_n - T_n$ deterministisch für jedes $n \in \mathbb{N}$, womit die Annahme, dass (iii) nicht gilt ad absurdum geführt ist.

Nachdem wir (iii) schon gezeigt haben, folgt mit Übung 7.1.1, dass $\sum_{n=1}^{\infty} (Y_n - \mathbf{E}[Y_n])$ fast sicher konvergiert. Wegen (15.9) folgt (ii). \square

Als Ergänzung bringen wir ohne Beweis eine Abschätzung für die Konvergenzgeschwindigkeit im Zentralen Grenzwertsatz (siehe beispielsweise [164, Kapitel III, §11] für einen Beweis), die mit anderen Konstanten (statt 0.8) unabhängig von Berry [13] und Esseen [51] gefunden wurde.

Satz 15.51 (Berry-Esseen). Seien X_1, X_2, \dots unabhängig und identisch verteilt mit $\mathbf{E}[X_1] = 0$, $\mathbf{E}[X_1^2] = \sigma^2 \in (0, \infty)$ und $\gamma := \mathbf{E}[|X_1|^3] < \infty$. Seien $S_n^* := \frac{1}{\sqrt{n}\sigma^2}(X_1 + \dots + X_n)$ und $\Phi : x \mapsto \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$ die Verteilungsfunktion der Standardnormalverteilung. Dann gilt für jedes $n \in \mathbb{N}$

$$\sup_{x \in \mathbb{R}} |\mathbf{P}[S_n^* \leq x] - \Phi(x)| \leq \frac{0.8 \gamma}{\sigma^3 \sqrt{n}}.$$

Beispiel 15.52. Sei $\alpha \in (0, 1)$. Wir betrachten die Verteilung μ_α auf \mathbb{R} mit Dichte

$$f_\alpha(x) = \frac{1}{2\alpha} |x|^{-1-1/\alpha} \mathbb{1}_{\{|x| \geq 1\}}.$$

Seien X_1, X_2, \dots , u.i.v. mit $X_i \sim \mu_\alpha$. Dann ist $\mathbf{E}[X_1] = 0$ und $\sigma^2 := \mathbf{Var}[X_1] = 1/(1-2\alpha)$, falls $\alpha < 1/2$. Sei nun F_n die Verteilungsfunktion von S_n^* und F_Φ die Verteilungsfunktion der Standardnormalverteilung.

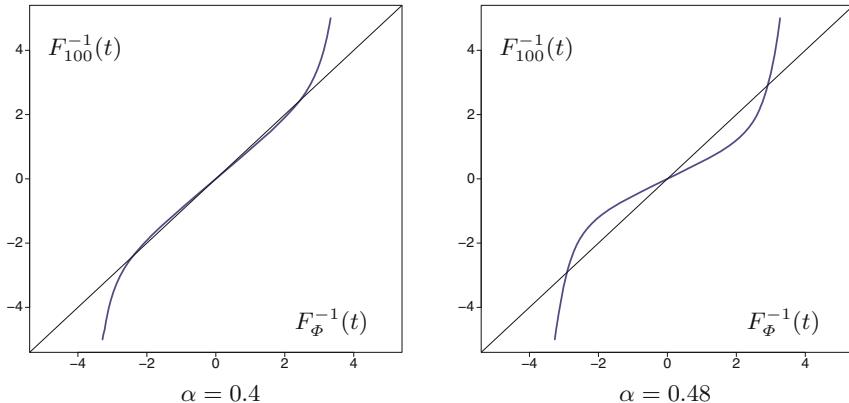


Abb. 15.2. Q-Q-Plots für S_{100}^* aus Beispiel 15.52 mit $\alpha = 0.4$ (links) und $\alpha = 0.48$ (rechts), jeweils gegen die Standardnormalverteilung. Zur Orientierung ist auch die Diagonale eingezeichnet.

Je näher die Verteilungsfunktionen F_n und F_Φ einander sind, desto besser liegen die Punkte $(F_\Phi^{-1}(t), F_n^{-1}(t))$ auf der Diagonalen $\{(x, x) : x \in \mathbb{R}\}$. Eine grafische

Darstellung der Punkte $(F_\phi^{-1}(t), F_n^{-1}(t))$, $t \in \mathbb{R}$ heißt **Q-Q-Plot**, da hier jeweils die Quantile der beiden Verteilungen gegeneinander aufgetragen werden.

Je näher α dem Wert $1/2$ kommt, desto weniger Momente besitzt die Verteilung μ_α . Entsprechend erwarten wir für festes n eine größere Abweichung für größeres α . Die beiden Grafiken in Abb. 15.2 illustrieren dies. \diamond

Übung 15.5.1. Die Argumentation aus Bemerkung 15.38 ist etwas direkter als die Argumentation mit dem Lévy'schen Stetigkeitssatz, allerdings etwas weniger robust: Man gebe eine Folge X_1, X_2, \dots von unabhängigen, reellen Zufallsvariablen an mit $\mathbf{E}[|X_n|] = \infty$ für jedes $n \in \mathbb{N}$, aber mit

$$\frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0,1}. \quad \clubsuit$$

Übung 15.5.2. Seien Y_1, Y_2, \dots u.i.v. mit $\mathbf{E}[Y_i] = 0$ und $\mathbf{E}[Y_i^2] = 1$. Davon unabhängig seien Z_1, Z_2, \dots unabhängige Zufallsvariablen mit

$$\mathbf{P}[Z_i = i] = \mathbf{P}[Z_i = -i] = \frac{1}{2}(1 - \mathbf{P}[Z_i = 0]) = \frac{1}{2} \frac{1}{i^2}.$$

Setze $X_i := Y_i + Z_i$ und $S_n = X_1 + \dots + X_n$ für $i, n \in \mathbb{N}$.

Man zeige: $n^{-1/2} S_n \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0,1}$, aber $(X_i)_{i \in \mathbb{N}}$ erfüllt keine Lindeberg-Bedingung.

Hinweis: Möglichst nicht direkt ausrechnen! \clubsuit

Übung 15.5.3. Seien X_1, X_2, \dots u.i.v. Zufallsvariablen mit Dichte

$$f(x) = \frac{1}{|x|^3} \mathbb{1}_{\mathbb{R} \setminus [-1, 1]}(x).$$

Dann ist $\mathbf{E}[X_1^2] = \infty$, aber es gibt Zahlen A_1, A_2, \dots , sodass

$$\frac{X_1 + \dots + X_n}{A_n} \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0,1}.$$

Man gebe die Folge $(A_n)_{n \in \mathbb{N}}$ explizit an. \clubsuit

15.6 Mehrdimensionaler Zentraler Grenzwertsatz

Wir kommen zu einer mehrdimensionalen Variante des zentralen Grenzwertsatzes.

Definition 15.53. Sei C eine (strikt) positiv definite symmetrische reelle $d \times d$ Matrix und $\mu \in \mathbb{R}^d$. Ein Zufallsvektor $X = (X_1, \dots, X_d)^T$ heißt **d -dimensional normalverteilt** mit Erwartungswert μ und Kovarianzmatrix C , falls X die Dichte

$$f_{\mu, C}(x) = \frac{1}{\sqrt{(2\pi)^d \det(C)}} \exp\left(-\frac{1}{2} \langle x - \mu, C^{-1}(x - \mu) \rangle\right) \quad (15.10)$$

für $x \in \mathbb{R}^d$ hat. Wir schreiben $X \sim \mathcal{N}_{\mu, C}$.

Satz 15.54. Sei $\mu \in \mathbb{R}^d$ und C eine positiv definite symmetrische reelle $d \times d$ -Matrix. Ist $X \sim \mathcal{N}_{\mu, C}$, dann gelten:

- (i) $\mathbf{E}[X_i] = \mu_i$ für jedes $i = 1, \dots, d$.
- (ii) $\text{Cov}[X_i, X_j] = C_{i,j}$ für alle $i, j = 1, \dots, d$.
- (iii) $\langle \lambda, X \rangle \sim \mathcal{N}_{\langle \lambda, \mu \rangle, \langle \lambda, C \lambda \rangle}$ für jedes $\lambda \in \mathbb{R}^d$.
- (iv) $\varphi(t) := \mathbf{E}[e^{it \cdot X}] = e^{i\langle t, \mu \rangle} e^{-\frac{1}{2}\langle t, Ct \rangle}$ für jedes $t \in \mathbb{R}^d$.

Es gilt sogar $X \sim \mathcal{N}_{\mu, C} \iff (\text{iii}) \iff (\text{iv})$.

Beweis. (i) und (ii) sind einfache Rechnungen, ebenso (iii) und (iv). Die Implikation (iii) \implies (iv) ist simpel. Die Familie $\{f_t : x \mapsto e^{it \cdot x}, t \in \mathbb{R}^d\}$ ist trennend für $\mathcal{M}_1(\mathbb{R}^d)$ nach dem Satz von Stone–Weierstraß. Also legt φ die Verteilung von X eindeutig fest. \square

Bemerkung 15.55. Für eindimensionale Normalverteilungen liegt es nahe, $\mathcal{N}_{\mu, 0}$ als δ_μ zu definieren. Einen so einfachen Begriff können wir bei mehrdimensionalen Normalverteilungen nicht mehr erwarten (außer für den Fall $C = 0$), wenn eine Entartung nur in einigen Richtungen auftritt, also C nur noch positiv semidefinit und symmetrisch ist. In diesem Fall definieren wir $\mathcal{N}_{\mu, C}$ als diejenige Verteilung auf \mathbb{R}^n mit charakteristischer Funktion $\varphi(t) = e^{i\langle t, \mu \rangle} e^{-\frac{1}{2}\langle t, Ct \rangle}$. \diamond

Satz 15.56 (Cramér-Wold Device). Es seien $X_n = (X_{n,1}, \dots, X_{n,d})^T \in \mathbb{R}^d$, $n \in \mathbb{N}$, Zufallsvektoren. Dann sind äquivalent:

- (i) Es gibt einen Zufallsvektor X mit $X_n \xrightarrow{n \rightarrow \infty} X$.
- (ii) Für jedes $\lambda \in \mathbb{R}^d$ gibt es eine Zufallsvariable X^λ mit $\langle \lambda, X_n \rangle \xrightarrow{n \rightarrow \infty} X^\lambda$.

Gelten (i) und (ii), so ist $X^\lambda \stackrel{\mathcal{D}}{=} \langle \lambda, X \rangle$ für jedes $\lambda \in \mathbb{R}^d$.

Beweis. Gelte (i). Sei $\lambda \in \mathbb{R}^d$ und $s \in \mathbb{R}$. Die Abbildung $\mathbb{R}^d \rightarrow \mathbb{C}$, $x \mapsto e^{is\langle \lambda, x \rangle}$ ist stetig und beschränkt, also gilt $\mathbf{E}[e^{is\langle \lambda, X_n \rangle}] \xrightarrow{n \rightarrow \infty} \mathbf{E}[e^{is\langle \lambda, X \rangle}]$. Damit gilt (ii) für $X^\lambda := \langle \lambda, X \rangle$.

Gelte nun (ii). Dann ist $(\mathbf{P}_{X_{n,l}})_{n \in \mathbb{N}}$ straff, $l = 1, \dots, d$. Also ist $(\mathbf{P}_{X_n})_{n \in \mathbb{N}}$ straff und damit relativ folgenkompakt (Satz von Prohorov). Für jeden schwachen Häufungspunkt Q von $(\mathbf{P}_{X_n})_{n \in \mathbb{N}}$ ist für jedes $\lambda \in \mathbb{R}^d$

$$\int Q(dx) e^{i\langle \lambda, x \rangle} = \mathbf{E}[e^{i\lambda X}].$$

Also ist der Häufungspunkt Q eindeutig und damit $(\mathbf{P}_{X_n})_{n \in \mathbb{N}}$ schwach konvergent. Es gilt also (i).

Gilt (ii), so sind die schwachen Limiten X^λ in Verteilung eindeutig und nach dem Gezeigten ist $X^\lambda = \langle \lambda, X \rangle$ eine mögliche Wahl. Es gilt also stets $X^\lambda \xrightarrow{\mathcal{D}} \langle \lambda, X \rangle$. \square

Satz 15.57 (Zentraler Grenzwertsatz im \mathbb{R}^d). Seien $(X_n)_{n \in \mathbb{N}}$ u.i.v. Zufallsvektoren mit $\mathbf{E}[X_{n,i}] = 0$ und $\mathbf{E}[X_{n,i}X_{n,j}] = C_{ij}$, $i, j = 1, \dots, d$. Sei $S_n^* := \frac{X_1 + \dots + X_n}{\sqrt{n}}$. Dann gilt

$$\mathbf{P}_{S_n^*} \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0,C} \text{ schwach.}$$

Beweis. Sei $\lambda \in \mathbb{R}^d$. Setze $X_n^\lambda = \langle \lambda, X_n \rangle$, $S_n^\lambda = \langle \lambda, S_n^* \rangle$ und $S_\infty \sim \mathcal{N}_{0,C}$. Dann ist $\mathbf{E}[X_n^\lambda] = 0$ und $\mathbf{Var}[X_n^\lambda] = \langle \lambda, C\lambda \rangle$. Nach dem eindimensionalen Zentralen Grenzwertsatz gilt $\mathbf{P}_{S_n^\lambda} \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0, \langle \lambda, C\lambda \rangle} = \mathbf{P}_{\langle \lambda, S_\infty \rangle}$. Nach Satz 15.56 zeigt dies die Aussage. \square

Übung 15.6.1. Sei $\mu \in \mathbb{R}^d$, C eine symmetrische positiv semidefinite reelle $d \times d$ Matrix und $X \sim \mathcal{N}_{\mu,C}$ (im Sinne von Bemerkung 15.55). Man zeige: Für jedes $m \in \mathbb{N}$ und jede reelle $m \times d$ Matrix A gilt $AX \sim \mathcal{N}_{A\mu, ACAT}$. 

Übung 15.6.2 (Cholesky-Faktorisierung). Sei C eine positiv definite symmetrische reelle $d \times d$ Matrix. Dann existiert eine reelle $d \times d$ Matrix $A = (a_{kl})$ mit $A \cdot A^T = C$. Man kann A sogar als untere Dreiecksmatrix wählen. Sei $W := (W_1, \dots, W_d)^T$, wo W_1, \dots, W_d unabhängig und $\mathcal{N}_{0,1}$ verteilt sind. Wir setzen $X := AW + \mu$. Man zeige: $X \sim \mathcal{N}_{\mu,C}$. 

16 Unbegrenzt teilbare Verteilungen

Die Normalverteilung $\mathcal{N}_{\mu, \sigma^2}$ lässt sich für jedes $n \in \mathbb{N}$ als n -te Faltungspotenz eines W-Maßes schreiben (nämlich von $\mathcal{N}_{\mu/n, \sigma^2/n}$). Die selbe Eigenschaft, die wir unbegrenzte Teilbarkeit nennen, hat die Poisson-Verteilung. Im ersten Abschnitt untersuchen wir, welche W-Maße auf \mathbb{R} unbegrenzt teilbar sind und geben eine erschöpfende Beschreibung der Klasse dieser Maße durch die Lévy-Khinchin Formel.

Die Normalverteilung hat (im Gegensatz zur Poisson-Verteilung) die Eigenschaft, dass sie als Grenzwert reskaliertes Summen von u.i.v. Zufallsvariablen auftritt (Zentraler Grenzwertsatz). Im zweiten Abschnitt untersuchen wir knapp die Teilkasse unbegrenzt teilbarer Maße auf \mathbb{R} , die diese Eigenschaft haben.

16.1 Die Lévy-Khinchin Formel

Zur Abkürzung verwenden wir in diesem Abschnitt die Bezeichnung „CFW“ für „charakteristische Funktion eines W-Maßes auf \mathbb{R} “.

Definition 16.1. Ein Maß $\mu \in \mathcal{M}_1(\mathbb{R})$ heißt **unbegrenzt teilbar**, falls es für jedes $n \in \mathbb{N}$ ein $\mu_n \in \mathcal{M}_1(\mathbb{R})$ mit der Eigenschaft $\mu_n^{*n} = \mu$ gibt. Analog nennen wir eine CFW φ unbegrenzt teilbar, falls es zu jedem $n \in \mathbb{N}$ eine CFW φ_n gibt mit $\varphi = \varphi_n^n$. Eine reelle Zufallsvariable X heißt unbegrenzt teilbar, falls es zu jedem $n \in \mathbb{N}$ u.i.v. Zufallsvariablen $X_{n,1}, \dots, X_{n,n}$ gibt mit $X \stackrel{D}{=} X_{n,1} + \dots + X_{n,n}$.

Offenbar sind alle drei Begriffe der unendlichen Teilbarkeit äquivalent, und wir wollen sie synonym verwenden. Man beachte, dass die Eindeutigkeit von μ_n bezüglich φ_n keineswegs evident ist. Tatsächlich folgt aus der n -fachen Teilbarkeit noch nicht die Eindeutigkeit der n -ten Faltungswurzel $\mu^{*1/n} := \mu_n$ bezüglich von φ_n . Um dies für gerades n einzusehen, wähle man etwa eine reelle CFW φ , für die $|\varphi| \neq \varphi$ ebenfalls eine CFW ist (siehe Beispiel 15.16 und 15.17). Dann ist $\varphi^n = |\varphi|^n$ n -fach teilbar, jedoch sind die Faktoren nicht eindeutig.

Mit Hilfe des Lévy'schen Stetigkeitssatzes kann man zeigen (siehe Übung 16.1.1), dass $\varphi(t) \neq 0$ für alle $t \in \mathbb{R}$ gilt, falls φ unbegrenzt teilbar ist. Die probabilistische Bedeutung dieser Aussage liegt darin, dass $\log(\varphi(t))$ als stetige Funktion eindeutig

definiert ist und damit auch nur genau eine stetige Funktion $\varphi^{1/n} = \exp(\log(\varphi)/n)$ existiert. Die n -ten Faltungswurzeln sind also eindeutig definiert, falls die Verteilung unbegrenzt teilbar ist.

Beispiele 16.2. (i) δ_x ist unbegrenzt teilbar mit $\delta_{x/n}^{*n} = \delta_x$ für jedes $n \in \mathbb{N}$.

(ii) Die Normalverteilung ist unbegrenzt teilbar mit $\mathcal{N}_{m,\sigma^2} = \mathcal{N}_{m/n,\sigma^2/n}^{*n}$.

(iii) Die Cauchy-Verteilung Cau_a mit Dichte $x \mapsto (a\pi)^{-1}(1 + (x/a)^2)^{-1}$ ist unbegrenzt teilbar mit $\text{Cau}_a = \text{Cau}_{a/n}^{*n}$. In der Tat: Cau_a hat CFW $\varphi_a(t) = e^{-a|t|}$, also ist $\varphi_{a/n}^{*n} = \varphi_a$.

(iv) Jede symmetrische stabile Verteilung mit Index $\alpha \in (0, 2]$ und Größenparameter $\gamma > 0$, also mit CFW $\varphi_{\alpha,\gamma}(t) = e^{-|\gamma t|^\alpha}$, ist unbegrenzt teilbar. In der Tat ist $\varphi_{\alpha,\gamma/n^{1/\alpha}}^{*n} = \varphi_{\alpha,\gamma}$. (Genau genommen haben wir bislang erst für $\alpha \in (0, 1]$ (in Korollar 15.25) und für $\alpha = 2$ (Normalverteilung) gezeigt, dass $\varphi_{\alpha,\gamma}$ überhaupt eine CFW ist. In Abschnitt 16.2 zeigen wir, dass dies tatsächlich für alle $\alpha \in (0, 2]$ richtig ist. Für $\alpha > 2$ ist $\varphi_{\alpha,\gamma}$ hingegen keine CFW, siehe Übung 15.4.3.)

(v) Die Gamma-Verteilung $\Gamma_{\theta,r}$ mit CFW $\varphi_{\theta,r}(t) = \exp(r\psi_\theta(t))$, wo $\psi_\theta(t) = \log(1 - it/\theta)$ ist, ist unbegrenzt teilbar mit $\Gamma_{\theta,r} = \Gamma_{\theta,r/n}^{*n}$.

(vi) Die Poisson-Verteilung ist unbegrenzt teilbar mit $\text{Poi}_\lambda = \text{Poi}_{\lambda/n}^{*n}$.

(vii) Die negative Binomialverteilung $b_{r,p}^-(\{k\}) = \binom{-r}{k}(-1)^k p^r (1-p)^k$, $k \in \mathbb{N}_0$, mit Parametern $r > 0$ und $p \in (0, 1)$ ist unbegrenzt teilbar mit $b_{r,p}^- = (b_{r/n,p}^-)^{*n}$. In der Tat ist $\varphi_{r,p}(t) = e^{r\psi_p(t)}$, wo

$$\psi_p(t) = \log(p) - \log(1 - (1-p)e^{it}).$$

(viii) Seien X und Y unabhängig und $X \sim \mathcal{N}_{0,\sigma^2}$ sowie $Y \sim \Gamma_{\theta,r}$, wobei $\sigma^2, \theta, r > 0$ sind. Man kann zeigen, dass die Zufallsvariable $Z := X/\sqrt{Y}$ unbegrenzt teilbar ist (siehe [71] oder [140]). Insbesondere ist die Student'sche t -Verteilung mit $k \in \mathbb{N}$ Freiheitsgraden unbegrenzt teilbar (dieses ist der Fall $\sigma^2 = 1$ und $\theta = r = k/2$).

(ix) Die Binomialverteilung $b_{n,p}$ ist für $n \in \mathbb{N}$ und $p \in (0, 1)$ nicht unbegrenzt teilbar (warum?).

(x) Etwas allgemeiner ist außer der trivialen Verteilung keine Verteilung unbegrenzt teilbar, die auf ein endliches Intervall konzentriert ist. \diamond

Ein Hauptziel dieses Abschnitts ist es zu zeigen, dass sich jede unbegrenzt teilbare Verteilung aus drei generischen zusammensetzt:

- den Punktverteilungen δ_x mit $x \in \mathbb{R}$,
- den Normalverteilungen $\mathcal{N}_{\mu, \sigma^2}$ mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$,
- (Grenzwerten von) Faltungen von Poisson-Verteilungen.

Da die Faltungen von Poisson-Verteilungen eine besondere Rolle spielen, wollen wir sie hier gesondert betrachten.

Ist $\nu \in \mathcal{M}_1(\mathbb{R})$ mit CFW φ_ν und ist $\lambda > 0$, so kann man leicht nachrechnen, dass $\varphi(t) = \exp(\lambda(\varphi_\nu(t) - 1))$ die CFW von $\mu_\lambda = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \nu^{*k}$ ist. Formal können wir also $\mu_\lambda = e^{*\lambda(\nu-\delta_0)}$ schreiben. Tatsächlich ist μ_λ unbegrenzt teilbar mit $\mu_\lambda = \mu_{\lambda/n}^{*n}$. Wir wollen nun die Parameter λ und ν zu $\lambda\nu$ zusammenfassen. Für $\nu \in \mathcal{M}_f(\mathbb{R})$ können wir $\nu^{*n} = \nu(\mathbb{R})^n (\nu/\nu(\mathbb{R}))^{*n}$ setzen, beziehungsweise $\nu^{*n} = 0$, falls $\nu = 0$. In beiden Fällen ist $\nu^{*0} := \delta_0$. Wir treffen daher die folgende Definition.

Definition 16.3. Die **zusammengesetzte Poissonverteilung** (compound Poisson distribution) mit Intensitätsmaß $\nu \in \mathcal{M}_f(\mathbb{R})$ ist das folgende W-Maß auf \mathbb{R} :

$$\text{CPoi}_\nu := e^{*(\nu-\nu(\mathbb{R})\delta_0)} := e^{-\nu(\mathbb{R})} \sum_{n=0}^{\infty} \frac{\nu^{*n}}{n!}.$$

Die CFW von CPoi_ν ist gegeben durch

$$\varphi_\nu(t) = \exp \left(\int (e^{itx} - 1) \nu(dx) \right). \quad (16.1)$$

Speziell ist $\text{CPoi}_{\mu+\nu} = \text{CPoi}_\mu * \text{CPoi}_\nu$, also ist CPoi_ν unbegrenzt teilbar.

Beispiel 16.4. Für jede messbare Menge $A \subset \mathbb{R} \setminus \{0\}$ und jedes $r > 0$ ist

$$r^{-1} \text{CPoi}_{r\nu}(A) = e^{-r\nu(\mathbb{R})} \nu(A) + e^{-r\nu(\mathbb{R})} \sum_{k=2}^{\infty} \frac{r^{k-1} \nu^{*k}(A)}{k!} \xrightarrow{r \downarrow 0} \nu(A).$$

Wir wollen dies benutzen um zu zeigen, dass $b_{r,p}^- = \text{CPoi}_{r\nu}$ für ein gewisses $\nu \in \mathcal{M}_f(\mathbb{N})$. Wir berechnen dazu für $k \in \mathbb{N}$

$$r^{-1} b_{r,p}^-(\{k\}) = \frac{r(r+1) \cdots (r+k-1)}{r k!} p^r (1-p)^k \xrightarrow{r \downarrow 0} \frac{(1-p)^k}{k}.$$

Wenn $b_{r,p}^- = \text{CPoi}_{r\nu}$ für ein $\nu \in \mathcal{M}_f(\mathbb{N})$ ist, ist also $\nu(\{k\}) = (1-p)^k/k$. Wir berechnen die CFW von $\text{CPoi}_{r\nu}$ für dieses ν

$$\varphi_{r\nu}(t) = \exp \left(r \sum_{k=1}^{\infty} \frac{((1-p)e^{it})^k - (1-p)^k}{k} \right) = p^r (1 - (1-p)e^{it})^{-r}.$$

Dies ist aber die CFW von $b_{r,p}^-$, also ist tatsächlich $b_{r,p}^- = \text{CPoi}_{r\nu}$. \diamond

Nicht jede unbegrenzt teilbare Verteilung ist vom Typ CPoi $_{\nu}$, allerdings gilt:

Satz 16.5. Ein W-Maß μ auf \mathbb{R} ist genau dann unbegrenzt teilbar, wenn es eine Folge $(\nu_n)_{n \in \mathbb{N}}$ in $\mathcal{M}_f(\mathbb{R} \setminus \{0\})$ gibt mit $\text{CPoi}_{\nu_n} \xrightarrow{n \rightarrow \infty} \mu$.

Da jedes CPoi $_{\nu_n}$ unbegrenzt teilbar ist, müssen wir einerseits zeigen, dass diese Eigenschaft unter schwachen Limiten erhalten bleibt. Andererseits zeigen wir, dass für unbegrenzt teilbares μ die Folge $\nu_n = n\mu^{*1/n}$ das Gewünschte leistet. Wir bereiten den Beweis mit einem weiteren Satz vor.

Satz 16.6. Sei $(\varphi_n)_{n \in \mathbb{N}}$ eine Folge von CFWs. Dann sind äquivalent:

- (i) Für jedes $t \in \mathbb{R}$ existiert $\varphi(t) = \lim_{n \rightarrow \infty} \varphi_n^n(t)$, und φ ist stetig in 0.
- (ii) Für jedes $t \in \mathbb{R}$ existiert $\psi(t) = \lim_{n \rightarrow \infty} n(\varphi_n(t) - 1)$, und ψ ist stetig in 0.

Gelten (i) und (ii), so ist $\varphi = e^{\psi}$ eine CFW.

Beweis. Der Beweis beruht auf der Taylor-Entwicklung des Logarithmus'

$$|\log(z) - (z - 1)| \leq \frac{1}{2}|z - 1|^2 \quad \text{für } z \in \mathbb{C} \text{ mit } |z - 1| < \frac{1}{2}.$$

Speziell gilt für $(z_n)_{n \in \mathbb{N}}$ in \mathbb{C}

$$\limsup_{n \rightarrow \infty} n|z_n - 1| < \infty \iff \limsup_{n \rightarrow \infty} |n \log(z_n)| < \infty \quad (16.2)$$

und $\lim_{n \rightarrow \infty} n(z_n - 1) = \lim_{n \rightarrow \infty} n \log(z_n)$, falls einer der Limiten existiert.

Wenden wir dies auf $z_n = \varphi_n(t)$ an, so folgt (i) aus (ii). Andererseits folgt (ii) aus (i), wenn $\liminf_{n \rightarrow \infty} n \log(|\varphi_n(t)|) > -\infty$, also wenn $\varphi(t) \neq 0$ für jedes $t \in \mathbb{R}$.

Da φ stetig in 0 ist und $\varphi(0) = 1$ gilt, gibt es ein $\varepsilon > 0$ mit $|\varphi(t)| > \frac{1}{2}$ für jedes $t \in [-\varepsilon, \varepsilon]$. Da φ und φ_n CFWs sind, sind auch $|\varphi|^2$ und $|\varphi_n|^2$ CFWs. Aus der punktweisen Konvergenz von $|\varphi_n(t)|^{2n}$ gegen $|\varphi(t)|^2$ folgt nach dem Lévy'schen Stetigkeitssatz also die gleichmäßige Konvergenz auf kompakten Mengen. Wende nun (16.2) mit $z_n = |\varphi_n(t)|^2$ an. Für $t \in [-\varepsilon, \varepsilon]$ ist daher $(n(1 - |\varphi_n(t)|^2))_{n \in \mathbb{N}}$ beschränkt. Nach Lemma 15.11(v) ist dann aber auch $n(1 - |\varphi_n(2t)|^2) \leq 4n(1 - |\varphi_n(t)|^2)$ beschränkt, also

$$|\varphi(2t)|^2 \geq \liminf_{n \rightarrow \infty} \exp(4n(|\varphi_n(t)|^2 - 1)) = (|\varphi(t)|^2)^4.$$

Iterativ erhalten wir $|\varphi(t)| \geq 2^{-(4^k)}$ für $|t| \leq 2^k \varepsilon$. Es gibt also ein $\gamma > 0$, sodass

$$|\varphi(t)| > \frac{1}{2} e^{-\gamma t^2} \quad \text{für jedes } t \in \mathbb{R}. \quad (16.3)$$

Gelten (i) und (ii), so ist

$$\log \varphi(t) = \lim_{n \rightarrow \infty} n \log(\varphi_n(t)) = \lim_{n \rightarrow \infty} n(\varphi_n(t) - 1) = \psi(t).$$

Nach dem Lévy'schen Stetigkeitssatz ist φ als stetiger Limes von CFWs selbst eine CFW. \square

Korollar 16.7. Gelten die Bedingungen von Satz 16.6, so ist φ^r eine CFW für jedes $r > 0$. Insbesondere ist $\varphi = (\varphi^{1/n})^n$ unbegrenzt teilbar.

Beweis. Ist φ_n die CFW von $\mu_n \in \mathcal{M}_1(\mathbb{R})$, so ist $e^{rn(\varphi_n-1)}$ die CFW von $\text{CPoi}_{rn\mu_n}$. Als in 0 stetiger Limes von CFWs ist $\varphi^r = e^{r\psi} = \lim_{n \rightarrow \infty} e^{rn(\varphi_n-1)}$ nach dem Lévy'schen Stetigkeitssatz eine CFW. Mit $r = \frac{1}{n}$ folgt, dass $\varphi = (\varphi^{1/n})^n$ unbegrenzt teilbar ist. \square

Korollar 16.8. Eine in 0 stetige Funktion $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ ist genau dann eine unbegrenzt teilbare CFW, wenn es eine Folge $(\varphi_n)_{n \in \mathbb{N}}$ von CFWs gibt mit $\varphi_n(t) \rightarrow \varphi(t)$ für jedes $t \in \mathbb{R}$.

Beweis. Die eine Richtung ist schon in Korollar 16.7 gezeigt worden. Sei also φ eine unbegrenzt teilbare CFW. Dann leistet $\varphi_n = \varphi^{1/n}$ das Gewünschte. \square

Korollar 16.9. Ist $(\mu_n)_{n \in \mathbb{N}}$ eine (schwach) konvergente Folge unbegrenzt teilbarer W-Maße auf \mathbb{R} , so ist $\mu = \lim_{n \rightarrow \infty} \mu_n$ unbegrenzt teilbar.

Beweis. Wende Satz 16.6 an mit φ_n die CFW von $\mu_n^{*1/n}$. \square

Korollar 16.10. Ist $\mu \in \mathcal{M}_1(\mathbb{R})$ unbegrenzt teilbar, so existiert eine stetige Faltungshalbgruppe $(\mu_t)_{t \geq 0}$ mit $\mu_1 = \mu$ und ein stochastischer Prozess $(X_t)_{t \geq 0}$ mit unabhängigen, stationären Zuwächsen $X_t - X_s \sim \mu_{t-s}$ für $t > s$.

Beweis. Sei φ die CFW von μ . Die Existenz der Faltungshalbgruppe folgt aus Korollar 16.8 und 16.7, indem wir μ_r durch φ^r definieren. Die Stetigkeit der Halbgruppe folgt, da $\varphi^r \rightarrow 1$ für $r \rightarrow 0$ (weil $\varphi^r(t) \neq 0$ für alle $t \in \mathbb{R}$). Schließlich folgt die Existenz des Prozesses X aus Satz 14.47. \square

Korollar 16.11. Ist φ eine unbegrenzt teilbare CFW, so existiert ein $\gamma > 0$ mit $|\varphi(t)| \geq \frac{1}{2}e^{-\gamma t^2}$ für jedes $t \in \mathbb{R}$. Speziell ist $t \mapsto e^{-|t|^\alpha}$ für kein $\alpha > 2$ eine CFW.

Beweis. Dies folgt direkt aus (16.3). \square

Beweis (von Satz 16.5). Da CPoi_{ν_n} unbegrenzt teilbar ist, ist nach Korollar 16.9 auch der schwache Limes unbegrenzt teilbar.

Sei nun μ unbegrenzt teilbar mit CFW φ . Wähle W-Maße μ_n mit CFW φ_n wie in Korollar 16.8. Nach Satz 16.6 gilt $e^{n(\varphi_n - 1)} \xrightarrow{n \rightarrow \infty} \varphi$, also auch $\text{CPoi}_{n\mu_n} \xrightarrow{n \rightarrow \infty} \nu$.

□

Ohne Beweis bringen wir die folgende Verschärfung von Korollar 16.8, die auf einer feineren Analyse mit den Argumenten aus Satz 16.6 beruht.

Satz 16.12. Sei $(\varphi_{n,l}; l = 1, \dots, k_n, n \in \mathbb{N})$ ein Schema von CFWs mit der Eigenschaft

$$\sup_{L > 0} \limsup_{n \rightarrow \infty} \sup_{t \in [-L, L]} \sup_{l=1, \dots, k_n} |\varphi_{n,l}(t) - 1| = 0. \quad (16.4)$$

Existiert für jedes $t \in \mathbb{R}$ der Limes $\varphi(t) := \lim_{n \rightarrow \infty} \prod_{l=1}^{k_n} \varphi_{n,l}(t)$, und ist φ stetig in 0, so ist φ eine unbegrenzt teilbare CFW.

Beweis. Siehe etwa [59, Kapitel XV.7]. □

In dem Fall, wo für jedes n die $\varphi_{n,l}$ alle gleich sind und $k_n \xrightarrow{n \rightarrow \infty} \infty$, gilt (16.4) automatisch, wenn das Produkt gegen eine stetige Funktion konvergiert. Der Satz liefert also tatsächlich eine Verbesserung von Korollar 16.8.

Der Wert des Satzes liegt in der folgenden Beobachtung. Sei $(X_{n,l}; l = 1, \dots, k_n, n \in \mathbb{N})$ ein Schema reeller Zufallsvariablen mit CFWs $\varphi_{n,l}$. Genau dann ist das Schema asymptotisch vernachlässigbar, wenn (16.4) gilt: Gilt $P[|X_{n,l}| > \varepsilon] < \delta$, so ist für jedes $t \in [-1/\varepsilon, 1/\varepsilon]$ schon $|\varphi_{n,l}(t) - 1| \leq 2\varepsilon + \delta$. Also folgt (16.4) aus der Vernachlässigbarkeit des Schemas. Andererseits impliziert (16.4), dass für jede Folge (l_n) mit $l_n \leq k_n$ gilt $\varphi_{n,l_n} \xrightarrow{n \rightarrow \infty} 1$, also $X_{n,l_n} \xrightarrow{n \rightarrow \infty} 0$ stochastisch.

Aus diesen Betrachtungen und Satz 16.12 erhalten wir den folgenden Satz.

Satz 16.13. Sei $(X_{n,l}; l = 1, \dots, k_n, n \in \mathbb{N})$ ein unabhängiges, asymptotisch vernachlässigbares Schema reeller Zufallsvariablen. Gibt es ein S mit $X_{n,1} + \dots + X_{n,k_n} \xrightarrow{n \rightarrow \infty} S$, so ist S unbegrenzt teilbar.

Wir wollen als eine direkte Anwendung von Satz 16.5 die unbegrenzt teilbaren W-Maße auf $[0, \infty)$ durch Angabe ihrer Laplace-Transformierten vollständig beschreiben. Der folgende Satz ist von eigenständigem Interesse, dient aber hauptsächlich zur Einstimmung auf die etwas technischere Beschreibung aller unbegrenzt teilbaren W-Maße auf \mathbb{R} .

Satz 16.14 (Lévy-Khinchin Formel für $[0, \infty)$). Sei $\mu \in \mathcal{M}_1([0, \infty))$ und $u : [0, \infty) \rightarrow [0, \infty)$, $t \mapsto -\log \int e^{-tx} \mu(dx)$ die log-Laplace Transformierte von μ . Genau dann ist μ unbegrenzt teilbar, wenn es ein $\alpha \geq 0$ und ein σ -endliches Maß $\nu \in \mathcal{M}((0, \infty))$ mit

$$\int (1 \wedge x) \nu(dx) < \infty \quad (16.5)$$

gibt, sodass

$$u(t) = \alpha t + \int (1 - e^{-tx}) \nu(dx) \quad \text{für } t \geq 0. \quad (16.6)$$

Das Paar (α, ν) ist dann eindeutig. Wir nennen ν das kanonische Maß oder Lévy-Maß von μ und α den deterministischen Anteil.

Beweis. „ \Rightarrow “ Sei zunächst μ unbegrenzt teilbar. Der Fall $\mu = \delta_0$ ist trivial. Sei nun $\mu \neq \delta_0$, also $u(1) > 0$.

Nach Satz 16.5 existieren $\nu_1, \nu_2, \dots \in \mathcal{M}_f(\mathbb{R} \setminus \{0\})$ mit $\text{CPoi}_{\nu_n} \xrightarrow{n \rightarrow \infty} \mu$. Offenbar können wir $\nu_n((-\infty, 0)) = 0$ annehmen. Setzen wir $u_n(t) := \int (1 - e^{-tx}) \nu_n(dx)$, so gilt (ähnlich wie in (16.1)) $u_n(t) \xrightarrow{n \rightarrow \infty} u(t)$ für jedes $t \geq 0$. Speziell ist $u_n(1) > 0$ für große n . Definiere $\tilde{\nu}_n \in \mathcal{M}_1([0, \infty))$ durch $\tilde{\nu}_n(dx) := \frac{1-e^{-x}}{u_n(1)} \nu_n(dx)$. Für jedes $t \geq 0$ gilt dann

$$\int e^{-tx} \tilde{\nu}_n(dx) = \frac{u_n(t+1) - u_n(t)}{u_n(1)} \xrightarrow{n \rightarrow \infty} \frac{u(t+1) - u(t)}{u(1)}.$$

Also existiert $\tilde{\nu} := \text{w-lim } \tilde{\nu}_n$ (in $\mathcal{M}_1([0, \infty))$) und ist eindeutig durch u festgelegt. Wir setzen $\alpha := \tilde{\nu}(\{0\}) u(1)$ und definieren $\nu \in \mathcal{M}((0, \infty))$ durch

$$\nu(dx) = u(1)(1 - e^{-x})^{-1} \mathbb{1}_{(0, \infty)}(x) \tilde{\nu}(dx).$$

Wegen $1 \wedge x \leq 2(1 - e^{-x})$ für alle $x \geq 0$ ist dann offenbar

$$\int (1 \wedge x) \nu(dx) \leq 2 \int (1 - e^{-x}) \nu(dx) \leq 2u(1) < \infty.$$

Für jedes $t \geq 0$ ist die Funktion (vergleiche (15.8))

$$f_t : [0, \infty) \rightarrow [0, \infty), \quad x \mapsto \begin{cases} \frac{1-e^{-tx}}{1-e^{-x}}, & \text{falls } x > 0, \\ t, & \text{falls } x = 0, \end{cases}$$

stetig und beschränkt (durch $t \wedge 1$), also gilt

$$\begin{aligned} u(t) &= \lim_{n \rightarrow \infty} u_n(t) = \lim_{n \rightarrow \infty} u_n(1) \int f_t d\tilde{\nu}_n \\ &= u(1) \int f_t d\tilde{\nu} = \alpha t + \int (1 - e^{-tx}) \nu(dx). \end{aligned}$$

„ \Leftarrow “ Seien nun α und ν gegeben. Sei $I_0 = [1, \infty)$ und $I_k = [1/(k+1), 1/k]$ für $k \in \mathbb{N}$. Seien X_0, X_1, \dots unabhängige Zufallsvariablen mit $\mathbf{P}_{X_k} = \text{CPoi}(\nu|_{I_k})$ für $k = 0, 1, \dots$, und sei $X := \alpha + \sum_{k=0}^{\infty} X_k$. Für jedes $k \in \mathbb{N}$ ist $\mathbf{E}[X_k] = \int_{I_k} x \nu(dx)$, also ist $\sum_{k=1}^{\infty} \mathbf{E}[X_k] = \int_{(0,1)} x \nu(dx) < \infty$. Mithin gilt $X < \infty$ fast sicher und $\alpha + \sum_{k=0}^n X_k \xrightarrow{n \rightarrow \infty} X$. Daher ist

$$-\log \mathbf{E}[e^{-tX}] = \alpha t - \sum_{k=0}^{\infty} \log \mathbf{E}[e^{-tX_k}] = \alpha t + \int (1 - e^{-tx}) \nu(dx). \quad \square$$

Beispiel 16.15. Wir können zu einer unbegrenzt teilbaren Verteilung μ auf $[0, \infty)$ das Lévy-Maß ν ausrechnen durch

$$\nu = \text{v-lim}_{n \rightarrow \infty} n \mu^{*1/n} \Big|_{(0, \infty)}. \quad (16.7)$$

Oft ist auch α leicht direkt zugänglich, etwa über die Darstellung aus Übung 16.1.3. Wir erhalten so beispielsweise für die Gamma-Verteilung $\alpha = 0$ und

$$n \Gamma_{\theta, 1/n}(A) = \frac{\theta^{1/n}}{\Gamma(1/n)/n} \int_A x^{(1/n)-1} e^{-\theta x} dx \xrightarrow{n \rightarrow \infty} \int_A x^{-1} e^{-\theta x} dx,$$

also $\nu(dx) = x^{-1} e^{-\theta x} dx$. \diamond

Für unbegrenzt teilbare Verteilungen auf \mathbb{R} wollen wir eine ähnliche Darstellung angeben wie im vorigen Satz. Nun ist aber eine unbegrenzt teilbare, reelle Zufallsvariable X nicht einfach die Differenz zweier unbegrenzt teilbarer, nichtnegativer Zufallsvariablen, wie etwa die Normalverteilung zeigt. Außerdem haben wir nun mehr Freiheiten, wenn wir X wie im letzten Beweis als Summe unabhängiger Zufallsvariablen X_k darstellen wollen.

Wir definieren also eine reelle Zufallsvariable X als Summe unabhängiger Zufallsvariablen

$$X = b + X^N + X_0 + \sum_{k=1}^{\infty} (X_k - \alpha_k), \quad (16.8)$$

wobei $b \in \mathbb{R}$, $X^N = \mathcal{N}_{0, \sigma^2}$ für ein $\sigma^2 \geq 0$ und $\mathbf{P}_{X_k} = \text{CPoi}_{\nu_k}$ mit Intensitätsmaß ν_k , das auf $I_k := (-1/k, -1/(k+1)] \cup [1/(k+1), 1/k]$ (mit $1/0 = \infty$), $k \in \mathbb{N}_0$, konzentriert liegt. Ferner ist $\alpha_k = \mathbf{E}[X_k] = \int x \nu_k(dx)$ für $k \geq 1$. Für die fast sichere Konvergenz der Reihe ist hinreichend (und, wie eine einfache Anwendung des Kolmogorov'schen Dreireihensatzes zeigt, notwendig), dass

$$\sum_{k=1}^{\infty} \mathbf{Var}[X_k] < \infty. \quad (16.9)$$

(Anders als in Satz 16.14 ist nicht notwendig, dass $\sum_{k=1}^{\infty} \mathbf{E}[|X_k - \alpha_k|] < \infty$. Dies ergibt eine größere Freiheit bei der Wahl von ν als im Fall nichtnegativer Zufallsvariablen.) Nun ist $\mathbf{Var}[X_k] = \int x^2 \nu_k(dx)$. Setzen wir also $\nu = \sum_{k=0}^{\infty} \nu_k$, so ist

(16.9) äquivalent zu der Bedingung $\int_{(-1,1)} x^2 \nu(dx) < \infty$. Da ν_0 stets endlich ist, ist dies wiederum äquivalent zu $\int (x^2 \wedge 1) \nu(dx) < \infty$.

Definition 16.16. Ein σ -endliches Maß ν auf \mathbb{R} mit $\nu(\{0\}) = 0$ und

$$\int (x^2 \wedge 1) \nu(dx) < \infty \quad (16.10)$$

heißt **kanonisches Maß**. Sind $\sigma^2 \geq 0$ und $b \in \mathbb{R}$, so heißt (σ^2, b, ν) ein **kanonisches Tripel**.

Zu jedem kanonischen Tripel gehört über die Konstruktion (16.8) eine unbegrenzt teilbare Zufallsvariable. Wir setzen

$$\psi_0(t) = \log \mathbf{E}[e^{itX_0}] = \int_{I_0} (e^{itx} - 1) \nu(dx)$$

und für $k \in \mathbb{N}$

$$\psi_k(t) = \log \mathbf{E}[e^{it(X_k - \alpha_k)}] = \int_{I_k} (e^{itx} - 1 - itx) \nu(dx).$$

Also genügt

$$\psi(t) := \log \mathbf{E}[e^{itX}] = -\frac{\sigma^2}{2} t^2 + ibt + \sum_{k=0}^{\infty} \psi_k(t)$$

der Lévy-Khinchin Formel

$$\psi(t) = -\frac{\sigma^2}{2} t^2 + ibt + \int (e^{itx} - 1 - itx \mathbb{1}_{\{|x|<1\}}) \nu(dx). \quad (16.11)$$

Satz 16.17 (Lévy-Khinchin Formel). Sei $\mu \in \mathcal{M}_1(\mathbb{R})$ und

$$\psi(t) := \log \int e^{itx} \mu(dx).$$

μ ist genau dann unbegrenzt teilbar, wenn es ein kanonisches Tripel (σ^2, b, ν) gibt, sodass (16.11) gilt. Dieses Tripel ist durch (16.11) eindeutig festgelegt.

Wir nennen ν wiederum das Lévy-Maß von μ , σ^2 den Gauß'schen Koeffizienten und b die Zentrierungskonstante.

Beweis. Wir haben schon gezeigt, dass es zu jedem kanonischen Tripel (σ^2, b, ν) eine unbegrenzt teilbare Verteilung μ gibt, sodass (16.11) gilt. Es bleibt zu zeigen:

- (i) Durch (16.11) wird das kanonische Tripel eindeutig bestimmt.
- (ii) Zu jeder unbegrenzt teilbaren Verteilung existiert ein kanonisches Tripel, so dass (16.11) gilt.

(i) Eindeutigkeit. Wir setzen $g_t(x) = e^{itx} - 1 - itx \mathbb{1}_{\{|x|<1\}}$. Für jedes $x \neq 0$ ist

$$2 \geq \left| \frac{g_t(x)}{t^2(1 \wedge x^2)} \right| \xrightarrow{t \rightarrow \infty} 0.$$

Wegen (16.10) gilt nach dem Satz über majorisierte Konvergenz

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\psi(t)}{t^2} &= -\frac{\sigma^2}{2} + \lim_{t \rightarrow \infty} \frac{ib}{t} + \lim_{t \rightarrow \infty} \int_{-\infty}^{\infty} \left(\frac{g_t(x)}{t^2(1 \wedge x^2)} \right) (1 \wedge x^2) \nu(dx) \\ &= -\frac{\sigma^2}{2}. \end{aligned} \quad (16.12)$$

Da σ^2 damit eindeutig ist, können wir im Folgenden $\sigma^2 = 0$ annehmen. Definiere

$$\bar{\psi}(t) = \psi(t) - \frac{1}{2} \int_{t-1}^{t+1} \psi(s) ds. \quad (16.13)$$

Dann ist

$$\bar{\psi}(t) = \int_{\mathbb{R}} e^{itx} \left(1 - \frac{1}{2} \int_{-1}^1 e^{isx} ds \right) \nu(dx) = \int_{\mathbb{R}} e^{itx} h(x) \nu(dx), \quad (16.14)$$

wobei $h(x) = 1 - \frac{\sin(x)}{x}$ für $x \neq 0$ und $h(0) = 0$ ist. Offenbar sind h und $\hat{h}(x) = h(x)/(1 \wedge x^2)$ mit der Festsetzung $\hat{h}(0) = 1/6$ stetig und beschränkt und

$$0 < 1 - \sin(1) \leq \hat{h}(x) \leq \frac{3}{2} \quad \text{für jedes } x \in \mathbb{R}.$$

$\bar{\psi}$ ist die charakteristische Funktion von $\tilde{\nu} \in \mathcal{M}_f(\mathbb{R})$, wo $\tilde{\nu}(dx) = h(x)\nu(dx)$. Damit ist $\tilde{\nu}$ durch ψ eindeutig festgelegt und wegen $\nu(dx) = (\mathbb{1}_{\{x \neq 0\}}/h(x))\tilde{\nu}(dx)$ auch ν . Die Zahl b ergibt sich nun als Differenz der restlichen Terme.

(ii) Existenz eines kanonischen Tripels. Sei μ unbegrenzt teilbar und $\psi(t) = \log \int e^{itx} \mu(dx)$. Offenbar ist $\text{Im}(\psi)$ ungerade und $\text{Re}(\psi(t)) \leq 0$ für jedes $t \in \mathbb{R}$, also ist $\bar{\psi}(0) \geq 0$ (mit $\bar{\psi}$ aus (16.13)), und es gilt $\bar{\psi}(0) = 0$ genau dann, wenn $\text{Re}(\psi(t)) = 0$ für alle $t \in [-1, 1]$. Nach Übung 15.2.4 gilt dies genau dann, wenn $\mu = \delta_b$ gilt für ein $b \in \mathbb{R}$. In diesem Fall ist $(0, b, 0)$ das zugehörige kanonische Tripel.

Sei nun $\bar{\psi}(0) > 0$. Nach Satz 16.5 existiert eine Folge $(\nu_n)_{n \in \mathbb{N}}$ in $\mathcal{M}_f(\mathbb{R})$ mit $\text{CPoi}_{\nu_n} \xrightarrow{n \rightarrow \infty} \mu$ und $\nu_n(\{0\}) = 0$ für jedes $n \in \mathbb{N}$. Setze

$$b_n = \int x \mathbb{1}_{\{|x|<1\}} \nu_n(dx).$$

Dann ist nach (16.1) und mit g_t aus (i)

$$\psi_n(t) := \log \int e^{itx} \text{CPoi}_{\nu_n}(dx) = \int (e^{itx} - 1) \nu_n(dx) = \int g_t d\nu_n + ib_n t.$$

Wie in (16.14) ist

$$\overline{\psi}_n(t) := \psi_n(t) - \frac{1}{2} \int_{t-1}^{t+1} \psi_n(s) ds = \int e^{itx} h(x) \nu_n(dx).$$

Da $\psi_n \xrightarrow{n \rightarrow \infty} \psi$ gleichmäßig auf kompakten Mengen konvergiert (Satz 15.23(i)), und weil ψ stetig und damit lokal beschränkt ist, gilt $\overline{\psi}_n \xrightarrow{n \rightarrow \infty} \overline{\psi}$ punktweise, also

$$\int e^{itx} h(x) \nu_n(dx) \xrightarrow{n \rightarrow \infty} \overline{\psi}(t). \quad (16.15)$$

Speziell ist $\overline{\psi}_n(0) > 0$ für große n . Setzen wir $\tilde{\nu}_n(dx) = (h(x)/\overline{\psi}_n(0))\nu_n(dx) \in \mathcal{M}_1(\mathbb{R})$, so gilt $\int e^{itx} \tilde{\nu}_n(dx) \xrightarrow{n \rightarrow \infty} \overline{\psi}(t)/\overline{\psi}(0)$, wobei die rechte Seite stetig ist. Nach dem Lévy'schen Stetigkeitssatz gibt es also ein $\tilde{\nu} \in \mathcal{M}_1(\mathbb{R})$ mit $\tilde{\nu}_n \xrightarrow{n \rightarrow \infty} \tilde{\nu}$ und

$$\overline{\psi}(t) = \overline{\psi}(0) \int e^{itx} \tilde{\nu}(dx).$$

Wir setzen $\sigma^2 := 6\overline{\psi}(0)\tilde{\nu}(\{0\})$ und definieren ein kanonisches Maß ν durch

$$\nu(dx) = \frac{\overline{\psi}(0)}{h(x)} \mathbb{1}_{\{x \neq 0\}} \tilde{\nu}(dx).$$

Die Abbildung (vergleiche (15.8))

$$f_t : \mathbb{R} \rightarrow \mathbb{C}, \quad x \mapsto \begin{cases} \frac{g_t(x)}{h(x)}, & \text{falls } x \neq 0, \\ -3t^2, & \text{falls } x = 0, \end{cases}$$

ist stetig und beschränkt. Nach Konstruktion gilt

$$\int g_t d\nu_n = \overline{\psi}_n(0) \int f_t d\tilde{\nu}_n \xrightarrow{n \rightarrow \infty} \overline{\psi}(0) \int f_t d\tilde{\nu} = -\frac{\sigma^2}{2} t^2 + \int g_t d\nu.$$

Also existiert auch

$$itb := \lim_{n \rightarrow \infty} itb_n = \lim_{n \rightarrow \infty} \left(\psi_n(t) - \int g_t d\nu_n \right) = \psi(t) + \frac{\sigma^2}{2} t^2 - \int g_t d\nu,$$

und es ist

$$\psi(t) = -\frac{\sigma^2}{2} t^2 + ibt + \int g_t d\nu. \quad \square$$

Bemerkung 16.18. Es gibt mehrere Varianten der Lévy-Khinchin Formel

$$\psi(t) = -\frac{\sigma^2}{2} t^2 + ibt + \int (e^{itx} - 1 - it f(x)) \nu(dx),$$

die sich in der Funktion $it f(x)$ unterscheiden, die als Zentrierung im Integral abgezogen wird. Wir haben $f(x) = x \mathbb{1}_{\{|x| < 1\}}$ gewählt, weil dies am besten zur Konstruktion mit den Zufallsvariablen X_k passte. Jedoch ist bei gegebenem kanonischen Maß ν jede Funktion \tilde{f} möglich, für die $\int |f - \tilde{f}| d\nu < \infty$ ist, also allgemein jedes \tilde{f} , für das $|f(x) - \tilde{f}(x)|/(1 \wedge x^2)$ beschränkt ist. Eine übliche Funktion ist beispielsweise $\tilde{f}(x) = \sin(x)$. Das Lévy-Maß und der Gauß'sche Anteil σ^2 sind bei allen diesen Darstellungen gleich, jedoch ändert sich das b :

$$\tilde{b} - b = \int (f - \tilde{f}) d\nu.$$

Wenn ν ein Maß ist, das auf $(0, \infty)$ konzentriert liegt und $\int (1 \wedge x) \nu(dx) < \infty$ erfüllt, ist das hier gewählte f bezüglich ν integrierbar und kann daher durch $\tilde{f} = 0$ ersetzt werden. Wir erhalten so Satz 16.14 als Spezialfall zurück. Da jedoch die Klasse der kanonischen Maße, die auf $(0, \infty)$ konzentriert liegen, auf Grund der schwächeren Bedingung (16.10) größer ist als die in Satz 16.14 betrachtete, folgt, dass eine Verteilung mit kanonischem Tripel $(b, 0, \nu)$, wo $\nu((-\infty, 0)) = 0$ und $\int (1 \wedge x) \nu(dx) = \infty$ ist, zu einer unbegrenzt teilbaren Verteilung μ gehört, die nicht auf $[0, \infty)$ konzentriert liegt – egal wie b gewählt wird. \diamond

Wir können zu gegebener unbegrenzt teilbarer Verteilung μ wieder das kanonische Maß ν ausrechnen als den vagen Limes

$$\nu = \text{v-lim}_{n \rightarrow \infty} n \mu^{*1/n} \Big|_{(0, \infty)}. \quad (16.16)$$

Beispiel 16.19. Für die Cauchy-Verteilung Cau_a mit $\psi(t) = -a |t|$ bekommen wir auf Grund der Symmetrie $b = 0$ und nach (16.12) $\sigma^2 = -2 \lim_{t \rightarrow \infty} \psi(t)/t^2 = 0$. Schließlich ist für $A \subset \mathbb{R}$ mit $(-\varepsilon, \varepsilon) \cap A = \emptyset$ für ein $\varepsilon > 0$

$$n \text{Cau}_{1/n}(A) = \frac{1}{\pi} \int_A \frac{n^2}{1 + (nx)^2} dx \xrightarrow{n \rightarrow \infty} \frac{1}{\pi} \int_A \frac{1}{x^2} dx.$$

Also hat Cau_1 das kanonische Tripel $(0, 0, (\pi x^2)^{-1} dx)$. \diamond

Übung 16.1.1. Man zeige mit einem Varianzargument, dass eine unbegrenzt teilbare Verteilung, die auf einem endlichen Intervall konzentriert liegt, schon in einem einzigen Punkt konzentriert ist. \clubsuit

Übung 16.1.2. Sei φ unbegrenzt teilbar, und für jedes $n \in \mathbb{N}$ sei φ_n eine CFW mit $\varphi_n^n = \varphi$. Man zeige mit Hilfe des Lévy'schen Stetigkeitssatzes, dass gleichmäßig auf kompakten Mengen $\varphi_n \xrightarrow{n \rightarrow \infty} 1$ gilt und folgere hieraus, dass $\varphi(t) \neq 0$ für jedes $t \in \mathbb{R}$. \clubsuit

Übung 16.1.3. Man zeige, dass unter den Bedingungen von Satz 16.14 gilt:

$$\alpha = \sup \{x \geq 0 : \mu([0, x)) = 0\}.$$



16.2 Stabile Verteilungen

Eine Verteilung μ auf den reellen Zahlen heißt stabil, wenn für jedes $n \in \mathbb{N}$ die n -fache Faltung μ^{*n} bis auf eine lineare Größenänderung und eine Verschiebung der Lage wieder gleich μ ist. Stabilität ist also eine Selbstähnlichkeitseigenschaft. Wir zeigen zunächst, dass die Klasse stabiler Verteilungen recht übersichtlich ist und leicht parametrisiert werden kann. Danach zitieren wir Ergebnisse, nach denen stabile Verteilungen genau diejenigen Verteilungen sind, die als Grenzwerte von Summen von u.i.v. Zufallsvariablen auftreten können.

Symmetrische stabile Verteilungen

Für $\alpha \in (0, 2)$ sei

$$\theta_\alpha := \int_{\mathbb{R}} (1 - \cos(x)) |x|^{-\alpha-1} dx = \begin{cases} -2\Gamma(-\alpha) \cos(\alpha\pi/2), & \text{falls } \alpha \neq 1, \\ \pi, & \text{falls } \alpha = 1. \end{cases}$$

(Beachte, dass das Integral für $\alpha \in \mathbb{R} \setminus (0, 2)$ divergiert.) Dann ist das Maß ν_α mit Dichte $\nu_\alpha(dx)/dx = \theta_\alpha^{-1} |x|^{-\alpha-1}$ ein kanonisches Maß, denn

$$\int (1 \wedge x^2) \nu_\alpha(dx) = 2\theta_\alpha^{-1} (\alpha^{-1} + (2-\alpha)^{-1}) < \infty.$$

Sei ψ_α die logarithmierte charakteristische Funktion, die zum unbegrenzt teilbaren Maß μ_α mit kanonischem Tripel $(0, 0, \nu_\alpha)$ gehört. Nach der Lévy-Khintchin Formel ist

$$\begin{aligned} \psi_\alpha(t) &= \int_{-\infty}^{\infty} (e^{itx} - 1 - itx \mathbb{1}_{\{|x|<1\}}) \theta_\alpha^{-1} |x|^{-\alpha-1} dx \\ &= -\theta_\alpha^{-1} \int_{-\infty}^{\infty} (1 - \cos(tx)) |x|^{-\alpha-1} dx \\ &= -|t|^\alpha. \end{aligned}$$

Also ist $\varphi_\alpha(t) := e^{-|t|^\alpha}$ die charakteristische Funktion des unbegrenzt teilbaren Maßes μ_α , das wir symmetrische **stabile Verteilung** mit Index α nennen. Die Namensgebung leitet sich aus der Eigenschaft ab, dass für u.i.v. nach μ_α verteilte Zufallsvariablen X_1, X_2, \dots gilt

$$X_1 + \dots + X_n \stackrel{\mathcal{D}}{=} n^{1/\alpha} X_n \quad \text{für jedes } n \in \mathbb{N}. \quad (16.17)$$

Allgemeine stabile Verteilungen

Motiviert durch Gleichung (16.17) fassen wir nun den Begriff der stabilen Verteilung etwas allgemeiner.

Definition 16.20 (Stabile Verteilung). Sei $\mu \in \mathcal{M}_1(\mathbb{R})$ nicht in einem Punkte konzentriert, und seien X_1, X_2, \dots u.i.v. Zufallsvariablen mit Verteilung μ . Die Verteilung μ heißt **stabil im weiteren Sinne**, falls es nichtnegative Zahlen a_1, a_2, \dots und reelle Zahlen d_1, d_2, \dots gibt, sodass gilt

$$X_1 + \dots + X_n \stackrel{\mathcal{D}}{=} a_n X_1 + d_n \quad \text{für jedes } n \in \mathbb{N}. \quad (16.18)$$

μ heißt **stabil (im engeren Sinne)**, falls (16.18) gilt mit $d_1 = d_2 = \dots = 0$.

μ heißt **stabil im weiteren Sinne mit Index $\alpha \in (0, 2]$** , falls $a_n = n^{1/\alpha}$, $n \in \mathbb{N}$, gewählt werden kann und stabil (im engeren Sinne) mit Index $\alpha \in (0, 2]$, falls zudem $d_1 = d_2 = \dots = 0$ gewählt werden kann.

Bemerkung 16.21. Ist μ stabil im weiteren Sinne, so ist μ offenbar unbegrenzt teilbar. \diamond

Satz 16.22. Sei μ stabil im weiteren Sinne.

- (i) Es existiert ein $\alpha \in (0, 2]$, so dass μ stabil im weiteren Sinne mit Index α ist.
- (ii) Ist $\alpha = 2$, so ist μ eine Normalverteilung.
- (iii) Ist $\alpha \in (0, 2)$, so hat das Lévy-Maß ν von μ die Dichte

$$\frac{\nu(dx)}{dx} = \begin{cases} c^- (-x)^{-\alpha-1}, & \text{falls } x < 0, \\ c^+ x^{-\alpha-1}, & \text{falls } x > 0, \end{cases} \quad (16.19)$$

für gewisse $c^-, c^+ \geq 0$, $c^- + c^+ > 0$.

- (iv) Ist $\alpha \neq 1$, dann gibt es ein $b \in \mathbb{R}$, sodass $\mu * \delta_{-b}$ stabil mit Index α ist.
- (v) Ist $\alpha = 1$, so ist $d_n = (c^+ - c^-) n \log(n)$, $n \in \mathbb{N}$. Ist $c^- = c^+$, so ist μ eine Cauchy-Verteilung.

Bemerkung 16.23. Ist μ unbegrenzt teilbar mit Lévy-Maß ν gegeben durch (16.19), so ist $\psi(t) := \log \int e^{itx} \mu(dx)$ gegeben durch

$$\psi(t) = \begin{cases} |t|^\alpha \Gamma(-\alpha) \left[(c^+ + c^-) \cos\left(\frac{\pi\alpha}{2}\right) + i(c^+ - c^-) \sin\left(\frac{\pi\alpha}{2}\right) \right], & \alpha \neq 1, \\ -|t|(c^+ + c^-) \left[\frac{\pi}{2} + i \operatorname{sign}(t)(c^+ - c^-) \log(|t|) \right], & \alpha = 1. \end{cases} \quad (16.20)$$

\diamond

Lemma 16.24. Sei μ unbegrenzt teilbar mit kanonischem Tripel (σ^2, b, ν) , also mit log-charakteristischer Funktion $\psi(t) := \log(\int e^{itx} \mu(dx))$ gegeben durch

$$\psi(t) = -\frac{\sigma^2}{2}t^2 + ibt + \int (e^{itx} - 1 - itx \mathbb{1}_{\{|x|<1\}}) \nu(dx).$$

Seien ferner $a > 0$, $d \in \mathbb{R}$, $n \in \mathbb{N}$ und X, X_1, \dots, X_n u.i.v. Zufallsvariablen mit Verteilung μ .

- (i) $X_1 + \dots + X_n$ hat das kanonische Tripel $(n\sigma^2, nb, n\nu)$.
- (ii) $aX + d$ hat das kanonische Tripel $(a^2\sigma^2, \tilde{b}, \nu \circ m_a^{-1})$, wobei $m_a : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto ax$ die Multiplikation mit a ist und

$$\tilde{b} := ab + d + a \int (\mathbb{1}_{\{|x|<1/a\}} - \mathbb{1}_{\{|x|<1\}}) x \nu(dx). \quad (16.21)$$

Beweis. (i) Die log-charakteristische Funktion von $X_1 + \dots + X_n$ ist $n\psi$.

(ii) Die log-charakteristische Funktion von $aX + d$ ist

$$\begin{aligned} \psi_{aX+d}(t) &= \psi(at) + idt \\ &= -\frac{a^2\sigma^2}{2}t^2 + i(ab + d)t + \int (e^{iatx} - 1 - iatx \mathbb{1}_{\{|x|<1\}}) \nu(dx) \\ &= -\frac{a^2\sigma^2}{2}t^2 + i\tilde{b}t + \int (e^{iatx} - 1 - iatx \mathbb{1}_{\{|x|<1/a\}}) \nu(dx) \\ &= -\frac{a^2\sigma^2}{2}t^2 + i\tilde{b}t + \int (e^{itx} - 1 - itx \mathbb{1}_{\{|x|<1\}}) \nu \circ m_a^{-1}(dx). \end{aligned} \quad \square$$

Lemma 16.25 (Skalierung des kanonischen Tripels). Unter den Bedingungen von Satz 16.22 sei (σ^2, b, ν) das kanonische Tripel von μ .

(i) Es gelten

$$(a_n^2 - n)\sigma^2 = 0 \quad \text{für alle } n \in \mathbb{N} \quad (16.22)$$

und (mit m_{a_n} wie in Lemma 16.24)

$$n\nu = \nu \circ m_{a_n}^{-1} \quad \text{für alle } n \in \mathbb{N}. \quad (16.23)$$

(ii) Ist $\nu = 0$, so gilt $a_n = n^{1/2}$ für alle $n \in \mathbb{N}$ und

$$d_n = b(n - n^{1/2}). \quad (16.24)$$

(iii) Ist $\alpha \in (0, 2)$, ν durch (16.19) gegeben und $a_n = n^{1/\alpha}$, so gilt

$$d_n = \left(b + \frac{c^+ - c^-}{\alpha - 1} \right) (n - n^{1/\alpha}), \quad \text{falls } \alpha \neq 1, \quad (16.25)$$

und

$$d_n = (c^+ - c^-) n \log(n), \quad \text{falls } \alpha = 1. \quad (16.26)$$

Beweis. (i) Sei $(a_n^2 \sigma^2, \tilde{b}_n, \nu \circ m_{a_n}^{-1})$ das kanonische Tripel von $a_n X + d_n$, wie im vorangehenden Lemma bestimmt, und $(n\sigma^2, nb, n\nu)$ dasjenige von $X_1 + \dots + X_n$. Nach (16.18) und wegen der Eindeutigkeit des kanonischen Tripels (Satz 16.17), folgt $a_n^2 \sigma^2 = n\sigma^2$, $\tilde{b}_n = nb$, und $\nu \circ m_{a_n}^{-1} = n\nu$.

(ii) Ist $\nu = 0$, so ist $\sigma^2 > 0$, weil μ nach Voraussetzung nicht in einem Punkte konzentriert ist. Nach (16.22) ist also $a_n = n^{1/2}$. Nach Lemma 16.24(ii) ist $nb = \tilde{b}_n = bn^{1/2} + d_n$, also gilt (16.24).

(iii) Wir berechnen \tilde{b}_n mit (16.21) explizit:

$$\begin{aligned} nb = \tilde{b}_n &= bn^{1/\alpha} + d_n - n^{1/\alpha} \int \mathbb{1}_{\{n^{-1/\alpha} \leq |x| < 1\}} x \nu(dx) \\ &= bn^{1/\alpha} + d_n - n^{1/\alpha} (c^+ - c^-) \int_{n^{-1/\alpha}}^1 x^{-\alpha} dx \\ &= bn^{1/\alpha} + d_n - (c^+ - c^-) \begin{cases} (1-\alpha)^{-1} (n^{1/\alpha} - n), & \text{falls } \alpha \neq 1, \\ n \log(n), & \text{falls } \alpha = 1. \end{cases} \end{aligned}$$

Auflösen nach d_n ergibt (16.25) und (16.26). \square

Beweis (von Satz 16.22). Wir unterscheiden die Fälle, wo $a_n n^{-1/2} \xrightarrow{n \rightarrow \infty} \infty$ gilt, und wo dies nicht gilt.

1. Fall. Es gelte $\liminf_{n \rightarrow \infty} a_n n^{-1/2} < \infty$. Sei $C \in [1, \infty)$ und $(n_k)_{k \in \mathbb{N}}$ eine Teilfolge, so dass $a_{n_k} n_k^{-1/2} \leq C$ für jedes $k \in \mathbb{N}$ gilt. Dann gilt für jedes $x \in \mathbb{R} \setminus \{0\}$

$$C^2 \geq n_k^{-1} (1 \vee a_{n_k}^2) \geq \frac{n_k^{-1} (1 \wedge a_{n_k}^2 x^2)}{1 \wedge x^2} \xrightarrow{k \rightarrow \infty} 0.$$

Wegen (16.23) und (16.10) folgt mit dem Satz über majorisierte Konvergenz

$$\int_{-\infty}^{\infty} (1 \wedge x^2) \nu(dx) = \int_{-\infty}^{\infty} \frac{n_k^{-1} (1 \wedge a_{n_k}^2 x^2)}{1 \wedge x^2} (1 \wedge x^2) \nu(dx) \xrightarrow{k \rightarrow \infty} 0.$$

Das heißt, es gilt $\nu = 0$. Nach Lemma 16.25(ii) ist $\mu * \delta_{-b}$ stabil mit Index 2. Damit ist (i) gezeigt.

2. Fall. Es gelte

$$a_n n^{-1/2} \xrightarrow{n \rightarrow \infty} \infty. \quad (16.27)$$

Nach (16.22) ist $\sigma^2 = 0$, also $\nu \neq 0$. Wir definieren die Funktion

$$F(x) = \begin{cases} \nu([x, \infty)), & \text{falls } x > 0, \\ \nu((-\infty, x]), & \text{falls } x < 0. \end{cases}$$

Da $\nu \neq 0$ ist, gibt es ein $x_0 \in \mathbb{R} \setminus \{0\}$ mit $F(x_0) > 0$. Aus Symmetriegründen können wir annehmen, dass $x_0 > 0$ ist. Aus (16.23) folgt

$$n F(x) = F(x/a_n) \quad \text{für jedes } x \in \mathbb{R} \setminus \{0\}, n \in \mathbb{N},$$

und damit

$$F\left(\left(\frac{a_{n+1}}{a_n}\right)^k x_0\right) = \left(\frac{n}{n+1}\right)^k F(x_0) \quad \text{für jedes } k \in \mathbb{Z}.$$

Mit anderen Worten heißt dies

$$F(x) = (x/x_0)^{-\alpha_n} F(x_0) \quad \text{für jedes } x \in \{(a_{n+1}/a_n)^k x_0 : k \in \mathbb{Z}\},$$

wobei $\alpha_n := \log((n+1)/n)/\log(a_{n+1}/a_n)$ ist. Da F monoton fallend ist und $F(x) \xrightarrow{x \rightarrow \infty} 0$ gilt, haben wir $\alpha_n > 0$ für jedes $n \in \mathbb{N}$ und

$$\left(\frac{m}{m+1}\right) \left(\frac{x}{x_0}\right)^{-\alpha_m} \leq \frac{F(x)}{F(x_0)} \leq \left(\frac{n+1}{n}\right) \left(\frac{x}{x_0}\right)^{-\alpha_n} \quad \text{für } x > 0, m, n \in \mathbb{N}.$$

Lassen wir $x \rightarrow \infty$ gehen, folgt $\alpha_m \geq \alpha_n$ und aus Symmetriegründen auch $\alpha_m \leq \alpha_n$. Wir setzen also $\alpha := \alpha_1 > 0$ und erhalten $a_n = n^{1/\alpha}$ für jedes $n \in \mathbb{N}$ (beachte, dass $a_1 = 1$ aus (16.18) folgt). Nach der Annahme (16.27) ist $\alpha < 2$. Damit ist (i) gezeigt.

Es gilt $F(1) = x_0^\alpha F(x_0) > 0$ und $F(x) = x^{-\alpha} F(1)$ für jedes $x > 0$. Analog erhalten wir $F'(x) = (-x)^{-\alpha} F(-1)$ für $x < 0$ (mit dem selben $\alpha \in (0, 2)$), denn das ist ja schon durch die Folge $(a_n)_{n \in \mathbb{N}}$ festgelegt). Setzen wir nun $c^+ = \alpha \nu([1, \infty))$ und $c^- := \alpha \nu((-\infty, -1])$, so folgt (16.19) und damit (iii).

Die Aussagen (iv) und (v) folgen direkt aus Lemma 16.25. \square

Konvergenz gegen stabile Verteilungen

Zur Abrundung des Bildes zitieren wir aus [59, Kapitel XVII.5] (siehe auch [68] und [137]) Sätze darüber, dass nur stabile Verteilungen als Grenzverteilungen reskalierter Summen von u.i.v. Zufallsvariablen X_1, X_2, \dots auftreten können, wie die genauen Skalierungen aussehen, und welche Verteilungen P_{X_1} zu welchen Grenzverteilungen führen.

Seien im Folgenden X, X_1, X_2, \dots u.i.v. Zufallsvariablen und $S_n = X_1 + \dots + X_n$ für $n \in \mathbb{N}$.

Definition 16.26 (Anziehungsbereich einer Verteilung). Sei $\mu \in \mathcal{M}_1(\mathbb{R})$ nicht auf einen Punkt konzentriert. Der **Anziehungsbereich** (domain of attraction) $\text{Dom}(\mu) \subset \mathcal{M}_1(\mathbb{R})$ ist die Menge aller Verteilungen P_X mit der Eigenschaft, dass es Folgen reeller Zahlen $(a_n)_{n \in \mathbb{N}}$ und $(d_n)_{n \in \mathbb{N}}$ gibt mit

$$\frac{S_n - d_n}{a_n} \xrightarrow{n \rightarrow \infty} \mu.$$

Ist μ stabil (im weiteren Sinne) mit Index $\alpha \in (0, 2]$, so liegt P_X im **normalen Anziehungsbereich** (domain of normal attraction), falls $a_n = n^{1/\alpha}$ gewählt werden kann.

Satz 16.27. Sei $\mu \in \mathcal{M}_1(\mathbb{R})$ nicht auf einen Punkt konzentriert. Genau dann ist $\text{Dom}(\mu) \neq \emptyset$, wenn μ stabil (im weiteren Sinne) ist. Es gilt dann $\mu \in \text{Dom}(\mu)$.

Eine wichtige Rolle spielt im Folgenden die Funktion

$$U(x) := \mathbf{E}[X^2 \mathbb{1}_{\{|X| \leq x\}}]. \quad (16.28)$$

Eine Funktion $H : (0, \infty) \rightarrow (0, \infty)$ heißt **langsam variierend** bei ∞ , falls

$$\lim_{x \rightarrow \infty} \frac{H(\gamma x)}{H(x)} = 1 \quad \text{für alle } \gamma > 0.$$

Wir nehmen im Folgenden an, dass es ein $\alpha \in (0, 2]$ gibt, mit der Eigenschaft:

$$U(x) x^{\alpha-2} \text{ ist langsam variierend bei } \infty. \quad (16.29)$$

Satz 16.28. (i) Liegt \mathbf{P}_X im Anziehungsbereich einer Verteilung, dann existiert ein $\alpha \in (0, 2]$, sodass (16.29) gilt.

(ii) Im Falle $\alpha = 2$ gilt: Ist \mathbf{P}_X nicht in einem Punkt konzentriert, so ist (16.29) hinreichend dafür, dass \mathbf{P}_X im Anziehungsbereich einer Verteilung liegt.

(iii) Im Falle $\alpha \in (0, 2)$ gilt: Genau dann liegt \mathbf{P}_X im Anziehungsbereich einer Verteilung, wenn (16.29) gilt und

$$p := \lim_{x \rightarrow \infty} \frac{\mathbf{P}[X \geq x]}{\mathbf{P}[|X| \geq x]} \quad \text{existiert.} \quad (16.30)$$

Satz 16.29. Es sei \mathbf{P}_X im Anziehungsbereich einer α -stabilen Verteilung – es gelte also Bedingung (ii) oder (iii) aus Satz 16.28 – und es sei $(a_n)_{n \in \mathbb{N}}$ so gewählt, dass

$$C := \lim_{n \rightarrow \infty} \frac{n U(a_n)}{a_n^2} \in (0, \infty)$$

existiert. Es sei ferner μ diejenige stabile Verteilung mit Index α , deren charakteristische Funktion durch (16.20) gegeben ist mit $c^+ = Cp$ und $c^- = C(1 - p)$.

(i) Im Falle $\alpha \in (0, 1)$ sei $d_n \equiv 0$.

(ii) Im Falle $\alpha = 2$ und $\mathbf{Var}[X] < \infty$ sei $\mathbf{E}[X] = 0$.

(iii) Im Falle $\alpha \in (1, 2]$ sei $d_n = n \mathbf{E}[X]$ für jedes $n \in \mathbb{N}$.

(iv) Im Falle $\alpha = 1$ sei $d_n = n a_n \mathbf{E}[\sin(X/a_n)]$ für jedes $n \in \mathbb{N}$.

Dann gilt

$$\frac{S_n - d_n}{a_n} \xrightarrow{n \rightarrow \infty} \mu.$$

Korollar 16.30. Liegt \mathbf{P}_X im Anziehungsbereich einer stabilen Verteilung mit Index α , so gilt $\mathbf{E}[|X|^\beta] < \infty$ für alle $\beta \in (0, \alpha)$ und $\mathbf{E}[|X|^\beta] = \infty$, falls $\beta > \alpha$ und $\alpha < 2$.

Übung 16.2.1. Sei μ eine α -stabile Verteilung und φ die zugehörige charakteristische Funktion.

- (i) Man zeige durch eine direkte Rechnung, die nur die Definition der stabilen Verteilung verwendet, dass $|\varphi(t) - 1| \leq C|t|^\alpha$ gilt für t nahe bei 0 (für ein $C < \infty$).
- (ii) Man folgere mit Hilfe von Übung 15.3.2, dass $\mu = \delta_0$ gilt, falls $\alpha > 2$ ist.
- (iii) Man modifiziere das Argument, um zu zeigen, dass für $\alpha > 2$ auch α -stabile Verteilungen im weiteren Sinne notwendigerweise trivial sind. \clubsuit

Übung 16.2.2. Man zeige: Die Verteilung auf \mathbb{R} mit Dichte $f(x) = \frac{1 - \cos(x)}{\pi x^2}$ ist nicht unbegrenzt teilbar. \clubsuit

Übung 16.2.3. Sei Φ die Verteilungsfunktion der Standardnormalverteilung $\mathcal{N}_{0,1}$ und $F : \mathbb{R} \rightarrow [0, 1]$ definiert durch

$$F(x) = \begin{cases} 2(1 - \Phi(x^{-1/2})), & \text{falls } x > 0, \\ 0, & \text{sonst.} \end{cases}$$

Man zeige:

- (i) F ist die Verteilungsfunktion einer $\frac{1}{2}$ -stabilen Verteilung.
- (ii) Sind X_1, X_2, \dots u.i.v. mit Verteilungsfunktion F , so divergiert $\frac{1}{n} \sum_{k=0}^n X_k$ für $n \rightarrow \infty$ fast sicher. \clubsuit

Hinweis: Man bestimme die Dichte von F und zeige, dass die Laplace Transformierte gegeben ist durch $\lambda \mapsto e^{-\sqrt{2\lambda}}$. \clubsuit

Übung 16.2.4. Welche der folgenden Verteilungen liegen im Anziehungsbereich einer stabilen Verteilung und gegebenenfalls zu welchem Parameter?

- (i) Die Verteilung auf \mathbb{R} mit Dichte

$$f(x) = \begin{cases} \varrho \frac{1}{1+\alpha} |x|^\alpha, & \text{falls } x < -1, \\ (1 - \varrho) \frac{1}{1+\beta} x^\beta, & \text{falls } x > 1, \\ 0, & \text{sonst.} \end{cases}$$

Dabei sind $\alpha, \beta < -1$ und $\varrho \in [0, 1]$.

- (ii) Die Exponentialverteilung \exp_θ für $\theta > 0$.
- (iii) Die Verteilung auf \mathbb{N} mit Gewichten $c n^\alpha$ falls n gerade ist und $c n^\beta$, falls n ungerade ist. Dabei sind $\alpha, \beta < -1$, und $c = (2^\alpha \zeta(-\alpha) + (1 - 2^\beta) \zeta(-\beta))^{-1}$ (ζ ist die Riemann'sche Zetafunktion) ist die Normierungskonstante. ♣

17 Markovketten

Markovprozesse mit abzählbarem Zustandsraum (und diskreter Zeit) sind trotz ihrer Simplizität interessante mathematische Objekte, mit denen sich eine Vielzahl von Phänomenen modellieren lässt. Wir bringen hier einen Einblick in die grundlegenden Begriffe und schauen dann Beispiele etwas detaillierter an. Der Zusammenhang mit der (diskreten) Potentialtheorie wird erst in Kapitel 19 untersucht. Beim ersten Lesen kann in Abschnitt 17.1 die (etwas abstrakte) Konstruktion von allgemeinen Markovprozessen übersprungen werden. Als weitere Lektüre empfehlen sich beispielsweise [132, 72, 153, 24, 98, 32, 70, 162, 126, 133].

17.1 Begriffsbildung und Konstruktion

Im Folgenden sei E stets ein polnischer Raum mit Borel'scher σ -Algebra $\mathcal{B}(E)$, $I \subset \mathbb{R}$ und $(X_t)_{t \in I}$ ein stochastischer Prozess mit Werten in E . Wir nehmen an, dass $(\mathcal{F}_t)_{t \in I} = \mathbb{F} = \sigma(X)$ die von X erzeugte Filtration ist.

Definition 17.1. Wir sagen, dass X die **elementare Markoveigenschaft (ME)** hat, falls für jedes $A \in \mathcal{B}(E)$ und je zwei $s, t \in I$ mit $s \leq t$ gilt

$$\mathbf{P}[X_t \in A | \mathcal{F}_s] = \mathbf{P}[X_t \in A | X_s].$$

Bemerkung 17.2. Ist E abzählbar, so hat X genau dann die elementare Markoveigenschaft, wenn für jedes $n \in \mathbb{N}$ und alle $s_1 < \dots < s_n < t$ und $i_1, \dots, i_n, i \in E$ mit $\mathbf{P}[X_{s_1} = i_1, \dots, X_{s_n} = i_n] > 0$ gilt

$$\mathbf{P}[X_t = i | X_{s_1} = i_1, \dots, X_{s_n} = i_n] = \mathbf{P}[X_t = i | X_{s_n} = i_n]. \quad (17.1)$$

In der Tat impliziert (17.1) natürlich sofort die elementare Markoveigenschaft. Habe nun X die elementare Markoveigenschaft. Wir bemerken (siehe (8.6)), dass für fast alle $\omega \in \{X_{s_n} = i_n\}$ gilt $\mathbf{P}[X_t = i | X_{s_n}](\omega) = \mathbf{P}[X_t = i | X_{s_n} = i_n]$. Also gilt mit $A := \{X_{s_1} = i_1, \dots, X_{s_n} = i_n\}$

$$\begin{aligned} \mathbf{P}[X_t = i, X_{s_1} = i_1, \dots, X_{s_n} = i_n] \\ = \mathbf{E}[\mathbf{E}[\mathbb{1}_{\{X_t=i\}} | \mathcal{F}_{s_n}] \mathbb{1}_A] = \mathbf{E}[\mathbf{E}[\mathbb{1}_{\{X_t=i\}} | X_{s_n}] \mathbb{1}_A] \\ = \mathbf{E}[\mathbf{P}[X_t = i | X_{s_n} = i_n] \mathbb{1}_A] = \mathbf{P}[X_t = i | X_{s_n} = i_n] \mathbf{P}[A], \end{aligned}$$

wobei wir in der zweiten Gleichheit die elementare Markoveigenschaft verwendet haben. Teilen wir jetzt auf beiden Seiten durch $\mathbf{P}[A]$, so folgt (17.1). \diamond

Definition 17.3. Sei $I \subset [0, \infty)$ abgeschlossen unter Addition, und sei $0 \in I$. Ein stochastischer Prozess $X = (X_t)_{t \in I}$ heißt **Markovprozess** mit Verteilungen $(\mathbf{P}_x)_{x \in E}$ auf dem Raum (Ω, \mathcal{A}) , falls gilt:

- (i) Für jedes $x \in E$ ist X ein stochastischer Prozess auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P}_x)$ mit $\mathbf{P}_x[X_0 = x] = 1$.
- (ii) Die Abbildung $\kappa : E \times \mathcal{B}(E)^{\otimes I} \rightarrow [0, 1]$, $(x, B) \mapsto \mathbf{P}_x[X \in B]$ ist ein stochastischer Kern.
- (iii) Es gilt die (**schwache**) **Markoveigenschaft (ME)**: Für jedes $A \in \mathcal{B}(E)$, jedes $x \in E$ und je zwei $s, t \in I$ gilt

$$\mathbf{P}_x[X_{t+s} \in A | \mathcal{F}_s] = \kappa_t(X_s, A) \quad \mathbf{P}_x\text{-f.s.}$$

Hierbei definiert für jedes $t \in I$ und $x \in E$ sowie $A \in \mathcal{B}(E)$

$$\kappa_t(x, A) := \kappa(x, \{y \in E^I : y(t) \in A\}) = \mathbf{P}_x[X_t \in A]$$

den stochastischen Kern $\kappa_t : E \times \mathcal{B}(E) \rightarrow [0, 1]$ der **Übergangswahrscheinlichkeiten** von X zur Zeitdifferenz t .

Wir schreiben stets \mathbf{E}_x für die Erwartungswerte bezüglich \mathbf{P}_x und $\mathcal{L}_x[X] = \mathbf{P}_x$ sowie $\mathcal{L}_x[X | \mathcal{F}] = \mathbf{P}_x[X \in \cdot | \mathcal{F}]$ (für eine reguläre Version der bedingten Verteilung von X gegeben \mathcal{F}) und so fort.

Ist E höchstens abzählbar, so heißt X **diskreter Markovprozess**.

Im Spezialfall $I = \mathbb{N}_0$ heißt X **Markovkette**. Es heißt dann κ_n auch die Familie der n -Schritt-Übergangswahrscheinlichkeiten.

Bemerkung 17.4. Die schwache Markoveigenschaft impliziert die elementare Markoveigenschaft. In der Tat ist „schwache ME = elementare ME + zeitliche Homogenität“. \diamond

Wir verwenden im Folgenden die etwas nachlässige Bezeichnung $\mathbf{P}_{X_s}[X \in \cdot] := \kappa(X_s, \cdot)$. Wir verstehen also X_s als Startwert eines zweiten Markovprozesses mit denselben Verteilungen $(\mathbf{P}_x)_{x \in E}$.

Beispiel 17.5. Seien Y_1, Y_2, \dots u.i.v. \mathbb{R}^d -wertige Zufallsvariablen und

$$S_n^x = x + \sum_{i=1}^n Y_i \quad \text{für } x \in \mathbb{R}^d \text{ und } n \in \mathbb{N}_0.$$

Definiere W-Maße \mathbf{P}_x auf $((\mathbb{R}^d)^{\mathbb{N}_0}, (\mathcal{B}(\mathbb{R}^d))^{\otimes \mathbb{N}_0})$ durch $\mathbf{P}_x = \mathbf{P} \circ (S^x)^{-1}$. Dann ist der kanonische Prozess $X_n : (\mathbb{R}^d)^{\mathbb{N}_0} \rightarrow \mathbb{R}^d$ eine Markovkette mit Verteilungen $(\mathbf{P}_x)_{x \in \mathbb{R}^d}$. Der Prozess X heißt Irrfahrt auf \mathbb{R}^d mit Startwert x . \diamond

Beispiel 17.6. Wir können das vorangehende Beispiel leicht auf die Situation stetiger Zeit, also $I = [0, \infty)$, übertragen. Sei hierzu $(\nu_t)_{t \geq 0}$ eine Faltungshalbgruppe auf \mathbb{R}^d und $\kappa_t(x, dy) = \delta_x * \nu_t(dy)$. Zu jedem $x \in \mathbb{R}^d$ sei \mathbf{P}_x das in Satz 14.47 konstruierte Maß auf $((\mathbb{R}^d)^{[0, \infty)}, \mathcal{B}(\mathbb{R}^d)^{\otimes [0, \infty)})$ mit

$$\mathbf{P}_x \circ (X_0, X_{t_1}, \dots, X_{t_n})^{-1} = \delta_x \otimes \bigotimes_{i=0}^{n-1} \kappa_{t_{i+1}-t_i}$$

für je endlich viele Punkte $0 = t_0 < t_1 < \dots < t_n$. Man prüft leicht, dass die Abbildung $\kappa : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)^{\otimes [0, \infty)}, (x, A) \mapsto \mathbf{P}_x[A]$ ein stochastischer Kern ist. Die schwache Markoveigenschaft folgt direkt aus der Tatsache, dass die Zuwächse unabhängig und stationär sind. \diamond

Beispiel 17.7. (Siehe Beispiel 9.5 und Satz 5.36.) Sei $\theta > 0$ und $\nu_t^\theta(\{k\}) = e^{-\theta t} \frac{t^k \theta^k}{k!}$, $k \in \mathbb{N}_0$, die Faltungshalbgruppe der Poisson-Verteilung. Der Markovprozess X auf \mathbb{N}_0 mit dieser Halbgruppe heißt **Poissonprozess** mit Rate θ . \diamond

Wir wollen, ähnlich wie in Beispiel 17.6, nun etwas allgemeiner zu einer Markov'schen Halbgruppe von stochastischen Kernen einen Markovprozess herstellen.

Satz 17.8. Sei $I \subset [0, \infty)$ abgeschlossen unter Addition, und sei $(\kappa_t)_{t \in I}$ eine Markov'sche Halbgruppe stochastischer Kerne von E nach E . Dann gibt es einen Messraum (Ω, \mathcal{A}) und einen Markovprozess $((X_t)_{t \in I}, (\mathbf{P}_x)_{x \in E})$ auf (Ω, \mathcal{A}) mit Übergangswahrscheinlichkeiten

$$\mathbf{P}_x[X_t \in A] = \kappa_t(x, A) \quad \text{für alle } x \in E, A \in \mathcal{B}(E), t \in I. \quad (17.2)$$

Umgekehrt definiert für jeden Markovprozess X die Gleichung (17.2) eine Halbgruppe stochastischer Kerne. Durch (17.2) sind die endlichdimensionalen Verteilungen von X eindeutig bestimmt.

Beweis. „ \Rightarrow “ Wir konstruieren X als kanonischen Prozess. Sei $\Omega = E^{[0, \infty)}$ und $\mathcal{A} = \mathcal{B}(E)^{\otimes [0, \infty)}$. Ferner sei X_t die Projektion auf die t -te Koordinate. Für $x \in E$ definieren wir (siehe Korollar 14.43) auf (Ω, \mathcal{A}) das W-Maß \mathbf{P}_x , sodass für endlich viele Zeitpunkte $0 = t_0 < t_1 < \dots < t_n$ gilt

$$\mathbf{P}_x \circ (X_{t_0}, \dots, X_{t_n})^{-1} = \delta_x \otimes \bigotimes_{i=0}^{n-1} \kappa_{t_{i+1}-t_i}.$$

Dann ist

$$\begin{aligned} \mathbf{P}_x[X_{t_0} \in A_0, \dots, X_{t_n} \in A_n] \\ = \int_{A_{n-1}} \mathbf{P}_x[X_{t_0} \in A_0, \dots, X_{t_{n-2}} \in A_{n-2}, X_{t_{n-1}} \in dx_{n-1}] \\ \quad \kappa_{t_n-t_{n-1}}(x_{n-1}, A_n), \end{aligned}$$

also $\mathbf{P}_x[X_{t_n} \in A_n | \mathcal{F}_{t_{n-1}}] = \kappa_{t_n - t_{n-1}}(X_{t_{n-1}}, A_n)$. Damit ist X als Markovprozess erkannt. Ferner ist $\mathbf{P}_x[X_t \in A] = (\delta_x \cdot \kappa_t)(A) = \kappa_t(x, A)$.

„ \Leftarrow “ Sei nun $(X, (\mathbf{P}_x)_{x \in E})$ ein Markovprozess. Dann definiert

$$\kappa_t(x, A) := \mathbf{P}_x[X_t \in A] \quad \text{für alle } x \in E, A \in \mathcal{B}(E), t \in I,$$

einen stochastischen Kern κ_t . Nach der Markoveigenschaft ist

$$\begin{aligned} \kappa_{t+s}(x, A) &= \mathbf{P}_x[X_{t+s} \in A] = \mathbf{E}_x[\mathbf{P}_{X_s}[X_t \in A]] \\ &= \int \mathbf{P}_x[X_s \in dy] \mathbf{P}_y[X_t \in A] \\ &= \int \kappa_s(x, dy) \kappa_t(y, A) = (\kappa_s \cdot \kappa_t)(x, A). \end{aligned}$$

Also ist $(\kappa_t)_{t \in I}$ eine Markov'sche Halbgruppe. \square

Satz 17.9. Ein stochastischer Prozess $X = (X_t)_{t \in I}$ ist genau dann ein Markovprozess, wenn es einen stochastischen Kern $\kappa : E \times \mathcal{B}(E)^{\otimes I} \rightarrow [0, 1]$ gibt, sodass für jede $\mathcal{B}(E)^{\otimes I} - \mathcal{B}(\mathbb{R})$ messbare, beschränkte Funktion $f : E^I \rightarrow \mathbb{R}$ und für jedes $s \geq 0$ und $x \in E$ gilt:

$$\mathbf{E}_x[f((X_{t+s})_{t \in I}) | \mathcal{F}_s] = \mathbf{E}_{X_s}[f(X)] := \int_{E^I} \kappa(X_s, dy) f(y). \quad (17.3)$$

Beweis. „ \Leftarrow “ Die schwache Markoveigenschaft folgt aus (17.3) mit der Funktion $f(y) = \mathbb{1}_A(y(t))$, denn $\mathbf{P}_{X_s}[X_t \in A] = \mathbf{P}_x[X_{t+s} \in A | \mathcal{F}_s] = \kappa_t(X_s, A)$.

„ \Rightarrow “ Nach den üblichen Approximationsargumenten reicht es, Funktionen f zu betrachten, die nur von endlich vielen Koordinaten $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ abhängen. Wir führen den Beweis per Induktion über n .

Für $n = 1$ und f eine Indikatorfunktion ist dies die (schwache) Markoveigenschaft. Für allgemeines, messbares f folgt die Aussage nun aus den üblichen Approximationsargumenten.

Sei nun die Aussage für $n \in \mathbb{N}$ bereits gezeigt. Es reicht wiederum, für f eine Indikatorfunktion der Art $f(x) = \mathbb{1}_{B_1 \times \dots \times B_{n+1}}(x_{t_1}, \dots, x_{t_{n+1}})$ (mit $B_1, \dots, B_{n+1} \in \mathcal{B}(E)$) zu betrachten. Zusammen mit der Markoveigenschaft (dritte und fünfte Gleichheit in der folgenden Gleichungskette) und der Induktionsvoraussetzung (vierte Gleichheit) erhalten wir

$$\begin{aligned}
& \mathbf{E}_x \left[f((X_{t+s})_{t \geq 0}) \mid \mathcal{F}_s \right] \\
&= \mathbf{E}_x \left[\mathbf{E}_x \left[f((X_{t+s})_{t \geq 0}) \mid \mathcal{F}_{t_n+s} \right] \mid \mathcal{F}_s \right] \\
&= \mathbf{E}_x \left[\mathbf{E}_x \left[\mathbb{1}_{\{X_{t_{n+1}+s} \in B_{n+1}\}} \mid \mathcal{F}_{t_n+s} \right] \mathbb{1}_{B_1}(X_{t_1+s}) \cdots \mathbb{1}_{B_n}(X_{t_n+s}) \mid \mathcal{F}_s \right] \\
&= \mathbf{E}_x \left[\mathbf{P}_{X_{t_n+s}} \left[X_{t_{n+1}-t_n} \in B_{n+1} \right] \mathbb{1}_{B_1}(X_{t_1+s}) \cdots \mathbb{1}_{B_n}(X_{t_n+s}) \mid \mathcal{F}_s \right] \\
&= \mathbf{E}_{X_s} \left[\mathbf{P}_{X_{t_n}} \left[X_{t_{n+1}-t_n} \in B_{n+1} \right] \mathbb{1}_{B_1}(X_{t_1}) \cdots \mathbb{1}_{B_n}(X_{t_n}) \right] \\
&= \mathbf{E}_{X_s} \left[\mathbf{P}_{X_0} \left[X_{t_{n+1}} \in B_{n+1} \mid \mathcal{F}_{t_n} \right] \mathbb{1}_{B_1}(X_{t_1}) \cdots \mathbb{1}_{B_n}(X_{t_n}) \right] \\
&= \mathbf{E}_{X_s} \left[\mathbf{P}_{X_0} \left[X_{t_1} \in B_1, \dots, X_{t_{n+1}} \in B_{n+1} \mid \mathcal{F}_{t_n} \right] \right] \\
&= \mathbf{E}_{X_s} [f(X)]. \quad \square
\end{aligned}$$

Korollar 17.10. Ein stochastischer Prozess $(X_n)_{n \in \mathbb{N}_0}$ ist genau dann eine Markovkette, wenn

$$\mathcal{L}_x \left[(X_{n+k})_{n \in \mathbb{N}_0} \mid \mathcal{F}_k \right] = \mathcal{L}_{X_k} \left[(X_n)_{n \in \mathbb{N}_0} \right] \quad \text{für jedes } k \in \mathbb{N}_0. \quad (17.4)$$

Beweis. Wenn die bedingten Verteilungen existieren, ist nach Satz 17.9 die Gleichung (17.4) äquivalent dazu, dass X eine Markovkette ist. Zu zeigen ist also nur, dass die bedingten Verteilungen auch existieren.

Da E polnisch ist, ist $E^{\mathbb{N}_0}$ polnisch und $\mathcal{B}(E^{\mathbb{N}_0}) = \mathcal{B}(E)^{\otimes \mathbb{N}_0}$ (siehe Satz 14.8). Nach Satz 8.37 existiert also eine reguläre Version der bedingten Verteilung von $(X_{n+k})_{n \in \mathbb{N}_0}$ gegeben \mathcal{F}_k . \square

Satz 17.11. Sei $I = \mathbb{N}_0$. Ist $(X_n)_{n \in \mathbb{N}_0}$ ein stochastischer Prozess mit Verteilungen $(\mathbf{P}_x, x \in E)$, so folgt die schwache Markoveigenschaft in Definition 17.3(iii) schon aus der Existenz eines stochastischen Kerns $\kappa_1 : E \times \mathcal{B}(E) \rightarrow [0, 1]$ mit der Eigenschaft: Für jedes $A \in \mathcal{B}(E)$, jedes $x \in E$ und jedes $s \in I$ gilt

$$\mathbf{P}_x \left[X_{s+1} \in A \mid \mathcal{F}_s \right] = \kappa_1(X_s, A). \quad (17.5)$$

In diesem Fall erhält man die n -Schritt Übergangskerne κ_n induktiv durch

$$\kappa_n = \kappa_{n-1} \cdot \kappa_1 = \int_E \kappa_{n-1}(\cdot, dx) \kappa_1(x, \cdot).$$

Speziell ist die Familie $(\kappa_n)_{n \in \mathbb{N}}$ eine Markov'sche Halbgruppe, und die Verteilung von X ist durch die Angabe von κ_1 eindeutig festgelegt.

Beweis. Setze in Satz 17.9 $t_i = i$ für $i \in \mathbb{N}_0$. Für den Beweis des Satzes wurde nur (17.5) ausgenutzt. \square

Die (schwache) Markoveigenschaft eines Prozesses besagt, dass zu fester Zeit t die Zukunft (nach t) von der Vergangenheit (bis t) nur durch die Gegenwart (also den Wert zur Zeit t) abhängt. Wir können diesen Begriff verallgemeinern, indem wir statt fester Zeiten auch Stoppzeiten zulassen.

Definition 17.12. Sei $I \subset [0, \infty)$ abgeschlossen unter Addition. Ein Markovprozess $(X_t)_{t \in I}$ mit Verteilungen $(\mathbf{P}_x, x \in E)$ hat die **starke Markoveigenschaft**, falls für jede f.s. endliche Stoppzeit τ und jede $\mathcal{B}(E)^{\otimes I} - \mathcal{B}(\mathbb{R})$ messbare, beschränkte Funktion $f : E^I \rightarrow \mathbb{R}$ sowie jedes $x \in E$ gilt:

$$\mathbf{E}_x [f((X_{\tau+t})_{t \in I}) | \mathcal{F}_\tau] = \mathbf{E}_{X_\tau} [f(X)] := \int_{E^I} \kappa(X_\tau, dy) f(y). \quad (17.6)$$

Bemerkung 17.13. Ist I höchstens abzählbar, so ist die starke Markoveigenschaft äquivalent dazu, dass für jede fast sicher endliche Stoppzeit τ gilt

$$\mathcal{L}_x [(X_{\tau+t})_{t \in \mathbb{N}_0} | \mathcal{F}_\tau] = \mathcal{L}_{X_\tau} [(X_t)_{t \in \mathbb{N}_0}] := \kappa(X_\tau, \cdot). \quad (17.7)$$

Dies folgt genau wie in Korollar 17.10. \diamond

Die meisten relevanten Markovprozesse besitzen auch die starke Markoveigenschaft. Statt hier den Begriff der Relevanz zu diskutieren, was sich wohl kaum erschöpfend machen ließe, wollen wir lieber zeigen, dass für abzählbare Zeitmenge die starke Markoveigenschaft aus der schwachen folgt. In zeitstetigen Situationen hingegen muss man im Allgemeinen mehr arbeiten, um die starke Markoveigenschaft zu etablieren.

Satz 17.14. Ist $I \subset [0, \infty)$ höchstens abzählbar und abgeschlossen unter Addition, so hat jeder Markovprozess $(X_n)_{n \in I}$ mit Verteilungen $(\mathbf{P}_x)_{x \in E}$ die starke Markoveigenschaft.

Beweis. Sei $f : E^I \rightarrow \mathbb{R}$ messbar und beschränkt. Dann ist für jedes $s \in I$ die Zufallsvariable $\mathbb{1}_{\{\tau=s\}} \mathbf{E}_x [f((X_{s+t})_{t \in I}) | \mathcal{F}_\tau]$ messbar bezüglich \mathcal{F}_s . Mit der Turmeigenschaft der bedingten Erwartung und Satz 17.9 in der dritten Gleichheit erhalten wir daher

$$\begin{aligned} \mathbf{E}_x [f((X_{\tau+t})_{t \in I}) | \mathcal{F}_\tau] &= \sum_{s \in I} \mathbb{1}_{\{\tau=s\}} \mathbf{E}_x [f((X_{s+t})_{t \in I}) | \mathcal{F}_\tau] \\ &= \sum_{s \in I} \mathbf{E}_x [\mathbb{1}_{\{\tau=s\}} \mathbf{E}_x [f((X_{s+t})_{t \in I}) | \mathcal{F}_s] | \mathcal{F}_\tau] \\ &= \sum_{s \in I} \mathbf{E}_x [\mathbb{1}_{\{\tau=s\}} \mathbf{E}_{X_s} [f((X_t)_{t \in I})] | \mathcal{F}_\tau] \\ &= \mathbf{E}_{X_\tau} [f((X_t)_{t \in I})]. \end{aligned} \quad \square$$

Wir bringen eine einfache Anwendung der starken Markoveigenschaft.

Satz 17.15 (Reflexionsprinzip). Seien Y_1, Y_2, \dots u.i.v. reelle Zufallsvariablen mit symmetrischer Verteilung $\mathcal{L}[Y_1] = \mathcal{L}[-Y_1]$. Setze $X_0 = 0$ und $X_n := Y_1 + \dots + Y_n$ für $n \in \mathbb{N}$. Dann gilt für jedes $n \in \mathbb{N}_0$ und $a > 0$

$$\mathbf{P} \left[\sup_{m \leq n} X_m \geq a \right] \leq 2 \mathbf{P}[X_n \geq a] - \mathbf{P}[X_n = a]. \quad (17.8)$$

Gilt $\mathbf{P}[Y_1 \in \{-1, 0, 1\}] = 1$, so gilt für $a \in \mathbb{N}$ in (17.8) sogar Gleichheit.

Beweis. Sei $a > 0$ und $n \in \mathbb{N}$. Definiere die bei $(n+1)$ abgeschnittene Zeit des ersten Überschreitens von a

$$\tau := \inf\{m \geq 0 : X_m \geq a\} \wedge (n+1).$$

Dann ist τ eine beschränkte Stopzeit und

$$\sup_{m \leq n} X_m \geq a \iff \tau \leq n.$$

Setze $f(m, X) = \mathbb{1}_{\{m \leq n\}} (\mathbb{1}_{\{X_{n-m} > a\}} + \frac{1}{2} \mathbb{1}_{\{X_{n-m} = a\}})$. Dann ist

$$f(\tau, (X_{\tau+m})_{m \in \mathbb{N}_0}) = \mathbb{1}_{\{\tau \leq n\}} (\mathbb{1}_{\{X_n > a\}} + \frac{1}{2} \mathbb{1}_{\{X_n = a\}}).$$

Die starke Markoveigenschaft von X liefert

$$\mathbf{E}_0 \left[f \left(\tau, (X_{\tau+m})_{m \geq 0} \right) \mid \mathcal{F}_\tau \right] = \varphi(\tau, X_\tau),$$

wobei $\varphi(m, x) = \mathbf{E}_x[f(m, X)]$. (Hierbei bezeichnet \mathbf{E}_x die Erwartung für X , falls $X_0 = x$.) Wegen der Symmetrie der Y_i ist

$$\varphi(m, x) \begin{cases} \geq \frac{1}{2}, & \text{falls } m \leq n \text{ und } x \geq a, \\ = \frac{1}{2}, & \text{falls } m \leq n \text{ und } x = a, \\ = 0, & \text{falls } m > n. \end{cases}$$

Also gilt

$$\begin{aligned} \{\tau \leq n\} &= \{\tau \leq n\} \cap \{X_\tau \geq a\} \subset \left\{ \varphi(\tau, X_\tau) \geq \frac{1}{2} \right\} \cap \{\tau \leq n\} \\ &= \{\varphi(\tau, X_\tau) > 0\} \cap \{\tau \leq n\}. \end{aligned}$$

Nun folgt (17.8) aus

$$\begin{aligned} \mathbf{P}[X_n > a] + \frac{1}{2} \mathbf{P}[X_n = a] &= \mathbf{E} [f(\tau, (X_{\tau+m})_{m \geq 0})] \\ &= \mathbf{E}_0 [\varphi(\tau, X_\tau) \mathbb{1}_{\{\tau \leq n\}}] \geq \frac{1}{2} \mathbf{P}_0 [\tau \leq n]. \end{aligned} \quad (17.9)$$

Gilt $\mathbf{P}[Y_1 \in \{-1, 0, 1\}] = 1$, und ist $a \in \mathbb{N}$, so ist $X_\tau = a$, falls $\tau \leq n$. Also ist

$$\{\varphi(\tau, X_\tau) > 0\} \cap \{\tau \leq n\} = \left\{ \varphi(\tau, X_\tau) = \frac{1}{2} \right\} \cap \{\tau \leq n\}.$$

Daher gilt Gleichheit im letzten Schritt von (17.9) und damit auch in (17.8). \square

Übung 17.1.1. Sei $I \subset \mathbb{R}$ $X = (X_t)_{t \in I}$ ein stochastischer Prozess. Definiere für $t \in I$ die σ -Algebren, die die Vergangenheit bis und die Zukunft ab t kodieren:

$$\mathcal{F}_{\leq t} := \sigma(X_s : s \in I, s \leq t) \quad \text{und} \quad \mathcal{F}_{\geq t} := \sigma(X_s : s \in I, s \geq t).$$

Man zeige: X hat genau dann die elementare Markoveigenschaft, wenn für jedes $t \in I$ die σ -Algebren $\mathcal{F}_{\leq t}$ und $\mathcal{F}_{\geq t}$ unabhängig sind gegeben $\sigma(X_t)$ (vergleiche Definition 12.20).

Mit anderen Worten: Ein Prozess hat die elementare Markoveigenschaft genau dann, wenn Vergangenheit und Zukunft unabhängig sind gegeben die Gegenwart. \clubsuit

17.2 Diskrete Markovketten, Beispiele

Sei E höchstens abzählbar und $I = \mathbb{N}_0$. Ein Markovprozess $X = (X_n)_{n \in \mathbb{N}_0}$ auf E ist nach Definition 17.3 eine diskrete Markovkette (oder Markovkette mit diskretem Zustandsraum).

Ist X eine diskrete Markovkette, so ist $(\mathbf{P}_x)_{x \in E}$ festgelegt durch die Angabe von der **Übergangsmatrix**

$$p = (p(x, y))_{x, y \in E} := (\mathbf{P}_x[X_1 = y])_{x, y \in E}.$$

Die n -Schrittübergangswahrscheinlichkeiten

$$p^{(n)}(x, y) := \mathbf{P}_x[X_n = y]$$

ergeben sich nämlich als n -faches Matrixprodukt

$$p^{(n)}(x, y) = p^n(x, y),$$

wobei

$$p^n(x, y) = \sum_{z \in E} p^{n-1}(x, z)p(z, y)$$

und $p^0 = I$ die Einheitsmatrix ist.

Durch Iteration folgt die **Chapman-Kolmogorov'sche Gleichung** (siehe (14.14)) für alle $m, n \in \mathbb{N}_0$ und $x, y \in E$

$$p^{(m+n)}(x, y) = \sum_{z \in E} p^{(m)}(x, z)p^{(n)}(z, y). \quad (17.10)$$

Definition 17.16. Eine Matrix $(p(x, y))_{x, y \in E}$ mit nichtnegativen Einträgen und

$$\sum_{y \in E} p(x, y) = 1 \quad \text{für jedes } x \in E$$

heißt **stochastische Matrix** auf E .

Nun ist eine stochastische Matrix im Wesentlichen ein stochastischer Kern von E nach E . In Satz 17.8 hatten wir gesehen, dass zu der Halbgruppe von Kernen $(p^n)_{n \in \mathbb{N}}$ genau eine diskrete Markovkette existiert, deren Übergangswahrscheinlichkeiten durch p gegeben sind. Die dort angegebenen Argumente waren eher abstrakter Natur. Wir wollen hier eine Konstruktion von X angeben, mit der man beispielsweise auch eine Computersimulation bauen kann.

Sei $(R_n)_{n \in \mathbb{N}_0}$ eine unabhängige Familie von Zufallsvariablen mit Werten in E^E mit der Eigenschaft, dass

$$\mathbf{P}[R_n(x) = y] = p(x, y) \quad \text{für alle } x, y \in E. \quad (17.11)$$

Beispielsweise wähle man $(R_n(x), x \in E, n \in \mathbb{N})$ als eine unabhängige Familie von Zufallsvariablen mit Werten in E und Verteilungen

$$\mathbf{P}[R_n(x) = y] = p(x, y) \quad \text{für alle } x, y \in E \text{ und } n \in \mathbb{N}_0.$$

Man beachte aber, dass wir in (17.11) weder die Unabhängigkeit der Zufallsvariablen $(R_n(x), x \in E)$ gefordert haben, noch, dass alle R_n die selbe Verteilung haben (lediglich die eindimensionalen Randverteilungen sind festgelegt). In der Tat möchte man in vielen wichtigen Anwendungen wohldefinierte Abhängigkeitsstrukturen haben, um Markovketten mit unterschiedlichen Startzuständen miteinander zu *koppeln*. Diesen Faden verfolgen wir erst in Abschnitt 18.2 weiter.

Für $x \in E$ definiere

$$X_0^x = x \quad \text{und} \quad X_n^x = R_n(X_{n-1}^x) \quad \text{für } n \in \mathbb{N}.$$

Schließlich definieren wir $\mathbf{P}_x := \mathcal{L}[X^x]$ als die Verteilung von X^x , also als ein W-Maß auf dem Folgenraum $(E^{\mathbb{N}_0}, \mathcal{B}(E)^{\otimes \mathbb{N}_0})$.

Satz 17.17. (i) Der kanonische Prozess X auf $(E^{\mathbb{N}_0}, \mathcal{B}(E)^{\otimes \mathbb{N}_0})$ ist bezüglich der Verteilung $(\mathbf{P}_x)_{x \in E}$ eine Markovkette mit Übergangsmatrix p .
(ii) Insbesondere gehört zu jeder stochastischen Matrix p genau eine diskrete Markovkette X mit Übergangswahrscheinlichkeiten p .

Beweis. „(ii)“ folgt aus (i), da Satz 17.11 die Eindeutigkeit von X liefert.

„(i)“ Für $n \in \mathbb{N}_0$ und $x, y, z \in E$ gilt nach Konstruktion

$$\begin{aligned}
\mathbf{P}_x[X_{n+1} = z \mid \mathcal{F}_n, X_n = y] &= \mathbf{P}[X_{n+1}^x = z \mid \sigma(R_m, m \leq n), X_n^x = y] \\
&= \mathbf{P}[R_{n+1}(X_n^x) = z \mid \sigma(R_m, m \leq n), X_n^x = y] \\
&= \mathbf{P}[R_{n+1}(y) = z] \\
&= p(y, z).
\end{aligned}$$

Nach Satz 17.11 ist X also eine Markovkette mit Übergangsmatrix p . \square

Beispiel 17.18 (Irrfahrt auf \mathbb{Z}). Sei $E = \mathbb{Z}$, und gelte

$$p(x, y) = p(0, y - x) \quad \text{für alle } x, y \in \mathbb{Z}.$$

Wir sagen in diesem Fall, dass p **translationsinvariant** ist. Eine diskrete Markovkette X mit Übergangsmatrix p ist eine Irrfahrt auf \mathbb{Z} . Es ist nämlich $X_n \stackrel{\mathcal{D}}{=} X_0 + Z_1 + \dots + Z_n$, wo $(Z_n)_{n \in \mathbb{N}}$ u.i.v. sind mit $\mathbf{P}[Z_n = x] = p(0, x)$.

Die R_n aus der expliziten Konstruktion erhalten wir durch $R_n(x) := x + Z_n$. \diamond

Beispiel 17.19 (Simulation am Computer). Wir betrachten die Situation, wo $E = \{1, \dots, k\}$ sogar endlich ist und wollen eine Markovkette X mit Übergangsmatrix p am Computer simulieren. Wir nehmen an, dass der Computer einen Zufallszahlengenerator bereitstellt, der eine Folge $(U_n)_{n \in \mathbb{N}}$ unabhängiger uniform auf $[0, 1]$ verteilter Zufallsvariablen erzeugt.

Wir setzen $r(i, 0) = 0$, $r(i, j) = p(i, 1) + \dots + p(i, j)$ für $i, j \in E$, und definieren Y_n durch

$$R_n(i) = j \iff U_n \in [r(i, j - 1), r(i, j)).$$

Per Konstruktion ist dann $\mathbf{P}[R_n(i) = j] = r(i, j) - r(i, j - 1) = p(i, j)$. \diamond

Beispiel 17.20 (Verzweigungsprozess als Markovkette). Wir wollen den Galton-Watson Verzweigungsprozess (siehe Definition 3.9) als Markovkette auf $E = \mathbb{N}_0$ auffassen.

Sei hierzu $(q_k)_{k \in \mathbb{N}_0}$ ein Wahrscheinlichkeitsvektor, den wir als Verteilung der Nachkommenschaft *eines* Individuums auffassen. Definiere $q_k^{*0} = \mathbb{1}_{\{0\}}(k)$ und

$$q_k^{*n} = \sum_{l=0}^k q_{k-l}^{*(n-1)} q_l \quad \text{für } n \in \mathbb{N}$$

als n -fache Faltung von q sowie die Matrix p durch $p(x, y) = q_y^{*x}$ für $x, y \in \mathbb{N}_0$. Damit ist q_k^{*n} die Wahrscheinlichkeit für n Individuen, genau k Nachfahren zu haben.

Seien nun $(Y_{n,i}, n \in \mathbb{N}_0, i \in \mathbb{N}_0)$ u.i.v. mit $\mathbf{P}[Y_{n,i} = k] = q_k$. Für $x \in \mathbb{N}_0$ definieren wir den Verzweigungsprozess X mit x Urahnen und Nachkommenverteilung q durch $X_0 = x$ und $X_n := \sum_{i=1}^{X_{n-1}} Y_{n-1,i}$. Um zu zeigen, dass X eine Markovkette ist, berechnen wir

$$\begin{aligned}
\mathbf{P}[X_n = x_n \mid X_0 = x, X_1 = x_1, \dots, X_{n-1} = x_{n-1}] \\
&= \mathbf{P}[Y_{n-1,1} + \dots + Y_{n-1,x_{n-1}} = x_n] \\
&= \mathbf{P}_{Y_{1,1}}^{*x_{n-1}}(\{x_n\}) = q_{x_n}^{*x_{n-1}} = p(x_{n-1}, x_n).
\end{aligned}$$

Also ist X eine Markovkette auf \mathbb{N}_0 mit Übergangsmatrix p . \diamond

Beispiel 17.21 (Wright'sches Evolutionsmodell). In der Biologie beschreibt das Wright'sche Evolutionsmodell ([182]) die Vererbung eines genetischen Merkmals mit zwei möglichen Ausprägungen, etwa A und B, (zum Beispiel Resistenz/keine Resistenz gegen ein bestimmtes Antibiotikum) in einer Population konstanter Größe $N \in \mathbb{N}$ mit diskreter Generationenfolge. Die Individuen werden dabei als **haploid** angenommen, die Chromosomen liegen also einfach vor (wie etwa bei gewissen Einzellern) und nicht als Paare (wie etwa bei Säugetieren).

Wir betrachten hier den Fall, wo keines der beiden Merkmale einen Selektionsvorteil bietet. Es wird also angenommen, dass sich jedes Individuum der neuen Generation zufällig (gleichverteilt) eines der Individuen der vorangehenden Generation als Ahn (oder Vorgänger) aussucht und dessen komplettes Erbgut übernimmt. Die Wahl wird für jedes Individuum unabhängig getroffen, wobei mehrere Individuen auf den selben Ahn zurückgehen können. Beträgt die Anzahl der Individuen vom Typ A in der Elterngeneration $k \in \{0, \dots, N\}$, so ist die Anzahl solcher Individuen in der Kindergeneration zufällig und binomialverteilt mit Parametern N und k/N .

Wir können die Genfrequenzen (also die relativen Anteile k/N) in diesem Modell offenbar durch eine Markovkette X auf $E = \{0, 1/N, \dots, (N-1)/N, 1\}$ mit Übergangsmatrix $p(x, y) = b_{N,x}(\{Ny\})$ beschreiben. Man beachte, dass X ein (beschränktes) Martingal ist. Nach dem Martingalkonvergenzsatz konvergiert X also \mathbf{P}_x -fast sicher gegen eine Zufallsvariable X_∞ mit $\mathbf{E}_x[X_\infty] = \mathbf{E}_x[X_0] = x$. Ähnlich wie beim Wählermodell (siehe Beispiel 11.16), das in der Tat sehr eng verwandt mit diesem Modell ist, können wir argumentieren, dass X_∞ nur die stabilen Randwerte 0 und 1 annehmen kann. Es gilt also $\mathbf{P}_x[\lim_{n \rightarrow \infty} X_n = 1] = x = 1 - \mathbf{P}_x[\lim_{n \rightarrow \infty} X_n = 0]$. \diamond

Beispiel 17.22 (Diskretes Moran-Modell). Wir wollen ein dem Wright'schen Evolutionsmodell verwandtes Modell mit Überlappung der Generationen betrachten. Die Situation ist wie beim Wright'schen Modell, jedoch soll jetzt pro Zeitschritt immer nur genau ein Individuum durch ein neues ersetzt werden, dessen Typ durch eine zufällige Wahl aus der Elterngeneration bestimmt wird.

Da die Typen des zu ersetzenen und des neuen Individuums unabhängig sind, erhalten wir als Modell für die Genfrequenzen eine Markovkette X auf $E = \{0, \frac{1}{N}, \dots, 1\}$ mit Übergangsmatrix

$$p(x, y) = \begin{cases} x(1-x), & \text{falls } y = x + 1/N, \\ x^2 + (1-x)^2, & \text{falls } y = x, \\ x(1-x), & \text{falls } y = x - 1/N, \\ 0, & \text{sonst.} \end{cases}$$

Auch hier ist X wieder ein beschränktes Martingal, und wir können den quadratischen Variationsprozess ausrechnen:

$$\langle X \rangle_n = \sum_{i=1}^n \mathbf{E}[(X_i - X_{i-1})^2 | X_{i-1}] = \frac{2}{N^2} \sum_{i=0}^{n-1} X_i(1 - X_i). \quad (17.12) \diamond$$

Übung 17.2.1 (Diskretes Martingalproblem). Sei $E \subset \mathbb{R}$ höchstens abzählbar und X eine Markovkette auf E mit Übergangsmatrix p und der Eigenschaft, dass es für jedes x eine höchstens dreielementige Menge $A_x \subset E$ gibt mit $p(x, y) = 0$ für jedes $y \in E \setminus A_x$. Sei $d(x) := \sum_{y \in E} (y - x) p(x, y)$ für $x \in E$.

- (i) Man zeige: Durch $M_n := X_n - \sum_{k=0}^{n-1} d(X_k)$ wird ein Martingal M definiert mit quadratischem Variationsprozess $\langle M \rangle_n = \sum_{i=0}^{n-1} f(X_i)$ für eine eindeutig bestimmte Funktion $f : E \rightarrow [0, \infty)$.
- (ii) Man zeige: Die Übergangsmatrix p ist durch Angabe von f und d eindeutig bestimmt.
- (iii) Man berechne für das Moran-Modell (Beispiel 17.22) die Übergangsmatrix aus der expliziten Form (17.12) des quadratischen Variationsprozesses. ♣

17.3 Diskrete Markovprozesse in stetiger Zeit

Sei E abzählbar und $(X_t)_{t \in [0, \infty)}$ ein Markovprozess auf E mit Übergangswahrscheinlichkeiten $p_t(x, y) = \mathbf{P}_x[X_t = y]$ (für $x, y \in E$). (Manche Autoren nennen solch einen Prozess auch Markovkette in stetiger Zeit.)

Sind $x, y \in E$ mit $x \neq y$, so sagen wir, dass X mit Rate $q(x, y)$ von x nach y springt, falls der folgende Limes existiert

$$q(x, y) := \lim_{t \downarrow 0} \frac{1}{t} \mathbf{P}_x[X_t = y].$$

Wir nehmen nun an, dass $q(x, y)$ für alle $y \neq x$ existiert, und dass

$$\sum_{y \neq x} q(x, y) < \infty \quad \text{für jedes } x \in E \quad (17.13)$$

gilt. Wir setzen dann

$$q(x, x) = - \sum_{y \neq x} q(x, y). \quad (17.14)$$

Schließlich fordern wir (was zusammen mit (17.14) äquivalent zur Vertauschung von Limes uns Summation über $y \neq x$ in der Zeile über (17.13) ist)

$$\lim_{t \downarrow 0} \frac{1}{t} (\mathbf{P}_x[X_t = y] - \mathbb{1}_{\{x=y\}}) = q(x, y) \quad \text{für alle } x, y \in E. \quad (17.15)$$

Definition 17.23. Gelten (17.13), (17.14) und (17.15), so heißt q die **Q -Matrix** von X . Manchmal wird q auch der **Generator** der Halbgruppe $(p_t)_{t \geq 0}$ genannt.

Beispiel 17.24 (Poissonprozess). Der Poissonprozess mit Rate $\alpha > 0$ (vergleiche Kapitel 5.5) hat die Q -Matrix $q(x, y) = \alpha(\mathbb{1}_{\{y=x+1\}} - \mathbb{1}_{\{y=x\}})$. \diamond

Satz 17.25. Gilt $q(x, y) \geq 0$ für alle $x, y \in E$ mit $x \neq y$, gelten (17.13), (17.14), und ist

$$\lambda := \sup_{x \in E} |q(x, x)| < \infty, \quad (17.16)$$

so ist q die Q -Matrix eines eindeutig bestimmten Markovprozesses X .

Ganz naiv betrachtet legt (17.15) nahe, dass man $p_t = e^{tq}$ in einem geeigneten Sinne definiert. Dann wäre rein formal $q = \frac{d}{dt} p_t|_{t=0}$. Der folgende Beweis zeigt, dass diese formale Argumentation unter den angegebenen Bedingungen rigoros gemacht werden kann.

Beweis. Sei I die Einheitsmatrix. Definiere

$$p(x, y) = \frac{1}{\lambda} q(x, y) + I(x, y) \quad \text{für } x, y \in E,$$

falls $\lambda > 0$ und sonst $p = I$. Dann ist p eine stochastische Matrix und $q = \lambda(p - I)$. Sei $((Y_n)_{n \in \mathbb{N}_0}, (\mathbf{P}_x^Y)_{x \in E})$ eine diskrete Markovkette mit Übergangsmatrix p , und sei $((T_t)_{t \geq 0}, (\mathbf{P}_n^T)_{n \in \mathbb{N}_0})$ ein Poissonprozess mit Rate λ . Sei $X_t := Y_{T_t}$ und $\mathbf{P}_x = \mathbf{P}_x^Y \otimes \mathbf{P}_0^T$. Dann ist $\mathfrak{X} := ((X_t)_{t \geq 0}, (\mathbf{P}_x)_{x \in E})$ ein Markovprozess und

$$\begin{aligned} p_t(x, y) &:= \mathbf{P}_x [X_t = y] = \sum_{n=0}^{\infty} \mathbf{P}_0^T [T_t = n] \mathbf{P}_x^Y [Y_n = y] \\ &= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{\lambda^n t^n}{n!} p^n(x, y). \end{aligned}$$

Diese Potenzreihe (in t) ist überall konvergent (da der lineare Operator p endliche Norm $\|p\|_2 \leq 1$ hat) gegen die Matrix-Exponentialfunktion $e^{\lambda t p}(x, y)$, und es gilt

$$p_t(x, y) = e^{-\lambda t} e^{\lambda t p}(x, y) = e^{\lambda t(p-I)}(x, y) = e^{tq}(x, y).$$

Durchgliedweise Differentiation der Potenzreihe erhalten wir $\frac{d}{dt} p_t(x, y)|_{t=0} = q(x, y)$. Damit ist \mathfrak{X} der gewünschte Markovprozess.

Wir nehmen nun an, dass $(\tilde{p}_t)_{t \geq 0}$ die Übergangswahrscheinlichkeiten eines weiteren Markovprozesses $\tilde{\mathfrak{X}}$ sind, mit dem selben Generator q , also mit

$$\lim_{s \downarrow 0} \frac{1}{s} (\tilde{p}_s(x, y) - I(x, y)) = q(x, y).$$

Man prüft leicht nach, dass

$$\lim_{s \downarrow 0} \frac{1}{s} (p_{t+s}(x, y) - p_t(x, y)) = (q \cdot p_t)(x, y)$$

gilt, das heißt, es gilt $(d/dt)p_t(x, y) = (q \cdot p_t)(x, y)$. Analog gilt ebenfalls $(d/dt)\tilde{p}_t = (q \cdot \tilde{p}_t)(x, y)$. Damit gilt aber auch

$$p_t(x, y) - \tilde{p}_t(x, y) = \int_0^t (q(p_s - \tilde{p}_s))(x, y) ds.$$

Setzen wir $r_s = p_s - \tilde{p}_s$, so ist $\|r_s\|_2 \leq 2$ und $\|q\|_2 \leq 2\lambda$, also

$$\sup_{s \leq t} \|r_s\|_2 \leq \sup_{s \leq t} \int_0^s \|qr_u\|_2 du \leq \|q\|_2 \sup_{s \leq t} \int_0^s \|r_u\|_2 du \leq 2\lambda t \sup_{s \leq t} \|r_s\|_2.$$

Für $t < 1/(2\lambda)$ folgt hieraus $r_t = 0$, also $\tilde{p}_t = p_t$. Für allgemeines $t > 0$ wähle $n \in \mathbb{N}$, so dass $t/n < 1/(2\lambda)$. Dann ist $\tilde{p}_t = (\tilde{p}_{t/n})^n = (p_{t/n})^n = p_t$. \square

Bemerkung 17.26. Auf die Bedingung (17.16) kann nicht ersatzlos verzichtet werden, wie dieses Beispiel zeigt: Sei $E = \mathbb{N}$ und

$$q(x, y) = \begin{cases} x^2, & \text{falls } y = x + 1, \\ -x^2, & \text{falls } y = x, \\ 0, & \text{sonst.} \end{cases}$$

Wir stellen uns einen Kandidaten X für einen Markovprozess mit Q -Matrix q explizit her. Seien T_1, T_2, \dots unabhängige, exponentialverteilte Zufallsvariablen mit $\mathbf{P}_{T_n} = \exp_{n^2}$. Setze $S_n = T_1 + \dots + T_{n-1}$ und $X_t = \sup\{n \in \mathbb{N}_0 : S_n \leq t\}$. Dann macht X zu jedem Zeitpunkt höchstens einen Schritt nach rechts, und es gilt aufgrund der Gedächtnislosigkeit der Exponentialverteilung (siehe Übung 8.1.1)

$$\begin{aligned} \mathbf{P}[X_{t+s} \geq n+1 | X_t = n] &= \mathbf{P}[S_{n+1} \leq t+s | S_n \leq t, S_{n+1} > t] \\ &= \mathbf{P}[T_n \leq s+t-S_n | S_n \leq t, T_n > t-S_n] = \mathbf{P}[T_n \leq s] \\ &= 1 - \exp(-n^2 s). \end{aligned}$$

Es folgt

$$\lim_{s \downarrow 0} s^{-1} \mathbf{P}[X_{t+s} = n+1 | X_t = n] = n^2$$

und

$$\lim_{s \downarrow 0} s^{-1} (\mathbf{P}[X_{t+s} = n | X_t = n] - 1) = -n^2,$$

also

$$\lim_{s \downarrow 0} s^{-1} (\mathbf{P}[X_{t+s} = m | X_t = n] - I(m, n)) = q(m, n) \quad \text{für alle } m, n \in \mathbb{N}.$$

Schreiben wir

$$\tau^n = \inf\{t \geq 0 : X_t = n\} = S_n \quad \text{für } n \in \mathbb{N},$$

so gilt $\mathbf{E}_1[\tau^n] = \sum_{k=1}^{n-1} \frac{1}{k^2}$. Nach dem Satz von der monotonen Konvergenz ist

$$\mathbf{E}_1\left[\sup_{n \in \mathbb{N}} \tau^n\right] = \lim_{n \rightarrow \infty} \mathbf{E}_1[\tau^n] = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

Das heißt, X überschreitet in endlicher Zeit alle Schranken. Wir sagen, dass X explodiert. \diamond

Beispiel 17.27 (Eine Variante des Pólya'schen Urnenmodells). Wir betrachten eine Variante des Pólya'schen Urnenmodells mit schwarzen und roten Kugeln (vergleiche Beispiel 12.29), wo nicht jeweils einfach nur *eine* weitere Kugel der selben Farbe zurückgelegt wird, sondern für die k -te Kugel, die von einer Farbe gezogen wird, werden r_k weitere Kugeln zurückgelegt. Dabei sind die Zahlen $r_1, r_2, \dots \in \mathbb{N}$ die Parameter des Modells. Der Fall $1 = r_1 = r_2 = \dots$ entspricht dem klassischen Pólya'schen Urnenmodell. Sei

$$X_n := \begin{cases} 1, & \text{falls die } n\text{-te Kugel schwarz ist,} \\ 0, & \text{sonst.} \end{cases}$$

Beim klassischen Modell hatten wir gesehen (Beispiel 12.29), dass der Anteil der schwarzen Kugeln gegen eine betaverteilte Zufallsvariable Z konvergiert, und dass gegeben Z die Folge X_1, X_2, \dots unabhängig und Ber_Z verteilt ist. Ganz ähnliche Aussagen bekommen wir in dem Fall, wo $r = r_1 = r_2 = \dots$ ist für ein $r \in \mathbb{N}$. In der Tat ändern sich hier nur die Parameter der Betaverteilung. Insbesondere (da die Betaverteilung keine Atome in 0 und 1 hat), werden von jeder Farbe fast sicher unendlich viele Kugeln gezogen. Es gilt also $\mathbf{P}[B] = 0$, wo B das Ereignis ist, dass von einer der Farben nur endlich viele Kugeln gezogen werden.

Wir werden jetzt sehen, dass dies nicht so sein muss, wenn die Zahlen r_k nur rasch genug wachsen. Wir nehmen an, dass anfangs je eine rote und eine schwarze Kugel in der Urne liegen und schreiben $w_n = 1 + \sum_{k=1}^n r_k$ für die Gesamtzahl von Kugeln einer Farbe, nachdem die Farbe bereits n -mal gezogen wurde ($n \in \mathbb{N}_0$).

Wir betrachten zunächst eine extreme Situation, wo $w_n = 2^n$ für jedes $n \in \mathbb{N}$. Die Größe

$$S_n = 2(X_1 + \dots + X_n) - n$$

zählt, wie viel mehr schwarze Kugeln als rote Kugeln bis zum n -ten Schritt gezogen wurden. Dann ist für jedes $n \in \mathbb{N}_0$

$$\mathbf{P}[X_{n+1} = 1 | S_n] = \frac{2^{S_n}}{1 + 2^{S_n}} \quad \text{und} \quad \mathbf{P}[X_{n+1} = 0 | S_n] = \frac{2^{-S_n}}{1 + 2^{-S_n}}.$$

Zusammen erhalten wir, dass $(Z_n)_{n \in \mathbb{N}_0} := (|S_n|)_{n \in \mathbb{N}_0}$ eine Markovkette auf \mathbb{N}_0 ist mit Übergangsmatrix

$$p(z, z') = \begin{cases} 2^z/(1+2^z), & \text{falls } z' = z+1 > 1, \\ 1, & \text{falls } z' = z+1 = 1, \\ 1/(1+2^z), & \text{falls } z' = z-1, \\ 0, & \text{sonst.} \end{cases}$$

Das Ereignis B von oben können wir schreiben als

$$B = \{Z_{n+1} < Z_n \text{ nur endlich oft}\}.$$

Sei $A = \{Z_{n+1} > Z_n \text{ für alle } n \in \mathbb{N}_0\}$ das Ereignis, dass Z auf direktem Weg nach ∞ flieht und $\tau_z = \inf\{n \in \mathbb{N}_0 : Z_n \geq z\}$. Offenbar ist

$$\mathbf{P}_z[A] = \prod_{z'=z}^{\infty} p(z', z'+1) \geq 1 - \sum_{z'=z}^{\infty} \frac{1}{1+2^{z'}} \geq 1 - 2^{1-z}.$$

Man kann leicht zeigen, dass $\mathbf{P}_0[\tau_z < \infty] = 1$ ist für jedes $z \in \mathbb{N}_0$. Wir erhalten für jedes $z \in \mathbb{N}_0$ mit der starken Markoveigenschaft

$$\mathbf{P}_0[B] \geq \mathbf{P}_0[Z_{n+1} > Z_n \text{ für alle } n \geq \tau_z] = \mathbf{P}_z[A] \geq 1 - 2^{1-z}$$

und damit $\mathbf{P}_0[B] = 1$. Damit ist nachgewiesen, dass fast sicher irgendwann nur noch Kugeln einer Farbe gezogen werden.

Wir wollen nun von diesem extremen Beispiel weg und mit (noch) subtleren Methoden, die an das obige Beispiel mit der Explosion des Markovprozesses anknüpfen, arbeiten.

Wir wollen nun zeigen, dass $\mathbf{P}[B] = 1$, falls $\sum_{n=0}^{\infty} \frac{1}{w_n} < \infty$. Hierzu betrachten wir unabhängige Zufallsvariablen $T_1^r, T_1^s, T_2^r, T_2^s, \dots$ mit $\mathbf{P}_{T_n^r} = \mathbf{P}_{T_n^s} = \exp_{w_{n-1}}$. Ferner sei $T_{\infty}^r = \sum_{n=1}^{\infty} T_n^r$ und $T_{\infty}^s = \sum_{n=1}^{\infty} T_n^s$. Offenbar ist $\mathbf{E}[T_{\infty}^r] = \sum_{n=0}^{\infty} 1/w_n < \infty$, also ist insbesondere $\mathbf{P}[T_{\infty}^r < \infty] = 1$. Die analoge Aussage gilt für T_{∞}^s . Man beachte, dass T_{∞}^r und T_{∞}^s unabhängig sind und Dichten haben (weil T_1^r und T_1^s Dichten haben), also gilt $\mathbf{P}[T_{\infty}^r = T_{\infty}^s] = 0$.

Seien nun

$$R_t = \sup \{n \in \mathbb{N} : T_1^r + \dots + T_{n-1}^r \leq t\}$$

und

$$S_t = \sup \{n \in \mathbb{N} : T_1^s + \dots + T_{n-1}^s \leq t\}.$$

Seien $R := \{T_1^r + \dots + T_n^r, n \in \mathbb{N}\}$ und $S := \{T_1^s + \dots + T_n^s, n \in \mathbb{N}\}$ die Sprungzeitpunkte von (R_t) und (S_t) sowie $U := R \cup S = \{u_1, u_2, \dots\}$, wobei $u_1 < u_2 < \dots$ Sei

$$X_n = \begin{cases} 1, & \text{falls } u_n \in S, \\ 0, & \text{sonst.} \end{cases}$$

Sei $L_n = x_1 + \dots + x_n$. Dann ist

$$\begin{aligned} \mathbf{P}[X_{n+1} = 1 \mid X_1 = x_1, \dots, X_n = x_n] &= \mathbf{P}[u_{n+1} \in S \mid (u_k \in S \iff x_k = 1) \text{ für jedes } k \leq n] \\ &= \mathbf{P}[T_1^s + \dots + T_{L_n+1}^s < T_1^r + \dots + T_{n-L_n+1}^r \\ &\quad \mid T_1^s + \dots + T_{L_n+1}^s > T_1^r + \dots + T_{n-L_n}^r] \\ &= \mathbf{P}[T_{L_n+1}^s < T_{n-L_n+1}^r] = \frac{w_{L_n}}{w_{L_n} + w_{n-L_n}}. \end{aligned}$$

Also ist $(X_n)_{n \in \mathbb{N}_0}$ das erweiterte Urnenmodell mit Gewichten $(w_n)_{n \in \mathbb{N}_0}$. Wir betrachten nun das Ereignis B^c , dass von jeder Farbe unendlich viele Kugeln gezogen werden. Offenbar ist $\{X_n = 1 \text{ unendlich oft}\} = \{\sup S = \sup U\}$ und $\{X_n = 0 \text{ unendlich oft}\} = \{\sup R = \sup U\}$. Wegen $\sup S = T_\infty^s$ und $\sup R = T_\infty^r$ ist also $\mathbf{P}[B^c] = \mathbf{P}[T_\infty^r = T_\infty^s] = 0$. \diamond

Übung 17.3.1. Seien $r, s, R, S \in \mathbb{N}$. Man betrachte das Pólya'sche Urnenmodell $(X_n)_{n \in \mathbb{N}_0}$ mit $r_k = r$ und $s_k = s$ für alle $k \in \mathbb{N}$ und anfänglich R roten Kugeln und S schwarzen Kugeln. Man zeige, dass der Anteil der schwarzen Kugeln fast sicher gegen eine Zufallsvariable Z mit Beta-Verteilung konvergiert und bestimme die Parameter. Man zeige, dass $(X_n)_{n \in \mathbb{N}_0}$ u.i.v. ist gegeben Z und $X_i \sim \text{Ber}_Z$ für jedes $i \in \mathbb{N}_0$. \clubsuit

Übung 17.3.2. Man zeige, dass fast sicher unendlich viele Kugeln jeder Farbe gezogen werden, falls $\sum_{n=0}^{\infty} \frac{1}{w_n} = \infty$. \clubsuit

17.4 Diskrete Markovketten, Rekurrenz und Transienz

Sei im Folgenden $X = (X_n)_{n \in \mathbb{N}_0}$ eine Markovkette auf dem abzählbaren Raum E mit Übergangsmatrix p .

Definition 17.28. Für jedes $x \in E$ sei $\tau_x := \tau_x^1 := \inf\{n > 0 : X_n = x\}$ und

$$\tau_x^k = \inf \{n > \tau_x^{k-1} : X_n = x\} \quad \text{für } k \in \mathbb{N}, k \geq 2.$$

τ_x^k heißt k -te **Eintrittszeit** von X in x . Für $x, y \in E$ sei

$$F(x, y) := \mathbf{P}_x[\tau_y^1 < \infty] = \mathbf{P}_x[\text{es gibt ein } n \geq 1 \text{ mit } X_n = y]$$

die Wahrscheinlichkeit jemals von x nach y zu gehen. Speziell ist $F(x, x)$ die Rückkehrswahrscheinlichkeit (nach dem ersten Sprung) von x nach x .

Man beachte, dass $\tau_x^1 > 0$ selbst bei Start in $X_0 = x$ gilt.

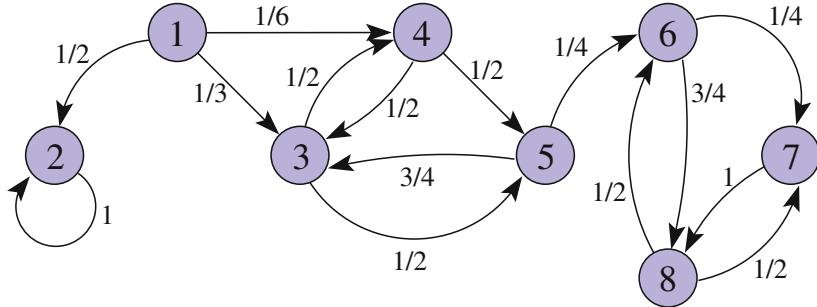


Abb. 17.1. Markovkette mit acht Zuständen. Die Zahlen sind die Übergangswahrscheinlichkeiten für die entsprechenden Pfeile.

Der Zustand 2 ist absorbierend, die Zustände 1, 3, 4 und 5 transient, die Zustände 6, 7 und 8 (positiv) rekurrent.

Satz 17.29. Für alle $x, y \in E$ und $k \in \mathbb{N}$ gilt $\mathbf{P}_x[\tau_y^k < \infty] = F(x, y) F(y, y)^{k-1}$.

Beweis. Wir führen den Beweis per Induktion über k . Für $k = 1$ ist die Aussage per Definition richtig. Sei nun $k \geq 2$. Dann ist wegen der starken Markoveigenschaft von X (siehe Satz 17.14)

$$\begin{aligned} \mathbf{P}_x[\tau_y^k < \infty] &= \mathbf{E}_x \left[\mathbf{P}_x \left[\tau_y^k < \infty \mid \mathcal{F}_{\tau_y^{k-1}} \right] \mathbb{1}_{\{\tau_y^{k-1} < \infty\}} \right] \\ &= \mathbf{E}_x \left[F(y, y) \cdot \mathbb{1}_{\{\tau_y^{k-1} < \infty\}} \right] \\ &= F(y, y) \cdot F(x, y) F(y, y)^{k-2} = F(x, y) F(y, y)^{k-1}. \quad \square \end{aligned}$$

Definition 17.30. Ein Zustand $x \in E$ heißt

- **rekurrent**, falls $F(x, x) = 1$,
- **positiv rekurrent**, falls $\mathbf{E}_x[\tau_x^1] < \infty$,
- **nullrekurrent**, falls x rekurrent ist, aber nicht positiv rekurrent,
- **transient**, falls $F(x, x) < 1$,
- **absorbierend**, falls $p(x, x) = 1$.

Die Markovkette X heißt (positiv / null-) rekurrent, falls jeder Zustand $x \in E$ (positiv / null-) rekurrent ist und transient, falls jeder rekurrente Zustand absorbierend ist.

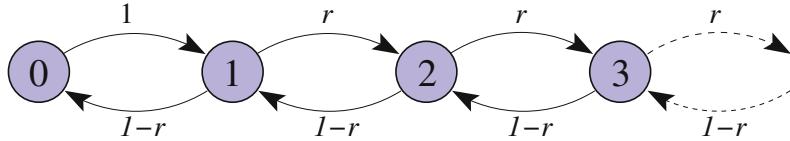


Abb. 17.2. Markovkette auf \mathbb{N}_0 mit Parameter $r \in (0, 1)$. Die Kette ist positiv rekurrent, wenn $r \in (0, 1/2)$, nullrekurrent, wenn $r = 1/2$ und transient für $r \in (1/2, 1)$.

Bemerkung 17.31. Offenbar gilt:

„absorbierend“ \implies „positiv rekurrent“ \implies „rekurrent“.

◇

Beispiel 17.32. (i) In Abb. 17.1 ist offenbar der Zustand 2 absorbierend. Wird die Kette nicht in 2 gefangen, so wird sie irgendwann einmal den Sprung von 5 nach 6 machen und danach nicht wieder zurückkehren. Also sind 1, 3, 4 und 5 transient. Die Zustände 6, 7 und 8 sind positiv rekurrent. Man kann zeigen (siehe Übung 17.6.1), dass $\mathbf{E}_6[\tau_6] = \frac{17}{4}$, $\mathbf{E}_7[\tau_7] = \frac{17}{5}$ und $\mathbf{E}_8[\tau_8] = \frac{17}{8}$.

(ii) In Abb. 17.2 hat die Kette eine Tendenz, nach rechts auszuwandern, falls $r > \frac{1}{2}$ ist. Daher ist in diesem Fall jeder Zustand transient. Ist hingegen $r \in (0, \frac{1}{2})$, so hat die Kette einen Drall nach links (außer in 0) und besucht daher jeden Zustand immer wieder, ist also rekurrent. Mit etwas mehr Nachdenken kann man zeigen (siehe Übung 17.6.4), dass die Kette in diesem Fall sogar positiv rekurrent ist und im verbleibenden Fall $r = \frac{1}{2}$ nullrekurrent.

◇

Definition 17.33. Mit $N(y) = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=y\}}$ bezeichnen wir die Gesamtzahl der Besuche von X in y und mit

$$G(x, y) = \mathbf{E}_x[N(y)] = \sum_{n=0}^{\infty} p^n(x, y)$$

die **Greenfunktion** von X .

Satz 17.34. (i) Für alle $x, y \in E$ gilt (mit $1/0 := \infty$, $0/0 := 0$ und $0 \cdot \infty := 0$)

$$G(x, y) = \begin{cases} \frac{F(x, y)}{1 - F(y, y)}, & \text{falls } x \neq y \\ \frac{1}{1 - F(y, y)}, & \text{falls } x = y \end{cases} = F(x, y) G(y, y) + \mathbb{1}_{\{x=y\}}. \quad (17.17)$$

(ii) Ein nichtabsorbierender Zustand $x \in E$ ist genau dann rekurrent, wenn $G(x, x) = \infty$ gilt.

Beweis. (ii) folgt aus (i). Wir müssen also noch (17.17) zeigen. Nach Satz 17.29 ist

$$\begin{aligned} G(x, y) &= \mathbf{E}_x[N(y)] = \sum_{k=1}^{\infty} \mathbf{P}_x[N(y) \geq k] \\ &= \mathbb{1}_{\{x=y\}} + \sum_{k=1}^{\infty} \mathbf{P}_x[\tau_y^k < \infty] = \mathbb{1}_{\{x=y\}} + \sum_{k=1}^{\infty} F(x, y) F(y, y)^{k-1} \\ &= \begin{cases} \frac{F(x, y)}{1 - F(y, y)}, & \text{falls } x \neq y, \\ \frac{1}{1 - F(x, x)}, & \text{falls } x = y. \end{cases} \end{aligned}$$

Die zweite Gleichheit in (17.17) folgt hieraus direkt. \square

Satz 17.35. Ist x rekurrent und $F(x, y) > 0$, dann ist auch y rekurrent, und es gilt $F(x, y) = F(y, x) = 1$.

Beweis. Seien $x, y \in E$, $x \neq y$, mit $F(x, y) > 0$. Dann gibt es ein $k \in \mathbb{N}$ und Punkte $x_1, \dots, x_k \in E$ mit $x_k = y$ und $x_i \neq x$ für jedes $i = 1, \dots, k$ sowie

$$\mathbf{P}_x[X_i = x_i \text{ für jedes } i = 1, \dots, k] > 0.$$

Speziell ist $p^k(x, y) > 0$. Nach der Markoveigenschaft ist

$$\begin{aligned} 1 - F(x, x) &= \mathbf{P}_x[\tau_x^1 = \infty] \geq \mathbf{P}_x[X_1 = x_1, \dots, X_k = x_k, \tau_x^1 = \infty] \\ &= \mathbf{P}_x[X_1 = x_1, \dots, X_k = x_k] \cdot \mathbf{P}_y[\tau_x^1 = \infty] \\ &= \mathbf{P}_x[X_1 = x_1, \dots, X_k = x_k](1 - F(y, x)). \end{aligned}$$

Ist nun $F(x, x) = 1$, dann ist auch $F(y, x) = 1$. Wegen $F(y, x) > 0$ existiert ein $l \in \mathbb{N}$ mit $p^l(y, x) > 0$. Also ist für $n \in \mathbb{N}_0$

$$p^{l+n+k}(y, y) \geq p^l(y, x)p^n(x, x)p^k(x, y).$$

Mithin ist, falls x rekurrent ist,

$$G(y, y) \geq \sum_{n=0}^{\infty} p^{l+n+k}(y, y) \geq p^l(y, x)p^k(x, y)G(x, x) = \infty.$$

Folglich ist auch y rekurrent. Wenn wir jetzt im Argument x und y vertauschen, dann erhalten wir noch $F(x, y) = 1$. \square

Definition 17.36. Eine diskrete Markovkette heißt

- **irreduzibel**, falls $F(x, y) > 0$ für alle $x, y \in E$ gilt, oder äquivalent $G(x, y) > 0$,
- **schwach irreduzibel**, falls $F(x, y) + F(y, x) > 0$ für alle $x, y \in E$ gilt.

Satz 17.37. Eine irreduzible diskrete Markovkette ist entweder rekurrent oder transient. Ist $|E| \geq 2$, so gibt es keine absorbierenden Zustände.

Beweis. Das folgt direkt aus Satz 17.35. □

Satz 17.38. Ist E endlich und X irreduzibel, so ist X rekurrent.

Beweis. Offenbar ist für jedes $x \in E$

$$\sum_{y \in E} G(x, y) = \sum_{n=0}^{\infty} \sum_{y \in E} p^n(x, y) = \sum_{n=0}^{\infty} 1 = \infty.$$

Da E endlich ist, gibt es ein $y \in E$ mit $G(x, y) = \infty$. Wegen $F(y, x) > 0$ existiert ein $k \in \mathbb{N}$ mit $p^k(y, x) > 0$, also ist $p^{n+k}(x, x) \geq p^n(x, y) p^k(y, x)$ und

$$G(x, x) \geq \sum_{n=0}^{\infty} p^n(x, y) p^k(y, x) = p^k(y, x) G(x, y) = \infty. \quad \square$$

Übung 17.4.1. Sei x positiv rekurrent und $F(x, y) > 0$. Man zeige, dass auch y positiv rekurrent ist. ♣

17.5 Anwendung: Rekurrenz und Transienz von Irrfahrten

Wir wollen in diesem Abschnitt die Rekurrenz- und Transienzeigenschaften von Irrfahrten auf \mathbb{Z}^D , $D = 1, 2, \dots$ untersuchen. Eine ausführlichere Behandlung findet der Leser im Buch von Spitzer [168].

Wir wollen untersuchen, ob die symmetrische einfache Irrfahrt X auf \mathbb{Z}^D , die in jedem Schritt mit gleicher Wahrscheinlichkeit zu einem der $2D$ nächsten Nachbarn springt, rekurrent oder transient ist. Sei also $E = \mathbb{Z}^D$ und

$$p(x, y) = \begin{cases} \frac{1}{2D}, & \text{falls } |x - y| = 1, \\ 0, & \text{sonst.} \end{cases}$$

Der zentrale Grenzwertsatz legt nahe zu vermuten, dass

$$p^n(0, 0) \approx C_D n^{-D/2} \quad \text{für } n \rightarrow \infty$$

für eine Konstante C_D , die von der Dimension abhängt. Wir müssen hier jedoch zunächst einmal den Fall ausschließen, wo n ungerade ist, denn für ungerades n ist

offenbar $p^n(0, 0) = 0$. Seien also Y_1, Y_2, \dots unabhängige \mathbb{Z}^D -wertige Zufallsvariablen mit $\mathbf{P}[Y_i = x] = p^2(0, x)$. Dann ist $X_{2n} \stackrel{\mathcal{D}}{=} S_n := Y_1 + \dots + Y_n$ für $n \in \mathbb{N}_0$, also $G(0, 0) = \sum_{n=0}^{\infty} \mathbf{P}[S_n = 0]$. Offenbar hat $Y_1 = (Y_1^1, \dots, Y_1^D)$ die Kovarianzmatrix $C_{i,j} := \mathbf{E}[Y_1^i \cdot Y_1^j] = \frac{2}{D} \mathbb{1}_{\{i=j\}}$. Nach dem lokalen zentralen Grenzwertsatz (siehe etwa [23, Seite 224ff] für eine eindimensionale Version dieses Satzes oder Übung 17.5.1 für eine analytische Herleitung) gilt

$$n^{D/2} p^{2n}(0, 0) = n^{D/2} \mathbf{P}[S_n = 0] \xrightarrow{n \rightarrow \infty} 2(4\pi/D)^{-D/2}. \quad (17.18)$$

Nun ist genau dann $\sum_{n=1}^{\infty} n^{-\alpha} < \infty$, wenn $\alpha > 1$ ist, also ist $G(0, 0) < \infty$ genau dann, wenn $D > 2$ ist. Wir haben damit einen Satz von Pólya gezeigt:

Satz 17.39 (Pólya (1921)). *Die symmetrische einfache Irrfahrt auf \mathbb{Z}^D ist genau dann rekurrent, wenn $D \leq 2$.*

Das hier verwendete Vorgehen hat den Nachteil, dass wir den lokalen zentralen Grenzwertsatz bemüht haben, den wir nicht bewiesen haben. Wir wollen daher weitere Ansätze betrachten, die ohne dieses Hilfsmittel auskommen und auch an sich von Interesse sind.

Betrachten wir zunächst die eindimensionale einfache Irrfahrt, die mit Wahrscheinlichkeit p einen Schritt nach rechts macht und mit Wahrscheinlichkeit $1 - p$ einen Schritt nach links. Dann ist

$$G(0, 0) = \sum_{n=0}^{\infty} \binom{2n}{n} (p(1-p))^n = \sum_{n=0}^{\infty} \binom{-1/2}{n} (-4p(1-p))^n.$$

Unter Benutzung des verallgemeinerten binomischen Lehrsatzes (Lemma 3.5) folgt (wegen $(1 - 4p(1 - p))^{1/2} = |2p - 1|$)

$$G(0, 0) = \begin{cases} \frac{1}{|2p-1|}, & \text{falls } p \neq \frac{1}{2}, \\ \infty, & \text{falls } p = \frac{1}{2}. \end{cases} \quad (17.19)$$

Wir erhalten also, dass die einfache Irrfahrt auf \mathbb{Z} genau dann rekurrent ist, wenn sie symmetrisch ist, also falls $p = \frac{1}{2}$ gilt.

Die Transienz im Falle $p \neq \frac{1}{2}$ folgt natürlich auch direkt aus dem starken Gesetz der großen Zahl, denn $\lim_{n \rightarrow \infty} \frac{1}{n} X_n = \mathbf{E}_0[X_1] = 2p - 1$ fast sicher. Tatsächlich haben wir bei diesem Argument nur benutzt, dass die einzelnen Schritte von X einen Erwartungswert haben, der ungleich Null ist.

Betrachten wir nun die allgemeinere Situation, wo X nicht notwendigerweise nur zu den nächsten Nachbarn springt, wo aber immer noch $\mathbf{E}_0[|X_1|] < \infty$ und $\mathbf{E}_0[X_1] = 0$ gelten. Das starke Gesetz der großen Zahl liefert hier nicht direkt die gewünschte Aussage, sondern wir müssen etwas sorgfältiger argumentieren.

Die Markoveigenschaft liefert für jedes $N \in \mathbb{N}$ und $y \neq x$

$$G_N(x, y) := \sum_{k=0}^N \mathbf{P}_x[X_k = y] = \sum_{k=0}^N \mathbf{P}_x[\tau_y^1 = k] \sum_{l=0}^{N-k} \mathbf{P}_y[X_l = y] \leq G_N(y, y).$$

Hieraus folgt für jedes $L \in \mathbb{N}$

$$\begin{aligned} G_N(0, 0) &\geq \frac{1}{2L+1} \sum_{|y| \leq L} G_N(0, y) \\ &= \frac{1}{2L+1} \sum_{k=0}^N \sum_{|y| \leq L} p^k(0, y) \\ &\geq \frac{1}{2L+1} \sum_{k=1}^N \sum_{y: |y/k| \leq L/N} p^k(0, y). \end{aligned}$$

Nach dem schwachen Gesetz der großen Zahl ist $\liminf_{k \rightarrow \infty} \sum_{|y| \leq \varepsilon k} p^k(0, y) = 1$ für jedes $\varepsilon > 0$, also folgt, wenn wir $L = \varepsilon N$ setzen

$$\liminf_{N \rightarrow \infty} G_N(0, 0) \geq \frac{1}{2\varepsilon} \quad \text{für jedes } \varepsilon > 0.$$

Wir haben damit $G(0, 0) = \infty$ und folglich die Rekurrenz von X gezeigt.

Zusammen mit der vorangehenden, einfachen Richtung haben wir gezeigt:

Satz 17.40. Eine Irrfahrt auf \mathbb{Z} mit $\sum_{x=-\infty}^{\infty} |x| p(0, x) < \infty$ ist genau dann rekurrent, wenn $\sum_{x=-\infty}^{\infty} x p(0, x) = 0$ gilt.

Wie steht es nun für symmetrische einfache Irrfahrten in Dimension $D = 2$ und höheren Dimensionen? Damit die Irrfahrt nach $2n$ Schritten wieder im Ursprung ist, muss sie k_i Schritte in die i -te Richtung machen und k_i Schritte in die Gegenrichtung, wobei $k_1 + \dots + k_D = n$ ist. Wir erhalten also

$$p^{2n}(0, 0) = (2D)^{-2n} \sum_{k_1+\dots+k_D=n} \binom{2n}{k_1, k_1, \dots, k_D, k_D}, \quad (17.20)$$

wobei $\binom{N}{l_1, \dots, l_r} = \frac{N!}{l_1! \cdots l_r!}$ der Multinomialkoeffizient ist. Speziell ist für $D = 2$

$$\begin{aligned} p^{2n}(0, 0) &= 4^{-2n} \sum_{k=0}^n \frac{(2n)!}{(k!)^2 ((n-k)!)^2} \\ &= 4^{-2n} \binom{2n}{n} \sum_{k=0}^n \binom{n}{k} \binom{n}{n-k} = \left(2^{-2n} \binom{2n}{n} \right)^2, \end{aligned}$$

wobei wir im letzten Schritt eine einfache kombinatorische Identität benutzt haben, die beispielsweise direkt aus der Faltungsformel $(b_{n,p} * b_{n,p})(\{n\}) = b_{2n,p}(\{n\})$ folgt. Nach der Stirling'schen Formel gilt nun

$$\lim_{n \rightarrow \infty} \sqrt{n} 2^{-2n} \binom{2n}{n} = \frac{1}{\sqrt{\pi}},$$

also $\lim_{n \rightarrow \infty} np^{2n}(0, 0) = \frac{1}{\pi}$. Insbesondere ist also $\sum_{n=1}^{\infty} p^{2n}(0, 0) = \infty$, das heißt, die zweidimensionale, symmetrische einfache Irrfahrt ist rekurrent.

Für $D \geq 3$ lässt sich die Summe über die Multinomialkoeffizienten nicht mehr in befriedigender Weise ausrechnen. Man kann allerdings immer noch obere Abschätzungen angeben, die zeigen, dass es ein $c = c_D$ gibt, sodass $p^{2n}(0, 0) \leq c n^{-D/2}$ gilt, woraus dann $G(0, 0) \leq c \sum_{n=1}^{\infty} n^{-D/2} < \infty$ folgt (siehe etwa [65, Beispiel 6.31] oder [58, Seite 361]). Wir wollen hier jedoch eine andere Argumentation verfolgen.

Die Sache wäre ganz einfach, wenn die einzelnen Koordinaten der Kette unabhängig wären. Dann wäre ja die Wahrscheinlichkeit, dass zur Zeit $2n$ alle Koordinaten gleich Null sind, gleich der D -ten Potenz der Wahrscheinlichkeit, dass etwa die erste Koordinate gleich Null ist. Für eine Koordinate ist aber (weil sich eine einzelne Koordinate ja nur mit Wahrscheinlichkeit $1/D$ bewegt, also nur Varianz $1/D$ hat) die Wahrscheinlichkeit, von der 0 aus startend nach $2n$ Schritten wieder in 0 zu sein ungefähr $(n \pi/D)^{-1/2}$. Bis auf einen Faktor erhielte man so (17.18), ohne dass man den mehrdimensionalen lokalen Grenzwertsatz direkt bemüht hätte.

Eine Möglichkeit, die Koordinaten tatsächlich unabhängig zu machen, besteht darin, die zeitdiskrete Markovkette in einen zeitstetigen Markovprozess auf \mathbb{Z}^D zu wandeln, der die gleiche Greenfunktion hat.

Wir betrachten also D unabhängige Poissonprozesse $(T_t^i)_{t \geq 0}$, $i = 1, \dots, D$ mit Rate $1/D$ und D unabhängige, symmetrische einfache Irrfahrten Z^1, \dots, Z^D auf \mathbb{Z} . Wir setzen $T := T^1 + \dots + T^D$, $Y_t^i := Z_{T_t^i}^i$ für $i = 1, \dots, D$ und $Y_t = (Y_t^1, \dots, Y_t^D)$. Dann ist Y eine Markovkette in stetiger Zeit mit Q -Matrix $q(x, y) = p(x, y) - \mathbf{1}_{\{x=y\}}$. Da T ein Poissonprozess mit Rate 1 ist, ist auch $(X_{T_t})_{t \geq 0}$ ein Markovprozess mit Q -Matrix q . Es folgt $(X_{T_t})_{t \geq 0} \stackrel{D}{=} (Y_t)_{t \geq 0}$. Wir berechnen nun

$$\begin{aligned} G_Y := \int_0^\infty \mathbf{P}_0[Y_t = 0] dt &= \int_0^\infty \sum_{n=0}^{\infty} \mathbf{P}_0[X_{2n} = 0, T_t = 2n] dt \\ &= \sum_{n=0}^{\infty} p^{2n}(0, 0) \int_0^\infty e^{-t} \frac{t^{2n}}{(2n)!} dt = G(0, 0). \end{aligned}$$

Die beiden Prozesse $(X_n)_{n \in \mathbb{N}_0}$ und $(Y_t)_{t \in [0, \infty)}$ haben also die selbe Greenfunktion. Nun sind aber die Koordinaten von Y tatsächlich unabhängig, also ist

$$G_Y = \int_0^\infty \mathbf{P}_0[Y_t^1 = 0]^D dt.$$

Wir müssen also nur noch $\mathbf{P}_0[Y_t^1 = 0]$ für große t berechnen. Wir können so argumentieren: nach dem Gesetz der großen Zahl ist $T_t^1 \approx t/D$ für große t . Außerdem gilt $\mathbf{P}_0[Y_t^1 \text{ ist gerade}] \approx \frac{1}{2}$. Es gilt also, mit $n_t = \lfloor t/2D \rfloor$ für $t \rightarrow \infty$ (vergleiche Übung 17.5.2)

$$\mathbf{P}_0[Y_t^1 = 0] \sim \frac{1}{2} \mathbf{P}[Z_{2n_t}^1 = 0] = \frac{1}{2} \binom{2n_t}{n_t} 4^{-n_t} \sim (2\pi/D)^{-1/2} t^{-1/2}. \quad (17.21)$$

Da genau dann $\int_1^\infty t^{-\alpha} dt < \infty$ gilt, wenn $\alpha > 1$ ist, so gilt auch $G_Y < \infty$ genau dann, wenn $D > 2$ ist. Dies ist aber gerade die Aussage des Satzes von Pólya.

Schließlich stellen wir noch eine dritte Methode vor, um Rekurrenz und Transienz von Irrfahrten zu untersuchen, die unabhängig von den euklidischen Eigenschaften des D -dimensionalen Gitters ist und auf der Fourier-Inversionsformel beruht.

Wir betrachten zunächst eine allgemeine (zeitdiskrete) irreduzible Irrfahrt mit Übergangsmatrix p auf \mathbb{Z}^D . Mit $\phi(t) = \sum_{x \in \mathbb{Z}^D} e^{i\langle t, x \rangle} p(0, x)$ bezeichnen wir die charakteristische Funktion eines einzelnen Übergangs. Die Faltung der Übergangswahrscheinlichkeiten überträgt sich in Potenzen der charakteristischen Funktion, also ist

$$\phi^n(t) = \sum_{x \in \mathbb{Z}^D} e^{i\langle t, x \rangle} p^n(0, x).$$

Nach der Fourier-Inversionsformel (Satz 15.10) erhalten wir aus ϕ^n die n -Schritt Übergangswahrscheinlichkeiten zurück durch

$$p^n(0, x) = (2\pi)^{-D} \int_{[-\pi, \pi)^D} e^{-i\langle t, x \rangle} \phi^n(t) dt.$$

Speziell ist für $\lambda \in (0, 1)$

$$\begin{aligned} R_\lambda &:= \sum_{n=0}^{\infty} \lambda^n p^n(0, 0) = (2\pi)^{-D} \sum_{n=0}^{\infty} \int_{[-\pi, \pi)^D} \lambda^n \phi^n(t) dt \\ &= (2\pi)^{-D} \int_{[-\pi, \pi)^D} \frac{1}{1 - \lambda \phi(t)} dt. \\ &= (2\pi)^{-D} \int_{[-\pi, \pi)^D} \operatorname{Re} \left(\frac{1}{1 - \lambda \phi(t)} \right) dt. \end{aligned}$$

Nun ist $G(0, 0) = \lim_{\lambda \uparrow 1} R_\lambda$, also

$$X \text{ ist rekurrent} \iff \lim_{\lambda \uparrow 1} \int_{[-\pi, \pi)^D} \operatorname{Re} \left(\frac{1}{1 - \lambda \phi(t)} \right) dt = \infty. \quad (17.22)$$

Wäre $\phi(t) = 1$ für ein $t \in (-2\pi, 2\pi)^D \setminus \{0\}$, so wäre $\phi^n(t) = 1$ für jedes $n \in \mathbb{N}$ und damit nach Übung 15.2.1 $\mathbf{P}_0[\langle X_n, t/(2\pi) \rangle \in \mathbb{Z}] = 1$, also wäre X nicht irreduzibel, im Widerspruch zur Annahme. Wegen der Stetigkeit von ϕ ist also für jedes $\varepsilon > 0$

$$\inf \{|\phi(t) - 1| : t \in [-\pi, \pi]^D \setminus (-\varepsilon, \varepsilon)^D\} > 0.$$

Es gilt also der folgende Satz.

Satz 17.41 (Chung-Fuchs (1951)). Eine irreduzible Irrfahrt auf \mathbb{Z}^D mit charakteristischer Funktion ϕ ist genau dann rekurrent, wenn für jedes $\varepsilon > 0$ gilt:

$$\lim_{\lambda \uparrow 1} \int_{(-\varepsilon, \varepsilon)^D} \operatorname{Re} \left(\frac{1}{1 - \lambda \phi(t)} \right) dt = \infty. \quad (17.23)$$

Betrachten wir nun die symmetrische einfache Irrfahrt, so ist $\phi(t) = \frac{1}{D} \sum_{i=1}^D \cos(t_i)$. Entwickeln wir die Kosinusfunktion in eine Taylorreihe um 0, so erhalten wir $\cos(t_i) = 1 - \frac{1}{2} t_i^2 + O(t_i^4)$, also $1 - \phi(t) = \frac{1}{2D} \|t\|_2^2 + O(\|t\|_2^4)$. Es folgt, dass X genau dann rekurrent ist, wenn $\int_{\|t\|_2 < \varepsilon} \|t\|_2^{-2} dt = \infty$. Wir berechnen dieses Integral in Polarkoordinaten (mit C_D die Oberfläche der Einheitskugel in \mathbb{R}^D)

$$\int_{\|t\|_2 < \varepsilon} \|t\|_2^{-2} dt = C_D \int_0^\varepsilon r^{D-1} r^{-2} dr = \infty \iff D \leq 2.$$

Also ist X genau dann rekurrent, wenn $D \leq 2$ ist.

Wir werden in Kapitel 19.3 noch eine weitere, strukturell völlig andere Methode kennenlernen, um den Satz von Pólya zu beweisen, die auf dem Zusammenhang von Irrfahrten mit elektrischen Netzwerken beruht.

Tatsächlich können wir mit Hilfe des Satzes von Chung-Fuchs die Greenfunktion $G_D(0,0)$ der symmetrischen einfachen Irrfahrt auf \mathbb{Z}^D berechnen, indem wir das so genannte **Watson Integral**

$$G_D(0,0) = (2\pi)^{-D} \int_{[-\pi, \pi]^D} \frac{D}{D - (\cos(x_1) + \dots + \cos(x_D))} dx \quad (17.24)$$

numerisch berechnen. Hierzu folgen wir [88] (wo sich auch noch eine Verfeinerung des Verfahrens findet) und führen das D -fache Integral auf ein zweifaches zurück. Indem wir die Gleichung $\frac{1}{\lambda} = \int_0^\infty e^{-\lambda t} dt$ für den Integranden benutzen und den Satz von Fubini verwenden, erhalten wir

$$G_D(0,0) = \frac{D}{(2\pi)^D} \int_0^\infty e^{-Dt} \left(\int_{[-\pi, \pi]^D} e^{t(\cos(x_1) + \dots + \cos(x_D))} dx \right) dt$$

und damit

$$G_D(0,0) = D \int_0^\infty e^{-Dt} I_0(t)^D dt, \quad (17.25)$$

wobei $I_0(t) := \frac{1}{\pi} \int_0^\pi e^{t \cos(\theta)} d\theta$ die modifizierte Bessel-Funktion erster Art bezeichnet. Die Darstellung (17.25) lässt sich mittels numerischer Integration in guter Genauigkeit schnell berechnen (siehe Tabelle 17.1).

Für den Fall $D = 3$ hat Watson [178] die Darstellung

$$G_3(0,0) = 12 \frac{18 + 12\sqrt{2} - 10\sqrt{3} - 7\sqrt{6}}{\pi^2} K \left((2 - \sqrt{3})(\sqrt{3} - \sqrt{2}) \right)^2$$

angegeben, wobei $K(m) = \int_0^1 ((1-t^2)(1-mt^2))^{-1/2} dt$ das vollständige elliptische Integral erster Ordnung mit Modul $m \in (-1, 1)$ ist und eine Darstellung als schnell konvergierende Reihe besitzt

$$K(m) = \frac{\pi}{2} \left(1 + \sum_{n=1}^{\infty} \left(\frac{(2n)!}{4^n (n!)^2} \right)^2 m^{2n} \right).$$

Glasser und Zucker [67] haben einen Ausdruck als Produkt von Gammafunktionen gefunden:

$$G_3(0, 0) = \frac{\sqrt{6}}{32\pi^3} \Gamma\left(\frac{1}{24}\right) \Gamma\left(\frac{5}{24}\right) \Gamma\left(\frac{7}{24}\right) \Gamma\left(\frac{11}{24}\right) = 1.5163860591519780181\dots$$

D	$G_D(0, 0)$	$F_D(0, 0)$
2	∞	1
3	1.51638605915	0.34053732955
4	1.23946712185	0.19320167322
5	1.15630812484	0.13517860982
6	1.11696337322	0.10471549562
7	1.09390631559	0.08584493411
8	1.07864701202	0.07291264996
9	1.06774608638	0.06344774965
10	1.05954374789	0.05619753597
11	1.05313615291	0.05045515982
12	1.04798637482	0.04578912090
13	1.04375406289	0.04191989708
14	1.04021240323	0.03865787709
15	1.03720412092	0.03586962312
16	1.03461657857	0.03345836447
17	1.03236691238	0.03135214040
18	1.03039276285	0.02949628913
19	1.02864627888	0.02784852234
20	1.02709011674	0.02637559869

Tabelle 17.1. Greenfunktion $G_D(0, 0)$ und Rückkehrwahrscheinlichkeit $F_D(0, 0)$ der einfachen symmetrischen Irrfahrt auf \mathbb{Z}^D , numerisch berechnet mit Formel (17.25).

Übung 17.5.1. Für $n \in \mathbb{N}_0$ sei p^n die Matrix der n -Schritt Übergangswahrscheinlichkeiten der einfachen symmetrischen Irrfahrt auf \mathbb{Z}^D . Man leite für $n \in \mathbb{N}$ die Formel

$$p^{2n}(0,0) = (2\pi)^{-D} \int_{[-\pi,\pi]^D} D^{-2n} (\cos(t_1) + \dots + \cos(t_D))^{2n} dt$$

her (siehe Satz 15.10). Man schließe durch geeignete Abschätzung des Integrals die Konvergenz $n^{D/2} p^{2n}(0,0) \xrightarrow{n \rightarrow \infty} 2(4\pi/D)^{-D/2}$ (siehe (17.18)). 

Übung 17.5.2. Man zeige (17.21) formal. 

Übung 17.5.3. Man zeige mit Hilfe von Satz 17.41, dass eine Irrfahrt auf \mathbb{Z}^2 mit $\sum_{x \in \mathbb{Z}^2} x p(0,x) = 0$ rekurrent ist, falls $\sum_{x \in \mathbb{Z}^2} \|x\|_2^2 p(0,x) < \infty$. 

Übung 17.5.4. Man zeige mit Hilfe von Satz 17.41, dass für $D \geq 3$ jede irreduzible Irrfahrt auf \mathbb{Z}^D transient ist. 

Übung 17.5.5. Man zeige (17.25) für $G_D(0,0)$ direkt mit den $p^{2n}(0,0)$ aus (17.20) und mit der Reihendarstellung $I_0(t) = \sum_{k=0}^{\infty} (k!)^{-2} (t/2)^k$. 

17.6 Invariante Verteilungen

Sei im Folgenden stets p eine stochastische Matrix auf dem diskreten Raum E sowie $(X_n)_{n \in \mathbb{N}_0}$ eine zugehörige Markovkette.

In diesem Abschnitt interessieren wir uns dafür, welche Wahrscheinlichkeitsverteilungen unter der Dynamik der Markovkette erhalten bleiben. In Kapitel 18 werden wir unter einfachen Bedingungen zeigen, wann die Verteilung einer Markovkette für lange Zeiten gegen eine solche Verteilung konvergiert.

Definition 17.42. Ist μ ein Maß auf E und $f : E \rightarrow \mathbb{R}$ eine Abbildung, so schreiben wir $\mu p(\{x\}) = \sum_{y \in E} \mu(\{y\})p(y,x)$ und $\mu f(x) = \sum_{y \in E} p(x,y)f(y)$, falls die Summen konvergieren.

Definition 17.43. (i) Ein σ -endliches Maß μ auf E heißt **invariantes Maß**, falls

$$\mu p = \mu.$$

μ heißt **invariante Verteilung**, falls zudem $\mu(E) = 1$ gilt. Mit \mathcal{I} bezeichnen wir die Menge der invarianten Verteilungen.

(ii) Eine Funktion $f : E \rightarrow \mathbb{R}$ heißt **subharmonisch**, falls μf existiert und $f \leq \mu f$ gilt. f heißt **superharmonisch**, falls $f \geq \mu f$ gilt und **harmonisch**, falls $f = \mu f$.

Bemerkung 17.44. Im Sinne der linearen Algebra ist ein invariantes Maß ein Links-Eigenvektor von p und eine harmonische Funktion ein Rechts-Eigenvektor, jeweils zum Eigenwert 1. \diamond

Lemma 17.45. Ist f beschränkt und (sub-, super-) harmonisch, so ist $(f(X_n))_{n \in \mathbb{N}_0}$ ein (Sub-, Super-) Martingal bezüglich der erzeugten Filtration $\mathbb{F} = \sigma(X)$.

Beweis. Sei f beschränkt und subharmonisch. Dann ist

$$\begin{aligned}\mathbf{E}_x[f(X_n) | \mathcal{F}_{n-1}] &= \mathbf{E}_{X_{n-1}}[f(X_1)] = \sum_{y \in E} p(X_{n-1}, y) f(y) \\ &= p f(X_{n-1}) \geq f(X_{n-1}).\end{aligned}$$

□

Satz 17.46. Ist jeder Zustand transient, so existiert keine invariante Verteilung.

Beweis. Nach Voraussetzung und Satz 17.34 ist $G(x, y) = \sum_{n=0}^{\infty} p^n(x, y) < \infty$ für alle $x, y \in E$, also $p^n(x, y) \xrightarrow{n \rightarrow \infty} 0$. Für jedes W-Maß μ auf E ist daher $\mu p^n(\{x\}) \xrightarrow{n \rightarrow \infty} 0$. Wäre μ invariant, so wäre aber $\mu p^n(\{x\}) = \mu(\{x\})$ für jedes $n \in \mathbb{N}$. \square

Satz 17.47. Sei x ein rekurrenter Zustand und $\tau_x^1 = \inf\{n \geq 1 : X_n = x\}$. Dann wird ein invariantes Maß μ_x definiert durch

$$\mu_x(\{y\}) = \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbb{1}_{\{X_n=y\}} \right] = \sum_{n=0}^{\infty} \mathbf{P}_x [X_n = y, \tau_x^1 > n].$$

Beweis. Zunächst müssen wir zeigen, dass $\mu_x(\{y\}) < \infty$ ist für jedes $y \in E$. Für $y = x$ ist offenbar $\mu_x(\{x\}) = 1$. Für $y \neq x$ und $F(x, y) = 0$ ist $\mu_x(\{y\}) = 0$. Sei nun $y \neq x$ und $F(x, y) > 0$. Da x rekurrent ist, ist $F(x, y) = F(y, x) = 1$, und y ist rekurrent (Satz 17.35). Sei

$$\widehat{F}(x, y) = \mathbf{P}_x [\tau_x^1 > \tau_y^1].$$

Dann ist $\widehat{F}(x, y) > 0$ (sonst würde y nicht getroffen) und nach Vertauschung der Rollen von x und y auch $\widehat{F}(y, x) > 0$.

Nach der starken Markoveigenschaft (Satz 17.14) ist

$$\begin{aligned}\mathbf{E}_y \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbb{1}_{\{X_n=y\}} \right] &= 1 + \mathbf{E}_y \left[\sum_{n=\tau_x^1}^{\tau_x^1 - 1} \mathbb{1}_{\{X_n=y\}} ; \tau_x^1 > \tau_y^1 \right] \\ &= 1 + \left(1 - \widehat{F}(y, x) \right) \mathbf{E}_y \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbb{1}_{\{X_n=y\}} \right].\end{aligned}$$

Also ist

$$\mathbf{E}_y \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbb{1}_{\{X_n=y\}} \right] = \frac{1}{\widehat{F}(y, x)}.$$

Mithin ist

$$\mu_x(\{y\}) = \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbb{1}_{\{X_n=y\}} \right] = \mathbf{E}_x \left[\sum_{n=\tau_y^1}^{\tau_x^1 - 1} \mathbb{1}_{\{X_n=y\}}; \tau_x^1 > \tau_y^1 \right] = \frac{\widehat{F}(x, y)}{\widehat{F}(y, x)} < \infty.$$

Definiere $\bar{p}_n(x, y) = \mathbf{P}_x [X_n = y; \tau_x^1 > n]$. Dann ist für jedes $z \in E$

$$\mu_x p(\{z\}) = \sum_{y \in E} \mu_x(\{y\}) p(y, z) = \sum_{n=0}^{\infty} \sum_{y \in E} \bar{p}_n(x, y) p(y, z).$$

1. Fall: $x \neq z$. Dann ist

$$\begin{aligned} \sum_{y \in E} \bar{p}_n(x, y) p(y, z) &= \sum_{y \in E} \mathbf{P}_x [X_n = y, \tau_x^1 > n, X_{n+1} = z] \\ &= \mathbf{P}_x [\tau_x^1 > n + 1; X_{n+1} = z] = \bar{p}_{n+1}(x, z). \end{aligned}$$

Also ist (wegen $\bar{p}_0(x, z) = 0$)

$$\mu_x p(\{z\}) = \sum_{n=0}^{\infty} \bar{p}_{n+1}(x, z) = \sum_{n=1}^{\infty} \bar{p}_n(x, z) = \sum_{n=0}^{\infty} \bar{p}_n(x, z) = \mu_x(\{z\}).$$

2. Fall: $x = z$. Jetzt ist

$$\sum_{y \in E} \bar{p}_n(x, y) p(y, x) = \sum_{y \in E} \mathbf{P}_x [X_n = y; \tau_x^1 > n; X_{n+1} = x] = \mathbf{P}_x [\tau_x^1 = n + 1].$$

Also ist (wegen $\mathbf{P}_x [\tau_x^1 = 0] = 0$)

$$\mu_x p(\{x\}) = \sum_{n=0}^{\infty} \mathbf{P}_x [\tau_x^1 = n + 1] = 1 = \mu_x(\{x\}). \quad \square$$

Korollar 17.48. Ist X positiv rekurrent, so ist $\pi := \frac{\mu_x}{\mathbf{E}_x [\tau_x^1]}$ für jedes $x \in E$ eine invariante Verteilung.

Satz 17.49. Ist X irreduzibel, so hat X höchstens eine invariante Verteilung.

Bemerkung 17.50. (i) Man kann auch zeigen: Ist X irreduzibel und rekurrent, so ist ein invariantes Maß von X bis auf einen Faktor eindeutig. Der Beweis

ist allerdings aufwändiger als der für invariante Verteilungen. Weil die Aussage hier nicht benötigt wird, lassen wir den Beweis als Übungsaufgabe (siehe Übung 17.6.6), vergleiche [44, Kapitel 6.5]).

(ii) Für transientes X kann es im Allgemeinen mehrere invariante Maße geben (nicht aber Verteilungen). Beispielsweise sind für die Irrfahrt auf \mathbb{Z} mit Übergangswahrscheinlichkeiten $p(x, x+1) = r$ und $p(x, x-1) = 1-r$ (für ein $r \in (0, 1)$) die invarianten Maße die nichtnegativen Linearkombinationen der Maße μ_1 und μ_2 , die durch $\mu_1(\{x\}) \equiv 1$ und $\mu_2(\{x\}) = (r/(1-r))^x$, $x \in \mathbb{Z}$ gegeben sind. Ist X transient, also $r \neq 1/2$, so ist $\mu_1 \neq \mu_2$. \diamond

Beweis. Seien π und ν invariante Verteilungen. Wähle einen beliebigen Wahrscheinlichkeitsvektor $(g_n)_{n \in \mathbb{N}}$ mit $g_n > 0$ für jedes $n \in \mathbb{N}$. Definiere die stochastische Matrix $\tilde{p}(x, y) = \sum_{n=1}^{\infty} g_n p^n(x, y)$. Dann ist $\tilde{p}(x, y) > 0$ für alle $x, y \in E$ und $\pi \tilde{p} = \pi$ sowie $\nu \tilde{p} = \nu$.

Betrachte nun das signierte Maß $\mu = \pi - \nu$. Es gilt $\mu \tilde{p} = \mu$. Wäre nun $\mu \neq 0$, so gäbe es (wegen $\mu(E) = 0$) Punkte $x_1, x_2 \in E$ mit $\mu(\{x_1\}) > 0$ und $\mu(\{x_2\}) < 0$. Offensichtlich wäre für jedes $y \in E$ dann $|\mu(\{x_1\}) \tilde{p}(x_1, y) + \mu(\{x_2\}) \tilde{p}(x_2, y)| < |\mu(\{x_1\}) \tilde{p}(x_1, y)| + |\mu(\{x_2\}) \tilde{p}(x_2, y)|$, also

$$\begin{aligned} \|\mu \tilde{p}\|_{TV} &= \sum_{y \in E} \left| \sum_{x \in E} \mu(\{x\}) \tilde{p}(x, y) \right| \\ &< \sum_{y \in E} \sum_{x \in E} |\mu(\{x\})| \tilde{p}(x, y) = \sum_{x \in E} |\mu(\{x\})| = \|\mu\|_{TV}. \end{aligned}$$

Da dies widersprüchlich ist, gilt $\mu = 0$. \square

Es sei daran erinnert, dass \mathcal{I} die Menge der invarianten Verteilungen von X ist.

Satz 17.51. Sei X irreduzibel. X ist genau dann positiv rekurrent, wenn $\mathcal{I} \neq \emptyset$ ist. In diesem Fall ist $\mathcal{I} = \{\pi\}$ mit

$$\pi(\{x\}) = \frac{1}{\mathbf{E}_x[\tau_x^1]} > 0 \quad \text{für jedes } x \in E.$$

Beweis. Ist X positiv rekurrent, so ist $\mathcal{I} \neq \emptyset$ nach Korollar 17.48. Sei nun $\mathcal{I} \neq \emptyset$ und $\pi \in \mathcal{I}$. Da X irreduzibel ist, ist $\pi(\{x\}) > 0$ für jedes $x \in E$. Sei $\mathbf{P}_{\pi} = \sum_{x \in E} \pi(\{x\}) \mathbf{P}_x$. Sei $x \in E$ fest und für $n \in \mathbb{N}_0$

$$\sigma_x^n = \sup \{m \leq n : X_m = x\} \in \mathbb{N}_0 \cup \{-\infty\}$$

die letzte Eintrittszeit in x bis zur Zeit n . (Man bemerke, dass dies keine Stoppzeit ist.) Nach der Markoveigenschaft gilt dann für $k \leq n$

$$\begin{aligned}
\mathbf{P}_\pi[\sigma_x^n = k] &= \mathbf{P}_\pi[X_k = x, X_{k+1} \neq x, \dots, X_n \neq x] \\
&= \mathbf{P}_\pi[X_{k+1} \neq x, \dots, X_n \neq x | X_k = x] \mathbf{P}_\pi[X_k = x] \\
&= \pi(\{x\}) \mathbf{P}_x[X_1, \dots, X_{n-k} \neq x] \\
&= \pi(\{x\}) \mathbf{P}_x[\tau_x^1 \geq n - k + 1].
\end{aligned}$$

Also ist für jedes $n \in \mathbb{N}_0$ (wegen $\mathbf{P}_y[\tau_x^1 < \infty] = 1$ für jedes $y \in E$)

$$\begin{aligned}
1 &= \sum_{k=0}^n \mathbf{P}_\pi[\sigma_x^n = k] + \mathbf{P}_\pi[\sigma_x^n = -\infty] \\
&= \pi(\{x\}) \sum_{k=0}^n \mathbf{P}_x[\tau_x^1 \geq n - k + 1] + \mathbf{P}_\pi[\tau_x^1 \geq n + 1] \\
&\xrightarrow{n \rightarrow \infty} \pi(\{x\}) \sum_{k=1}^{\infty} \mathbf{P}_x[\tau_x^1 \geq k] = \pi(\{x\}) \mathbf{E}_x[\tau_x^1].
\end{aligned}$$

Mithin ist $\mathbf{E}_x[\tau_x^1] = \frac{1}{\pi(\{x\})} < \infty$, und damit ist X positiv rekurrent. \square

Beispiel 17.52. Seien $(p_x)_{x \in \mathbb{N}_0}$ Zahlen aus $(0, 1]$, und sei X eine irreduzible Markovkette auf \mathbb{N}_0 mit Übergangsmatrix

$$p(x, y) = \begin{cases} p_x, & \text{falls } y = x + 1, \\ 1 - p_x, & \text{falls } y = 0, \\ 0, & \text{sonst.} \end{cases}$$

Ist μ ein invariantes Maß, also $\mu p = \mu$, dann gilt

$$\begin{aligned}
\mu(\{n\}) &= p_{n-1} \mu(\{n-1\}) \quad \text{für } n \in \mathbb{N}, \\
\mu(\{0\}) &= \sum_{n=0}^{\infty} \mu(\{n\})(1 - p_n).
\end{aligned}$$

Das heißt,

$$\mu(\{n\}) = \mu(\{0\}) \prod_{k=0}^{n-1} p_k$$

und (beachte, dass die Reihe eine Teleskopreihe ist)

$$\mu(\{0\}) = \mu(\{0\}) \sum_{n=0}^{\infty} (1 - p_n) \prod_{k=0}^{n-1} p_k = \mu(\{0\}) \left(1 - \prod_{n=0}^{\infty} p_n \right).$$

Daher existiert genau dann ein nichttriviales invariantes Maß μ (kann also $\mu(\{0\}) > 0$ gewählt werden), wenn $\prod_{n=0}^{\infty} p_n = 0$. Dies ist wiederum äquivalent zu der Bedingung $\sum_{n=0}^{\infty} (1 - p_n) = \infty$. Man zeigt leicht, etwa mit einem Borel-Cantelli Argument, dass dies in der Tat genau die Bedingung für Rekurrenz von X ist.

Gilt $\mu \neq 0$, so ist μ genau dann ein endliches Maß, wenn

$$M := \sum_{n=0}^{\infty} \prod_{k=0}^{n-1} p_k < \infty.$$

Also ist X genau dann positiv rekurrent wenn $M < \infty$. In der Tat kann man leicht zeigen, dass M die erwartete Rückkehrzeit in 0 ist. Das Kriterium für positive Rekurrenz kann in diesem Beispiel also auch mit Hilfe von Satz 17.51 hergeleitet werden.

Eine notwendige Bedingung für $M < \infty$ ist offenbar $\sum_{n=0}^{\infty} (1 - p_n) = \infty$, dass also X rekurrent ist. Eine einfache hinreichende Bedingung für $M < \infty$ ist

$$\sum_{n=0}^{\infty} \exp\left(-\sum_{k=0}^{n-1} (1 - p_k)\right) < \infty. \quad \diamond$$

Übung 17.6.1. Betrachte die Markovkette aus Abb. 17.1 (Seite 374). Man bestimme die Menge aller invarianten Verteilungen und zeige, dass die Zustände 6, 7 und 8 positiv rekurrent sind mit erwarteten Eintrittszeiten

$$\mathbf{E}_6[\tau_6] = \frac{17}{4}, \quad \mathbf{E}_7[\tau_7] = \frac{17}{5} \quad \text{und} \quad \mathbf{E}_8[\tau_8] = \frac{17}{8}. \quad \clubsuit$$

Übung 17.6.2. Sei $X = (X_t)_{t \geq 0}$ eine Markovkette auf E in stetiger Zeit mit Q -Matrix q . Man zeige: Ein Wahrscheinlichkeitsmaß π auf E ist genau dann eine invariante Verteilung für X , wenn $\sum_{x \in E} \pi(\{x\})q(x, y) = 0$ für alle $y \in E$. \clubsuit

Übung 17.6.3. Sei G eine abzählbare, abelsche Gruppe und p die Übergangsmatrix einer irreduziblen Irrfahrt X auf G , das heißt, es gilt $p(hg, hf) = p(g, f)$ für alle $h, g, f \in G$. (Dies verallgemeinert den Begriff der Irrfahrt auf \mathbb{Z}^D .) Man zeige mit Hilfe von Satz 17.51: X ist genau dann positiv rekurrent, wenn G endlich ist. \clubsuit

Übung 17.6.4. Sei $r \in [0, 1]$ und X die Markovkette auf \mathbb{N}_0 mit Übergangsmatrix (siehe Abb. 17.2 auf Seite 375)

$$p(x, y) = \begin{cases} 1, & \text{falls } x = 0 \text{ und } y = 1, \\ r, & \text{falls } y = x + 1 \geq 2, \\ 1 - r, & \text{falls } y = x - 1, \\ 0, & \text{sonst.} \end{cases}$$

Man bestimme das invariante Maß und zeige mit Hilfe von Satz 17.51:

- (i) Ist $r \in (0, \frac{1}{2})$, so ist X positiv rekurrent.
- (ii) Ist $r = \frac{1}{2}$, so ist X nullrekurrent.
- (iii) Ist $r \in \{0\} \cup (\frac{1}{2}, 1]$, so ist X transient. ♣

- Übung 17.6.5.** (i) Man zeige mit einem direkten Argument, dass die Markovkette in Beispiel 17.52 genau dann rekurrent ist, wenn $\sum_{n=0}^{\infty} (1 - p_n) = \infty$ gilt.
(ii) Man zeige, dass M die erwartete Rückkehrzeit zu 0 ist und folgere, dass die Kette genau dann positiv rekurrent ist, wenn $M < \infty$ gilt.
(iii) Man gebe Beispiele für Folgen $(p_x)_{x \in \mathbb{N}_0}$ an, sodass die Kette (a) transient ist, (b) nullrekurrent ist, (c) positiv rekurrent ist, (d) positiv rekurrent ist, aber

$$\sum_{n=0}^{\infty} \exp\left(-\sum_{k=0}^{n-1} (1 - p_k)\right) = \infty. \quad \clubsuit$$

Übung 17.6.6. Sei X eine irreduzible rekurrente Markovkette mit Übergangsmatrix p . Man zeige die Behauptung von Bemerkung 17.50(i), nämlich, dass das invariante Maß eindeutig ist bis auf konstante Vielfache.

Anleitung: Sei $\pi \neq 0$ ein invariantes Maß für X . Wir schreiben kurz $\mathbf{P}_{\pi} = \sum_{x \in E} \pi(\{x\}) \mathbf{P}_x$. (Bemerke, dass dieses Maß im allgemeinen nicht endlich ist.) Für $x, y \in E$ mit $x \neq y$ leite man induktiv her, dass

$$\pi(\{y\}) = \mathbf{P}_{\pi}[\tau_x^1 \geq n, X_0 \neq x, X_n = y] + \sum_{k=1}^n \mathbf{P}_{\pi}[\tau_x^1 \geq k, X_0 = x, X_k = y].$$

Hieraus folgere man

$$\pi(\{y\}) \geq \sum_{k=1}^{\infty} \mathbf{P}_{\pi}[\tau_x^1 \geq k, X_0 = x, X_k = y] = \pi(\{x\}) \mu_x(\{y\}),$$

wobei μ_x das invariante Maß aus Satz 17.47 ist. Schließlich benutze man die Tatsache, dass $\pi p^n = \pi$ und $\mu_x p^n = \mu_x$ für jedes $n \in \mathbb{N}$ gilt, um $\pi(\{y\}) = \pi(\{x\}) \mu_x(\{y\})$ zu zeigen. ♣

17.7 Anwendung: Stochastische Ordnung und Kopplung

Es ist oftmals nützlich, einen gemeinsamen Wahrscheinlichkeitsraum für zwei Verteilungen anzugeben, sodass die jeweiligen Verteilungen sich als die Randverteilungen ergeben. Wir stellen zunächst das Prinzip der Kopplung abstrakt vor und geben dann Beispiele an.

Es gibt mehrere Konzepte, um Wahrscheinlichkeitsmaße auf \mathbb{R} oder \mathbb{R}^d so zu ordnen, dass das „größere“ der beiden tendenziell größere Werte stärker bevorzugt. Eine der prominentesten Ordnungen stellt die stochastische Ordnung dar, die wir hier vorstellen und den Zusammenhang mit Kopplungen erläutern wollen. Schließlich, und dies dient als Entschuldigung dafür, diesen Abschnitt im Kapitel über Markovketten anzusiedeln, benutzen wir eine einfache Markovkette, um einen Satz zur stochastischen Ordnung von Binomialverteilungen zu beweisen.

Definition 17.53. Seien $(E_1, \mathcal{E}_1, \mu_1)$ und $(E_2, \mathcal{E}_2, \mu_2)$ Wahrscheinlichkeitsräume. Ein W-Maß μ auf $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ mit $\mu(\cdot \times E_2) = \mu_1$ und $\mu(E_1 \times \cdot) = \mu_2$ heißt **Kopplung** von μ_1 und μ_2 .

Natürlich ist das Produktmaß $\mu = \mu_1 \otimes \mu_2$ (siehe Satz 14.14) eine Kopplung, aber es gibt auch spannendere Beispiele.

Beispiel 17.54. Seien X eine reelle Zufallsvariable und $f, g : \mathbb{R} \rightarrow \mathbb{R}$ monoton wachsende Funktionen mit $\mathbf{E}[f(X)^2] < \infty$ und $\mathbf{E}[g(X)^2] < \infty$. Wir wollen zeigen, dass die Zufallsvariablen $f(X)$ und $g(X)$ nichtnegativ korreliert sind.

Sei dazu Y eine **unabhängige Kopie** von X , also eine von X unabhängige Zufallsvariable mit $\mathbf{P}_Y = \mathbf{P}_X$. Speziell ist $\mathbf{E}[f(X)] = \mathbf{E}[f(Y)]$ und $\mathbf{E}[g(X)] = \mathbf{E}[g(Y)]$. Für alle Zahlen $x, y \in \mathbb{R}$ ist $(f(x) - f(y))(g(x) - g(y)) \geq 0$. Also ist

$$\begin{aligned} 0 &\leq \mathbf{E}[(f(X) - f(Y))(g(X) - g(Y))] \\ &= \mathbf{E}[f(X)g(X)] - \mathbf{E}[f(X)]\mathbf{E}[g(Y)] + \mathbf{E}[f(Y)g(Y)] - \mathbf{E}[f(Y)]\mathbf{E}[g(X)] \\ &= 2 \mathbf{Cov}[f(X), g(X)]. \end{aligned} \quad \diamond$$

Beispiel 17.55. Sei (E, ϱ) ein polnischer Raum. Für zwei W-Maße P und Q auf $(E, \mathcal{B}(E))$ schreiben wir $K(P, Q) \subset \mathcal{M}_1(E \times E)$ für die Menge der Kopplungen von P und Q . Wir können dann einen Abstand, die so genannte **Wasserstein Metrik**, auf $\mathcal{M}_1(E)$ definieren durch

$$d_W(P, Q) := \inf \left\{ \int \varrho(x, y) \varphi(d(x, y)) : \varphi \in K(P, Q) \right\}. \quad (17.26)$$

Man kann zeigen (Satz von Kantorovich-Rubinstein [92], siehe auch [42, Seite 420ff]), dass

$$d_W(P, Q) = \sup \left\{ \int f d(P - Q) : f \in \text{Lip}_1(E; \mathbb{R}) \right\}. \quad (17.27)$$

Man vergleiche diese Darstellung der Wasserstein Metrik mit derjenigen der Totalvariationsnorm:

$$\|P - Q\|_{TV} = \sup \left\{ \int f d(P - Q) : f \in \mathcal{L}^\infty(E) \text{ mit } \|f\|_\infty \leq 1 \right\}. \quad (17.28)$$

Tatsächlich können wir auch hier eine Definition durch eine Kopplung angeben: Sei $D := \{(x, x) : x \in E\}$ die Diagonale in $E \times E$. Dann ist

$$\|P - Q\|_{TV} = \inf \{\varphi((E \times E) \setminus D) : \varphi \in K(P, Q)\}. \quad (17.29)$$

Siehe [66] für einen Vergleich verschiedener Metriken auf $\mathcal{M}_1(E)$. \diamond

Ein weiteres Beispiel für eine komplexere Kopplung liefert der folgende Satz von Skorohod, den wir hier nur zitieren.

Satz 17.56 (Skorohod Kopplung). *Es seien μ, μ_1, μ_2, \dots W-Maße auf einem polnischen Raum E mit $\mu_n \xrightarrow{n \rightarrow \infty} \mu$. Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$ mit Zufallsvariablen X, X_1, X_2, \dots mit $\mathbf{P}_X = \mu$ und $\mathbf{P}_{X_n} = \mu_n$ für jedes $n \in \mathbb{N}$ sowie $X_n \xrightarrow{n \rightarrow \infty} X$ fast sicher.*

Beweis. Siehe etwa [91, Seite 79]. \square

Wir kommen nun zum Begriff der stochastischen Ordnung.

Definition 17.57. *Seien $\mu_1, \mu_2 \in \mathcal{M}_1(\mathbb{R}^d)$. Wir schreiben $\mu_1 \leq_{st} \mu_2$, falls*

$$\int f d\mu_1 \leq \int f d\mu_2$$

für jede monoton wachsende, beschränkte Funktion $f : \mathbb{R}^d \rightarrow \mathbb{R}$ gilt. Wir sagen dann, dass μ_2 **stochastisch größer** als (oder gleich) μ_1 ist.

Offenbar ist \leq_{st} eine Halbordnung auf $\mathcal{M}_1(\mathbb{R}^d)$. Die stochastische Ordnung gehört zur Klasse der so genannten Integralordnungen, da für eine gewisse Klasse von Funktionen (hier: beschränkt und monoton wachsend) gefordert wird, dass die Integrale geordnet sind. Man kann stattdessen etwa konvexe Funktionen betrachten, oder Indikatorfunktionen auf unteren oder oberen Orthanten.

Seien F_1 und F_2 die Verteilungsfunktionen von μ_1 und μ_2 . Gilt $\mu_1 \leq_{st} \mu_2$, so gilt $F_1(x) \geq F_2(x)$ für jedes $x \in \mathbb{R}^d$. Ist $d = 1$, so sind die beiden Aussagen äquivalent. Für $d \geq 2$ hingegen ist die Bedingung $F_1 \geq F_2$ schwächer als die Bedingung $\mu_1 \leq_{st} \mu_2$. Man betrachte hierzu etwa das Beispiel $d = 2$ sowie

$$\mu_1 = \frac{1}{2} \delta_{(0,0)} + \frac{1}{2} \delta_{(1,1)} \quad \text{und} \quad \mu_2 = \frac{1}{2} \delta_{(1,0)} + \frac{1}{2} \delta_{(0,1)}.$$

Die Halbordnung, die durch Vergleich der Verteilungsfunktionen definiert wird, wird auch als (untere) Orthantenordnung bezeichnet.

Einen Überblick über verschiedene Ordnungen von Wahrscheinlichkeitsmaßen findet man beispielsweise in [130]. Der folgende Satz wurde von Strassen [169] in weit größerer Allgemeinheit für Integralordnungen gezeigt.

Satz 17.58 (Satz von Strassen). Sei

$$L := \{(x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d : x_1 \leq x_2\}.$$

Es gilt genau dann $\mu_1 \leq_{st} \mu_2$, wenn es eine Kopplung φ von μ_1 und μ_2 gibt mit $\varphi(L) = 1$.

Beweis. Sei φ eine solche Kopplung. Für jedes monoton wachsendes, beschränktes $f : \mathbb{R}^d \rightarrow \mathbb{R}$ ist $f(x_1) - f(x_2) \leq 0$ für jedes $x = (x_1, x_2) \in L$, also

$$\int f d\mu_1 - \int f d\mu_2 = \int_L (f(x_1) - f(x_2)) \varphi(dx) \leq 0$$

und damit $\mu_1 \leq_{st} \mu_2$.

Gelte nun $\mu_1 \leq_{st} \mu_2$. Wir betrachten nur $d = 1$ (siehe [130, Thm. 3.3.5] für $d \geq 2$). Durch $F((x_1, x_2)) := \min(F_1(x_1), F_2(x_2))$ wird eine Verteilungsfunktion auf $\mathbb{R} \times \mathbb{R}$ definiert (Übung 1.5.5), die zu einer Kopplung φ mit $\varphi(L) = 1$ gehört. Etwas expliziter lässt sich diese Kopplung durch Zufallsvariablen darstellen. Sei U eine auf $(0, 1)$ gleichverteilte Zufallsvariable und

$$X_i := F_i^{-1}(U) := \inf \{x \in \mathbb{R} : F_i(x) \geq U\}$$

eine nach μ_i verteilte Zufallsvariable (siehe Beweis von Satz 1.104). Dann gilt $X_1 \leq X_2$ fast sicher, das heißt $\mathbf{P}[(X_1, X_2) \in L] = 1$. Offenbar hat (X_1, X_2) die Verteilungsfunktion F . \square

Während der Satz von Strassen nur abstrakt die Existenz von Kopplungen liefert, gibt es in vielen Beispielen eine natürliche Kopplung, die oftmals auch als Hilfsmittel für den Beweis einer stochastischen Ordnung verwendet werden kann.

Beispiel 17.59. Sei $n \in \mathbb{N}$ und $0 \leq p_1 \leq p_2 \leq 1$ sowie Y_1, \dots, Y_n unabhängige uniform auf $[0, 1]$ verteilte Zufallsvariablen. Setze $X_i = \#\{k \leq n : Y_k \leq p_i\}$, $i = 1, 2$. Dann ist $X_i \sim b_{n, p_i}$ und $X_1 \leq X_2$ fast sicher. Diese Kopplung zeigt unmittelbar $b_{n, p_1} \leq_{st} b_{n, p_2}$.

Ähnlich simpel ist die Aussage $b_{m, p} \leq_{st} b_{n, p}$ für $m \leq n$ und $p \in [0, 1]$. \diamond

Satz 17.60. Seien $n_1, n_2 \in \mathbb{N}$ und $p_1, p_2 \in (0, 1)$. Dann gilt $b_{n_1, p_1} \leq_{st} b_{n_2, p_2}$ genau dann, wenn

$$(1 - p_1)^{n_1} \geq (1 - p_2)^{n_2} \tag{17.30}$$

und

$$n_1 \leq n_2. \tag{17.31}$$

Beweis. (Der Beweis folgt der Darstellung in [107, Section 3].)

Offenbar sind (17.30) und (17.31) wegen $b_{n_i, p_i}(\{0\}) = (1 - p_i)^{n_i}$ notwendig für

die stochastische Ordnung der beiden Verteilungen. Zu zeigen ist nur, dass diese Bedingungen auch hinreichend sind.

Es gelten nun also (17.30) und (17.31). Nach Beispiel 17.59 reicht es, das kleinste p_2 zu betrachten, das (17.30) erfüllt. Es gelte also $(1 - p_1)^{n_1} = (1 - p_2)^{n_2}$. Wir setzen $\lambda := -n_1 \log(1 - p_1) = -n_2 \log(1 - p_2)$. Wir wollen nun die Binomialverteilung konstruieren, indem wir eine Poi_λ -verteilte Anzahl von Kugeln in n_i Urnen verteilen und die Anzahl der nichtleeren Urnen bestimmen. Sei dazu $T \sim \text{Poi}_\lambda$ und seien X_1, X_2, \dots unabhängig und uniform verteilt auf $[0, 1]$ sowie unabhängig von T . Für $n \in \mathbb{N}$, $t \in \mathbb{N}_0$ und $l = 1, \dots, n$ sei

$$M_{n,t,l} = \#\{s \leq t : X_s \in ((l-1)/n, l/n]\}$$

und

$$N_{n,t} := \sum_{l=1}^n \mathbf{1}_{\{M_{n,t,l} > 0\}}$$

(die Anzahl der nichtleeren Urnen nachdem t Kugeln verteilt wurden). Nach Satz 5.35 sind $M_{n,T,1}, \dots, M_{n,T,n}$ unabhängig und $\text{Poi}_{\lambda/n}$ -verteilt, speziell also

$$\mathbf{P}[M_{n_i,T,l} > 0] = 1 - e^{-\lambda/n_i} = p_i$$

und damit $N_{n_i,T} \sim b_{n_i, p_i}$, $i = 1, 2$. Es reicht nun also zu zeigen, dass $N_{n_1,T} \leq_{\text{st}} N_{n_2,T}$ gilt. Hierfür wiederum reicht es zu zeigen, dass

$$N_{n_1,t} \leq_{\text{st}} N_{n_2,t} \quad \text{für alle } t \in \mathbb{N}_0. \quad (17.32)$$

gilt. In der Tat: Sei $f : \{0, \dots, n\} \rightarrow \mathbb{R}$ monoton wachsend. Dann ist

$$\begin{aligned} \mathbf{E}[f(N_{n_1,T})] &= \sum_{t=0}^{\infty} \mathbf{E}[f(N_{n_1,t})] \mathbf{P}[T = t] \\ &\leq \sum_{t=0}^{\infty} \mathbf{E}[f(N_{n_2,t})] \mathbf{P}[T = t] = \mathbf{E}[f(N_{n_2,T})]. \end{aligned}$$

Wir zeigen nun also (17.32) induktiv. Für $t = 0$ ist die Aussage trivial. Gelte nun (17.32) für t . Jetzt sind wir an der Stelle, wo wir Markovketten einsetzen. Man bemerke, dass (für festes n) $(N_{n,t})_{t=0,1,\dots}$ eine Markovkette ist mit Zustandsraum $\{0, \dots, n\}$ und Übergangsmatrix

$$p_n(k, l) = \begin{cases} k/n, & \text{falls } l = k, \\ 1 - k/n, & \text{falls } l = k + 1, \\ 0, & \text{sonst.} \end{cases}$$

Wir definieren nun noch

$$h_{n,l}(k) = \sum_{j=l}^n p_n(k, j) = \begin{cases} 0, & \text{falls } k < l - 1, \\ 1 - k/n, & \text{falls } k = l - 1, \\ 1, & \text{falls } k > l - 1. \end{cases}$$

Dann ist $\mathbf{P}[N_{n,t+1} \geq l] = \mathbf{E}[h_{n,l}(N_{n,t})]$ und $h_{n,l}(k)$ ist monoton wachsend in k und n . Also gilt nach Induktionsvoraussetzung

$$\begin{aligned}\mathbf{P}[N_{n_1,t+1} \geq l] &= \mathbf{E}[h_{n_1,l}(N_{n_1,t})] \leq \mathbf{E}[h_{n_1,l}(N_{n_2,t})] \\ &\leq \mathbf{E}[h_{n_2,l}(N_{n_2,t})] = \mathbf{P}[N_{n_2,t+1} \geq l].\end{aligned}$$

Mithin ist $N_{n_1,t+1} \leq_{st} N_{n_2,t+1}$, und damit ist die Induktion und der Beweis vollständig. \square

Übung 17.7.1. Man zeige die Aussage von Satz 17.60 für den Fall $n_2/n_1 \in \mathbb{N}$ mit einem direkten Kopplungsargument. ♣

Übung 17.7.2. Man zeige für die Poissonverteilung:

$$\text{Poi}_{\lambda_1} \leq_{st} \text{Poi}_{\lambda_2} \iff \lambda_1 \leq \lambda_2.$$



Übung 17.7.3. Seien $n \in \mathbb{N}$, $p \in (0, 1)$ und $\lambda > 0$. Man zeige:

$$b_{n,p} \leq_{st} \text{Poi}_\lambda \iff (1-p)^n \geq e^{-\lambda}.$$



18 Konvergenz von Markovketten

Wir betrachten eine Markovkette X mit invariante Verteilung π und untersuchen unter welchen Bedingungen die Verteilung von X_n für $n \rightarrow \infty$ gegen π konvergiert. Im Wesentlichen ist dafür notwendig und hinreichend, dass der Zustandsraum der Kette nicht in Unterräume zerfällt, die

- von der Kette nicht verlassen werden,
- oder von der Kette beispielsweise nur für ungerade n beziehungsweise gerade n besucht werden.

Im ersten Fall wäre die Kette reduzibel, im zweiten hingegen periodisch.

Wir untersuchen Periodizität von Ketten im ersten Abschnitt und zeigen im zweiten den Konvergenzsatz. Im dritten Abschnitt beschäftigen wir uns mit Anwendungen des Konvergenzsatzes für Computersimulationen mit der so genannten Markovketten Monte Carlo Methode. Im letzten Abschnitt beschreiben wir die Geschwindigkeit der Konvergenz gegen das Gleichgewicht mit Hilfe des Spektrums der Übergangsmatrix.

18.1 Periodizität von Markovketten

Wir untersuchen, unter welchen Bedingungen eine positiv rekurrente Markovkette X auf dem abzählbaren Raum E (und mit Übergangsmatrix p), die in einem beliebigen $\mu \in \mathcal{M}_1(E)$ gestartet wird, in Verteilung gegen eine invariante Verteilung π konvergiert, also $\mu p^n \xrightarrow{n \rightarrow \infty} \pi$ gilt. Sicherlich ist hierzu notwendig, dass π die einzige invariante Verteilung ist, und damit bis auf Vielfache der einzige Links-Eigenvektor von p zum Eigenwert 1. Hierfür ist ausreichend, dass die Kette irreduzibel ist (Satz 17.49).

Es sind gewisse Kontraktionseigenschaften von p notwendig, damit $\mu p^n \xrightarrow{n \rightarrow \infty} \pi$ für jedes $\mu \in \mathcal{M}_1(E)$ gelten kann. Offenbar ist 1 der betragsmäßig größte Eigenwert von p . Allerdings ist p nur dann (ausreichend) kontrahierend, wenn die Vielfachheit dieses Eigenwertes genau 1 ist und keine weiteren (komplexwertigen) Eigenwerte mit Betrag 1 existieren.

Für die letztgenannte Bedingung ist die Irreduzibilität der Kette nicht hinreichend, wie wir sehen, wenn wir auf $E = \{0, \dots, N-1\}$ die Markovkette mit Übergangsmatrix $p(x, y) = \mathbb{1}_{\{y=x+1(\text{mod } N)\}}$ betrachten. Der Eigenwert 1 hat die Vielfachheit 1. Jedoch sind alle N -ten Einheitswurzeln $e^{2\pi i k/N}$, $k = 0, \dots, N-1$, ebenfalls Eigenwerte mit Betrag 1. Offenbar ist die Gleichverteilung auf E invariant, jedoch existiert $\lim_{n \rightarrow \infty} \delta_x p^n$ für kein $x \in E$, denn jeder Punkt wird periodisch immer nur nach jeweils genau N Schritten besucht. Um Konvergenz zu erzielen, müssen wir also zunächst Periodizität untersuchen (und ausschließen). Hernach können wir für irreduzible *aperiodische* Markovketten einen Konvergenzsatz angeben.

Sind $m, n \in \mathbb{N}$, so schreiben wir $m|n$, falls m ein Teiler von n ist, also falls $\frac{n}{m} \in \mathbb{N}$. Ist $M \subset \mathbb{N}$, so schreiben wir $\text{ggT}(M)$ für den größten gemeinsamen Teiler aller $n \in M$. Sei im Folgenden stets X eine Markovkette auf dem abzählbaren Raum E mit Übergangsmatrix p .

Definition 18.1. (i) Für $x, y \in E$ schreiben wir

$$N(x, y) := \{n \in \mathbb{N}_0 : p^n(x, y) > 0\}.$$

Für jedes $x \in E$ heißt $d_x := \text{ggT}(N(x, x))$ die **Periode** des Punktes x .

(ii) Ist $d_x = d_y$ für alle $x, y \in E$, so heißt $d := d_x$ die Periode von X .

(iii) Ist $d_x = 1$ für jedes $x \in E$, so heißt X **aperiodisch**.

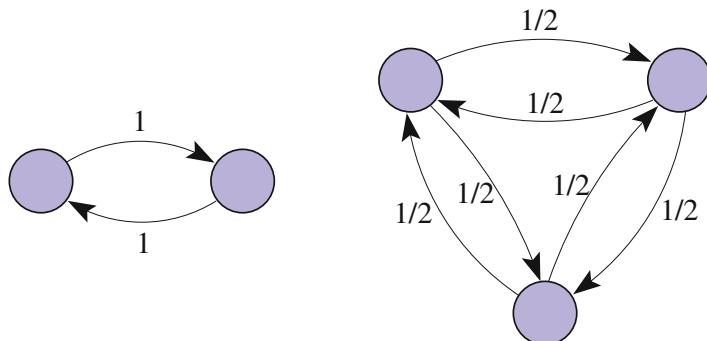


Abb. 18.1. Die linke Markovkette ist periodisch mit Periode 2, die rechte Markovkette ist aperiodisch.

Lemma 18.2. Für jedes $x \in E$ existiert ein $n_x \in \mathbb{N}$ mit

$$p^{nd_x}(x, x) > 0 \quad \text{für jedes } n \geq n_x. \quad (18.1)$$

Beweis. Seien $k_1, \dots, k_r \in N(x, x)$ mit $\text{ggT}(\{k_1, \dots, k_r\}) = d_x$. Dann ist für alle $m_1, \dots, m_r \in \mathbb{N}_0$ auch $\sum_{i=1}^r k_i m_i \in N(x, x)$. Elementare Zahlentheorie liefert

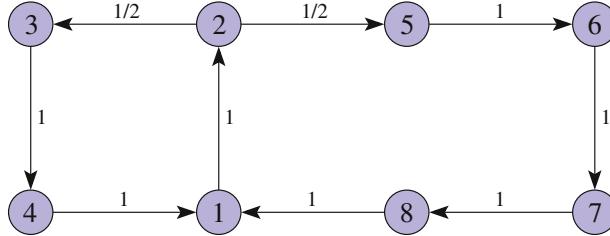


Abb. 18.2. Es ist $N(8, 8) = \{6, 10, 12, 14, 16, \dots\}$, also $d_8 := \text{ggT}(\{6, 10, 12, \dots\}) = 2$ und $n_8 = 5$. Die Kette hat also Periode 2. Hingegen ist $n_1 = 2$ und $n_4 = 4$.

uns nun, dass für jedes $n \geq n_x := r \cdot \prod_{i=1}^r (k_i/d_x)$ Zahlen $m_1, \dots, m_r \in \mathbb{N}_0$ existieren mit $n d_x = \sum_{i=1}^r k_i m_i$. Also gilt (18.1). \square

Das Problem, die *kleinste* Zahl N zu finden, sodass sich jedes $n d_x$, $n \geq N$ als nichtnegative ganzzahlige Linearkombination von k_1, \dots, k_r darstellen lässt, wird *Frobenius Problem* genannt. Die allgemeine Lösung ist unbekannt, allerdings hat Sylvester [172] für den Fall $r = 2$ gezeigt, dass $N = (k_1/d_x - 1)(k_2/d_x - 1)$ minimal ist. Im allgemeinen Fall ist als obere Schranken für N beispielsweise $2 \max\{k_i : i = 1, \dots, r\}^2 / (rd_x^2)$ bekannt, siehe etwa [50].

Lemma 18.3. Sei X irreduzibel. Dann gelten:

(i) $d := d_x = d_y$ für alle $x, y \in E$.

(ii) Für alle $x, y \in E$ existieren $n_{x,y} \in \mathbb{N}$ und $L_{x,y} \in \{0, \dots, d-1\}$ mit

$$nd + L_{x,y} \in N(x, y) \quad \text{für jedes } n \geq n_{x,y}. \quad (18.2)$$

$L_{x,y}$ ist eindeutig bestimmt, und es gilt

$$L_{x,y} + L_{y,z} + L_{z,x} = 0 \pmod{d} \quad \text{für alle } x, y, z \in E. \quad (18.3)$$

Beweis. (i) Seien $m, n \in \mathbb{N}_0$ mit $p^m(x, y) > 0$ und $p^n(y, z) > 0$. Dann ist

$$p^{m+n}(x, z) \geq p^m(x, y) p^n(y, z) > 0.$$

Also gilt

$$N(x, y) + N(y, z) := \{m + n : m \in N(x, y), n \in N(y, z)\} \subset N(x, z). \quad (18.4)$$

Sind speziell $m \in N(x, y)$, $n \in N(y, z)$ und $k \geq n_y$, so ist $kd_y \in N(y, y)$, also $m + kd_y \in N(x, y)$ und $m + n + kd_y \in N(x, z)$. Es folgt $d_x | (m + n + kd_y)$ für jedes $k \geq n_y$, also $d_x | d_y$. Analog erhalten wir $d_y | d_x$, also $d_x = d_y$.

(ii) Sei $m \in N(x, y)$. Dann ist $m + kd \in N(x, y)$ für jedes $k \geq n_x$. Also gilt (18.2) mit

$$n_{x,y} := n_x + \left\lfloor \frac{m}{d} \right\rfloor \quad \text{und} \quad L_{x,y} := m - d \left\lfloor \frac{m}{d} \right\rfloor.$$

Wegen (18.4) ist

$$(n_{x,y} + n_{y,z})d + L_{x,y} + L_{y,z} \in N(x, z).$$

Mit $z = x$ folgt: $d|(L_{x,y} + L_{y,x})$, also ist $L_{x,y}$ eindeutig in $\{0, \dots, d-1\}$ und $L_{x,y} = -L_{y,x} \pmod{d}$. Für allgemeines z folgt: $d|(L_{x,y} + L_{y,z} + L_{z,x})$, also gilt (18.3). \square

Satz 18.4. Sei X irreduzibel mit Periode d . Dann existiert eine disjunkte Zerlegung des Zustandsraums

$$E = \bigcup_{i=0}^{d-1} E_i \quad (18.5)$$

mit der Eigenschaft

$$p(x, y) > 0 \text{ und } x \in E_i \implies y \in E_{i+1 \pmod{d}}. \quad (18.6)$$

Bis auf zyklische Vertauschung ist diese Zerlegung eindeutig.

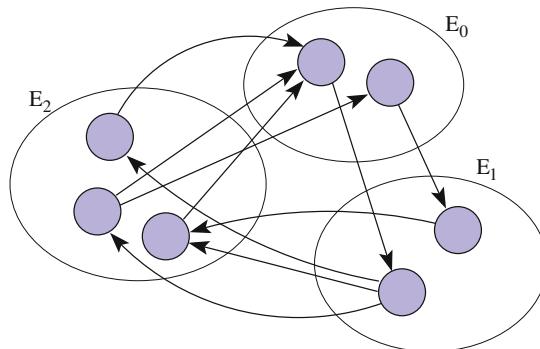


Abb. 18.3. Markovkette mit Periode $d = 3$.

Die Eigenschaft (18.6) besagt gerade, dass X die Mengen E_i nacheinander besucht und zu jedem Zeitschritt in das nächste E_i wechselt (siehe Abb. 18.3 oder Abb. 18.2, wo $d = 2$, $E_0 = \{1, 3, 5, 7\}$ und $E_1 = \{2, 4, 6, 8\}$ ist). Etwas formaler können wir dies schreiben als: Ist $x \in E_i$ für gewisses i , so ist $\mathbf{P}_x[X_n \in E_{i+n \pmod{d}}] = 1$.

Beweis. „Existenz“ Wähle ein beliebiges $x_0 \in E$ und setze

$$E_i := \{y \in E : L_{x_0, y} = i\} \quad \text{für } i = 0, \dots, d-1.$$

Offenbar gilt (18.5). Sei $i \in \{0, \dots, d-1\}$ und $x \in E_i$. Ist $y \in E$ mit $p(x, y) > 0$, so ist $L_{x,y} = 1$, also ist $L_{x_0,y} = L_{x_0,x} + L_{x,y} = i + 1 \pmod{d}$.

„Eindeutigkeit“ Sei $(\tilde{E}_i, i = 0, \dots, \tilde{d} - 1)$ eine weitere Zerlegung, die (18.5) und (18.6) erfüllt. Ohne Einschränkung sei $E_0 \cap \tilde{E}_0 \neq \emptyset$ (sonst vertausche die \tilde{E}_i zyklisch bis dies gilt) und $x_0 \in E_0 \cap \tilde{E}_0$ beliebig. Nach Voraussetzung impliziert $p(x_0, y) > 0$ nun $y \in E_1$ und $y \in \tilde{E}_1$, also $y \in E_1 \cap \tilde{E}_1$. Iterativ erhalten wir, dass $p^{nd+i}(x, y) > 0$ impliziert, dass $y \in E_i \cap \tilde{E}_i$ (für $n \in \mathbb{N}$ und $i = 0, \dots, d - 1$).

Da die Kette irreduzibel ist, existieren aber für jedes $y \in E$ Zahlen $n(y)$ und $i(y)$, sodass $p^{n(y)d+i(y)}(x_0, y) > 0$, also $y \in E_{i(y)} \cap \tilde{E}_{i(y)}$. Mithin gilt $E_i = \tilde{E}_i$ für jedes $i = 0, \dots, d - 1$. \square

18.2 Kopplung und Konvergenzsatz

Wir wollen die Kopplung diskreter Markovketten betrachten, die in unterschiedlichen Verteilungen μ und ν gestartet werden, und hiermit den Konvergenzsatz für Markovketten herleiten.

Im Folgenden sei E stets ein abzählbarer Raum und p eine stochastische Matrix auf E . Es sei an die allgemeine Definition 17.53 der Kopplung erinnert.

Definition 18.5. Ein bivariater Prozess $((X_n, Y_n))_{n \in \mathbb{N}_0}$ mit Werten in $E \times E$ heißt eine Kopplung, falls $(X_n)_{n \in \mathbb{N}_0}$ und $(Y_n)_{n \in \mathbb{N}_0}$ Markovketten mit Übergangsmatrix p sind.

Eine Kopplung heißt erfolgreich, falls $\mathbf{P}_{(x,y)}[\bigcup_{m \geq n} \{X_m \neq Y_m\}] \xrightarrow{n \rightarrow \infty} 0$ für alle $x, y \in E$.

Natürlich sind zwei unabhängig laufende Ketten eine Kopplung, allerdings vielleicht nicht die interessanteste.

Beispiel 18.6 (Unabhängiges Verschmelzen). Die wichtigste Kopplung sind die verschmelzenden Markovketten: Wir lassen X und Y unabhängig voneinander mit Übergangsmatrix p laufen, so lange bis sie sich das erste Mal treffen. Danach laufen die Ketten gemeinsam weiter. Diese Kopplung nennen wir **unabhängiges Verschmelzen**, sie hat die Übergangsmatrix

$$\bar{p}((x_1, y_1), (x_2, y_2)) = \begin{cases} p(x_1, x_2) \cdot p(y_1, y_2), & \text{falls } x_1 \neq y_1, \\ p(x_1, x_2), & \text{falls } x_1 = y_1, x_2 = y_2, \\ 0, & \text{falls } x_1 = y_1, x_2 \neq y_2. \end{cases}$$

Mit $\tau := \inf\{n \in \mathbb{N}_0 : X_n = Y_n\}$ bezeichnen wir den Verschmelzungszeitpunkt. Wir können die Kopplung tatsächlich aus zwei unabhängigen Ketten \tilde{X} und \tilde{Y} herstellen, indem wir $X := \tilde{X}$ setzen, $\tilde{\tau} := \inf\{n \in \mathbb{N}_0 : \tilde{X}_n = \tilde{Y}_n\}$ und

$$Y_n := \begin{cases} \tilde{Y}_n, & \text{falls } n < \tilde{\tau}, \\ X_n, & \text{falls } n \geq \tilde{\tau}. \end{cases}$$

Statt mit einer direkten Rechnung nachzuweisen, dass der so definierte Prozess (X, Y) eine Kopplung mit Übergangsmatrix \bar{p} ist, betrachten wir die Konstruktion von Markovketten aus Satz 17.17: Seien $(R_n(x) : n \in \mathbb{N}_0, x \in E)$ unabhängige Zufallsvariablen mit Verteilung $\mathbf{P}[R_n(x_1) = x_2] = p(x_1, x_2)$, und sei $\tilde{R}_n((x_1, y_1)) = (R_n(x_1), R_n(y_1))$. Dann ist $(\tilde{R}_n)_{n \in \mathbb{N}_0}$ unabhängig, und es gilt $\mathbf{P}[\tilde{R}_n((x_1, y_1)) = (x_2, y_2)] = \bar{p}((x_1, y_1), (x_2, y_2))$. Wie in Satz 17.17 gesehen, definiert $X_{n+1} := R_n(X_n)$ und $Y_{n+1} := R_n(Y_n)$ zwei Markovketten mit Übergangsmatrix p . Andererseits ist $(X_{n+1}, Y_{n+1}) = \tilde{R}_n((X_n, Y_n))$, also ist der bivariate Prozess tatsächlich eine Kopplung mit Übergangsmatrix \bar{p} . \diamond

Beispiel 18.7. Sei $E = \mathbb{Z}$ und $p(x, y) = 1/3$, falls $|x - y| \leq 1$ und 0 sonst. Offenbar gehört p zu einer aperiodischen, rekurrenten Irrfahrt auf \mathbb{Z} . Wir wollen zeigen, dass wir eine erfolgreiche Kopplung durch Verschmelzung unabhängiger Ketten erreichen können.

Seien also \tilde{X} und \tilde{Y} unabhängige Irrfahrten mit Übergangsmatrix p . Dann ist die Differenz $(Z_n)_{n \in \mathbb{N}_0} := (\tilde{X}_n - \tilde{Y}_n)_{n \in \mathbb{N}_0}$ eine symmetrische Irrfahrt mit endlichem Erwartungswert, also rekurrent. Zudem ist Z irreduzibel. Für je zwei Punkte $x, y \in \mathbb{Z}$ gilt also $\mathbf{P}_{(x,y)}[\tilde{\tau} < \infty] = \mathbf{P}_{x-y}[Z_n = 0 \text{ für ein } n \in \mathbb{N}_0] = 1$. Also verschmelzen X und Y fast sicher. \diamond

Rekurrenz, Irreduzibilität und Aperiodizität alleine reichen nicht aus, damit das unabhängige Verschmelzen eine erfolgreiche Kopplung ist. In Übung 18.2.4 wird an einem Beispiel gezeigt, dass wir auf die räumliche Homogenität nicht leicht verzichten können, wenn wir eine erfolgreiche Kopplung haben möchten. Der Verzicht auf Rekurrenz fällt leichter, wie der folgende Satz zeigt.

Satz 18.8. Sei X eine beliebige aperiodische und irreduzible Irrfahrt auf \mathbb{Z}^d mit Übergangsmatrix p . Dann existiert eine erfolgreiche Kopplung (X, Y) .

Beweis. 1. Schritt. Wir betrachten zunächst den Fall, wo $p(0, x) = 3^{-d}$ ist für jedes $x \in \{-1, 0, 1\}^d$. Die einzelnen Koordinaten $X^{(1)}, \dots, X^{(d)}$ von X sind unabhängige Irrfahrten auf \mathbb{Z} mit Übergangswahrscheinlichkeiten $\mathbf{P}_0[X_1^{(i)} = x_i] = 1/3$ für $x_i = -1, 0, 1$. Nach Beispiel 18.7 können wir nun unabhängige erfolgreiche Kopplungen $(X^{(i)}, Y^{(i)})$, $i = 1, \dots, d$, mit Verschmelzungszeiten $\tau^{(i)}$ konstruieren. Setzen wir $Y = (Y^{(1)}, \dots, Y^{(d)})$ und $\tau = \max\{\tau^{(1)}, \dots, \tau^{(d)}\} < \infty$, so ist (X, Y) eine erfolgreiche Kopplung und $X_n = Y_n$ für $n \geq \tau$.

2. Schritt. Wir betrachten nun den Fall, wo

$$\lambda' := \min \{p(0, x) : x \in \{-1, 0, 1\}^d\} > 0$$

ist. Wir definieren $\lambda = 3^d \lambda'$. Ist $\lambda = 1$, so liegt der Fall von Schritt 1 vor und wir sind fertig. Sei nun also $\lambda \in (0, 1)$. Wir definieren die Übergangsmatrix \hat{p} auf \mathbb{Z}^d durch $\hat{p}(x, y) = 3^{-d}$ für $y - x \in \{-1, 0, 1\}^d$. Dann ist auch $\check{p} := (p - \lambda \hat{p}) / (1 - \lambda)$ die Übergangsmatrix einer Irrfahrt auf \mathbb{Z}^d , und es gilt

$$p = \lambda \hat{p} + (1 - \lambda) \check{p}.$$

Seien nun \hat{X} und \check{X} unabhängige Irrfahrten mit Übergangsmatrizen \hat{p} und \check{p} und mit Start in $\hat{X}_0 = X_0$ und $\check{X}_0 = 0$. Seien ferner Z_1, Z_2, \dots u.i.v. Bernoulli Zufallsvariablen mit Parameter λ , die von \hat{X} und \check{X} unabhängig sind. Setze $S_n := Z_1 + \dots + Z_n$ für $n \in \mathbb{N}$ und

$$X_n := \hat{X}_{S_n} + \check{X}_{n-S_n}.$$

Mit anderen Worten: In jedem Schritt wird per Münzwurf ausgesucht, ob ein Sprung gemäß der Matrix \hat{p} oder \check{p} gemacht wird. Also ist X eine Irrfahrt mit Übergangsmatrix p .

Nach Schritt 1 existiert eine erfolgreiche Kopplung (\hat{X}, \hat{Y}) , so dass \hat{Y} unabhängig von \check{X} und Z_1, Z_2, \dots ist. Also ist

$$Y_n := \hat{Y}_{S_n} + \check{X}_{n-S_n}, \quad n \in \mathbb{N},$$

ebenfalls eine Irrfahrt mit Übergangsmatrix p . Da $S_n \rightarrow \infty$ fast sicher gilt, ist die Kopplung (X, Y) erfolgreich.

3. Schritt. Schließlich betrachten wir den allgemeinen Fall. Da X irreduzibel und aperiodisch ist, gibt es nach Lemma 18.3(ii) ein $N \in \mathbb{N}$, so dass für die N -Schritt Übergangsmatrix gilt

$$p^N(0, x) > 0 \quad \text{für alle } x \in \{-1, 0, 1\}^d.$$

Die Irrfahrt $X' = (X'_n)_{n \in \mathbb{N}} := (X_{nN})_{n \in \mathbb{N}}$ erfüllt also die Bedingungen aus Schritt 2. Sei (X', Y') die dort konstruierte Kopplung, und sei

$$\tau := \inf \{n \in \mathbb{N}_0 : X'_m = Y'_m \text{ für alle } m \geq n\}.$$

Dann ist Y' eine Irrfahrt mit Übergangsmatrix p^N . Wir definieren $Y_{nN} := Y'_n$ für $n \in \mathbb{N}_0$ und müssen nun noch die Lücken zwischen den Punkten $\{0, N, 2N, \dots\}$ geeignet auffüllen, so dass Y eine Irrfahrt wird, und (X, Y) eine erfolgreiche Kopplung.

Sei $(U^{x,y,n} : x, y \in \mathbb{Z}^d, n \in \mathbb{N}_0)$ eine unabhängige Familie von $(\mathbb{Z}^d)^{N-1}$ -wertigen Zufallsvariablen $U^{x,y,n} = (U_1^{x,y,n}, \dots, U_{N-1}^{x,y,n})$, so dass

$$\mathbf{P}[(X_1, \dots, X_{N-1}) \in \cdot | X_0 = x, X_N = y] = \mathbf{P}_{U^{x,y,n}}$$

gilt für alle $x, y \in \mathbb{Z}^d$ mit $p^N(x, y) > 0$ und alle $n \in \mathbb{N}_0$. Wir nehmen nun an, dass die $U^{x,y,n}$ auch noch unabhängig von X und Y' sind. Wir definieren für $k \in \{nN + 1, \dots, (n + 1)N - 1\}$

$$Y_k := \begin{cases} U_{k-nN}^{Y'_n, Y'_{n+1}, n}, & \text{falls } n < \tau, \\ X_k, & \text{sonst.} \end{cases}$$

Man prüft leicht nach, dass Y tatsächlich eine Irrfahrt mit Übergangsmatrix p ist. Nach Konstruktion ist die Kopplung (X, Y) erfolgreich. \square

Satz 18.9. Sei X eine Markovkette auf E mit Übergangsmatrix p . Existiert eine erfolgreiche Kopplung, so ist jede beschränkte, harmonische Funktion konstant.

Beweis. Sei $f : E \rightarrow \mathbb{R}$ beschränkt und harmonisch, also $pf = f$. Seien $x, y \in E$, und sei (X, Y) eine erfolgreiche Kopplung. Nach Lemma 17.45 sind $(f(X_n))_{n \in \mathbb{N}_0}$ und $(f(Y_n))_{n \in \mathbb{N}_0}$ Martingale, also gilt

$$|f(x) - f(y)| = |\mathbf{E}_{(x,y)}[f(X_n) - f(Y_n)]| \leq 2\|f\|_\infty \mathbf{P}_{(x,y)}[X_n \neq Y_n] \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Korollar 18.10. Ist X eine irreduzible Irrfahrt auf \mathbb{Z}^d , so ist jede beschränkte, harmonische Funktion konstant.

Diese Aussage gilt allgemeiner, wenn wir \mathbb{Z}^d durch eine lokalkompakte, abelsche Gruppe ersetzen und geht in dieser Form auf Choquet und Deny [29] zurück, siehe auch [153].

Beweis. Ist p die Übergangsmatrix von X , so sei \tilde{X} eine Markovkette mit Übergangsmatrix $\bar{p}(x, y) = \frac{1}{2}p(x, y) + \frac{1}{2}\mathbb{1}_{\{x\}}(y)$. Offenbar haben X und \tilde{X} die selben harmonischen Funktionen. Nun ist aber \tilde{X} eine aperiodische, irreduzible Irrfahrt, besitzt also nach Satz 18.8 eine erfolgreiche Kopplung für alle Startpunkte. \square

Satz 18.11. Sei p die Übergangsmatrix einer irreduziblen, positiv rekurrenten, aperiodischen Kette auf E . Dann ist die unabhängig verschmelzende Kette eine erfolgreiche Kopplung.

Beweis. Seien \tilde{X} und \tilde{Y} zwei unabhängige Markovketten auf E mit Übergangsmatrix p . Dann hat die bivariate Markovkette $Z := ((\tilde{X}_n, \tilde{Y}_n))_{n \in \mathbb{N}_0}$ die Übergangsmatrix \tilde{p} , die durch

$$\tilde{p}((x_1, y_1), (x_2, y_2)) = p(x_1, x_2) \cdot p(y_1, y_2)$$

definiert wird. Wir zeigen zunächst, dass die Matrix \tilde{p} irreduzibel ist. Nur an dieser Stelle benötigen wir die Aperiodizität von p . Seien also $(x_1, y_1), (x_2, y_2) \in E \times E$ gegeben. Dann existiert nach Lemma 18.2 ein $m_0 \in \mathbb{N}$ mit

$$p^n(x_1, x_2) > 0 \quad \text{und} \quad p^n(y_1, y_2) > 0 \quad \text{für jedes } n \geq m_0.$$

Für $n \geq m_0$ ist daher $\tilde{p}^n((x_1, y_1), (x_2, y_2)) > 0$. Also ist \tilde{p} irreduzibel.

Wir definieren nun die Stoppzeit τ des ersten Eintreffens von (\tilde{X}, \tilde{Y}) in der Diagonalen $D := \{(x, x) : x \in E\}$ durch $\tau := \inf \{n \in \mathbb{N}_0 : \tilde{X}_n = \tilde{Y}_n\}$. Sei π die invariante Verteilung von \tilde{X} . Offenbar ist dann das Produktmaß $\pi \otimes \pi \in \mathcal{M}_1(E \times E)$ eine (und damit die) invariante Verteilung von (\tilde{X}, \tilde{Y}) . Nach Satz 17.51 ist daher (\tilde{X}, \tilde{Y}) positiv rekurrent, also insbesondere rekurrent. Mithin gilt $\mathbf{P}_{(x,y)}[\tau < \infty] = 1$ für alle Startpunkte $(x, y) \in E \times E$ von Z . \square

Satz 18.12. Sei X eine Markovkette mit Übergangsmatrix p , zu der eine erfolgreiche Kopplung existiert. Für alle $\mu, \nu \in \mathcal{M}_1(E)$ gilt dann $\|(\mu - \nu)p^n\|_{TV} \xrightarrow{n \rightarrow \infty} 0$.

Ist speziell X aperiodisch und positiv rekurrent mit invariante Verteilung π , so gilt $\|\mathcal{L}_\mu[X_n] - \pi\|_{TV} \xrightarrow{n \rightarrow \infty} 0$ für jedes $\mu \in \mathcal{M}_1(E)$.

Beweis. Es reicht, den Fall $\mu = \delta_x, \nu = \delta_y$ für gewisse $x, y \in E$ zu betrachten. Summation über x und y liefert dann den allgemeinen Fall. Sei $(X_n, Y_n)_{n \in \mathbb{N}_0}$ eine erfolgreiche Kopplung. Dann ist

$$\|(\delta_x - \delta_y)p^n\|_{TV} \leq 2 \mathbf{P}_{(x,y)}[X_n \neq Y_n] \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Wir fassen den Zusammenhang von Aperiodizität und Verteilungskonvergenz von X im folgenden Satz zusammen.

Satz 18.13 (Konvergenzsatz für Markovketten). Sei X eine irreduzible, positiv rekurrente Markovkette auf E mit invariante Verteilung π . Dann sind äquivalent:

(i) X ist aperiodisch.

(ii) Für jedes $x \in E$ gilt

$$\|\mathcal{L}_x[X_n] - \pi\|_{TV} \xrightarrow{n \rightarrow \infty} 0. \quad (18.7)$$

(iii) Für ein $x \in E$ gilt (18.7).

(iv) Für jedes $\mu \in \mathcal{M}_1(E)$ gilt $\|\mu p^n - \pi\|_{TV} \xrightarrow{n \rightarrow \infty} 0$.

Beweis. Die Implikationen (iv) \iff (ii) \implies (iii) sind klar. Die Implikation (i) \implies (ii) wurde in Satz 18.12 gezeigt. Wir zeigen also (iii) \implies (i).

„(iii) \implies (i)“ Wir nehmen an, dass (i) nicht gilt. Hat X die Periode $d \geq 2$, und ist $n \in \mathbb{N}$ kein Vielfaches von d , so ist nach Satz 17.51

$$\|\delta_x p^n - \pi\|_{TV} \geq |p^n(x, x) - \pi(\{x\})| = \pi(\{x\}) > 0.$$

Für jedes $x \in E$ gilt daher $\limsup_{n \rightarrow \infty} \|\delta_x p^n - \pi\|_{TV} > 0$, folglich gilt (iii) nicht. \square

Übung 18.2.1. Sei d_P die Prohorov-Metrik (siehe (13.3) und Übung 13.2.1). Man zeige: $d_P(P, Q) \leq \sqrt{d_W(P, Q)}$ für alle $P, Q \in \mathcal{M}_1(E)$. Hat E endlichen Durchmesser $\text{diam}(E)$, so ist $d_W(P, Q) \leq (\text{diam}(E) + 1)d_P(P, Q)$ für alle $P, Q \in \mathcal{M}_1(E)$. \clubsuit

Übung 18.2.2. Man zeige durch eine direkte Rechnung, dass der in Beispiel 18.6 aus \tilde{X} und \tilde{Y} hergestellte Prozess (X, Y) eine Kopplung mit Übergangsmatrix \bar{p} ist. \clubsuit

Übung 18.2.3. Sei X eine beliebige aperiodische, irreduzible, rekurrente Irrfahrt auf \mathbb{Z}^d . Man zeige, dass dann zu je zwei Startpunkten die unabhängige Verschmelzung eine erfolgreiche Kopplung ist.

Hinweis: Man zeige, dass die Differenz zweier rekurrenter Irrfahrten stets wieder rekurrent ist. ♣

Übung 18.2.4. Sei X eine Markovkette auf \mathbb{Z}^2 mit Übergangsmatrix

$$p((x_1, x_2), (y_1, y_2)) = \begin{cases} \frac{1}{4}, & \text{falls } x_1 = 0, \|y - x\|_2 = 1, \\ \frac{1}{4}, & \text{falls } x_1 \neq 0 \text{ und } y_1 = x_1 \pm 1, x_2 = y_2, \\ \frac{1}{2}, & \text{falls } x_1 \neq 0 \text{ und } y_1 = x_1, x_2 = y_2, \\ 0, & \text{sonst.} \end{cases}$$

Anschaulich ist dies die symmetrische einfache Irrfahrt, bei der alle senkrechten Übergänge außerhalb der senkrechten Koordinatenachse blockiert werden. Man zeige, dass X nullrekurrent, irreduzibel und aperiodisch ist, und dass die unabhängige Verschmelzung keine erfolgreiche Kopplung ist. ♣

18.3 Markovketten Monte Carlo Methode

Es sei E eine endliche Menge und $\pi \in \mathcal{M}_1(E)$ mit $\pi(x) := \pi(\{x\}) > 0$ für jedes $x \in E$. Wir betrachten das Problem, eine Zufallsvariable Y mit Verteilung π mit dem Computer zu generieren. Dies ist etwa dann relevant, wenn E eine sehr große Menge ist und Summen vom Typ $\sum_{x \in E} f(x)\pi(x)$ numerisch approximiert werden sollen durch Schätzer $n^{-1} \sum_{i=1}^n f(Y_i)$ (siehe Beispiel 5.21).

Wir nehmen an, dass unser Computer in der Lage ist, Realisierungen von u.i.v. Zufallsvariablen U_1, U_2, \dots zu generieren, die uniform auf $[0, 1]$ verteilt sind. Die Verteilung π soll jedoch nicht leicht direkt herstellbar sein.

Metropolis-Algorithmus

Wir haben schon gesehen, wie man Markovketten mit dem Computer simulieren kann (Beispiel 17.19). Die Idee ist nun, eine Markovkette X zu erzeugen, deren Verteilung gegen π konvergiert. Wenn wir X lange genug laufen lassen, so wird X_n ungefähr wie π verteilt sein. Gleichzeitig sollte die Kette so gestaltet sein, dass in jedem Schritt immer nur wenige Übergänge wirklich möglich sind, sodass das in Beispiel 17.19 beschriebene Verfahren auch effizient umsetzbar ist. (Natürlich wäre eine Kette mit Übergangsmatrix $p(x, y) = \pi(y)$ gegen π konvergent, aber das Problem ließe sich hiermit nicht vereinfachen.) Die so beschriebene Methode des Ziehens von π -verteilten Stichproben wird **Markovketten Monte Carlo Methode** oder **MCMC** (für *Markov chain Monte Carlo*) genannt (siehe [18, 122, 129]).

Sei q die Übergangsmatrix einer beliebigen irreduziblen Markovkette auf E (mit $q(x, y) = 0$ für möglichst viele $y \in E$). Wir erstellen hieraus die Metropolis-Matrix (siehe [76, 124]).

Definition 18.14. Wir definieren eine stochastische Matrix p auf E durch

$$p(x, y) = \begin{cases} q(x, y) \min\left(1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\right), & \text{falls } x \neq y, q(x, y) > 0, \\ 0, & \text{falls } x \neq y, q(x, y) = 0, \\ 1 - \sum_{z \neq x} p(x, z), & \text{falls } x = y. \end{cases}$$

p heißt **Metropolis-Matrix** zu q und π .

Man sieht direkt, dass p **reversibel** ist (siehe Abschnitt 19.2), das heißt

$$\pi(x)p(x, y) = \pi(y)p(y, x) \quad \text{für alle } x, y \in E. \quad (18.8)$$

Speziell ist π invariant (Nachrechnen!). Wir erhalten sofort den folgenden Satz.

Satz 18.15. Sei q irreduzibel, und für alle $x, y \in E$ gelte $q(x, y) > 0$ genau dann, wenn $q(y, x) > 0$ ist. Dann ist die Metropolis-Matrix p zu q und π irreduzibel mit eindeutiger Gleichgewichtsverteilung π . Ist zudem q aperiodisch oder nicht selber reversibel bezüglich π , so ist p aperiodisch.

Zur Simulation einer Kette X , die gegen π konvergiert, können wir nun, ausgehend von einer Referenzkette, die Übergänge nach q macht, den **Metropolis-Algorithmus** verwenden: Schlägt die Kette mit Übergangsmatrix q einen Übergang vom aktuellen Zustand x nach y vor, so akzeptieren wir diesen Vorschlag mit Wahrscheinlichkeit

$$\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \wedge 1.$$

Ansonsten bleiben wir in x stehen.

In der Definition von p taucht π nur in der Form des Quotienten $\pi(y)/\pi(x)$ auf. In vielen Fällen von Interesse ist dieser Quotient relativ leicht berechenbar, auch wenn $\pi(x)$ und $\pi(y)$ selber nicht leicht zu bestimmen sind. Wir wollen dies an einem Beispiel erläutern.

Beispiel 18.16 (Ising Modell). Das Ising Modell ist ein thermodynamisches (und quantenmechanisches) Modell für Ferromagnetismus in Kristallen, das von folgenden Annahmen ausgeht:

- Atome sitzen auf den Punkten des Gitters Λ (zum Beispiel $\Lambda = \{0, \dots, N-1\}^2$),
- jedes Atom $i \in \Lambda$ hat ein magnetisches Moment (Spin): $x(i) \in \{-1, 1\}$, das entweder nach oben zeigt ($x(i) = +1$) oder nach unten ($x(i) = -1$),
- benachbarte Atome wechselwirken miteinander,

- auf Grund thermischer Schwankungen ist der Zustand des Systems zufällig und verteilt nach der so genannten **Boltzmann-Verteilung** π auf dem Zustandsraum $E := \{-1, 1\}^A$, abhängig von der inversen Temperatur $\beta = \frac{1}{T} \geq 0$.

Wir definieren die lokale Energie H^i , die das Energieniveau eines Atoms in $i \in A$ als Funktion des Zustands x des Gesamtsystems angibt

$$H^i(x) = \frac{1}{2} \sum_{j \in A: i \sim j} \mathbb{1}_{\{x(i) \neq x(j)\}}.$$

Hierbei bedeutet $i \sim j$, dass i und j Nachbarn sind in A (damit meinen wir koordinatenweise mod N , wir sprechen auch von *periodischen Randbedingungen*). Die Gesamtenergie (oder Hamiltonfunktion) des Systems im Zustand x ist die Summe der Einzelenergien,

$$H(x) = \sum_{i \in A} H^i(x) = \sum_{i \sim j} \mathbb{1}_{\{x(i) \neq x(j)\}}.$$

Die Boltzmann-Verteilung π auf $E := \{-1, 1\}^A$ zur inversen Temperatur $\beta \geq 0$ wird definiert durch

$$\pi(x) = Z_\beta^{-1} \exp(-\beta H(x)),$$

wobei die **Zustandssumme** $Z_\beta = \sum_{x \in E} \exp(-\beta H(x))$ (oder **Partitionsfunktion**) die Normierungskonstante ist, die π zu einem W-Maß macht.

Makroskopisch beobachtbar ist nicht jeder einzelne Spin, sondern nur die mittlere Magnetisierung, die sich als Betrag des Mittelwerts der einzelnen Spins ergibt

$$m_A(\beta) = \sum_{x \in E} \pi(x) \left| \frac{1}{\#A} \sum_{i \in A} x(i) \right|.$$

Wenn wir sehr große Systeme betrachten, sind wir nahe am so genannten thermodynamischen Limes

$$m(\beta) := \lim_{A \uparrow \mathbb{Z}^d} m_A(\beta).$$

Man kann mit einem Konturargument, ähnlich wie bei der Perkolation, zeigen (siehe [136]), dass (für $d \geq 2$) eine Zahl $\beta_c = \beta_c(d) \in (0, \infty)$ existiert, mit

$$m(\beta) \begin{cases} > 0, & \text{falls } \beta > \beta_c, \\ = 0, & \text{falls } \beta < \beta_c. \end{cases} \quad (18.9)$$

An einem ähnlichen Modell, dem Weiss'schen Ferromagneten, werden wir in Beispiel 23.20 die Existenz eines solchen **Phasenübergangs** rigoros nachweisen. In der Physik wird $T_c := 1/\beta_c$ die **Curie-Temperatur** für die spontane Magnetisierung genannt. Dies ist eine materialabhängige Konstante (Chrombromid (CrBr) 37 Kelvin, Nickel 645K, Eisen 1017K, Kobalt 1404 K). Unterhalb der Curie-Temperatur

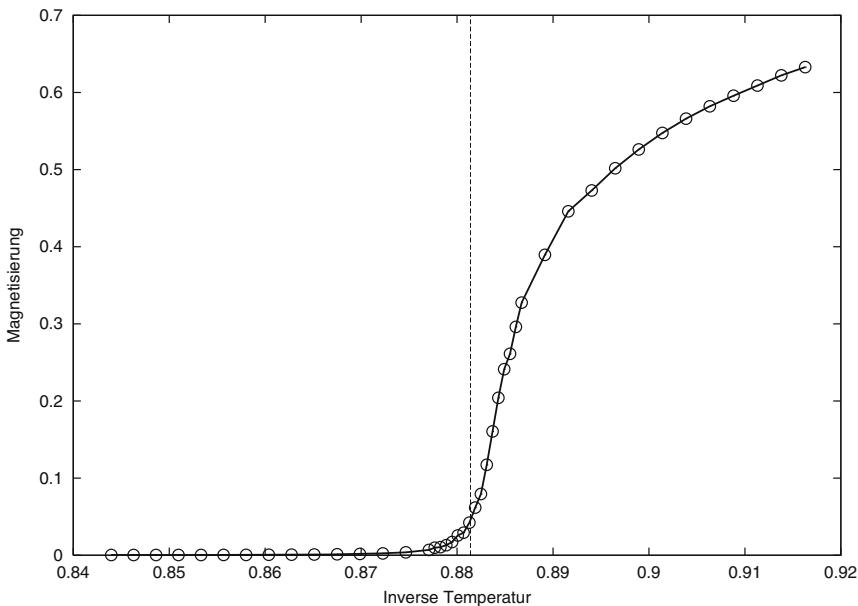


Abb. 18.4. Magnetisierungskurve im Ising-Modell auf einem 1000×1000 -Gitter, per Computersimulation berechnet. Die senkrechte Linie markiert die kritische Temperatur.

sind die Stoffe magnetisch, oberhalb sind sie es nicht. Dabei nimmt der Magnetisierungsgrad bei fallender Temperatur noch zu. Das Ising-Modell, das wir jetzt untersuchen, soll (zumindest in Computer-Simulationen) diesen Effekt einer kritischen Temperatur nachbilden.

Wir definieren den Zustand $x^{i,\sigma}$, bei dem an der Stelle i der Spin $\sigma \in \{-1, +1\}$ eingesetzt wird

$$x^{i,\sigma}(j) = \begin{cases} \sigma, & \text{falls } j = i, \\ x(j), & \text{falls } j \neq i. \end{cases}$$

Außerdem definieren wir den Zustand x^i , bei dem der Spin in i umgedreht wird $x^i := x^{i,-x(i)}$. Als vorschlagende Kette, oder Referenzkette, wählen wir nun eine Kette mit Übergangswahrscheinlichkeiten

$$q(x, y) = \begin{cases} \frac{1}{\#\Lambda}, & \text{falls } y = x^i \quad \text{für ein } i \in \Lambda, \\ 0, & \text{sonst.} \end{cases}$$

In Worten: Wir suchen einen Punkt $i \in \Lambda$ zufällig (uniform verteilt) aus und drehen den Spin an dieser Stelle um. Offenbar ist q irreduzibel.

Der Metropolis-Algorithmus zu dieser Kette akzeptiert den Vorschlag der Referenzkette sicher, falls $\pi(x^i) \geq \pi(x)$. Andernfalls wird der Vorschlag mit Wahrscheinlichkeit $\pi(x^i)/\pi(x)$ akzeptiert. Nun ist aber

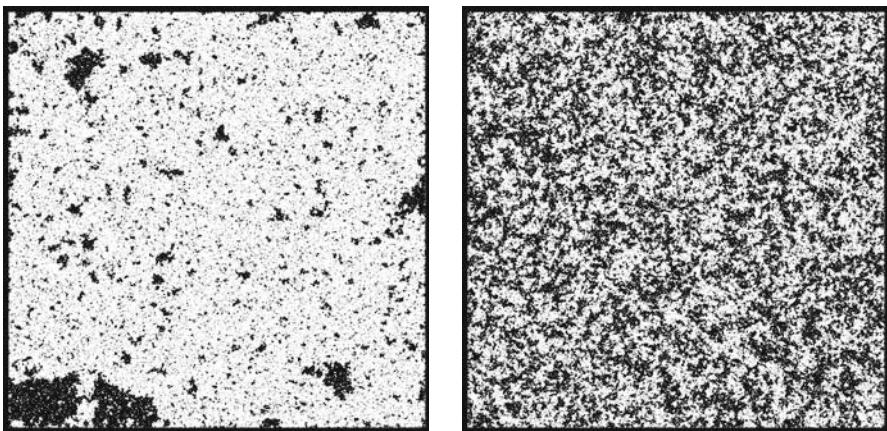


Abb. 18.5. Gleichgewichte des Ising-Modells für ein 800×800 Gitter. (schwarzer Punkt = spin +1) Links: kälter als die kritische Temperatur ($\beta > \beta_c$), rechts: wärmer.

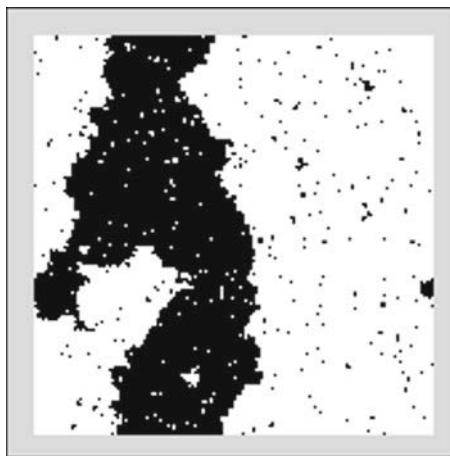


Abb. 18.6. Ising-Modell (150×150 Gitter) unterhalb der kritischen Temperatur. Die Computersimulation zeigt auch nach langer Laufzeit noch nicht das Gleichgewicht, sondern metastabile Zustände, in denen man die Weiss'schen Bezirke gut sehen kann.

$$\begin{aligned} H(x^i) - H(x) &= \sum_{j: j \sim i} \mathbb{1}_{\{x(j) \neq -x(i)\}} - \sum_{j: j \sim i} \mathbb{1}_{\{x(j) \neq x(i)\}} \\ &= -2 \sum_{j: j \sim i} \left(\mathbb{1}_{\{x(j) \neq x(i)\}} - \frac{1}{2} \right). \end{aligned}$$

Also ist $\pi(x^i)/\pi(x) = \exp(-2\beta \sum_{j \sim i} (\mathbb{1}_{\{x(j)=x(i)\}} - \frac{1}{2}))$, und dieser Ausdruck ist leicht zu berechnen, da er nur von den $2d$ Nachbarspins abhängt und zudem die

Kenntnis von Z_β nicht benötigt. Wir erhalten also als Metropolis-Übergangsmatrix

$$p(x, y) = \begin{cases} \frac{1}{\#\Lambda} \left(1 \wedge \exp \left[2\beta \sum_{j: j \sim i} (\mathbb{1}_{\{x(j) \neq x(i)\}} - \frac{1}{2}) \right] \right), & \text{falls } y = x^i \text{ für ein } i \in \Lambda, \\ 1 - \sum_{i \in \Lambda} p(x, x^i), & \text{falls } x = y, \\ 0, & \text{sonst.} \end{cases}$$

Praktisch wird man diese Kette simulieren, indem man sich unabhängige Zufallsvariablen I_1, I_2, \dots und U_1, U_2, \dots verschafft mit $I_n \sim \mathcal{U}_\Lambda$ und $U_n \sim \mathcal{U}_{[0,1]}$. Man setzt nun

$$F_n(x) = \begin{cases} x^{I_n}, & \text{falls } U_n \leq \exp \left[2\beta \sum_{j: j \sim i} (\mathbb{1}_{\{x(j) \neq x(i)\}} - \frac{1}{2}) \right], \\ x, & \text{sonst,} \end{cases}$$

und definiert die Markovkette $(X_n)_{n \in \mathbb{N}}$ durch $X_n = F_n(X_{n-1})$ für $n \in \mathbb{N}$. \diamond

Gibbs-Sampler

Wir betrachten eine Situation, in der, wie im obigen Beispiel, ein Zustand aus vielen Komponenten $x = (x_i)_{i \in \Lambda} \in E$ besteht, wobei Λ eine endliche Menge ist. Alternativ zur Metropolis-Kette betrachten wir ein weiteres Verfahren, um eine Markovkette mit gegebener invarianten Verteilung herzustellen. Beim so genannten **Gibbs-Sampler** oder *heat bath algorithm* ist die Idee, den Zustand lokal an die stationäre Verteilung anzupassen. Ist x der momentane Zustand, dann verfährt man wie folgt. Für $i \in \Lambda$ setze

$$x_{-i} := \{y \in E : y(j) = x(j) \text{ für } j \neq i\}.$$

Definition 18.17 (Gibbs-Sampler). Sei $q \in \mathcal{M}_1(\Lambda)$ mit $q(i) > 0$ für jedes $i \in \Lambda$. Die Übergangsmatrix p auf E mit

$$p(x, y) = \begin{cases} q_i \frac{\pi(x^{i,\sigma})}{\pi(x_{-i})}, & \text{falls } y = x^{i,\sigma} \text{ für ein } i \in \Lambda, \\ 0, & \text{sonst,} \end{cases}$$

heißt **Gibbs-Sampler** zur invarianten Verteilung π .

In Worten verfährt eine nach p konstruierte Kette in jedem Schritt wie folgt:

- (1) Wähle eine Komponente I gemäß einer Verteilung $(q_i)_{i \in \Lambda}$.
- (2) Ersetze in x durch $x^{I,\sigma}$ mit Wahrscheinlichkeit $\pi(x^{I,\sigma})/\pi(x_{-I})$.

Falls $I = i$ ist, dann hat der neue Zustand also die Verteilung $\mathcal{L}(X|X_{-i} = x_{-i})$, wobei X eine Zufallsvariable mit Verteilung π bezeichnet. Man beachte, dass man auch beim Gibbs-Sampler die Verteilung π nur bis auf die Normierungskonstante

zu kennen braucht (in einem etwas allgemeineren Rahmen lassen sich der Gibbs-Sampler und der Metropolis Algorithmus als Spezialfälle ein und desselben Verfahren auffassen). Für Zustände x und y , die sich nur in der i -ten Komponente unterscheiden, gilt (wegen $x_{-i} = y_{-i}$)

$$\pi(x) p(x, y) = \pi(x) q_i \frac{\pi(y)}{\pi(x_{-i})} = \pi(y) q_i \frac{\pi(x)}{\pi(y_{-i})} = \pi(y) p(y, x).$$

Der Gibbs-Sampler beschreibt also eine reversible Markovkette mit Gleichgewicht π . Die Irreduzibilität des Gibbs-Samplers ist von Fall zu Fall zu klären.

Beispiel 18.18 (Ising Modell). Im oben beschriebenen Ising-Modell ist $x_{-i} = \{x^{i,-1}, x^{i,+1}\}$. Daher ist für $i \in A$ und $\sigma \in \{-1, +1\}$

$$\begin{aligned} \pi(x^{i,\sigma} | x_{-i}) &= \frac{\pi(x^{i,\sigma})}{\pi(\{x^{i,-1}, x^{i,+1}\})} \\ &= \frac{e^{-\beta H(x^{i,\sigma})}}{e^{-\beta H(x^{i,-1})} + e^{-\beta H(x^{i,+1})}} \\ &= \left(1 + \exp \left[\beta(H(x^{i,\sigma}) - H(x^{i,-\sigma})) \right] \right)^{-1} \\ &= \left(1 + \exp \left[2\beta \sum_{j: j \sim i} (\mathbb{1}_{\{x(j) \neq \sigma\}} - \frac{1}{2}) \right] \right)^{-1}. \end{aligned}$$

Der Gibbs-Sampler des Ising-Modells ist also die Markovkette $(X_n)_{n \in \mathbb{N}_0}$ mit Werten in $E = \{-1, 1\}^A$ und mit Übergangsmatrix

$$p(x, y) = \begin{cases} \frac{1}{\#A} \left(1 + \exp \left[2\beta \sum_{j: j \sim i} (\mathbb{1}_{\{x(j) \neq y(j)\}} - \frac{1}{2}) \right] \right)^{-1}, & \text{falls } y = x^i \text{ für ein } i \in A, \\ 0, & \text{sonst.} \end{cases}$$

◇

Perfekte Simulation

Die bislang betrachtete MCMC Methode baut auf dem Prinzip Hoffnung: Wir lassen die Kette lange laufen und hoffen, dass sie sich in einem Zustand nahe dem Gleichgewicht befindet. Selbst wenn wir die Konvergenzgeschwindigkeit bestimmen können (und das ist oft nicht ganz leicht – wir kommen dazu in Abschnitt 18.4), werden wir doch nie einen Zustand bekommen, der *exakt* wie das Gleichgewicht verteilt ist.

Tatsächlich ist es, zumindest theoretisch, möglich, ein der MCMC Methode verwandtes Verfahren anzugeben, das *perfektes Ziehen von Stichproben* nach der Verteilung π ermöglicht, sogar, wenn wir über die Konvergenzgeschwindigkeit gar nichts wissen. Hierzu nehmen wir an, dass F_1, F_2, \dots u.i.v. zufällige Abbildungen $E \rightarrow E$ sind mit $\mathbf{P}[F(x) = y] = p(x, y)$ für alle $x, y \in E$. Wir hatten gesehen,

dass wir die Markovkette X mit Start in x durch $X_n = F_n \circ F_{n-1} \circ \dots \circ F_1(x)$ konstruieren können.

Nun gilt $F_1^n(x) := F_1 \circ \dots \circ F_n(x) \stackrel{\mathcal{D}}{=} F_n \circ \dots \circ F_1(x)$. Also gilt $\mathbf{P}[F_1^n(x) = y] \xrightarrow{n \rightarrow \infty} \pi(y)$ für jedes y . Ist nun aber F_1^n die konstante Abbildung, etwa $F_1^n \equiv x^*$ (für ein zufälliges x^*), so ist auch $F_1^m \equiv x^*$ für jedes $m \geq n$. Wenn man also durch geschickte Wahl der Verteilung der F_n erreichen kann, dass die Stoppzeit $T := \inf\{n \in \mathbb{N} : F_1^n \text{ ist konstant}\}$ fast sicher endlich ist (und das geht immer), so ist $\mathbf{P}[F_1^T(x) = y] = \pi(y)$ für alle $x, y \in E$. Ein einfacher Algorithmus für dieses Verfahren sieht so aus:

- (1) Setze $F \leftarrow \text{id}_E$ und $n \leftarrow 0$.
- (2) Setze $n \leftarrow n + 1$. Erzeuge F_n und setze $F \leftarrow F \circ F_n$.
- (3) Falls F nicht die konstante Abbildung ist, gehe zu (2).
- (4) Ausgabe $F(*)$.

Dieses Verfahren wird **Kopplung aus der Vergangenheit** (*coupling from the past*) genannt und geht auf Propp und Wilson [147] zurück (siehe auch [60, 61, 181, 146, 148, 99]). Interessante Simulationen sowie ein Forschungsüberblick finden sich im Internet unter <http://www.dbwilson.com/>. Einen guten Überblick über die MCMC Methode und die Kopplung aus der Vergangenheit gibt auch [72].

Praktisch ergeben sich zwei Probleme: Es muss die komplette Abbildung F_n erzeugt und mit F verknüpft werden. Die Rechenzeit dafür ist mindestens von der Ordnung der Größe des Raums E . Außerdem erfordert das Prüfen von F auf Konstanz einen Rechenaufwand von gleicher Größenordnung. Das Verfahren lässt sich effektiv nur durchführen, wenn man mehr Struktur zur Verfügung hat, etwa, wenn E eine Halbordnung mit einem kleinsten Element $\underline{0}$ und einem größten Element $\underline{1}$ besitzt (wie beim Ising-Modell) und man die Abbildungen F_n so wählen kann, dass sie fast sicher monoton wachsend sind. In diesem Fall braucht man immer nur $F(\underline{0})$ und $F(\underline{1})$ zu berechnen, und F ist konstant, falls $F(\underline{0}) = F(\underline{1})$.

18.4 Konvergenzgeschwindigkeit

Bei den bisherigen Betrachtungen ist die Frage nach der Geschwindigkeit der Konvergenz der Verteilung \mathbf{P}_{X_n} gegen π ignoriert worden. Für praktische Anwendungen ist aber dies genau die wichtigste Frage. Wir wollen hier nicht auf die Details eingehen, sondern das Thema nur kurz anreissen. Ohne Einschränkung sei $E = \{1, \dots, N\}$. Ist p reversibel (Gleichung (18.8)), so wird durch $f \mapsto pf$ ein symmetrischer linearer Operator auf $L^2(E, \pi)$ definiert (Übung!). Alle Eigenwerte $\lambda_1, \dots, \lambda_N$ (mit Mehrfachnennung je nach Vielfachheit) sind reell und dem Betrage nach nicht größer als 1, da p stochastisch ist. Wir können also die Eigenwerte dem

Betrage nach ordnen: $\lambda_1 = 1 \geq |\lambda_2| \geq \dots \geq |\lambda_N|$. Ist p irreduzibel und aperiodisch, so ist $|\lambda_2| < 1$. Sei $\mu_1 = \pi, \mu_2, \dots, \mu_N$ eine Orthonormalbasis aus Links-Eigenvektoren zu den Eigenwerten $\lambda_1, \dots, \lambda_N$. Für jedes $\mu = \alpha_1 \mu_1 + \dots + \alpha_N \mu_N$ ist dann $\mu p^n = \sum_{i=1}^N \lambda_i^n \alpha_i \mu_i$, also

$$\|\mu p^n - \pi\|_{TV} \leq C |\lambda_2|^n \quad (18.10)$$

für eine Konstante C (die nicht einmal von μ abhängt). Eine ähnliche Formel gilt für den Fall, wo p nicht reversibel ist, wobei Korrekturterme der Ordnung maximal n^{V-1} auftreten. Dabei ist V die Größe des größten Jordan-Kästchens zum Eigenwert λ_2 in der Jordan'schen Normalform von p , speziell also höchstens die Vielfachheit des betragsmäßig zweitgrößten Eigenwertes.

Die Konvergenzgeschwindigkeit ist also exponentiell mit einer Rate, die durch die **Spektrallücke** $1 - |\lambda_2|$ zum zweitgrößten Eigenwert von p bestimmt ist. Die analytische Bestimmung der Spektrallücke ist für große Räume E häufig extrem schwer.

Beispiel 18.19. Sei $r \in (0, 1)$ und $N \in \mathbb{N}$, $N \geq 2$, sowie $E = \{0, \dots, N-1\}$. Wir betrachten die Übergangsmatrix

$$p(i, j) = \begin{cases} r, & \text{falls } j = i+1 \pmod{N}, \\ 1-r, & \text{falls } j = i-1 \pmod{N}, \\ 0, & \text{sonst.} \end{cases}$$

p ist die Übergangsmatrix der einfachen (asymmetrischen) Irrfahrt auf dem diskreten Torus $\mathbb{Z}/(N)$, die mit Wahrscheinlichkeit r einen Schritt nach rechts springt, mit Wahrscheinlichkeit $1-r$ hingegen einen Schritt nach links springt. Offenbar ist p irreduzibel, und p ist genau dann aperiodisch, wenn N ungerade ist. Offensichtlich ist die Gleichverteilung \mathcal{U}_E die eindeutige invariante Verteilung.

Fall 1: N ungerade. Man prüft leicht nach, dass p die Eigenwerte

$$\lambda_k := r \theta_k + (1-r) \bar{\theta}_k = \cos\left(\frac{2\pi k}{N}\right) + (2r-1)i \sin\left(\frac{2\pi k}{N}\right), \quad k = 0, \dots, N-1,$$

hat, wobei $\theta_k = e^{2\pi i k/N}$, $k = 0, \dots, N-1$, die N -ten Einheitswurzeln sind, und die zugehörigen (Rechts-) Eigenvektoren

$$x^k := (\theta_k^0, \theta_k^1, \dots, \theta_k^{N-1}).$$

Die Beträge der Eigenwerte bekommen wir durch $|\lambda_k| = f(2\pi k/N)$, wobei

$$f(\vartheta) = \sqrt{1 - 4r(1-r)\sin(\vartheta)^2} \quad \text{für } \vartheta \in \mathbb{R}.$$

Da N ungerade ist, ist $|\lambda_k|$ maximal (außer für $k=0$) für $k = \frac{N-1}{2}$ und $k = \frac{N+1}{2}$ mit dem Wert $\gamma := \sqrt{1 - 4r(1-r)\sin(\pi/N)^2}$. Da die Eigenwerte alle unterschiedlich sind, hat jeder Eigenwert die Vielfachheit 1, und es gibt ein $C < \infty$ mit

$$\|\mu p^n - \mathcal{U}_E\|_{TV} \leq C \gamma^n \quad \text{für alle } n \in \mathbb{N}, \mu \in \mathcal{M}_1(E).$$

Fall 2: N gerade. In diesem Fall ist p nicht aperiodisch, nichtsdestoweniger haben die Eigenwerte und Eigenvektoren die selbe Gestalt wie im ersten Fall. Um eine aperiodische Kette zu erhalten, bilden wir für $\varepsilon > 0$ die Übergangsmatrix

$$p_\varepsilon := (1 - \varepsilon)p + \varepsilon I,$$

wo I die Einheitsmatrix auf E ist. p_ε beschreibt die Irrfahrt auf E , die mit Wahrscheinlichkeit ε am Ort stehen bleibt und mit Wahrscheinlichkeit $1 - \varepsilon$ einen Sprung gemäß p macht. Offenbar ist p_ε irreduzibel und aperiodisch. Die Eigenwerte sind

$$\lambda_{\varepsilon,k} = (1 - \varepsilon)\lambda_k + \varepsilon, \quad k = 0, \dots, N - 1,$$

mit zugehörigen Eigenvektoren x^k wie oben. Offenbar ist $\lambda_{\varepsilon,0} = 1$, und $\lambda_{\varepsilon,N/2} = 2\varepsilon - 1$ ist der betragsmäßig zweitgrößte Eigenwert, falls $\varepsilon > 0$ sehr klein ist. Für größere ε ist $|\lambda_{\varepsilon,1}| > |\lambda_{\varepsilon,N/2}|$. Genauer gilt: Setzen wir

$$\varepsilon_0 := \frac{(1 - (2r - 1)^2) \sin(2\pi/N)^2}{(1 - (2r - 1)^2) \sin(2\pi/N)^2 + 2 \cos(2\pi/N)},$$

so ist der Betrag γ_ε des betragsmäßig zweitgrößten Eigenwertes

$$\gamma_\varepsilon = |\lambda_{\varepsilon,N/2}| = 1 - 2\varepsilon, \quad \text{falls } \varepsilon \leq \varepsilon_0,$$

und

$$\begin{aligned} \gamma_\varepsilon &= |\lambda_{\varepsilon,1}| \\ &= \sqrt{\left((1 - \varepsilon) \cos\left(\frac{2\pi}{N}\right) + \varepsilon\right)^2 + \left((1 - \varepsilon)(2r - 1) \sin\left(\frac{2\pi}{N}\right)\right)^2}, \quad \text{falls } \varepsilon \geq \varepsilon_0. \end{aligned}$$

Es ist nicht schwer zu zeigen, dass $\varepsilon \mapsto |\lambda_{\varepsilon,N/2}|$ monoton fallend ist und $\varepsilon \mapsto |\lambda_{\varepsilon,1}|$ monoton wachsend. Daher ist γ_ε minimal für $\varepsilon = \varepsilon_0$.

Es gibt also ein $C < \infty$ mit

$$\|\mu p_\varepsilon^n - \mathcal{U}_E\|_{TV} \leq C \gamma_\varepsilon^n \quad \text{für alle } n \in \mathbb{N}, \mu \in \mathcal{M}_1(E),$$

und die beste Konvergenzgeschwindigkeit (in dieser Klasse von Übergangsmatrizen) wird erreicht durch die Wahl $\varepsilon = \varepsilon_0$. \diamond

Beispiel 18.20 (Gambler's Ruin). Wir betrachten das Gambler's Ruin Problem aus Beispiel 10.19 mit Erfolgswahrscheinlichkeit $r \in (0, 1)$. Hier ist der Zustandsraum $E = \{0, \dots, N\}$, und die Übergangsmatrix hat die Gestalt

$$p(i,j) = \begin{cases} r, & \text{falls } j = i + 1 \in \{2, \dots, N\}, \\ 1 - r, & \text{falls } j = i - 1 \in \{0, \dots, N - 2\}, \\ 1, & \text{falls } j = i \in \{0, N\}, \\ 0, & \text{sonst.} \end{cases}$$

Diese Übergangsmatrix ist nicht irreduzibel, sondern hat die zwei absorbierenden Zustände 0 und N . In Beispiel 10.19 (Gleichung (10.5)), für den Fall $r \neq \frac{1}{2}$, und Beispiel 10.16, für den Fall $r = \frac{1}{2}$, wurde gezeigt, dass für jedes $\mu \in \mathcal{M}_1(E)$

$$\mu p^n \xrightarrow{n \rightarrow \infty} (1 - m(\mu))\delta_0 + m(\mu)\delta_N, \quad (18.11)$$

wobei $m(\mu) = \int p_N(x) \mu(dx)$ ist, und die Wahrscheinlichkeit $p_N(x)$, dass die in x gestartete Kette N trifft, gegeben ist durch

$$p_N(x) = \begin{cases} \frac{1 - (\frac{1-r}{r})^x}{1 - (\frac{1-r}{r})^N}, & \text{falls } r \neq \frac{1}{2}, \\ \frac{x}{N}, & \text{falls } r = \frac{1}{2}. \end{cases}$$

Wie schnell geht nun die Konvergenz in (18.11)? Auch hier ist die Konvergenz exponentiell schnell, und die Rate wird wieder durch den zweitgrößten Eigenwert von p bestimmt.

Wir wollen nun also das Spektrum von p bestimmen. Klar sind $x^0 = (1, 0, \dots, 0)$ und $x^N = (0, \dots, 0, 1)$ Links-Eigenvektoren zum Eigenwert 1. Damit nun $x = (x_0, \dots, x_N)$ ein Links-Eigenvektor zum Eigenwert λ ist, müssen die folgenden Gleichungen erfüllt sein:

$$\lambda x_k = rx_{k-1} + (1-r)x_{k+1} \quad \text{für } k = 2, \dots, N-2, \quad (18.12)$$

und

$$\lambda x_{N-1} = rx_{N-2}. \quad (18.13)$$

Gelten (18.12) und (18.13) für x_1, \dots, x_{N-1} , so setzen wir $x_0 := \frac{1-p}{\lambda-1}x_1$ und $x_N := \frac{p}{\lambda-1}x_{N-1}$ und erhalten dadurch tatsächlich $xp = \lambda x$. Wir machen den Ansatz

$$\lambda = (1-r)\rho(\theta + \bar{\theta}) \quad \text{und} \quad x_k = \varrho^k(\theta^k - \bar{\theta}^k) \quad \text{für } k = 1, \dots, N-1,$$

wobei

$$\rho = \sqrt{r/(1-r)} \quad \text{und} \quad \theta \in \mathbb{C} \setminus \{-1, +1\} \quad \text{mit} \quad |\theta| = 1.$$

Es gilt also $\theta\bar{\theta} = 1$ und $(1-r)\rho^{k+1} = r\rho^{k-1}$. Daher ist für jedes $k = 2, \dots, N-1$

$$\begin{aligned} \lambda x_k &= (1-r)\rho^{k+1}(\theta^k - \bar{\theta}^k)(\theta + \bar{\theta}) \\ &= (1-r)\rho^{k+1}[(\theta^{k+1} - \bar{\theta}^{k+1}) + \theta\bar{\theta}(\theta^{k-1} - \bar{\theta}^{k-1})] \\ &= r\rho^{k-1}(\theta^{k-1} - \bar{\theta}^{k-1}) + (1-r)\rho^{k+1}(\theta^{k+1} - \bar{\theta}^{k+1}) \\ &= r x_{k-1} + (1-r)x_{k+1}, \end{aligned}$$

das heißt, es gilt (18.12). Die selbe Rechnung mit $k = N-1$ zeigt, dass (18.13) genau dann gilt, wenn $\theta^N - \bar{\theta}^N = 0$ ist, also wenn $\theta^{2N} = 1$ gilt. Wir erhalten also für θ die $N-1$ unterschiedlichen Werte (man beachte, dass die komplex konjugierten der hier angegebenen Werte zu den selben λ_n führen)

$$\theta_n = e^{(n/N)\pi i} \quad \text{für } n = 1, \dots, N-1.$$

Die zugehörigen Eigenwerte sind

$$\lambda_n = \sigma \cos\left(\frac{n\pi}{N}\right) \quad \text{für } n = 1, \dots, N-1.$$

Dabei ist die Varianz des einzelnen Irrfahrt-Schrittes:

$$\sigma^2 := 4r(1-r). \quad (18.14)$$

Da alle Eigenwerte reell sind, sind die zugehörigen Eigenvektoren gegeben durch

$$x_k^n = 2 \left(\frac{r}{1-r} \right)^{n/2} \sin\left(\frac{n\pi}{N}\right), \quad k = 1, \dots, N-1.$$

Für $n = 1$ und $n = N-1$ ist $|\lambda_n| = \sigma \cos\left(\frac{\pi}{N}\right)$ der betragsmäßig zweitgrößte Eigenwert. Es folgt, dass es ein $C > 0$ gibt, sodass für jedes $\mu \in \mathcal{M}_1(E)$ gilt

$$\mu p^n(\{1, \dots, N-1\}) \leq C \left(\sigma \cos\left(\frac{\pi}{N}\right) \right)^n \quad \text{für jedes } n \in \mathbb{N}.$$

Mit anderen Worten: Die Wahrscheinlichkeit, dass das Spiel bis zur n -ten Runde noch nicht entschieden ist, ist maximal $C (\sigma \cos(\pi/N))^n$.

Ein alternativer Zugang zu den Eigenwerten geht über die Nullstellen des charakteristischen Polynoms

$$\chi_N(x) = \det(p - xI), \quad x \in \mathbb{R}.$$

Man sieht sofort, dass $\chi_1(x) = (1-x)^2$ und $\chi_2(x) = -x(1-x)^2$ gilt. Über die Entwicklungsformel der Determinante durch Streichen von Zeilen und Spalten erhalten wir die Rekursionsformel

$$\chi_N(x) = -x \chi_{N-1}(x) - r(1-r) \chi_{N-2}(x). \quad (18.15)$$

Wir erhalten als Lösung (Nachrechnen!)

$$\chi_N(x) = (-1)^{N-1} (\sigma/2)^{N-1} (1-x)^2 U_{N-1}(x/\sigma), \quad (18.16)$$

wobei

$$U_m(x) := \sum_{k=0}^{\lfloor m/2 \rfloor} (-1)^k \binom{m-k}{k} (2x)^{m-2k}$$

das m -te **Chebyshev Polynom zweiter Art** bezeichnet.

Für $x \in (-\sigma, \sigma)$ kann man mit Hilfe der de Moivre'schen Formel zeigen, dass

$$\begin{aligned} \chi_N(x) &= (-1)^{N-1} (\sigma/2)^{N-1} (1-x)^2 \frac{\sin(N \arccos(x/\sigma))}{\sqrt{1-(x/\sigma)^2}} \\ &= (1-x)^2 \prod_{k=1}^{N-1} \left(\sigma \cos\left(\frac{\pi k}{N}\right) - x \right). \end{aligned} \quad (18.17)$$

Neben der doppelten Nullstelle 1 erhalten wir als Nullstellen

$$\sigma \cos(\pi k/N), \quad k = 1, \dots, N-1.$$

◇

Übung 18.4.1. Man zeige (18.16). ♣

Übung 18.4.2. Man zeige (18.17). ♣

Übung 18.4.3. Sei $\nu(dx) = \frac{2}{\pi} \sqrt{1-x^2} \mathbb{1}_{[-1,1]}(x) dx$. Man zeige, dass die Chebyshev Polynome zweiter Art bezüglich ν orthonormal sind:

$$\int U_m U_n d\nu = \mathbb{1}_{\{m=n\}}.$$

♣

Übung 18.4.4. Sei $E = \{1, 2, 3\}$ und $p = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/3 & 1/3 & 1/3 \\ 0 & 3/4 & 1/4 \end{pmatrix}$. Man bestimme die

invariante Verteilung und die exponentielle Konvergenzrate. ♣

Übung 18.4.5. Sei $E = \{0, \dots, N-1\}$, $r \in (0, 1)$ und

$$p(i, j) = \begin{cases} r, & \text{falls } j = i + 1 \pmod{N}, \\ 1-r, & \text{falls } j = i \pmod{N}, \\ 0, & \text{sonst.} \end{cases}$$

Man zeige, dass p die Übergangsmatrix einer irreduziblen, aperiodischen Irrfahrt ist, bestimme die invariante Verteilung und bestimme die exponentielle Konvergenzgeschwindigkeit. ♣

Übung 18.4.6. Sei $N \in \mathbb{N}$ und $E = \{0, 1\}^N$ der N -dimensionale Hyperkubus, das heißt, zwei Punkte $x, y \in E$ sind genau dann durch eine Kante verbunden, wenn sie sich in genau einer Koordinate unterscheiden. Sei p die Übergangsmatrix der Irrfahrt auf E , die mit Wahrscheinlichkeit $\varepsilon > 0$ am Ort bleibt, mit Wahrscheinlichkeit $1-\varepsilon$ hingegen zu einem (uniform gewählten) zufälligen Nachbarpunkt springt.

Man beschreibe p formal, zeige dass p aperiodisch und irreduzibel ist, und bestimme die invariante Verteilung sowie die exponentielle Konvergenzgeschwindigkeit. ♣

19 Markovketten und elektrische Netzwerke

Wir betrachten eine symmetrische einfache Irrfahrt auf \mathbb{Z}^2 . Nach dem Satz von Pólya (Satz 17.39) ist diese Irrfahrt rekurrent. Was passiert aber, wenn wir eine einzelne Kante aus dem Gitter \mathbb{L}^2 von \mathbb{Z}^2 entfernen? Intuitiv sollte dies nichts an der Rekurrenz ändern. Die in Kapitel 17.5 verwendeten Rechnungen sind allerdings in dieser Hinsicht nicht sehr robust und können hier nicht mehr zum Beweis der Rekurrenz benutzt werden. Noch unübersichtlicher wird die Situation, wenn wir die Irrfahrt auf die obere Halbebene $\{(x, y) : x \in \mathbb{Z}, y \in \mathbb{N}_0\}$ von \mathbb{Z}^2 beschränken. Wie sieht es hier mit der Rekurrenz aus? Oder wir betrachten die Situation von **Kantenperkolation** auf \mathbb{Z}^2 . Wir fixieren einen Parameter $p \in [0, 1]$ und definieren jede Kante von \mathbb{L}^2 mit Wahrscheinlichkeit p als *offen* und mit Wahrscheinlichkeit $1 - p$ als *geschlossen*. Nachdem dies im ersten Schritt geschehen ist, wird die Irrfahrt auf dem zufälligen Teilgraphen der offenen Kanten betrachtet. Der Irrfahrer wählt in jedem Schritt mit gleicher Wahrscheinlichkeit eine der benachbarten *offenen* Kanten aus. Für $p > \frac{1}{2}$ existiert genau eine unendlich große Zusammenhangskomponente offener Kanten (Satz 2.47). Ist die Irrfahrt auf dieser (zufälligen) Komponente rekurrent oder transient?

Ziel dieses Kapitels ist es, einen Zusammenhang zwischen gewissen Markovketten und elektrischen Netzwerken herzustellen, der

- es in manchen Fällen erlaubt, zwischen Rekurrenz und Transienz anhand von leicht berechenbaren Größen zu entscheiden,
- in anderen Fällen ein Vergleichskriterium bietet, das besagt, dass eine Irrfahrt auf einem Teilgraphen rekurrent ist, wenn die Irrfahrt auf dem ursprünglichen Graphen rekurrent ist. Damit lässt sich für alle oben betrachteten Irrfahrten Rekurrenz nachweisen.

Dieses Kapitel lehnt sich an [121] und [41] an.

19.1 Harmonische Funktionen

Sei in diesem Kapitel stets E eine abzählbare Menge und X eine diskrete Markovkette auf E mit Übergangsmatrix p und Greenfunktion G . Schließlich sei $F(x, y)$

die Wahrscheinlichkeit, von x aus jemals nach y zu gelangen. Vergleiche Abschnitt 17.4, insbesondere Definitionen 17.28 und 17.33.

Definition 19.1. Sei $A \subset E$. Eine Funktion $f : E \rightarrow \mathbb{R}$ heißt **harmonisch auf $E \setminus A$** , falls $pf(x) = \sum_{y \in E} p(x, y)f(y)$ existiert und $pf(x) = f(x)$ für jedes $x \in E \setminus A$ gilt.

Satz 19.2 (Superpositionsprinzip). Sind f und g harmonisch auf $E \setminus A$ und $\alpha, \beta \in \mathbb{R}$, so ist auch $\alpha f + \beta g$ harmonisch auf $E \setminus A$.

Beweis. Trivial. □

Beispiel 19.3. Sei X transient und $a \in E$ ein transienter Zustand (also ein nicht absorbierender). Dann ist $f(x) := G(x, a)$ harmonisch auf $E \setminus \{a\}$: Für $x \neq a$ ist

$$pf(x) = p \sum_{n=0}^{\infty} p^n(x, a) = \sum_{n=1}^{\infty} p^n(x, a) = G(x, a) - \mathbb{1}_{\{a\}}(x) = G(x, a). \quad \diamond$$

Beispiel 19.4. Für jedes $x \in E$ sei $\tau_x := \inf\{n > 0 : X_n = x\}$. Für $A \subset E$ sei

$$\tau := \tau_A := \inf_{x \in A} \tau_x$$

die Zeit des ersten Eintritts in A . Wir nehmen an, dass A so gewählt ist, dass $\mathbf{P}_x[\tau_A < \infty] = 1$ für jedes $x \in E$. Sei $g : A \rightarrow \mathbb{R}$ eine beschränkte Funktion. Wir definieren

$$f(x) := \begin{cases} g(x), & \text{falls } x \in A, \\ \mathbf{E}_x[g(X_\tau)], & \text{falls } x \in E \setminus A. \end{cases} \quad (19.1)$$

Dann ist f harmonisch in $E \setminus A$. Wir geben hierfür zwei Beweise an.

1. Beweis. Nach der Markoveigenschaft ist für $x \notin A$ und $y \in E$

$$\mathbf{E}_x[g(X_\tau) \mid X_1 = y] = \begin{cases} g(y), & \text{falls } y \in A \\ \mathbf{E}_y[g(X_\tau)], & \text{falls } y \in E \setminus A \end{cases} = f(y).$$

Also ist für $x \in E \setminus A$

$$\begin{aligned} f(x) &= \mathbf{E}_x[g(X_\tau)] = \sum_{y \in E} \mathbf{E}_x[g(X_\tau); X_1 = y] \\ &= \sum_{y \in E} p(x, y) \mathbf{E}_x[g(X_\tau) \mid X_1 = y] = \sum_{y \in E} p(x, y) f(y) = pf(x). \end{aligned}$$

2. Beweis. Wir verändern die Markovkette, indem wir einen Zustand Δ als Falle hinzufügen. Es gelte also $\tilde{E} = E \cup \{\Delta\}$ und

$$\tilde{p}(x, y) = \begin{cases} p(x, y), & \text{falls } x \in E \setminus A, y \neq \Delta, \\ 0, & \text{falls } x \in E \setminus A, y = \Delta, \\ 1, & \text{falls } x \in A \cup \{\Delta\}, y = \Delta. \end{cases} \quad (19.2)$$

Die so erzeugte Markovkette \tilde{X} ist transient mit Δ als einzigm absorbierenden Zustand. Weiterhin ist genau dann $p f = f$ auf $E \setminus A$, wenn $\tilde{p} f = f$ auf $E \setminus A$ ist. Wegen $\tilde{G}(y, y) = 1$ für $y \in A$ ist (vergleiche Satz 17.34)

$$\mathbf{P}_x[X_\tau = y] = \mathbf{P}_x[\tilde{\tau}_y < \infty] = \tilde{F}(x, y) = \tilde{G}(x, y) \quad \text{für alle } x \in E \setminus A, y \in A.$$

Nun ist $x \mapsto \tilde{G}(x, y)$ harmonisch auf $E \setminus A$. Nach dem Superpositionsprinzip ist auch

$$f(x) = \sum_{y \in A} \tilde{G}(x, y) g(y) \quad (19.3)$$

harmonisch auf $E \setminus A$. Wegen dieser Darstellung heißtt, in Analogie zur kontinuierlichen Potentialtheorie, \tilde{G} die **Greenfunktion** für die Gleichung $(p - I)f = 0$ auf $E \setminus A$. \diamond

Definition 19.5. Wir nennen das *Gleichungssystem*

$$\begin{aligned} (p - I)f(x) &= 0, & \text{für } x \in E \setminus A, \\ f(x) &= g(x), & \text{für } x \in A, \end{aligned} \quad (19.4)$$

das zu $p - I$ gehörige **Dirichlet-Problem** auf $E \setminus A$ mit Randwerten g auf A .

Die Existenz einer Lösung des Dirichlet-Problems haben wir bereits in Beispiel 19.4 gezeigt. Um die Eindeutigkeit (unter gewissen Bedingungen) zu zeigen, leiten wir zunächst das Maximumsprinzip für harmonische Funktionen her.

Ist $p = I$, so kann die Lösung des Dirichlet-Problems offenbar nicht eindeutig sein. Selbst in weniger extremen Situationen kann die Lösung mehrdeutig sein, wenn $E \setminus A$ in Bereiche zerfällt, zwischen denen die in A gestoppte Markovkette nicht hindurchlaufen kann. Um eine entsprechende Irreduzibilitätsbedingung formal auszudrücken, definieren wir die Übergangsmatrix p_A der in A gestoppten Kette durch

$$p_A(x, y) := \begin{cases} p(x, y), & \text{falls } x \notin A, \\ \mathbb{1}_{\{x=y\}}, & \text{falls } x \in A. \end{cases}$$

Ferner sei F_A für p_A analog wie F für p definiert. Für $x \in E$ sei

$$S_A^n(x) = \{y \in E : (p_A)^n(x, y) > 0\}, \quad \text{für } n \in \mathbb{N}_0$$

und

$$S_A(x) = \bigcup_{n=0}^{\infty} S_A^n(x) = \{y \in E : F_A(x, y) > 0\}.$$

Satz 19.6 (Maximumprinzip). Sei f eine harmonische Funktion auf $E \setminus A$.

(i) Gibt es ein $x_0 \in E \setminus A$ mit

$$f(x_0) = \sup f(S_A(x_0)), \quad (19.5)$$

so ist $f(y) = f(x_0)$ für alle $y \in S_A(x_0)$.

(ii) Gilt speziell $F_A(x, y) > 0$ für alle $x, y \in E \setminus A$, und gibt es ein $x_0 \in E \setminus A$ mit $f(x_0) = \sup f(E)$, so ist $f(x_0) = f(y)$ für alle $y \in E \setminus A$.

Beweis. (i) Sei $m := \sup f(S_A(x_0))$. Da f harmonisch auf $E \setminus A$ ist, ist $p_A f = f$ auf ganz E . Daher gilt für jedes $n \in \mathbb{N}$

$$f(x_0) = (p_A)^n f(x_0) = \sum_{y \in S_A^n(x_0)} p_A^n(x_0, y) f(y) \leq m$$

mit Gleichheit genau dann, wenn $f(y) = m$ für alle $y \in S_A^n(x_0)$. Da nach Voraussetzung (19.5) aber Gleichheit gilt, folgt $f(x_0) = f(y)$ für alle $y \in S_A(x_0)$.

(ii) Dies folgt direkt aus (i), denn $S_A(x) \supset E \setminus A$ für jedes $x \in E \setminus A$. \square

Satz 19.7 (Eindeutigkeitssatz für harmonische Funktionen). Sei $F(x, y) > 0$ für alle $x, y \in E$. Sei $A \subset E$, $A \neq \emptyset$, sodass $E \setminus A$ endlich ist. Sind f_1 und f_2 harmonisch auf $E \setminus A$ und ist $f_1 = f_2$ auf A , dann ist $f_1 = f_2$.

Mit anderen Worten: Das Dirichlet-Problem (19.4) besitzt eine eindeutige Lösung, die durch (19.3) (oder äquivalent (19.1)) gegeben ist.

Beweis. Nach dem Superpositionsprinzip ist $f := f_1 - f_2$ harmonisch auf $E \setminus A$, und es gilt $f|_A \equiv 0$.

Wir zeigen, dass $f \leq 0$ gilt. Aus Symmetriegründen ist dann auch $f \geq 0$, also $f \equiv 0$. Zu diesem Zweck nehmen wir an, dass es ein $x \in E$ gibt mit $f(x) > 0$ und führen diese Annahme zum Widerspruch.

Da $f|_A \equiv 0$ gilt und $E \setminus A$ endlich ist, gibt es ein $x_0 \in E \setminus A$ mit $f(x_0) = \max f(E) \geq f(x) > 0$.

Wegen $F(x, y) > 0$ für alle $x, y \in E$ ist

$$n_0 := \min \{n \in \mathbb{N}_0 : p^n(x_0, y) > 0 \text{ für ein } y \in A\} < \infty.$$

Offenbar ist $p^{n_0}(x_0, y) = (p_A)^{n_0}(x_0, y)$ für alle $y \in A$, also gibt es ein $y \in A$ mit $(p_A)^{n_0}(x_0, y) > 0$, das heißt $y \in S_A(x_0)$. Nach Satz 19.6 folgt $f(x_0) = f(y) = 0$ im Widerspruch zur Annahme. \square

Übung 19.1.1. Sei \bar{p} die substochastische $E \times E$ Matrix, die durch $\bar{p}(x, y) = \tilde{p}(x, y)$, $x, y \in E$, (mit \tilde{p} aus (19.2)) definiert wird, also $\bar{p}(x, y) = p(x, y) \mathbb{1}_{x \in E \setminus A}$, und sei I die Einheitsmatrix auf E . Man zeige:

- (i) $I - \bar{p}$ ist invertierbar.
- (ii) Setzen wir $\bar{G} := (I - \bar{p})^{-1}$, so ist $\bar{G}(x, y) = \tilde{G}(x, y)$ für alle $x, y \in E \setminus A$ und $\bar{G}(x, y) = \mathbb{1}_{\{x=y\}}$, falls $x \in A$. Speziell ist

$$\bar{G}(x, y) = \mathbf{P}_x[X_{\tau_A} = y] \quad \text{für } x \in E \setminus A \text{ und } y \in A. \quad \clubsuit$$

19.2 Reversible Markovketten

Definition 19.8. Die Markovkette X heißt **reversibel** bezüglich des Maßes π , falls

$$\pi(\{x\}) p(x, y) = \pi(\{y\}) p(y, x) \quad \text{für alle } x, y \in E. \quad (19.6)$$

Die Gleichung (19.6) heißt auch die **Gleichung der detaillierten Balance** (detailed balance).

X heißt reversibel, falls es ein π gibt, bezüglich dessen X reversibel ist.

Bemerkung 19.9. Ist X reversibel bezüglich π , dann ist π ein invariantes Maß für X , denn

$$\pi p(\{x\}) = \sum_{y \in E} \pi(\{y\}) p(y, x) = \sum_{y \in E} \pi(\{x\}) p(x, y) = \pi(\{x\}).$$

Nach Bemerkung 17.50 ist π bis auf konstante Vielfache eindeutig, falls X irreduzibel und rekurrent ist. \diamond

Beispiel 19.10. Sei (E, K) ein Graph mit Eckenmenge (oder Menge der Knoten) E und Kantenmenge K (siehe Seite 67). Mit $\langle x, y \rangle = \langle y, x \rangle \in K$ bezeichnen wir eine (ungerichtete) Kante, die x und y verbindet. Sei $C := (C(x, y), x, y \in E)$ eine Familie von Gewichten mit $C(x, y) = C(y, x) \geq 0$ für alle $x, y \in E$ und

$$C(x) := \sum_{y \in E} C(x, y) < \infty \quad \text{für jedes } x \in E.$$

Setzen wir $p(x, y) := \frac{C(x, y)}{C(x)}$ für alle $x, y \in E$, so ist X reversibel bezüglich $\pi(\{x\}) = C(x)$. Es gilt nämlich

$$\begin{aligned} \pi(\{x\}) p(x, y) &= C(x) \frac{C(x, y)}{C(x)} = C(x, y) \\ &= C(y, x) = C(y) \frac{C(y, x)}{C(y)} = \pi(\{y\}) p(y, x). \end{aligned} \quad \diamond$$

Definition 19.11. Seien (E, K) , C und X wie in Beispiel 19.10. Dann heißt X Irrfahrt auf E mit Gewichten C . Ist speziell $C(x, y) = \mathbb{1}_{\{(x,y)\in K\}}$, dann heißt X einfache Irrfahrt auf (E, K) .

Die Irrfahrt mit Gewichten C ist also reversibel. Es gilt aber auch die Umkehrung.

Satz 19.12. Ist X eine reversible Markovkette, so ist X eine Irrfahrt auf E mit Gewichten $C(x, y) = p(x, y)\pi(\{x\})$, falls π ein invariantes Maß ist. Ist X irreduzibel und rekurrent, so ist π und damit C bis auf Vielfache eindeutig.

Beweis. Klar. □

Übung 19.2.1. Man zeige: p ist genau dann reversibel bezüglich π , wenn die lineare Abbildung $f \mapsto pf$ in $L^2(\pi)$ selbstadjungiert ist. ♣

Übung 19.2.2. Sei $K \in \mathbb{N}$ und Zahlen $W_1, \dots, W_K \in \mathbb{R}$ und $\beta > 0$ gegeben. Wir definieren

$$p(i, j) := \frac{1}{Z} \exp(-\beta W_j) \quad \text{für alle } i, j = 1, \dots, K,$$

wobei $Z := \sum_{j=1}^K \exp(-\beta W_j)$ die Normalisierungskonstante ist.

In K (nummerierten) Urnen befinden sich insgesamt N ununterscheidbare Kugeln. In jedem Zeitschritt wird (uniform) eine der N Kugeln zufällig ausgesucht. Ist i die Nummer der Urne, aus der die Kugel gezogen wurde, so wird die Kugel mit Wahrscheinlichkeit $p(i, j)$ in die Urne mit der Nummer j gelegt.

- (i) Man gebe eine formale Beschreibung als Markovkette an.
- (ii) Man bestimme den invarianten Zustand π und zeige, dass die Kette reversibel bezüglich π ist. ♣

19.3 Endliche Elektrische Netzwerke

Ein (endliches) elektrisches Netzwerk (E, C) ist ein (endliches) System E von Punkten, die paarweise mit Drähten der **Leitfähigkeit** (*conductivity*) $C(x, y) \in [0, \infty)$, $x, y \in E$ verbunden sind. Wir interpretieren $C(x, y) = 0$ so, dass es „keinen Draht zwischen x und y “ gibt. Symmetrie erfordert $C(x, y) = C(y, x)$. Mit

$$R(x, y) = \frac{1}{C(x, y)} \in (0, \infty]$$

bezeichnen wir den **Widerstand** der Verbindung $\langle x, y \rangle$. Ist (E, K) ein Graph und $C(x, y) = \mathbb{1}_{\{(x,y)\in K\}}$, so bezeichnen wir (E, C) als **Einheitsnetzwerk** auf (E, K) .

Sei im Folgenden stets (E, C) ein **endliches** elektrisches Netzwerk.

Sei nun $A \subset E$. Wir legen an den Punkten $x_0 \in A$ jeweils elektrische Spannungen $u(x_0)$ an (zum Beispiel durch Anschluss einer oder mehrerer Batterien). Wie groß ist dann die Spannung $u(x)$ in $x \in E \setminus A$?

Definition 19.13. Eine Abbildung $I : E \times E \rightarrow \mathbb{R}$ heißt ein **Fluss** auf $E \setminus A$, falls sie antisymmetrisch ist ($I(x, y) = -I(y, x)$) und das **Kirchhoff'sche Gesetz** erfüllt:

$$\begin{aligned} I(x) &= 0, \quad \text{für } x \in E \setminus A, \\ I(A) &= 0, \end{aligned} \tag{19.7}$$

wobei

$$I(x) := \sum_{y \in E} I(x, y) \quad \text{und} \quad I(A) := \sum_{x \in A} I(x).$$

Definition 19.14. Ein Fluss $I : E \times E \rightarrow \mathbb{R}$ auf $E \setminus A$ heißt **elektrischer Fluss**, falls es eine Funktion $u : E \rightarrow \mathbb{R}$ gibt, bezüglich der das **Ohm'sche Gesetz** gilt:

$$I(x, y) = \frac{u(x) - u(y)}{R(x, y)} \quad \text{für alle } x, y \in E, x \neq y.$$

Wir nennen dann $I(x, y)$ die **Stromstärke** von x nach y und $u(x)$ die **elektrische Spannung** in x .

Satz 19.15. Eine elektrische Spannung u in (E, C) ist harmonisch auf $E \setminus A$:

$$u(x) = \sum_{y \in E} \frac{1}{C(x)} C(x, y) u(y) \quad \text{für jedes } x \in E \setminus A.$$

Speziell ist die elektrische Spannung durch Angabe der Werte auf A festgelegt, wenn das Netzwerk irreduzibel ist.

Beweis. Nach dem Ohm'schen und dem Kirchhoff'schen Gesetz ist

$$u(x) - \sum_{y \in E} \frac{C(x, y)}{C(x)} u(y) = \sum_{y \in E} \frac{C(x, y)}{C(x)} (u(x) - u(y)) = \frac{1}{C(x)} \sum_{y \in E} I(x, y) = 0.$$

Nach dem Eindeutigkeitssatz für harmonische Funktionen (Satz 19.7) ist u hierdurch und durch die Werte auf A eindeutig festgelegt. \square

Korollar 19.16. Sei X eine Markovkette auf E mit Kantengewichten C . Dann ist $u(x) = \mathbf{E}_x[u(X_{\tau_A})]$.

Betrachte $A = \{x_0, x_1\}$, $x_0 \neq x_1$, und $u(x_0) = 0$, $u(x_1) = 1$. Dann ist $I(x_1)$ der gesamte Stromfluss in das Netzwerk und $-I(x_0)$ der gesamte Stromfluss aus dem

Netzwerk. Das Kirchhoff'sche Gesetz besagt, dass der Stromfluss divergenzfrei ist, und dass in Summe genauso viel Strom rein- wie rausfließt. Mit anderen Worten eben $I(x_0) + I(x_1) = 0$.

In Anlehnung an das Ohm'sche Gesetz definieren wir den *effektiven Widerstand* zwischen x_0 und x_1 durch

$$R_{\text{eff}}(x_0 \leftrightarrow x_1) = \frac{u(x_1) - u(x_0)}{I(x_1)} = \frac{1}{I(x_1)} = -\frac{1}{I(x_0)}$$

und die *effektive Leitfähigkeit* durch $C_{\text{eff}}(x_0 \leftrightarrow x_1) = \frac{1}{R_{\text{eff}}(x_0 \leftrightarrow x_1)}$. Da I und u eindeutig durch die Angabe von x_0 , x_1 und C festgelegt sind, sind $C_{\text{eff}}(x_0 \leftrightarrow x_1)$ und $R_{\text{eff}}(x_0 \leftrightarrow x_1)$ Größen, die sich aus C berechnen lassen.

Wir betrachten nun zwei Mengen $A_0, A_1 \subset E$ mit $A_0 \cap A_1 = \emptyset$, $A_0, A_1 \neq \emptyset$, und setzen $u(x) = 0$ für jedes $x \in A_0$ sowie $u(x) = 1$ für jedes $x \in A_1$. Sei I der zugehörige elektrische Fluss. In Analogie zu oben treffen wir die folgende Definition.

Definition 19.17. Wir nennen $C_{\text{eff}}(A_0 \leftrightarrow A_1) := I(A_1)$ die **effektive Leitfähigkeit** zwischen A_0 und A_1 und $R_{\text{eff}}(A_0 \leftrightarrow A_1) := \frac{1}{I(A_1)}$ den **effektiven Widerstand** zwischen A_0 und A_1 .

Beispiel 19.18. (i) Sei $E = \{0, 1, 2\}$ mit $C(0, 2) = 0$, und $A_0 = \{x_0\} = \{0\}$, $A_1 = \{x_1\} = \{2\}$. Wir setzen $u(0) = 0$ und $u(2) = 1$. Dann ist (mit $p(x, y) = C(x, y)/C(x)$)

$$\begin{aligned} u(1) &= 1 \cdot p(1, 2) + 0 \cdot p(1, 0) = \frac{C(1, 2)}{C(1, 2) + C(1, 0)} = \frac{R(1, 0)}{R(1, 0) + R(1, 2)} \\ &= \frac{R_{\text{eff}}(1 \leftrightarrow 0)}{R_{\text{eff}}(1 \leftrightarrow 0) + R_{\text{eff}}(1 \leftrightarrow 2)}. \end{aligned}$$

Der gesamte Fluss ist

$$I(\{2\}) = u(1) C(0, 1) = \frac{1}{R(0, 1) + R(1, 2)} = \frac{1}{\frac{1}{C(0, 1)} + \frac{1}{C(1, 2)}}.$$

Entsprechend ist $R_{\text{eff}}(0 \leftrightarrow 2) = \frac{1}{I(\{2\})} = R(0, 1) + R(1, 2)$ und $C_{\text{eff}}(0 \leftrightarrow 2) = \left(\frac{1}{C(0, 1)} + \frac{1}{C(1, 2)}\right)^{-1}$.

(ii) (Reihenschaltung) Sei $n \in \mathbb{N}, n \geq 2$ und $E = \{0, \dots, n\}$ mit Leitfähigkeiten $C(k-1, k) > 0$ und $C(k, l) = 0$, falls $|k-l| > 1$. Nach dem Kirchhoff'schen Gesetz ist $I(l, l+1) = -I(x_1)$ für jedes $l = 0, \dots, n-1$. Nach dem Ohm'schen Gesetz folgt $u(1) = u(0) + I(x_1) R(0, 1)$, $u(2) = u(1) + I(x_1) R(1, 2)$ und so fort, also

$$u(k) - u(0) = I(x_1) \sum_{l=0}^{k-1} R(l, l+1)$$

und damit

$$R_{\text{eff}}(0 \leftrightarrow k) = \frac{u(k) - u(0)}{I(x_1)} = \sum_{l=0}^{k-1} R(l, l+1).$$

Aus Symmetriegründen erhalten wir ebenfalls

$$R_{\text{eff}}(k \leftrightarrow n) = \sum_{l=k}^{n-1} R(l, l+1)$$

und damit insbesondere $R_{\text{eff}}(0 \leftrightarrow n) = R_{\text{eff}}(0 \leftrightarrow k) + R_{\text{eff}}(k \leftrightarrow n)$.

Schließlich folgt für $k \in \{1, \dots, n-1\}$

$$u(k) = \frac{R_{\text{eff}}(0 \leftrightarrow k)}{R_{\text{eff}}(0 \leftrightarrow k) + R_{\text{eff}}(k \leftrightarrow n)}.$$

Wir erhalten so eine Aussage über die Ruinwahrscheinlichkeit der korrespondierenden Markovkette X auf $\{0, \dots, n\}$ durch

$$\mathbf{P}_k[\tau_n < \tau_0] = u(k) = \frac{R_{\text{eff}}(0 \leftrightarrow k)}{R_{\text{eff}}(0 \leftrightarrow n)} = \sum_{l=0}^{k-1} R(l, l+1) \Bigg/ \sum_{l=0}^{n-1} R(l, l+1). \quad (19.8)$$

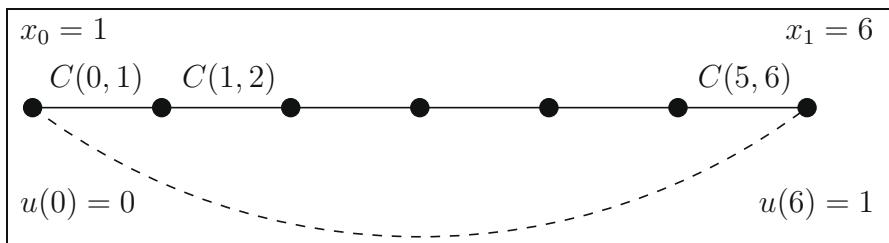


Abb. 19.1. Reihenschaltung von sechs Widerständen. Der effektive Gesamtwiderstand beträgt $R_{\text{eff}}(0 \leftrightarrow 6) = R(0, 1) + \dots + R(5, 6)$.

(iii) (Parallelschaltung) Sei $E = \{0, 1\}$. Wir wollen formal mehrere (parallele) Drähte zwischen 0 und 1 erlauben mit Leitfähigkeiten C_1, \dots, C_n . Dann ist nach dem Ohm'schen Gesetz der Strom entlang des i -ten Drahtes $I_i = \frac{u(1)-u(0)}{R_i} = \frac{1}{R_i}$, also der Gesamtstrom $I = \sum_{i=1}^n \frac{1}{R_i}$ und damit

$$C_{\text{eff}}(0 \leftrightarrow 1) = \sum_{i=1}^n C_i \quad \text{und} \quad R_{\text{eff}}(0 \leftrightarrow 1) = \left(\sum_{i=1}^n \frac{1}{R_i} \right)^{-1}. \quad \diamond$$

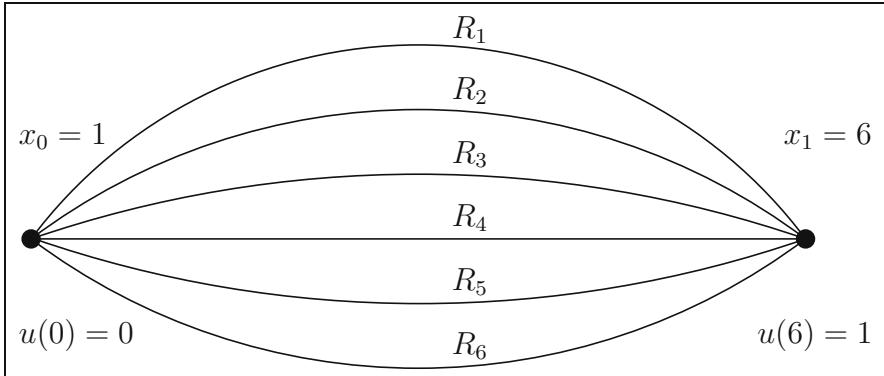


Abb. 19.2. Parallelschaltung von sechs Widerständen. Der effektive Gesamtwiderstand beträgt $R_{\text{eff}}(0 \leftrightarrow 1) = (R_1^{-1} + \dots + R_6^{-1})^{-1}$.

In allen drei Beispielen ist der effektive Widerstand eine monotone Funktion der einzelnen Widerstände. Dass dies allgemein gilt, ist die Aussage des Rayleigh'schen Monotonieprinzips.

Satz 19.19 (Rayleigh'sches Monotonieprinzip). Seien (E, C) und (E, C') elektrische Netzwerke mit $C(x, y) \geq C'(x, y)$ für alle $x, y \in E$.

Dann ist für $A_0, A_1 \subset E$ mit $A_0 \cap A_1 = \emptyset$, $A_0, A_1 \neq \emptyset$

$$C_{\text{eff}}(A_0 \leftrightarrow A_1) \geq C'_{\text{eff}}(A_0 \leftrightarrow A_1).$$

Der Rest dieses Abschnitts ist dem Beweis dieses Satzes gewidmet. Wir benötigen dazu den Energieerhaltungssatz sowie das Thomson'sche (oder Dirichlet'sche) Prinzip der Leistungsminimierung.

Satz 19.20 (Energieerhaltungssatz). Sei $A = A_0 \cup A_1$, und sei I ein Fluss (das heißt eine antisymmetrische Funktion, die dem Kirchhoff'schen Gesetz genügt, nicht aber notwendigerweise dem Ohm'schen Gesetz) auf $E \setminus A$. Ferner sei $w : E \rightarrow \mathbb{R}$ eine Funktion, die auf A_0 und A_1 jeweils konstant ist: $w|_{A_0} \equiv w_0$ und $w|_{A_1} \equiv w_1$. Dann gilt

$$(w_1 - w_0)I(A_1) = \frac{1}{2} \sum_{x, y \in E} (w(x) - w(y)) I(x, y).$$

Dies ist die diskrete Version des Satzes von Gauß für (wI) , wobei man beachte, dass das Kirchhoff'sche Gesetz besagt, dass I auf $E \setminus A$ divergenzfrei ist.

Beweis. Wir berechnen

$$\begin{aligned}
 \sum_{x,y \in E} (w(x) - w(y)) I(x,y) &= \sum_{x \in E} \left(w(x) \sum_{y \in E} I(x,y) \right) - \sum_{y \in E} \left(w(y) \sum_{x \in E} I(x,y) \right) \\
 &= \sum_{x \in A} \left(w(x) \sum_{y \in E} I(x,y) \right) - \sum_{y \in A} \left(w(y) \sum_{x \in E} I(x,y) \right) \\
 &= w_0 I(A_0) + w_1 I(A_1) - w_0 (-I(A_0)) - w_1 (-I(A_1)) \\
 &= 2(w_1 - w_0) I(A_1). \quad \square
 \end{aligned}$$

Definition 19.21. Sei I ein Fluss auf $E \setminus A$. Mit

$$L_I := L_I^C := \frac{1}{2} \sum_{x,y \in E} I(x,y)^2 R(x,y)$$

bezeichnen wir die **Leistung** von I im Netzwerk (E, C) .

Satz 19.22 (Thomson'sches oder Dirichlet'sches Prinzip der Leistungsminimierung). Seien I, J Einheitsflüsse von A_1 nach A_0 (das heißt $I(A_1) = J(A_1) = 1$). I sei zudem ein elektrischer Fluss (erfülle also das Ohm'sche Gesetz mit einer Spannungsfunktion u , die auf A_0 und A_1 jeweils konstant ist). Dann gilt

$$L_I \leq L_J$$

mit Gleichheit genau dann, wenn $I = J$ ist. Speziell ist der elektrische Einheitsfluss eindeutig festgelegt.

Beweis. Sei $D = J - I \not\equiv 0$ der Differenzfluss. Dann ist offenbar $D(A_0) = D(A_1) = 0$. Wir erhalten

$$\begin{aligned}
 &\sum_{x,y \in E} J(x,y)^2 R(x,y) \\
 &= \sum_{x,y \in E} (I(x,y) + D(x,y))^2 R(x,y) \\
 &= \sum_{x,y \in E} (I(x,y)^2 + D(x,y)^2) R(x,y) + 2 \sum_{x,y \in E} I(x,y) D(x,y) R(x,y) \\
 &= \sum_{x,y \in E} (I(x,y)^2 + D(x,y)^2) R(x,y) + 2 \sum_{x,y \in E} (u(x) - u(y)) D(x,y).
 \end{aligned}$$

Nach dem Energieerhaltungssatz ist der letzte Term

$$2 \sum_{x,y \in E} (u(x) - u(y)) D(x,y) = 4D(A_1)(u_1 - u_0) = 0.$$

Es folgt (wegen $D \not\equiv 0$)

$$L_J = L_I + \frac{1}{2} \sum_{x,y \in E} D(x,y)^2 R(x,y) > L_I. \quad \square$$

Beweis (Rayleigh'sches Monotonieprinzip, Satz 19.19). Seien I und I' die elektrischen Einheitsflüsse von A_1 nach A_0 bezüglich C beziehungsweise C' . Nach dem Thomson'schen Prinzip, dem Energieerhaltungssatz und der Voraussetzung $R(x, y) \leq R'(x, y)$ für alle $x, y \in E$ ist

$$\begin{aligned} R_{\text{eff}}(A_0 \leftrightarrow A_1) &= \frac{u(1) - u(0)}{I(A_1)} = u(1) - u(0) \\ &= \frac{1}{2} \sum_{x, y \in E} I(x, y)^2 R(x, y) \\ &\leq \frac{1}{2} \sum_{x, y \in E} I'(x, y)^2 R(x, y) \leq \frac{1}{2} \sum_{x, y \in E} I'(x, y)^2 R'(x, y) \\ &= u'(1) - u'(0) = R'_{\text{eff}}(A_0 \leftrightarrow A_1). \end{aligned} \quad \square$$

19.4 Rekurrenz und Transienz

Wir betrachten die Situation, wo E abzählbar ist und $A_1 = \{x_1\}$ für ein $x_1 \in E$. Sei X eine Irrfahrt auf E mit Gewichten $C = (C(x, y), x, y \in E)$, also mit Übergangswahrscheinlichkeiten $p(x, y) = C(x, y)/C(x)$ (vergleiche Definition 19.11).

Um die Ergebnisse über endliche elektrische Netzwerke aus dem letzten Abschnitt anwenden zu können, nehmen wir zudem immer an, dass $A_0 \subset E$ so gewählt ist, dass $E \setminus A_0$ endlich ist. Es sei dann stets $u = u_{x_1, A_0}$ die eindeutig bestimmte Spannungsfunktion auf E mit $u(x_1) = 1$ und $u(x) = 0$ für jedes $x \in A_0$. Nach Satz 19.7 ist u harmonisch und hat die Darstellung

$$\begin{aligned} u_{x_1, A_0}(x) &= \mathbf{E}_x \left[\mathbb{1}_{\{X_{\tau_{A_0} \cup \{x_1\}}} = x_1\}} \right] \\ &= \mathbf{P}_x [\tau_{x_1} < \tau_{A_0}] \quad \text{für jedes } x \in E \setminus (A_0 \cup \{x_1\}). \end{aligned}$$

Daher ist für den zu u gehörigen elektrischen Fluss I

$$\begin{aligned} -I(A_0) &= I(x_1) = \sum_{x \in E} I(x_1, x) = \sum_{x \in E} (u(x_1) - u(x)) C(x_1, x) \\ &= C(x_1) \sum_{x \in E} (1 - u(x)) p(x_1, x) \\ &= C(x_1) \left(\sum_{x \notin A_0 \cup \{x_1\}} p(x_1, x) \mathbf{P}_x [\tau_{A_0} < \tau_{x_1}] + \sum_{x \in A_0} p(x_1, x) \right) \\ &= C(x_1) \mathbf{P}_{x_1} [\tau_{A_0} < \tau_{x_1}]. \end{aligned}$$

Es folgt

$$\begin{aligned} p_F(x_1, A_0) &:= \mathbf{P}_{x_1} [\tau_{A_0} < \tau_{x_1}] \\ &= \frac{C_{\text{eff}}(x_1 \leftrightarrow A_0)}{C(x_1)} = \frac{1}{C(x_1)} \frac{1}{R_{\text{eff}}(x_1 \leftrightarrow A_0)}. \end{aligned} \quad (19.9)$$

Definition 19.23. Die **Fluchtwahrscheinlichkeit** von x_1 bezeichnen wir mit

$$p_F(x_1) = \mathbf{P}_{x_1} [\tau_{x_1} = \infty] = 1 - F(x_1, x_1).$$

Die **effektive Leitfähigkeit** von x_1 nach ∞ bezeichnen wir mit

$$\begin{aligned} C_{\text{eff}}(x_1 \leftrightarrow \infty) &:= C(x_1) \inf \{p_F(x_1, A_0) : A_0 \subset E \text{ mit } |E \setminus A_0| < \infty, A_0 \not\ni x_1\}. \end{aligned}$$

Lemma 19.24. Für jede absteigende Folge $A_0^n \downarrow \emptyset$ mit $|E \setminus A_0^n| < \infty$ und $x_1 \notin A_0^n$ für jedes $n \in \mathbb{N}$ ist

$$C_{\text{eff}}(x_1 \leftrightarrow \infty) = \lim_{n \rightarrow \infty} C_{\text{eff}}(x_1 \leftrightarrow A_0^n).$$

Beweis. Klar, weil

$$C_{\text{eff}}(x_1 \leftrightarrow \infty) = C(x_1) \inf \{p_F(x_1, A_0) : |E \setminus A_0| < \infty, A_0 \not\ni x_1\}, \quad (19.10)$$

und weil $p_F(x_1, A_0)$ monoton fallend in A_0 ist. \square

Satz 19.25. Es gilt

$$p_F(x_1) = \frac{1}{C(x_1)} C_{\text{eff}}(x_1 \leftrightarrow \infty). \quad (19.11)$$

Speziell gilt

$$x_1 \text{ ist rekurrent} \iff C_{\text{eff}}(x_1 \leftrightarrow \infty) = 0 \iff R_{\text{eff}}(x_1 \leftrightarrow \infty) = \infty.$$

Beweis. Sei $A_0^n \downarrow \emptyset$ eine absteigende Folge mit $|E \setminus A_0^n| < \infty$ und $x_1 \notin A_0^n$ für jedes $n \in \mathbb{N}$. Setze $F_n := \{\tau_{A_0^n} < \tau_{x_1}\}$. Für jedes $M \in \mathbb{N}$ ist

$$\mathbf{P}_{x_1} [\tau_{A_0^n} \leq M] \leq \sum_{k=0}^M \mathbf{P}_{x_1} [X_k \in A_0^n] \xrightarrow{n \rightarrow \infty} 0.$$

Also gilt $\tau_{A_0^n} \uparrow \infty$ fast sicher, und damit ist $F_n \downarrow \{\tau_{x_1} = \infty\}$ (bis auf eine Nullmenge). Wir erhalten

$$\frac{1}{C(x_1)} C_{\text{eff}}(x_1 \leftrightarrow \infty) = \lim_{n \rightarrow \infty} \mathbf{P}_{x_1} [F_n] = \mathbf{P}_{x_1} [\tau_{x_1} = \infty] = p_F(x_1). \quad \square$$

Beispiel 19.26. Die symmetrische einfache Irrfahrt auf $E = \mathbb{Z}$ ist rekurrent. Hier ist $C(x, y) = \mathbb{1}_{\{|x-y|=1\}}$. Der effektive Widerstand von 0 nach ∞ ist nach den Formeln über Parallel- und Reihenschaltung

$$R_{\text{eff}}(0 \leftrightarrow \infty) = \frac{1}{2} \sum_{i=0}^{\infty} R(i, i+1) = \infty. \quad \diamond$$

Beispiel 19.27. Die asymmetrische einfache Irrfahrt auf $E = \mathbb{Z}$ mit $p(x, x+1) = p \in (\frac{1}{2}, 1)$, $p(x, x-1) = 1-p$ ist transient. Hier ist

$$C(x, x+1) = \left(\frac{p}{1-p} \right)^x \quad \text{für } x \in \mathbb{Z},$$

und $C(x, y) = 0$, falls $|x-y| > 1$. Der effektive Widerstand von 0 nach ∞ ist nach dem Monotonieprinzip

$$\begin{aligned} R_{\text{eff}}(0 \leftrightarrow \infty) &= \lim_{n \rightarrow \infty} R_{\text{eff}}(0 \leftrightarrow \{-n, n\}) \\ &\leq \lim_{n \rightarrow \infty} R_{\text{eff}}(0 \leftrightarrow n) \\ &= \sum_{n=0}^{\infty} \left(\frac{1-p}{p} \right)^n = \frac{p}{2p-1} < \infty. \end{aligned} \quad \diamond$$

Beispiel 19.28. Die symmetrische einfache Irrfahrt auf $E = \mathbb{Z}^2$ ist rekurrent. Hier ist wieder $C(x, y) = \mathbb{1}_{\{|x-y|=1\}}$. Sei $B_n = \{-n, \dots, n\}^2$ und $\partial B_n = B_n \setminus B_{n-1}$. Wir stellen ein Netzwerk C' mit größeren Leitfähigkeiten her, indem wir ringförmige *Supraleiter* entlang ∂B einfügen. Wir ersetzen also $C(x, y)$ durch

$$C'(x, y) = \begin{cases} \infty, & \text{falls } x, y \in \partial B_n \quad \text{für ein } n \in \mathbb{N}, \\ C(x, y), & \text{sonst.} \end{cases}$$

Dann ist $R'_{\text{eff}}(B_n \leftrightarrow B_n^c) = \frac{1}{4(2n+1)}$ (merke: $4(2n+1)$ ist die Anzahl der Kanten, die B_n mit B_n^c verbinden), und daher ist

$$R'_{\text{eff}}(0 \leftrightarrow \infty) = \sum_{n=0}^{\infty} \frac{1}{4(2n+1)} = \infty.$$

Nach dem Monotonieprinzip ist daher $R_{\text{eff}}(0 \leftrightarrow \infty) \geq R'_{\text{eff}}(0 \leftrightarrow \infty) = \infty$. \diamond

Beispiel 19.29. Sei (E, K) ein beliebiger zusammenhängender Teilgraph des quadratischen Gitters $(\mathbb{Z}^2, \mathbb{L}^2)$. Dann ist die einfache Irrfahrt auf (E, K) (siehe Definition 19.11) rekurrent. Nach dem Monotonieprinzip ist nämlich

$$R_{\text{eff}}^{(E, K)}(0 \leftrightarrow \infty) \geq R_{\text{eff}}^{(\mathbb{Z}^2, \mathbb{L}^2)}(0 \leftrightarrow \infty) = \infty. \quad \diamond$$

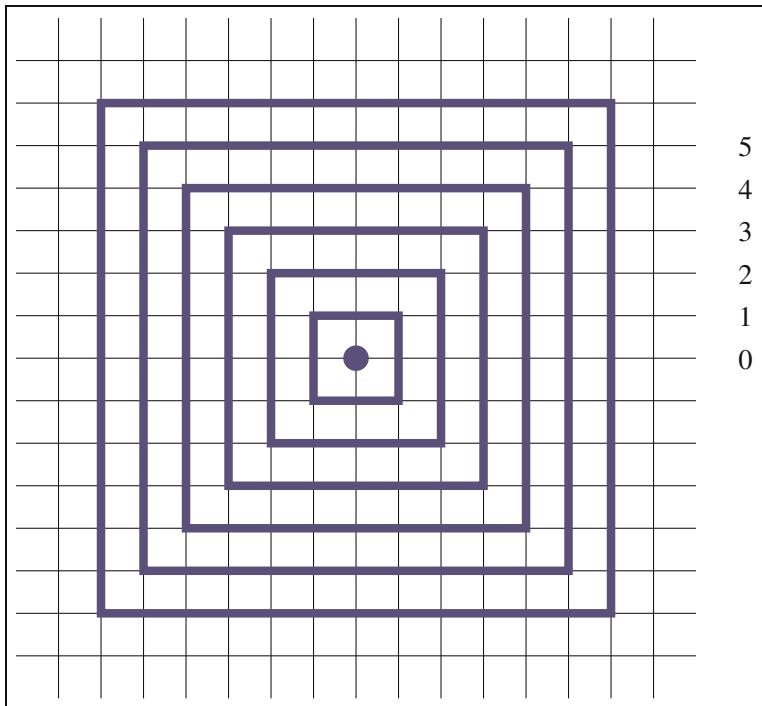


Abb. 19.3. Elektrisches Netzwerk auf \mathbb{Z}^2 . Die fetten Linien stellen *Supraleiter* dar. Zwischen dem n -ten und dem $(n + 1)$ -ten Supraleiter sind genau $4(2n + 1)$ Kanten.

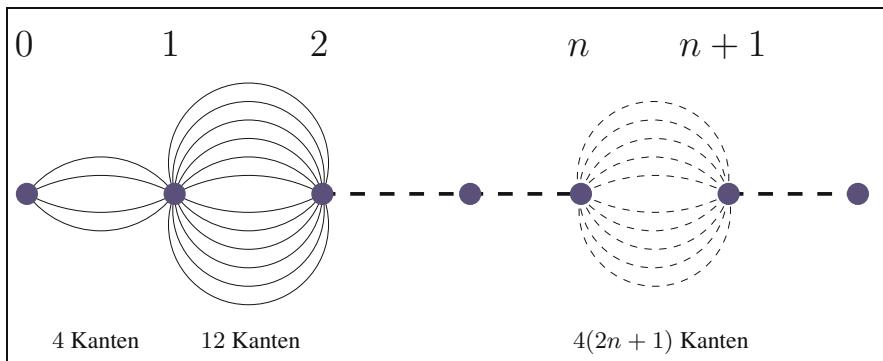


Abb. 19.4. Effektives Netzwerk, das aus \mathbb{Z}^2 durch Einfügen der Supraleiter entsteht. Die Ringe der Supraleiter sind hier zu einzelnen Punkten verschmolzen.

Wir formulieren das Vorgehen in den letzten Beispielen als Satz.

Satz 19.30. Seien C und C' Kantengewichte auf E mit $C'(x, y) \leq C(x, y)$ für alle $x, y \in E$. Ist die Markovkette X zu den Gewichten C rekurrent, so ist es auch die Markovkette X' zu den Gewichten C' .

Sei speziell (E, K) ein Graph und (E', K') ein Teilgraph. Ist die einfache Irrfahrt auf (E, K) rekurrent, so ist auch die einfache Irrfahrt auf (E', K') rekurrent.

Beweis. Das folgt direkt aus Satz 19.25 zusammen mit dem Rayleigh'schen Monotonieprinzip (Satz 19.19). \square

Beispiel 19.31. Die symmetrische einfache Irrfahrt auf \mathbb{Z}^3 ist transient. Zum Beweis konstruieren wir einen Teilgraphen, für den wir $R'_{\text{eff}}(0 \leftrightarrow \infty) < \infty$ ausrechnen können.

Skizze. Wir betrachten die Menge aller unendlichen Pfade, die in 0 starten und

- einen Schritt in x -Richtung, y -Richtung oder z -Richtung gehen (rechts, oben oder hinten, nicht links, unten oder vorne),
- eine eventuell andere Richtung x, y oder z wählen und dann zwei Schritte in diese Richtung gehen,
- in der n -ten Stufe eine der Richtungen x, y oder z wählen und 2^{n+1} Schritte in diese Richtung gehen.

Wir bezeichnen etwa mit $xyyxxxxzzzzzz\dots$ den Pfad, der zunächst die x -Richtung, dann y , dann x , dann z und so fort gewählt hat. Zwei Pfade benutzen offenbar nach dem Zeitpunkt, wo sich ihre Wege trennen, keine gemeinsamen Knoten mehr. Allerdings werden manche *Knoten* von mehreren Pfaden getroffen.

Wenn wir das elektrische Netzwerk mit Einheitswiderständen und Spannung 1 im Ursprung sowie Spannung 0 an allen Punkten von Pfaden nach der n -ten Stufe betrachten, so hängt aus Symmetriegründen die Spannung an jedem Knoten des Netzwerks nur vom Abstand (kürzester Weg entlang Pfaden) zum Ursprung ab. Wir erhalten also ein äquivalentes Netzwerk, wenn wir mehrfach benutzte Knoten durch entsprechend mehrere Knoten ersetzen (siehe Abb. 19.5). So erhalten wir ein Netzwerk, das eine Baumstruktur hat: jeweils nach 2^n Schritten verzweigt jeder Pfad in drei Pfade (siehe Abb. 19.6). Die 3^n Pfade von der n -ten Generation zur $(n+1)$ -ten Generation sind disjunkte Pfade der Länge 2^{n-1} . Sind $B(n)$ alle Punkte bis zur n -ten Generation, so ist

$$R'_{\text{eff}}(0 \leftrightarrow B(n+1)^c) = \sum_{k=0}^{n-1} R'_{\text{eff}}(B(k) \leftrightarrow B(k)^c) = \sum_{k=0}^{n-1} 2^k 3^{-k}.$$

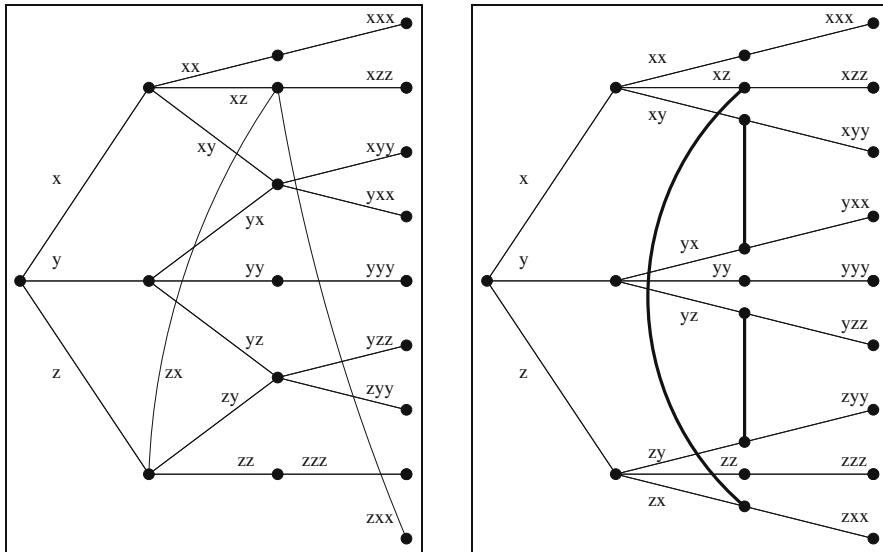


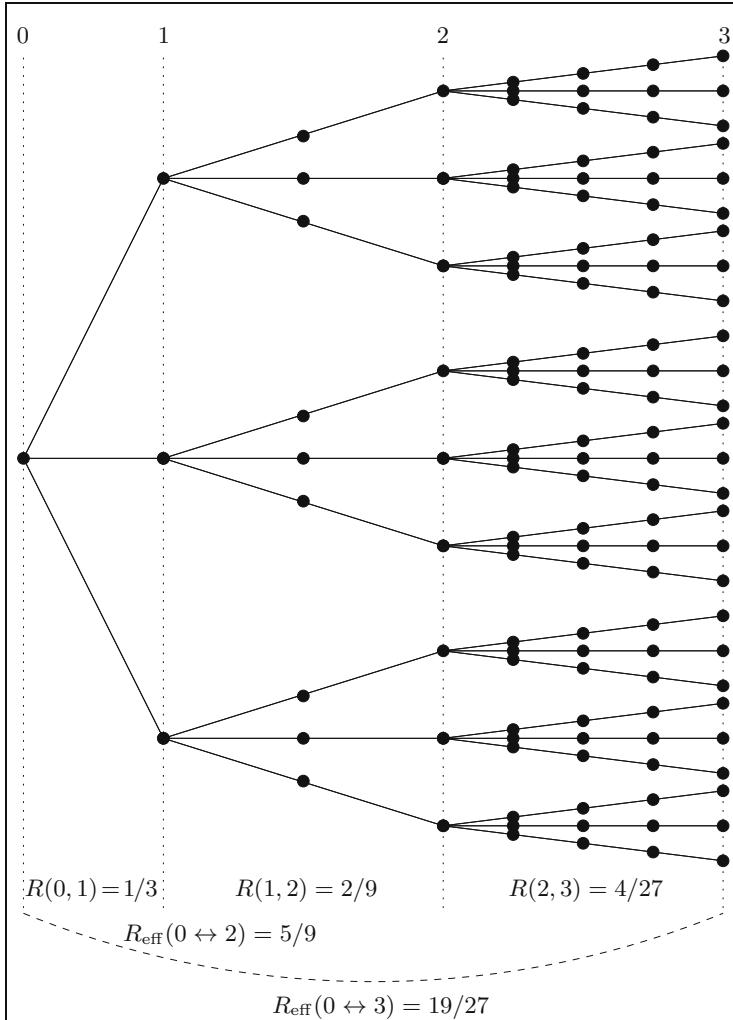
Abb. 19.5. Schema der ersten drei Schritte des Graphen von Beispiel 19.31. Links sind die tatsächlichen Kanten eingezeichnet, wobei beispielsweise xyy bedeutet, dass zunächst ein Schritt in x -Richtung gemacht wurde, dann einer in y -Richtung und jetzt die weiterführende Kante in y -Richtung betrachtet wird. Rechts sind die Knoten an den Enden von xz/zx , xy/yx und yz/zy jeweils in zwei Knoten aufgelöst und mit einem „Supraleiter“ (fette Linien) verbunden. Wenn wir die Supraleiter entfernen, so erhalten wir das Netzwerk aus Abb. 19.6, dessen effektiver Widerstand $R'_{\text{eff}}(0 \leftrightarrow \infty)$ nicht kleiner ist als derjenige in \mathbb{Z}^3 . (Wird an die Wurzel die Spannung 1 und an den rechten Punkten jeweils die Spannung 0 angelegt, so fließt aus Symmetriegründen durch die Supraleiter kein Strom. Das Netzwerk hier ist also sogar äquivalent zu dem in Abb. 19.6.)

Also ist $R'_{\text{eff}}(0 \leftrightarrow \infty) = \frac{1}{3} \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^k = 1 < \infty$. Für diesen Baum ist die Irrfahrt transient, nach Satz 19.30 also auch für \mathbb{Z}^3 . \diamond

Übung 19.4.1. Betrachte das elektrische Netzwerk auf \mathbb{Z}^d mit Einheitswiderständen zwischen benachbarten Punkten. Ferner sei X die symmetrische einfache Irrfahrt auf \mathbb{Z}^d . Schließlich seien $x_0, x_1 \in \mathbb{Z}^d$ zwei benachbarte Punkte. Man zeige:

- (i) Die effektive Leitfähigkeit zwischen x_0 und x_1 ist $C_{\text{eff}}(x_0 \leftrightarrow x_1) = d$.
- (ii) Ist $d \leq 2$, so ist $\mathbf{P}_{x_0}[\tau_{x_1} < \tau_{x_0}] = \frac{1}{2}$.
- (iii) Ist $d \geq 3$, so ist $\mathbf{P}_{x_0}[\tau_{x_1} < \tau_{x_0} \mid \tau_{x_0} < \infty] = \frac{1}{2}$.



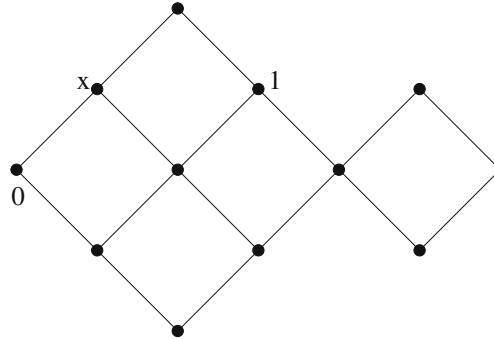
Abb. 19.6. Ein Baum als Teilgraph von \mathbb{Z}^3 , auf dem die Irrfahrt immer noch transient ist.

19.5 Netzwerkreduktion

Beispiel 19.32. Wir betrachten die Irrfahrt auf dem Graphen aus Abb. 19.7, die in x startet und an jedem Punkt mit gleicher Wahrscheinlichkeit zu einem der Nachbarpunkte springt. Mit welcher Wahrscheinlichkeit P trifft die Kette den Punkt 1 bevor sie den Punkt 0 trifft?

Wir fassen den Graphen als elektrisches Netzwerk auf mit Widerstand 1 an jeder Kante und Spannungen 0 in 0 sowie 1 in 1. Dann ist P die Spannung im Punkte x :

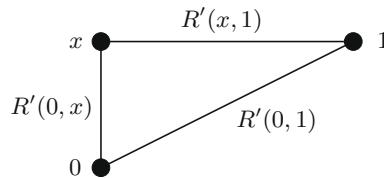
$$P = u(x).$$

**Abb. 19.7.** Ausgangssituation.

Zur Berechnung von $u(x)$ werden wir das Netzwerk schrittweise durch einfachere Netzwerke ersetzen, die zwischen den Punkten 0, 1, und x die gleichen effektiven Widerstände haben, so dass in x folglich die selbe Spannung $u(x)$ anliegt. \diamond

Reduziertes Netzwerk

Nehmen wir also an, dass wir das Netzwerk reduziert haben auf eines mit den drei Punkten 0, 1 und x und den Widerständen $R'(0, 1)$, $R'(0, x)$ und $R'(1, x)$.

**Abb. 19.8.** Reduziertes Netzwerk mit drei Knoten.

Dann ist offenbar

$$P = u(x) = \frac{R'(0, x)}{R'(0, x) + R'(1, x)}. \quad (19.12)$$

Wenn wir die effektiven Widerstände $R_{\text{eff}}(0 \leftrightarrow x)$, $R_{\text{eff}}(1 \leftrightarrow x)$ und $R_{\text{eff}}(0 \leftrightarrow 1)$ schon kennen, können wir uns die Arbeit der Netzwerkreduktion sparen und $u(x)$ direkt berechnen. Zur Herleitung der entsprechenden Formel stellen wir folgende Betrachtung an. Im reduzierten Netzwerk sind die effektiven Widerstände zwischen zwei Punkten leicht zu berechnen. Ist $\{a, b, c\} = \{0, 1, x\}$, so ist

$$R_{\text{eff}}(a \leftrightarrow b) = \left(\frac{1}{R'(a, b)} + \frac{1}{R'(a, c) + R'(b, c)} \right)^{-1}. \quad (19.13)$$

Auflösen dieser drei Gleichungen nach $R'(0, 1)$, $R'(0, x)$ sowie $R'(1, x)$ und Einsetzen in (19.12) liefert

$$P = u(x) = \frac{R_{\text{eff}}(0 \leftrightarrow 1) + R_{\text{eff}}(0 \leftrightarrow x) - R_{\text{eff}}(x \leftrightarrow 1)}{2 R_{\text{eff}}(0 \leftrightarrow 1)}. \quad (19.14)$$

Speziell gilt im Fall $R'(0, 1) = \infty$ (oder äquivalent: $R_{\text{eff}}(0 \leftrightarrow 1) = R_{\text{eff}}(0 \leftrightarrow x) + R_{\text{eff}}(x \leftrightarrow 1)$), dass $R_{\text{eff}}(0 \leftrightarrow x) = R'(0, x)$ und $R_{\text{eff}}(1 \leftrightarrow x) = R'(1, x)$, also

$$u(x) = \frac{R_{\text{eff}}(0 \leftrightarrow x)}{R_{\text{eff}}(0 \leftrightarrow x) + R_{\text{eff}}(x \leftrightarrow 1)}. \quad (19.15)$$

Da stets $u(x) \in [0, 1]$ ist, liefert Umstellen der Terme (wieder im allgemeinen Fall)

$$R_{\text{eff}}(1 \leftrightarrow x) \leq R_{\text{eff}}(0 \leftrightarrow 1) + R_{\text{eff}}(0 \leftrightarrow x). \quad (19.16)$$

Dies ist die Dreiecksungleichung für die effektiven Widerstände und zeigt, dass (in jedem elektrischen Netzwerk) die effektiven Widerstände eine Metrik liefern.

Schrittweise Reduktion von Netzwerken

Nachdem wir gesehen haben, wie man aus den effektiven Widerständen die Spannung $u(x)$ berechnet, wenden wir uns der Berechnung dieser effektiven Widerstände zu. Nach der Vorstellung der einzelnen Schritte wenden wir das Verfahren konkret für das eingangs betrachtete Beispiel an.

Es gibt vier elementare Vereinfachungsschritte für elektrische Netzwerke.

1. Entfernen von Schleifen. Die drei Punkte ganz rechts im Graphen bilden eine Schleife, die ohne Veränderungen im Rest des Netzwerks entfernt werden kann. Insbesondere kann jede Kante entfernt werden, die 0 und 1 direkt verbindet.

2. Zusammenfassen von seriellen Kannten. Zwei (oder mehr) Kanten, die seriell liegen, und deren dazwischen liegende Knoten keine weiteren Verbindungen haben, können durch ein Kante ersetzt werden, deren Widerstand die Summe der einzelnen Widerstände ist (siehe Abb. 19.1).

3. Zusammenfassen von parallelen Kannten. Zwei (oder mehr) Kanten mit Widerständen R_1, \dots, R_n , die die selben Knoten verbinden, können durch eine Kante mit Widerstand $R = (R_1^{-1} + \dots + R_n^{-1})^{-1}$ ersetzt werden (siehe Abb. 19.2).

4. Stern-Dreieck-Transformation. (Siehe Übung 19.5.1) Der sternförmige Ausschnitt eines Netzwerk links in Abb. 19.9 ist äquivalent zum dreieckigen Ausschnitt rechts, wenn die Widerstände $R_1, R_2, R_3, \tilde{R}_1, \tilde{R}_2, \tilde{R}_3$ die folgende Bedingung erfüllen

$$R_i \tilde{R}_i = \delta \quad \text{für jedes } i = 1, 2, 3, \quad (19.17)$$

wobei

$$\delta = R_1 R_2 R_3 (R_1^{-1} + R_2^{-1} + R_3^{-1}) = \frac{\tilde{R}_1 \tilde{R}_2 \tilde{R}_3}{\tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3}.$$

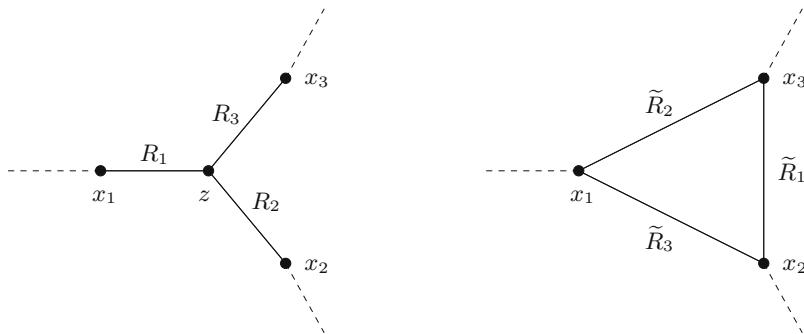


Abb. 19.9. Stern-Dreieck-Transformation.

Anwendung auf Beispiel 19.32

Wir lösen nun die Aufgabe aus Beispiel 19.32 konkret. Wir nehmen an, dass anfangs jede Kante den Widerstand 1 hat. Kanten, die im Verlauf der Reduktion andere Widerstände als 1 haben, werden mit dem entsprechenden Widerstand beschriftet.

Schritt 1. Die Schleife am rechten Rand wird entfernt (siehe Abb. 19.10).

Schritt 2. Die Serien an der oberen, rechten und unteren Ecke werden durch je einen Widerstand der Größe 2 ersetzt (siehe Abb. 19.10).

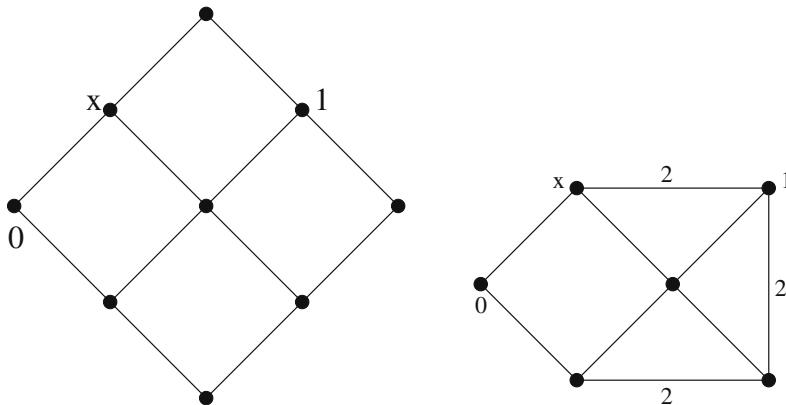


Abb. 19.10. Schritt 1 und 2.

Schritt 3. Der linke untere Knoten wird mit der Stern-Dreieck-Transformation entfernt. Hier ist $R_1 = 1$, $R_2 = 2$, $R_3 = 1$, $\delta = 5$, $\tilde{R}_1 = \delta/R_1 = 5$, $\tilde{R}_2 = \delta/R_2 = 5/2$ und $\tilde{R}_3 = \delta/R_3 = 5$ (siehe Abb. 19.11).

Schritt 4. Die parallelen Kanten mit Widerständen $R_1 = 5$ und $R_2 = 1$ werden ersetzt durch eine Kante mit $R = (\frac{1}{5} + 1)^{-1} = \frac{5}{6}$ (siehe Abb. 19.11).

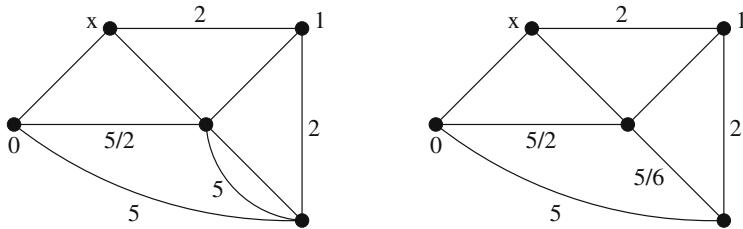


Abb. 19.11. Schritt 3 und 4.

Schritt 5. Der rechte untere Knoten wird mit der Stern-Dreieck-Transformation entfernt. Hier ist $R_1 = 5$, $R_2 = 2$, $R_3 = \frac{5}{6}$, $\delta = 95/6$, $\tilde{R}_1 = \delta/R_1 = 19/6$, $\tilde{R}_2 = \delta/R_2 = 95/12$ und $\tilde{R}_3 = \delta/R_3 = 19$ (siehe Abb. 19.12).

Schritt 6. Die parallelen Kanten werden durch je eine Kante ersetzt mit Widerstand $(\frac{12}{95} + \frac{2}{5})^{-1} = \frac{19}{10}$ beziehungsweise $(\frac{6}{19} + 1)^{-1} = \frac{19}{25}$ (siehe Abb. 19.12).

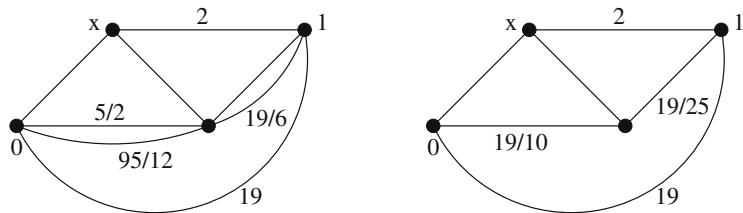


Abb. 19.12. Schritt 5 und 6.

Schritt 7. Der rechte untere Knoten wird mit der Stern-Dreieck-Transformation entfernt. Hier ist $R_1 = \frac{19}{10}$, $R_2 = \frac{19}{25}$, $R_3 = 1$, $\delta = \frac{513}{125}$, $\tilde{R}_1 = \delta/R_1 = \frac{54}{25}$, $\tilde{R}_2 = \delta/R_2 = \frac{27}{5}$ und $\tilde{R}_3 = \delta/R_3 = \frac{513}{125}$ (siehe Abb. 19.13).

Schritt 8. Die drei parallelen Kantenpaare werden durch je eine Kante ersetzt mit Widerständen $(\frac{5}{27} + 1)^{-1} = \frac{27}{32}$, $(\frac{25}{54} + \frac{1}{2})^{-1} = \frac{27}{26}$ und $(\frac{1}{19} + \frac{125}{513})^{-1} = \frac{27}{8}$.

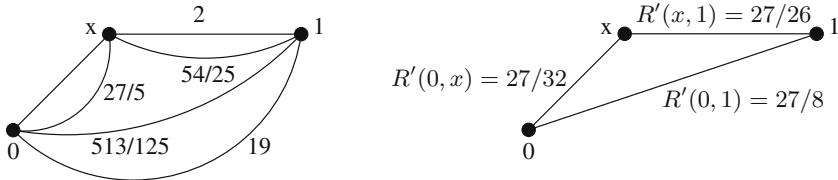


Abb. 19.13. Schritt 7 und 8.

Mit Gleichung (19.12) erhalten wir nun als Wahrscheinlichkeit, dass die Irrfahrt 1 erreicht, bevor sie 0 erreicht:

$$P = \frac{\frac{27}{32}}{\frac{27}{32} + \frac{27}{26}} = \frac{13}{29}.$$

Aus den Widerständen $R'(0, x)$, $R'(1, x)$ und $R'(0, 1)$ können wir mit Gleichung (19.13) die effektiven Widerstände (des reduzierten wie des ursprünglichen Netzwerkes, denn die effektiven Widerstände sind ja gleich geblieben) berechnen:

$$\begin{aligned} R_{\text{eff}}(0 \leftrightarrow x) &= \left(\frac{32}{27} + \frac{1}{\frac{27}{8} + \frac{27}{26}} \right)^{-1} = \frac{17}{24}, \\ R_{\text{eff}}(1 \leftrightarrow x) &= \left(\frac{26}{27} + \frac{1}{\frac{27}{32} + \frac{27}{8}} \right)^{-1} = \frac{5}{6}, \\ R_{\text{eff}}(0 \leftrightarrow 1) &= \left(\frac{8}{27} + \frac{1}{\frac{27}{26} + \frac{27}{32}} \right)^{-1} = \frac{29}{24}. \end{aligned}$$

Unter Verwendung von (19.14) können wir auch mit den effektiven Widerständen $u(x)$ berechnen:

$$P = u(x) = \frac{\frac{29}{24} + \frac{17}{24} - \frac{5}{6}}{2 \cdot \frac{29}{24}} = \frac{13}{29}.$$

Während diese Formel natürlich umständlicher ist als diejenige mit den Widerständen R' hat sie allerdings den Vorteil, dass sie auch anwendbar ist, wenn die effektiven Widerstände aus anderen Gründen bekannt sind. Beispielsweise könnte man das Netzwerk mit echten Widerständen löten und mit einem Messgerät die effektiven Widerstände ausmessen. ◇

Alternative Lösung

Wir können die Lösung des Problems aus Beispiel 19.32 auch ohne Netzwerke, alleine mit linearer Algebra angeben.

Welche Lösung eleganter ist, ist wohl Geschmackssache. Zunächst stellen wir die Übergangsmatrix p der Markovkette auf (hierfür werden die Knoten des Graphen

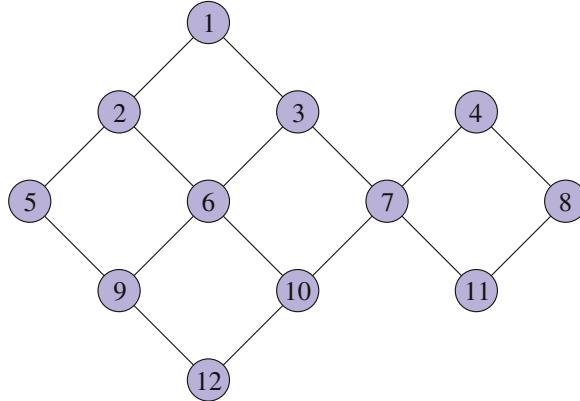


Abb. 19.14. Graph mit nummerierten Knoten.

von 1 bis 12 durchnummieriert wie in Abb. 19.14). Der Startpunkt ist die 2, das „Gewinnfeld“ ist die 3 und das „Verlustfeld“ die 5.

Nun wird die Matrix \bar{p} der in 3 und 5 getöteten Kette gebildet und $\bar{G} = (I - \bar{p})^{-1}$ berechnet. Nach Übung 19.1.1 (mit $A = \{3, 5\}$, $x = 2$ und $y = 3$) ist die Wahrscheinlichkeit dafür, 3 vor 5 zu treffen $P = \bar{G}(2, 3) = \frac{13}{29}$.

$$\bar{p} := \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{pmatrix}$$

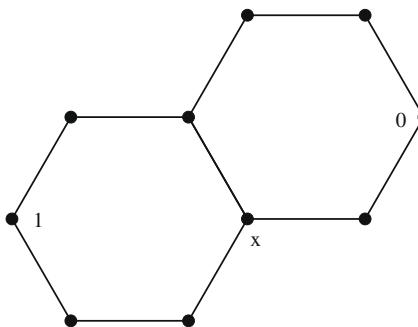
$$\overline{G} := (I - \bar{p})^{-1} = \begin{pmatrix} \frac{143}{116} & \frac{81}{116} & \frac{21}{29} & \frac{3}{58} & \frac{8}{29} & \frac{19}{58} & \frac{3}{29} & \frac{3}{58} & \frac{15}{116} & \frac{9}{58} & \frac{3}{58} & \frac{11}{116} \\ \frac{27}{58} & \frac{81}{58} & \frac{13}{29} & \frac{3}{29} & \frac{16}{29} & \frac{19}{29} & \frac{6}{29} & \frac{3}{29} & \frac{15}{58} & \frac{9}{29} & \frac{3}{29} & \frac{11}{58} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{3}{58} & \frac{9}{58} & \frac{24}{29} & \frac{165}{58} & \frac{5}{29} & \frac{15}{29} & \frac{78}{29} & \frac{68}{29} & \frac{21}{58} & \frac{30}{29} & \frac{107}{58} & \frac{27}{58} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{19}{116} & \frac{57}{116} & \frac{18}{29} & \frac{15}{58} & \frac{11}{29} & \frac{95}{58} & \frac{15}{29} & \frac{15}{58} & \frac{75}{116} & \frac{45}{58} & \frac{15}{58} & \frac{55}{116} \\ \frac{3}{58} & \frac{9}{58} & \frac{24}{29} & \frac{39}{29} & \frac{5}{29} & \frac{15}{29} & \frac{78}{29} & \frac{39}{29} & \frac{21}{58} & \frac{30}{29} & \frac{39}{29} & \frac{27}{58} \\ \frac{3}{58} & \frac{9}{58} & \frac{24}{29} & \frac{68}{29} & \frac{5}{29} & \frac{15}{29} & \frac{78}{29} & \frac{97}{29} & \frac{21}{58} & \frac{30}{29} & \frac{68}{29} & \frac{27}{58} \\ \frac{5}{58} & \frac{15}{58} & \frac{11}{29} & \frac{7}{29} & \frac{18}{29} & \frac{25}{29} & \frac{14}{29} & \frac{7}{29} & \frac{93}{58} & \frac{21}{29} & \frac{7}{29} & \frac{45}{58} \\ \frac{3}{29} & \frac{9}{29} & \frac{19}{29} & \frac{20}{29} & \frac{10}{29} & \frac{30}{29} & \frac{40}{29} & \frac{20}{29} & \frac{21}{29} & \frac{60}{29} & \frac{20}{29} & \frac{27}{29} \\ \frac{3}{58} & \frac{9}{58} & \frac{24}{29} & \frac{107}{58} & \frac{5}{29} & \frac{15}{29} & \frac{78}{29} & \frac{68}{29} & \frac{21}{58} & \frac{30}{29} & \frac{165}{58} & \frac{27}{58} \\ \frac{11}{116} & \frac{33}{116} & \frac{15}{29} & \frac{27}{58} & \frac{14}{29} & \frac{55}{58} & \frac{27}{29} & \frac{27}{58} & \frac{135}{116} & \frac{81}{58} & \frac{27}{58} & \frac{215}{116} \end{pmatrix}$$

Übung 19.5.1. Man zeige die Gültigkeit der Stern-Dreieck-Transformation. ♣

Übung 19.5.2. Man zeige für den unten stehenden hexagonalen Graphen, dass die Wahrscheinlichkeit, von x aus startend die 1 vor der 0 zu treffen, gleich $\frac{8}{17}$ ist

(i) mit der Methode der Netzwerkreduktion,

(ii) mit der Methode der Matrixinversion.



Übung 19.5.3. Man betrachte den Graphen aus Abb. 19.15.

(i) Zeige für die effektive Leitfähigkeit zwischen den Punkten a und z , dass $C_{\text{eff}}(a \longleftrightarrow z) = \sqrt{3}$.

(ii) Zeige, dass die Wahrscheinlichkeit $\mathbf{P}_a[\tau_z < \tau_a]$, dass eine in a gestartete Irrfahrt z trifft bevor sie nach a zurückkehrt, den Wert $\mathbf{P}_a[\tau_z < \tau_a] = 1/\sqrt{3}$ hat.



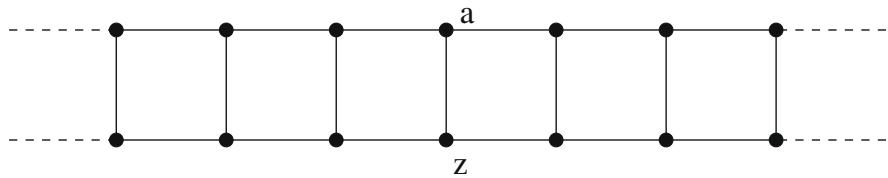


Abb. 19.15. Einfacher Leitergraph

Übung 19.5.4. Man bestimme $C_{\text{eff}}(a \longleftrightarrow z)$ und $\mathbf{P}_a[\tau_z < \tau_a]$ für den Graphen aus Abb. 19.16. (Das geht leichter als in Übung 19.5.3!) ♣

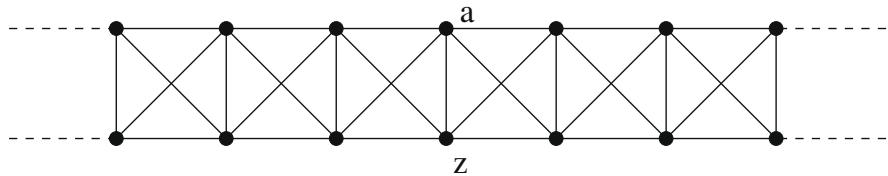


Abb. 19.16. Gekreuzter Leitergraph

Übung 19.5.5. Man bestimme die Wahrscheinlichkeit $\mathbf{P}_a[\tau_z < \tau_a]$ für den Graphen aus Abb. 19.17. ♣

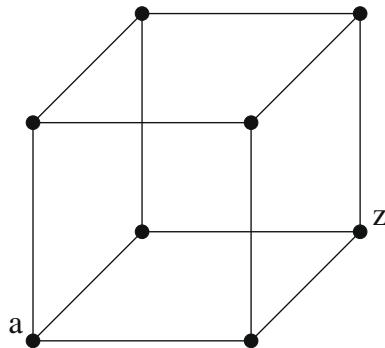


Abb. 19.17. Netzwerk Würfel

19.6 Irrfahrt in zufälliger Umgebung

(Vergleiche [186], [152] und [83, 84].) Wir betrachten eine Markovkette X auf \mathbb{Z} , die in jedem Schritt entweder einen Punkt nach links oder einen Punkt nach rechts springt, jeweils mit Wahrscheinlichkeit w_i^- beziehungsweise w_i^+ , falls X in $i \in \mathbb{Z}$ ist. Es seien also $w_i^- \in (0, 1)$ und $w_i^+ := 1 - w_i^-$ für $i \in \mathbb{Z}$. Dann ist X die Markovkette mit Übergangsmatrix

$$p_w(i, j) = \begin{cases} w_i^-, & \text{falls } j = i - 1, \\ w_i^+, & \text{falls } j = i + 1, \\ 0, & \text{sonst.} \end{cases}$$

Um X durch die Leitfähigkeiten eines elektrischen Netzwerks zu beschreiben, setzen wir $\varrho_i := w_i^- / w_i^+$ für $i \in \mathbb{Z}$ sowie $C_w(i, j) := 0$ falls $|i - j| \neq 1$ und

$$C_w(i+1, i) := C_w(i, i+1) := \begin{cases} \prod_{k=0}^i \varrho_k^{-1}, & \text{falls } i \geq 0, \\ \prod_{k=i}^{-1} \varrho_k, & \text{falls } i < 0. \end{cases}$$

Mit dieser Definition ist

$$\frac{C_w(i, i+1)}{C_w(i)} = \frac{1}{\varrho_i + 1} = w_i^+ \quad \text{und} \quad \frac{C_w(i, i-1)}{C_w(i)} = \frac{\varrho_i}{\varrho_i + 1} = w_i^-.$$

Die Übergangswahrscheinlichkeiten p_w werden also tatsächlich durch C_w beschrieben. Sei

$$R_w^+ := \sum_{i=0}^{\infty} R_w(i, i+1) = \sum_{i=0}^{\infty} \frac{1}{C_w(i, i+1)} = \sum_{i=0}^{\infty} \prod_{k=0}^i \varrho_k$$

und

$$R_w^- := \sum_{i=0}^{\infty} R_w(-i, -i-1) = \sum_{i=0}^{\infty} \frac{1}{C_w(-i, -i-1)} = \sum_{i=1}^{\infty} \prod_{k=-i}^{i-1} \varrho_k^{-1}.$$

R_w^+ und R_w^- sind die effektiven Widerstände von 0 nach $+\infty$ beziehungsweise nach $-\infty$. Also ist

$$R_{w,\text{eff}}(0 \leftrightarrow \infty) = \frac{1}{\frac{1}{R_w^-} + \frac{1}{R_w^+}}$$

genau dann endlich, wenn $R_w^- < \infty$ oder $R_w^+ < \infty$ ist. Nach Satz 19.25 gilt also

$$X \text{ ist transient} \iff R_w^- < \infty \text{ oder } R_w^+ < \infty. \quad (19.18)$$

In welche Richtung läuft X davon, wenn X transient ist?

Satz 19.33. (i) Gilt $R_w^- < \infty$ oder $R_w^+ < \infty$, so gilt (mit $\frac{\infty}{\infty} = 1$)

$$\mathbf{P}_0[X_n \xrightarrow{n \rightarrow \infty} -\infty] = \frac{R_w^+}{R_w^- + R_w^+} \quad \text{und} \quad \mathbf{P}_0[X_n \xrightarrow{n \rightarrow \infty} +\infty] = \frac{R_w^-}{R_w^- + R_w^+}.$$

(ii) Gilt $R_w^- = \infty$ und $R_w^+ = \infty$, so gilt $\liminf_{n \rightarrow \infty} X_n = -\infty$ und $\limsup_{n \rightarrow \infty} X_n = \infty$ fast sicher.

Beweis. (i) Sei $\tau_N := \inf \{n \in \mathbb{N}_0 : X_n \in \{-N, N\}\}$. Da X transient ist, ist $\mathbf{P}_0[\tau_N < \infty] = 1$ und (wie in (19.8))

$$\mathbf{P}_0[X_{\tau_N} = -N] = \frac{R_{w,\text{eff}}(0 \leftrightarrow N)}{R_{w,\text{eff}}(-N \leftrightarrow N)} = \frac{R_{w,\text{eff}}(0 \leftrightarrow N)}{R_{w,\text{eff}}(0 \leftrightarrow -N) + R_{w,\text{eff}}(0 \leftrightarrow N)}.$$

Es folgt, wiederum, weil X transient ist,

$$\begin{aligned} \mathbf{P}_0[X_n \xrightarrow{n \rightarrow \infty} -\infty] &= \mathbf{P}[\sup\{X_n : n \in \mathbb{N}_0\} < \infty] \\ &= \lim_{N \rightarrow \infty} \mathbf{P}[\sup\{X_n : n \in \mathbb{N}_0\} < N] \\ &\leq \limsup_{N \rightarrow \infty} \mathbf{P}[X_{\tau_N} = -N] \\ &= \frac{R_w^+}{R_w^- + R_w^+}. \end{aligned}$$

Aus Symmetriegründen (und da X transient ist), ist aber

$$\mathbf{P}_0[X_n \xrightarrow{n \rightarrow \infty} -\infty] = 1 - \mathbf{P}_0[X_n \xrightarrow{n \rightarrow \infty} \infty] \geq 1 - \frac{R_w^-}{R_w^- + R_w^+} = \frac{R_w^+}{R_w^- + R_w^+}.$$

(ii) Gilt $R_w^- = R_w^+ = \infty$, so ist X rekurrent, also wird *jeder* Punkt unendlich oft besucht, das heißt, es gilt $\limsup_{n \rightarrow \infty} X_n = \infty$ und $\liminf_{n \rightarrow \infty} X_n = -\infty$ f.s. \square

Wir wollen nun die Situation betrachten, wo die Zahlenfolge $w = (w_i^-)_{i \in \mathbb{Z}}$ zufällig ist. Das heißt, wir betrachten ein Zweistufen-Experiment: In der ersten Stufe werden u.i.v. Zufallsvariablen $W = (W_i^-)_{i \in \mathbb{Z}}$ auf $(0, 1)$ realisiert und $W_i^+ := 1 - W_i^-$ gesetzt. In der zweiten Stufe wird, gegeben W , eine Markovkette X auf \mathbb{Z} mit Übergangsmatrix p_W konstruiert, wobei

$$p_W(i, j) = \begin{cases} W_i^-, & \text{falls } j = i - 1, \\ W_i^+, & \text{falls } j = i + 1, \\ 0, & \text{sonst.} \end{cases}$$

Man beachte, dass X nur gegeben W eine Markovkette ist, also unter dem Wahrscheinlichkeitsmaß $\mathbf{P}[X \in \cdot | W]$, nicht aber bezüglich $\mathbf{P}[X \in \cdot]$. Ist nämlich W unbekannt, so kann man durch Beobachtung von X Rückschlüsse auf die tatsächliche Realisierung von W ziehen. Dies ist aber genau das, was Gedächtnis ist, steht also im Widerspruch zur Markoveigenschaft von X .

Definition 19.34. Der Prozess X heißt **Irrfahrt in der zufälligen Umgebung W** (*random walk in random environment*).

Seien $\varrho_i := W_i^- / W_i^+$ für $i \in \mathbb{Z}$ und R_W^- und R_W^+ wie oben definiert.

Satz 19.35 (Solomon (1975)). Es gelte $\mathbf{E}[|\log(\varrho_0)|] < \infty$.

- (i) Gilt $\mathbf{E}[\log(\varrho_0)] < 0$, so gilt $X_n \xrightarrow{n \rightarrow \infty} \infty$ f.s.
- (ii) Gilt $\mathbf{E}[\log(\varrho_0)] > 0$, so gilt $X_n \xrightarrow{n \rightarrow \infty} -\infty$ f.s.
- (iii) Gilt $\mathbf{E}[\log(\varrho_0)] = 0$, so gilt $\liminf_{n \rightarrow \infty} X_n = -\infty$ und $\limsup_{n \rightarrow \infty} X_n = \infty$ f.s.

Beweis. (i) und (ii) Aus Symmetriegründen reicht es, (ii) zu zeigen. Sei also $c := \mathbf{E}[\log(\varrho_0)] > 0$. Nach dem starken Gesetz der großen Zahl gibt es ein $n_0^- = n_0^-(\omega)$ mit

$$\prod_{k=-n}^1 \varrho_k^{-1} = \exp \left(- \sum_{k=-n}^1 \log(\varrho_k) \right) < e^{-cn/2} \quad \text{für alle } n \geq n_0^-.$$

Es folgt

$$R_W^- = \sum_{n=1}^{\infty} \prod_{k=-n}^1 \varrho_k^{-1} \leq \sum_{n=1}^{n_0^- - 1} \prod_{k=-n}^1 \varrho_k^{-1} + \sum_{n=n_0^-}^{\infty} e^{-cn/2} < \infty \quad \text{f.s.}$$

Analog gibt es ein $n_0^+ = n_0^+(\omega)$ mit

$$\prod_{k=0}^n \varrho_k > e^{cn/2} \quad \text{für alle } n \geq n_0^+.$$

Es folgt

$$R_W^+ = \sum_{n=0}^{\infty} \prod_{k=0}^n \varrho_k \geq \sum_{n=0}^{n_0^+ - 1} \prod_{k=0}^n \varrho_k + \sum_{n=n_0^+}^{\infty} e^{cn/2} = \infty \quad \text{f.s.}$$

Aus Satz 19.33 folgt nun $X_n \xrightarrow{n \rightarrow \infty} -\infty$ fast sicher.

(iii) Um zu zeigen, dass $R_W^- = R_W^+ = \infty$ fast sicher gilt, reicht es zu zeigen, dass $\limsup_{n \rightarrow \infty} \sum_{k=0}^n \log(\varrho_k) > -\infty$ und $\limsup_{n \rightarrow \infty} \sum_{k=-n}^1 \log(\varrho_k^{-1}) > -\infty$ fast sicher, wenn $\mathbf{E}[\log(\varrho_0)] = 0$ gilt. Wenn $\log(\varrho_0)$ von endlicher Varianz ist, folgt dies aus dem Zentralen Grenzwertsatz. Im allgemeinen Fall folgt dies aus Satz 20.21. \square

Übung 19.6.1. Man betrachte die Irrfahrt in zufälliger Umgebung aus Satz 19.35, aber eingeschränkt auf \mathbb{N}_0 . Damit ist gemeint, dass jeder Versuch der Irrfahrt, von 0 nach -1 zu springen verhindert wird. Man zeige, dass diese Irrfahrt in zufälliger Umgebung

- fast sicher transient ist, falls $E[\log(\varrho_0)] < \infty$,
- fast sicher nullrekurrent ist, falls $E[\log(\varrho_0)] = \infty$, und
- fast sicher positiv rekurrent ist, falls $E[\log(\varrho_0)] > \infty$.



20 Ergodentheorie

Gesetze der großen Zahl, zum Beispiel für u.i.v. Zufallsvariablen X_1, X_2, \dots besagen, dass $n^{-1} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_1]$ fast sicher konvergiert. Wir können also die Mittelung über die tatsächliche Realisierung vieler Zufallsvariablen mit der Mittelung über die möglichen Realisierungen eines X_i vertauschen. In der statistischen Physik spricht man von der Äquivalenz von *Zeitmittel* und *Scharmittel*, oder der Mittelung entlang einer Trajektorie (griechisch *odos*) des Systems gegenüber der Mittelung aller möglichen Zustände mit gleicher Energie (griechisch *ergon*). Hieraus leitet sich der Begriff der Ergodentheorie ab, die Gesetze der großen Zahl für Zufallsvariablen mit Abhängigkeiten, aber zeitlicher Stationarität liefert.

20.1 Begriffsbildung

Definition 20.1. Sei I eine unter Addition abgeschlossene Menge (die wichtigsten Beispiele für uns sind $I = \mathbb{N}_0$, $I = \mathbb{N}$, $I = \mathbb{Z}$, $I = \mathbb{R}$, $I = [0, \infty)$, $I = \mathbb{Z}^d$ usw.). Ein stochastischer Prozess $X = (X_t)_{t \in I}$ heißt **stationär**, falls

$$\mathcal{L}[(X_{t+s})_{t \in I}] = \mathcal{L}[(X_t)_{t \in I}] \quad \text{für jedes } s \in I. \quad (20.1)$$

Bemerkung 20.2. Ist $I = \mathbb{N}_0$, $I = \mathbb{N}$ oder $I = \mathbb{Z}$, so ist (20.1) äquivalent zu

$$\mathcal{L}[(X_{n+1})_{n \in I}] = \mathcal{L}[(X_n)_{n \in I}]. \quad \diamond$$

Beispiel 20.3. (i) Ist $X = (X_t)_{t \in I}$ u.i.v., so ist X stationär. Ist lediglich $\mathbf{P}_{X_t} = \mathbf{P}_{X_0}$ für jedes $t \in I$ (ohne die Unabhängigkeit), so ist X im Allgemeinen nicht stationär. Beispielsweise sei $I = \mathbb{N}_0$ und $X_1 = X_2 = X_3 = \dots$, jedoch $X_0 \neq X_1$. Dann ist X nicht stationär.

- (ii) Ist X eine Markovkette mit invarianter Verteilung π sowie $\mathcal{L}[X_0] = \pi$, dann ist X stationär.
- (iii) Sind $(Y_n)_{n \in \mathbb{Z}}$ u.i.v. und reell sowie $c_1, \dots, c_k \in \mathbb{R}$, dann definiert

$$X_n := \sum_{l=1}^k c_l Y_{n-l}$$

einen stationären Prozess X . X heißt manchmal auch **moving average** oder **gleitendes Mittel** mit Gewichten (c_1, \dots, c_k) . Eine genauere Betrachtung ergibt, dass X sogar dann stationär ist, wenn Y nur als stationär vorausgesetzt wird. \diamond

Lemma 20.4. Ist $(X_n)_{n \in \mathbb{N}_0}$ stationär, so lässt sich X zu einem stationären Prozess $(\tilde{X}_n)_{n \in \mathbb{Z}}$ fortsetzen.

Beweis. Sei \tilde{X} der kanonische Prozess auf $\Omega = E^{\mathbb{Z}}$. Sei $\tilde{\mathbf{P}}^{\{-n, -n+1, \dots\}} \in \mathcal{M}_1(E^{\{-n, -n+1, \dots\}})$ definiert durch

$$\begin{aligned} \tilde{\mathbf{P}}^{\{-n, -n+1, \dots\}}[\tilde{X}_{-n} \in A_{-n}, \tilde{X}_{-n+1} \in A_{-n+1}, \dots] \\ = \mathbf{P}[X_0 \in A_{-n}, X_1 \in A_{-n+1}, \dots]. \end{aligned}$$

Dann ist $(\tilde{\mathbf{P}}^{\{-n, -n+1, \dots\}}, n \in \mathbb{N})$ projektiv und $\{-n, -n+1, \dots\} \uparrow \mathbb{Z}$. Nach dem Satz von Ionescu-Tulcea (Satz 14.32) existiert der projektive Limes $\tilde{\mathbf{P}} := \lim_{\leftarrow n \rightarrow \infty} \tilde{\mathbf{P}}^{\{-n, -n+1, \dots\}}$. Per Konstruktion ist \tilde{X} stationär bezüglich $\tilde{\mathbf{P}}$ und

$$\tilde{\mathbf{P}} \circ ((\tilde{X}_n)_{n \in \mathbb{N}_0})^{-1} = \mathbf{P} \circ ((X_n)_{n \in \mathbb{N}_0})^{-1}. \quad \square$$

Im Folgenden sei stets $(\Omega, \mathcal{A}, \mathbf{P})$ ein W-Raum und $\tau : \Omega \rightarrow \Omega$ eine messbare Abbildung.

Definition 20.5. Ein Ereignis $A \in \mathcal{A}$ heißt **invariant**, falls $\tau^{-1}(A) = A$ und **quasi-invariant**, falls $\mathbb{1}_{\tau^{-1}(A)} = \mathbb{1}_A$ \mathbf{P} -f.s. Die σ -Algebra der invarianten Ereignisse bezeichnen wir mit

$$\mathcal{I} = \{A \in \mathcal{A} : \tau^{-1}(A) = A\}.$$

Zur Erinnerung: Eine σ -Algebra \mathcal{I} heißt \mathbf{P} -trivial, falls $\mathbf{P}[A] \in \{0, 1\}$ für jedes $A \in \mathcal{I}$ gilt.

Definition 20.6. (i) τ heißt **maßtreu**, falls

$$\mathbf{P}[\tau^{-1}(A)] = \mathbf{P}[A] \quad \text{für jedes } A \in \mathcal{A}.$$

In diesem Falle heißt $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ein **maßerhaltendes dynamisches System**.

(ii) Ist τ maßtreu und \mathcal{I} \mathbf{P} -trivial, so heißt $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ **ergodisch**.

Lemma 20.7. Sei $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ein maßerhaltendes dynamisches System.

- (i) Eine messbare Abbildung $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist genau dann \mathcal{I} -messbar, wenn $f \circ \tau = f$ ist.
- (ii) $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ist genau dann ergodisch, wenn jedes \mathcal{I} -messbare $f : (\Omega, \mathcal{I}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ \mathbf{P} -fast sicher konstant ist.

Beweis. (i) Für Indikatorfunktionen $f = \mathbb{1}_A$ ist dies klar. Der allgemeine Fall folgt mit den üblichen Approximationsargumenten (siehe Satz 1.96(i)).

(ii) „ \Rightarrow “ Sei $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ergodisch. Für jedes $c \in \mathbb{R}$ ist $f^{-1}((c, \infty)) \in \mathcal{I}$, also ist $\mathbf{P}[f^{-1}((c, \infty))] \in \{0, 1\}$ und damit

$$f = \inf \{c \in \mathbb{R} : \mathbf{P}[f^{-1}((c, \infty))] = 0\} \quad \mathbf{P}\text{-f.s.}$$

„ \Leftarrow “ Sei jede \mathcal{I} -messbare Abbildung \mathbf{P} -f.s. konstant. Ist $A \in \mathcal{I}$, so ist $\mathbb{1}_A$ \mathcal{I} -messbar, also \mathbf{P} -f.s. gleich 0 oder gleich 1 und damit $\mathbf{P}[A] \in \{0, 1\}$. \square

Beispiel 20.8. Sei $n \in \mathbb{N} \setminus \{1\}$, $\Omega = \mathbb{Z}/(n)$, $\mathcal{A} = 2^\Omega$ und \mathbf{P} die Gleichverteilung auf Ω . Sei $r \in \{1, \dots, n\}$ und

$$\tau : \Omega \rightarrow \Omega, \quad x \mapsto x + r \pmod{n}.$$

Dann ist τ maßtreu. Ist $d = \text{ggT}(n, r)$ und für $i = 0, \dots, d-1$

$$A_i = \{i, \tau(i), \tau^2(i), \dots, \tau^{n-1}(i)\} = i + \langle r \rangle,$$

so sind A_0, \dots, A_{d-1} die disjunkten Nebenklassen des Normalteilers $\langle r \rangle \trianglelefteq \Omega$. Also ist $A_i \in \mathcal{I}$ für $i = 0, \dots, d-1$, und jedes $A \in \mathcal{I}$ ist Vereinigung von gewissen A_i . Mithin gilt:

$$(\Omega, \mathcal{A}, \mathbf{P}, \tau) \text{ ist ergodisch} \iff \text{ggT}(r, n) = 1. \quad \diamond$$

Beispiel 20.9 (Rotation). Sei $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathbf{P} = \lambda$ das Lebesgue-Maß, $r \in (0, 1)$ und $\tau_r(x) = x + r \pmod{1}$. Offenbar ist $(\Omega, \mathcal{A}, \mathbf{P}, \tau_r)$ ein maßerhaltendes dynamisches System. Wir werden zeigen:

$$(\Omega, \mathcal{A}, \mathbf{P}, \tau_r) \text{ ist ergodisch} \iff r \text{ ist irrational.}$$

Sei $f : [0, 1] \rightarrow \mathbb{R}$ eine \mathcal{I} -messbare Funktion. Ohne Einschränkung sei f beschränkt also quadratisch integrierbar. Dann können wir f als Fourier-Reihe darstellen

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x} \quad \text{für } \mathbf{P}\text{-f.a. } x,$$

wobei die Reihe in L^2 konvergiert. Dabei ist die quadratsummierbare Folge $(a_n)_{n \in \mathbb{Z}}$ der Koeffizienten eindeutig bestimmt (vergleiche Übung 7.3.1 mit $a_n = (-i/2)a_n + (1/2)b_n$ und $c_{-n} = (i/2)a_n + (1/2)b_n$ für $n \in \mathbb{N}$ sowie $c_0 = b_0$). Es folgt

$$(f \circ \tau_r)(x) = \sum_{n=-\infty}^{\infty} (a_n e^{2\pi i n r}) e^{2\pi i n x} \quad \text{f.ü.}$$

Nach Lemma 20.7 ist f genau dann \mathcal{I} -messbar, wenn $f = f \circ \tau_r$ gilt, also wenn

$$a_n = a_n e^{2\pi i n r} \quad \text{für alle } n \in \mathbb{Z}.$$

Ist r irrational, so folgt $a_n = 0$ für $n \neq 0$, also ist f fast sicher konstant und damit ist $(\Omega, \mathcal{A}, \mathbf{P}, \tau_r)$ ergodisch.

Ist andererseits r rational, so gibt es ein $n \in \mathbb{Z} \setminus \{0\}$ mit $e^{2\pi i n r} = e^{-2\pi i n r} = 1$. Daher ist $x \mapsto e^{2\pi i n x} + e^{-2\pi i n x} = 2 \cos(2\pi n x)$ \mathcal{I} -messbar, jedoch nicht fast sicher konstant. In diesem Fall ist also $(\Omega, \mathcal{A}, \mathbf{P}, \tau_r)$ nicht ergodisch. \diamond

Beispiel 20.10. Sei $X = (X_n)_{n \in \mathbb{N}_0}$ ein stochastischer Prozess mit Werten in einem polnischen Raum E . Ohne Einschränkung können wir annehmen, dass X der kanonische Prozess auf dem W-Raum $(\Omega, \mathcal{A}, \mathbf{P}) = (E^{\mathbb{N}_0}, \mathcal{B}(E)^{\otimes \mathbb{N}_0}, \mathbf{P})$ ist. Definiere den **Shift**

$$\tau : \Omega \rightarrow \Omega, \quad (\omega_n)_{n \in \mathbb{N}_0} \mapsto (\omega_{n+1})_{n \in \mathbb{N}_0}.$$

Dann ist $X_n(\omega) = X_0(\tau^n(\omega))$. Also ist X genau dann stationär, wenn $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ein maßerhaltendes dynamisches System ist. \diamond

Definition 20.11. Der stochastische Prozess X (aus Beispiel 20.10) heißt **ergodisch**, falls $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ergodisch ist.

Beispiel 20.12. Seien $(X_n)_{n \in \mathbb{N}_0}$ u.i.v. und $X_n(\omega) = X_0(\tau^n(\omega))$. Ist $A \in \mathcal{I}$, so ist für $n \in \mathbb{N}$

$$A = \tau^{-n}(A) = \{\omega : \tau^n(\omega) \in A\} \in \sigma(X_n, X_{n+1}, \dots).$$

Also ist (mit \mathcal{T} die terminale σ -Algebra von $(X_n)_{n \in \mathbb{N}}$, siehe Definition 2.34)

$$\mathcal{I} \subset \mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots).$$

Nach dem Kolmogorov'schen 0-1 Gesetz (Satz 2.37) ist \mathcal{T} \mathbf{P} -trivial, also ist auch \mathcal{I} \mathbf{P} -trivial und damit $(X_n)_{n \in \mathbb{N}_0}$ ergodisch. \diamond

Übung 20.1.1. Sei G eine endliche Gruppe von maßtreuen messbaren Abbildungen auf $(\Omega, \mathcal{A}, \mathbf{P})$ und $\mathcal{A}_0 := \{A \in \mathcal{A} : g(A) = A \text{ für alle } g \in G\}$.

Man zeige: Für jedes $X \in \mathcal{L}^1(\mathbf{P})$ gilt

$$\mathbf{E}[X | \mathcal{A}_0] = \frac{1}{\#G} \sum_{g \in G} X \circ g.$$



20.2 Ergodensätze

In diesem Abschnitt ist stets $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ein maßerhaltendes dynamisches System. Ferner sei $f : \Omega \rightarrow \mathbb{R}$ messbar und

$$X_n(\omega) = f \circ \tau^n(\omega) \quad \text{für jedes } n \in \mathbb{N}_0.$$

Also ist $X = (X_n)_{n \in \mathbb{N}_0}$ ein stationärer, reeller stochastischer Prozess. Sei

$$S_n = \sum_{k=0}^{n-1} X_k$$

die n -te Partialsumme. Die Ergodensätze beschäftigen sich mit Gesetzen der großen Zahl für die (S_n) . Als Vorbereitung bringen wir ein Lemma.

Lemma 20.13 (Hopf'sches Maximal-Ergodenlemma). *Sei $X_0 \in \mathcal{L}^1(\mathbf{P})$. Setze $M_n = \max\{0, S_1, \dots, S_n\}$, $n \in \mathbb{N}$. Dann gilt*

$$\mathbf{E}[X_0 \mathbf{1}_{\{M_n > 0\}}] \geq 0 \quad \text{für jedes } n \in \mathbb{N}.$$

Beweis. Für $k \leq n$ ist $M_n(\tau(\omega)) \geq S_k(\tau(\omega))$. Also ist

$$X_0 + M_n \circ \tau \geq X_0 + S_k \circ \tau = S_{k+1}.$$

Daher ist $X_0 \geq S_{k+1} - M_n \circ \tau$ für $k = 1, \dots, n$. Offensichtlich ist $S_1 = X_0$ und $M_n \circ \tau \geq 0$, also auch (für $k = 0$) $X_0 \geq S_1 - M_n \circ \tau$ und damit auch

$$X_0 \geq \max\{S_1, \dots, S_n\} - M_n \circ \tau. \quad (20.2)$$

Außerdem ist offenbar

$$\{M_n > 0\}^c \subset \{M_n = 0\} \cap \{M_n \circ \tau \geq 0\} \subset \{M_n - M_n \circ \tau \leq 0\}. \quad (20.3)$$

Aus (20.2) und (20.3) und der Maßtreue von τ folgt

$$\begin{aligned} \mathbf{E}[X_0 \mathbf{1}_{\{M_n > 0\}}] &\geq \mathbf{E}[(\max\{S_1, \dots, S_n\} - M_n \circ \tau) \mathbf{1}_{\{M_n > 0\}}] \\ &= \mathbf{E}[(M_n - M_n \circ \tau) \mathbf{1}_{\{M_n > 0\}}] \\ &\geq \mathbf{E}[M_n - M_n \circ \tau] = \mathbf{E}[M_n] - \mathbf{E}[M_n] = 0. \quad \square \end{aligned}$$

Satz 20.14 (Individueller Ergodensatz, Birkhoff (1931)). *Sei $f = X_0 \in \mathcal{L}^1(\mathbf{P})$. Dann gilt*

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k = \frac{1}{n} \sum_{k=0}^{n-1} f \circ \tau^k \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_0 | \mathcal{I}] \quad \mathbf{P}\text{-f.s.}$$

Ist speziell τ ergodisch, so gilt $\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_0] \quad \mathbf{P}\text{-f.s.}$

Beweis. Ist τ ergodisch, so ist $\mathbf{E}[X_0|\mathcal{I}] = \mathbf{E}[X_0]$ und der Zusatz folgt aus der ersten Aussage.

Wir betrachten nun den allgemeinen Fall. Nach Lemma 20.7 ist $\mathbf{E}[X_0|\mathcal{I}] \circ \tau = \mathbf{E}[X_0|\mathcal{I}]$ \mathbf{P} -f.s. Wir können also $\tilde{X}_n := X_n - \mathbf{E}[X_0|\mathcal{I}]$ betrachten und daher ohne Beschränkung der Allgemeinheit $\mathbf{E}[X_0|\mathcal{I}] = 0$ annehmen. Setze

$$Z := \limsup_{n \rightarrow \infty} \frac{1}{n} S_n.$$

Sei $\varepsilon > 0$ und $F := \{Z > \varepsilon\}$. Zu zeigen ist, dass $\mathbf{P}[F] = 0$ gilt. Hieraus folgt dann $\mathbf{P}[Z > 0] = 0$ und analog mit $-X$ auch $\liminf_{n \rightarrow \infty} \frac{1}{n} S_n \geq 0$ fast sicher, also $\frac{1}{n} S_n \xrightarrow{n \rightarrow \infty} 0$ f.s.

Offenbar ist $Z \circ \tau = Z$, also $F \in \mathcal{I}$. Setze

$$\begin{aligned} X_n^\varepsilon &:= (X_n - \varepsilon) \mathbb{1}_F, & S_n^\varepsilon &:= X_0^\varepsilon + \dots + X_{n-1}^\varepsilon, \\ M_n^\varepsilon &:= \max\{0, S_1^\varepsilon, \dots, S_n^\varepsilon\}, & F_n &:= \{M_n^\varepsilon > 0\}. \end{aligned}$$

Dann ist $F_1 \subset F_2 \subset \dots$ und

$$\bigcup_{n=1}^{\infty} F_n = \left\{ \sup_{k \in \mathbb{N}} \frac{1}{k} S_k^\varepsilon > 0 \right\} = \left\{ \sup_{k \in \mathbb{N}} \frac{1}{k} S_k > \varepsilon \right\} \cap F = F,$$

also $F_n \uparrow F$. Majorisierte Konvergenz liefert $\mathbf{E}[X_0^\varepsilon \mathbb{1}_{F_n}] \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_0^\varepsilon]$.

Nach dem Maximal-Ergodenlemma (angewandt auf X^ε) ist $\mathbf{E}[X_0^\varepsilon \mathbb{1}_{F_n}] \geq 0$, also

$$0 \leq \mathbf{E}[X_0^\varepsilon] = \mathbf{E}[(X_0 - \varepsilon) \mathbb{1}_F] = \mathbf{E}[\mathbf{E}[X_0|\mathcal{I}] \mathbb{1}_F] - \varepsilon \mathbf{P}[F] = -\varepsilon \mathbf{P}[F].$$

Mithin ist $\mathbf{P}[F] = 0$. □

Als Folgerung erhält man den statistischen Ergodensatz oder L^p -Ergodensatz, den von Neumann 1931 vor Birkhoff gefunden, jedoch erst in [131] veröffentlicht hat. Zur Vorbereitung bringen wir ein elementares Lemma.

Lemma 20.15. Sei $p \geq 1$, und seien X_0, X_1, \dots identisch verteilte, reelle Zufallsvariablen mit $\mathbf{E}[|X_0|^p] < \infty$. Setzen wir $Y_n := \left| \frac{1}{n} \sum_{k=0}^{n-1} X_k \right|^p$ für $n \in \mathbb{N}$, so ist $(Y_n)_{n \in \mathbb{N}}$ gleichgradig integrierbar.

Beweis. Offenbar ist die einelementige Familie $\{|X_0|^p\}$ gleichgradig integrierbar. Nach Satz 6.19 existiert also eine monoton wachsende, konvexe Abbildung $f : [0, \infty) \rightarrow [0, \infty)$ mit $\frac{f(x)}{x} \rightarrow \infty$ für $x \rightarrow \infty$ und $C := \mathbf{E}[f(|X_0|^p)] < \infty$. Nach Satz 6.19 reicht es wiederum zu zeigen, dass $\mathbf{E}[f(Y_n)] \leq C$ für jedes $n \in \mathbb{N}$. Nach der Jensen'schen Ungleichung (für $x \mapsto |x|^p$) ist

$$Y_n \leq \frac{1}{n} \sum_{k=0}^{n-1} |X_k|^p.$$

Die Jensen'sche Ungleichung (diesmal auf f angewandt) liefert dann

$$f(Y_n) \leq f\left(\frac{1}{n} \sum_{k=0}^{n-1} |X_k|^p\right) \leq \frac{1}{n} \sum_{k=0}^{n-1} f(|X_k|^p),$$

$$\text{also } \mathbf{E}[f(Y_n)] \leq \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{E}[f(|X_k|^p)] = C.$$

□

Satz 20.16 (L^p -Ergodensatz, von Neumann 1931). Sei $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ein maßerhaltendes dynamisches System, $p \geq 1$, $X_0 \in \mathcal{L}^p(\mathbf{P})$ und $X_n = X_0 \circ \tau^n$. Dann gilt

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_0 | \mathcal{I}] \quad \text{in } L^p(\mathbf{P}).$$

Ist speziell τ ergodisch, so gilt $\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_0]$ in $L^p(\mathbf{P})$.

Beweis. Setze

$$Y_n := \left| \frac{1}{n} \sum_{k=0}^{n-1} X_k - \mathbf{E}[X_0 | \mathcal{I}] \right|^p \quad \text{für jedes } n \in \mathbb{N}.$$

Nach Lemma 20.15 ist $(Y_n)_{n \in \mathbb{N}}$ gleichgradig integrierbar, und nach dem Birkhoff'schen Ergodensatz gilt $Y_n \xrightarrow{n \rightarrow \infty} 0$ fast sicher. Nach Satz 6.25 gilt daher $\lim_{n \rightarrow \infty} \mathbf{E}[Y_n] = 0$.

Ist τ ergodisch, so ist $\mathbf{E}[X_0 | \mathcal{I}] = \mathbf{E}[X_0]$.

□

20.3 Beispiele

Beispiel 20.17. Sei $(X, (\mathbf{P}_x)_{x \in E})$ eine positiv rekurrente, irreduzible Markovkette auf dem abzählbaren Raum E mit invariante Verteilung π . Dann ist $\pi(\{x\}) > 0$ für jedes $x \in E$. Setze $\mathbf{P}_\pi = \sum_{x \in E} \pi(\{x\}) \mathbf{P}_x$. Dann ist X stationär auf $(\Omega, \mathcal{A}, \mathbf{P}_\pi)$. Wir schreiben τ für den Shift, also $X_n = X_0 \circ \tau^n$.

Sei nun $A \in \mathcal{I}$ invariant. Dann ist $A \in \mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots)$. Nach der starken Markoveigenschaft ist daher für jede endliche Stoppzeit σ (mit \mathcal{F}_σ die σ -Algebra der σ -Vergangenheit)

$$\mathbf{P}_\pi[X \in A | \mathcal{F}_\sigma] = \mathbf{P}_{X_\sigma}[X \in A]. \quad (20.4)$$

In der Tat ist $\{X \in A\} = \{X \in \tau^{-n}(A)\} = \{(X_n, X_{n+1}, \dots) \in A\}$. Für $B \in \mathcal{F}_\sigma$ erhalten wir mit der Markoveigenschaft (in der dritten Zeile)

$$\begin{aligned} \mathbf{E}_\pi[\mathbb{1}_{\{X \in B\}} \mathbb{1}_{\{X \in A\}}] &= \sum_{n=0}^{\infty} \sum_{x \in E} \mathbf{P}_\pi[X \in B, \sigma = n, X_n = x, X \in A] \\ &= \sum_{n=0}^{\infty} \sum_{x \in E} \mathbf{P}_\pi[X \in B, \sigma = n, X_n = x, X \circ \tau^n \in A] \\ &= \sum_{n=0}^{\infty} \sum_{x \in E} \mathbf{P}_\pi[X \in B, \sigma = n, X_n = x] \mathbf{P}_x[X \in A] \\ &= \mathbf{E}_\pi[\mathbb{1}_{\{X \in B\}} \mathbf{P}_{X_\sigma}[X \in A]]. \end{aligned}$$

Ist speziell $x \in E$ und $\sigma_x = \inf\{n \in \mathbb{N}_0 : X_n = x\}$, so ist $\sigma_x < \infty$, weil X rekurrent und irreduzibel ist. Es folgt aus (20.4) für jedes $x \in E$

$$\mathbf{P}_\pi[X \in A] = \mathbf{E}_\pi[\mathbf{P}_x[X \in A]] = \mathbf{P}_x[X \in A].$$

Also ist $\mathbf{P}_{X_n}[X \in A] = \mathbf{P}_\pi[X \in A]$ fast sicher und daher (mit $\sigma = n$ in (20.4))

$$\mathbf{P}_\pi[X \in A | X_0, \dots, X_n] = \mathbf{P}_{X_n}[X \in A] = \mathbf{P}_\pi[X \in A].$$

Nun ist $A \in \mathcal{I} \subset \sigma(X_1, X_2, \dots)$, also

$$\mathbf{P}_\pi[X \in A | X_0, \dots, X_n] \xrightarrow{n \rightarrow \infty} \mathbf{P}_\pi[X \in A | \sigma(X_0, X_1, \dots)] = \mathbb{1}_{\{X \in A\}}.$$

Damit folgt $\mathbf{P}_\pi[X \in A] \in \{0, 1\}$. Mithin ist X ergodisch.

Der Birkhoff'sche Ergodensatz liefert also für jedes $x \in E$

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k = x\}} \xrightarrow{n \rightarrow \infty} \pi(\{x\}) \quad \mathbf{P}_\pi\text{-f.s.}$$

In diesem Sinne ist $\pi(\{x\})$ die mittlere Aufenthaltsdauer von X in x . \diamond

Beispiel 20.18. Es seien P und Q W-Maße auf dem Messraum (Ω, \mathcal{A}) , und es seien $(\Omega, \mathcal{A}, P, \tau)$ und $(\Omega, \mathcal{A}, Q, \tau)$ ergodisch. Dann ist $P = Q$ oder $P \perp Q$. Ist nämlich $P \neq Q$, dann existiert f mit $|f| \leq 1$ und $\int f dP \neq \int f dQ$. Nach dem Birkhoff'schen Ergodensatz gilt aber

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ \tau^k \xrightarrow{n \rightarrow \infty} \begin{cases} \int f dP & P\text{-f.s.,} \\ \int f dQ & Q\text{-f.s.} \end{cases}$$

Setzen wir $A := \left\{ \frac{1}{n} \sum_{k=0}^{n-1} f \circ \tau^k \xrightarrow{n \rightarrow \infty} \int f dP \right\}$, so ist $P(A) = 1$ und $Q(A) = 0$. Also ist $P \perp Q$. \diamond

Übung 20.3.1. Sei (Ω, \mathcal{A}) ein Messraum und $\tau : \Omega \rightarrow \Omega$ eine messbare Abbildung.

- (i) Man zeige, dass die Menge $\mathcal{M} := \{\mu \in \mathcal{M}_1(\Omega) : \mu \circ \tau^{-1} = \mu\}$ der unter τ invarianten Maße eine konvexe Menge ist.
- (ii) Ein Element μ aus \mathcal{M} heißt extremal, wenn aus $\mu = \lambda\mu_1 + (1 - \lambda)\mu_2$ für gewisse $\mu_1, \mu_2 \in \mathcal{M}$ und $\lambda \in (0, 1)$ schon $\mu = \mu_1 = \mu_2$ folgt. Man zeige, dass $\mu \in \mathcal{M}$ genau dann extremal ist, wenn τ bezüglich μ ergodisch ist. ♣

Übung 20.3.2. Sei $p = 2, 3, 5, 6, 7, 10, \dots$ quadratfrei (das heißtt, es gibt keine Zahl $r = 2, 3, 4, \dots$, deren Quadrat ein Teiler von p ist) und $q \in \{2, 3, \dots, p - 1\}$. Für jedes $n \in \mathbb{N}$ sei a_n die führende Ziffer der p -adischen Entwicklung von q^n .

Man zeige die folgende Variante des Benford'schen Gesetzes: Für jedes $d \in \{1, \dots, p - 1\}$ gilt

$$\frac{1}{n} \# \{i \leq n : a_i = d\} \xrightarrow{n \rightarrow \infty} \frac{\log(d+1) - \log(d)}{\log(p)}. \quad \clubsuit$$

20.4 Anwendung: Rekurrenz von Irrfahrten

Sei $(X_n)_{n \in \mathbb{N}}$ ein stationärer Prozess mit Werten in \mathbb{R}^d . Setze $S_n := \sum_{k=1}^n X_k$ für jedes $n \in \mathbb{N}_0$. Ferner sei

$$R_n = \#\{S_1, \dots, S_n\}$$

die Anzahl der von S bis zur Zeit n besuchten Punkte (der so genannte *Range*). Außerdem sei $A := \{S_n \neq 0 \text{ für jedes } n \in \mathbb{N}\}$ das „Fluchtereignis“.

Satz 20.19. Es gilt $\lim_{n \rightarrow \infty} \frac{1}{n} R_n = \mathbf{P}[A | \mathcal{I}]$ fast sicher.

Beweis. Wir nehmen an, dass X der kanonische Prozess ist auf $(\Omega, \mathcal{A}, \mathbf{P}) = ((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes \mathbb{N}}, \mathbf{P})$, und dass $\tau : \Omega \rightarrow \Omega$ der Shift ist, also $X_n = X_0 \circ \tau^n$.

Offenbar ist

$$\begin{aligned} R_n &= \#\{k \leq n : S_l \neq S_k \text{ für jedes } l \in \{k+1, \dots, n\}\} \\ &\geq \#\{k \leq n : S_l \neq S_k \text{ für jedes } l > k\} \\ &= \sum_{k=1}^n \mathbb{1}_A \circ \tau^k. \end{aligned}$$

Der Birkhoff'sche Ergodensatz liefert nun

$$\liminf_{n \rightarrow \infty} \frac{1}{n} R_n \geq \mathbf{P}[A | \mathcal{I}] \quad \text{f.s.} \quad (20.5)$$

Für die andere Ungleichung betrachte $A_m = \{S_l \neq 0 \text{ für jedes } l = 1, \dots, m\}$. Dann ist für $n \geq m$

$$\begin{aligned} R_n &\leq m + \#\{k \leq n-m : S_l \neq S_k \text{ für jedes } l \in \{k+1, \dots, n\}\} \\ &\leq m + \#\{k \leq n-m : S_l \neq S_k \text{ für jedes } l \in \{k+1, \dots, k+m\}\} \\ &= m + \sum_{k=1}^{n-m} \mathbb{1}_{A_m} \circ \tau^k. \end{aligned}$$

Der Ergodensatz liefert wieder

$$\limsup_{n \rightarrow \infty} \frac{1}{n} R_n \leq \mathbf{P}[A_m | \mathcal{I}] \quad \text{f.s.} \quad (20.6)$$

Wegen $A_m \downarrow A$ und $\mathbf{P}[A_m | \mathcal{I}] \xrightarrow{n \rightarrow \infty} \mathbf{P}[A | \mathcal{I}]$ fast sicher (nach Satz 8.14(viii)) folgt aus (20.5) und (20.6) die Aussage. \square

Satz 20.20. Sei $X = (X_n)_{n \in \mathbb{N}}$ ein stationärer Prozess mit Werten in \mathbb{Z} und $\mathbf{E}[|X_1|] < \infty$ sowie $\mathbf{E}[X_1 | \mathcal{I}] = 0$ f.s. Sei $S_n = X_1 + \dots + X_n$, $n \in \mathbb{N}$. Dann gilt

$$\mathbf{P}[S_n = 0 \text{ für unendlich viele } n \in \mathbb{N}] = 1.$$

Speziell ist jede Irrfahrt auf \mathbb{Z} mit zentrierten Zuwächsen rekurrent (Satz von Chung-Fuchs [31], vergleiche Satz 17.40).

Beweis. Setze $A = \{S_n \neq 0 \text{ für jedes } n \in \mathbb{N}\}$.

1. Schritt. Wir zeigen $\mathbf{P}[A] = 0$. (Ist X u.i.v., so ist S eine Markovkette, und es folgt hieraus direkt die Rekurrenz von 0. Nur für den allgemeinen Fall stationärer Prozesse X brauchen wir einen weiteren Schritt.) Nach dem Ergodensatz gilt $\frac{1}{n} S_n \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_1 | \mathcal{I}] = 0$ f.s. Es folgt für jedes $m \in \mathbb{N}$

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left(\frac{1}{n} \max_{k=1, \dots, n} |S_k| \right) &= \limsup_{n \rightarrow \infty} \left(\frac{1}{n} \max_{k=m, \dots, n} |S_k| \right) \\ &\leq \max_{k \geq m} \frac{|S_k|}{k} \xrightarrow{m \rightarrow \infty} 0. \end{aligned}$$

Also ist

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \max_{k=1, \dots, n} S_k \right) = \lim_{n \rightarrow \infty} \left(\frac{1}{n} \min_{k=1, \dots, n} S_k \right) = 0.$$

Nun ist $R_n \leq 1 + \left(\max_{k=1, \dots, n} S_k \right) - \left(\min_{k=1, \dots, n} S_k \right)$, also $\frac{1}{n} R_n \xrightarrow{n \rightarrow \infty} 0$. Nach Satz 20.19 ist dann $\mathbf{P}[A] = 0$.

2. Schritt. Setze $\sigma_n := \inf\{m \in \mathbb{N} : S_{m+n} = S_n\}$ und $B_n := \{\sigma_n < \infty\}$ für $n \in \mathbb{N}_0$ und $B := \bigcap_{n=0}^{\infty} B_n$.

Wegen $\{\sigma_0 = \infty\} = A$ ist $\mathbf{P}[\sigma_0 < \infty] = 1$. Stationarität impliziert $\mathbf{P}[\sigma_n < \infty] = 1$ für jedes $n \in \mathbb{N}_0$, also $\mathbf{P}[B] = 1$.

Setze $\tau_0 = 0$ und iterativ $\tau_{n+1} = \tau_n + \sigma_{\tau_n}$ für $n \in \mathbb{N}_0$. Dann ist τ_n der Zeitpunkt der n -ten Rückkehr von S nach 0. Auf B ist $\tau_n < \infty$ für jedes $n \in \mathbb{N}_0$, also

$$\mathbf{P}[S_n = 0 \text{ unendlich oft}] = \mathbf{P}[\tau_n < \infty \text{ für alle } n \in \mathbb{N}] \geq \mathbf{P}[B] = 1. \quad \square$$

Wenn in Satz 20.20 die Zufallsvariablen X_n nicht ganzzahlig sind, kann man nicht hoffen, dass $S_n = 0$ für irgendein $n \in \mathbb{N}$ mit positiver Wahrscheinlichkeit gilt. Andererseits gilt auch hier eine Art Rekurrenzegenschaft, nämlich $S_n/n \xrightarrow{n \rightarrow \infty} 0$ fast sicher nach dem Ergodensatz. Damit ist allerdings noch nicht ausgeschlossen, dass vielleicht $S_n \xrightarrow{n \rightarrow \infty} \infty$ mit positiver Wahrscheinlichkeit gelten könnte, etwa, wenn S_n von der Größenordnung \sqrt{n} wächst. Der nächste Satz zeigt uns, dass der Partiellsummenprozess nur linear schnell nach ∞ gehen kann, wenn die X_n integrierbar sind.

Satz 20.21. Sei $(X_n)_{n \in \mathbb{N}}$ ein reeller ergodischer Prozess und jedes X_n integrierbar. Sei $S_n = X_1 + \dots + X_n$ für $n \in \mathbb{N}_0$. Dann sind die folgenden Aussagen äquivalent.

- (i) $S_n \xrightarrow{n \rightarrow \infty} \infty$ fast sicher.
- (ii) $\mathbf{P}\left[S_n \xrightarrow{n \rightarrow \infty} \infty\right] > 0$.
- (iii) $\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mathbf{E}[X_1] > 0$ fast sicher.

Sind die Zufallsvariablen X_1, X_2, \dots u.i.v. mit $\mathbf{E}[X_1] = 0$ und $\mathbf{P}[X_1 = 0] < 1$, so gilt $\liminf_{n \rightarrow \infty} S_n = -\infty$ und $\limsup_{n \rightarrow \infty} S_n = \infty$ fast sicher.

Beweis. „(i) \iff (ii)“ Offenbar ist $\{S_n \xrightarrow{n \rightarrow \infty} \infty\}$ ein invariantes Ereignis, hat also Wahrscheinlichkeit 0 oder 1.

„(iii) \implies (i)“ Dies ist trivial.

„(i) \implies (iii)“ Die Gleichheit folgt aus dem individuellen Ergodensatz. Es reicht also zu zeigen, dass $\liminf_{n \rightarrow \infty} S_n/n > 0$ fast sicher gilt.

Für $n \in \mathbb{N}_0$ und $\varepsilon > 0$ sei

$$A_n^\varepsilon := \{S_m > S_n + \varepsilon \text{ für alle } m \geq n + 1\}.$$

Sei $S^- := \inf\{S_n : n \in \mathbb{N}_0\}$. Nach Voraussetzung (i) ist $S^- > -\infty$ fast sicher und $\tau := \sup\{n \in \mathbb{N}_0 : S_n = S^-\}$ fast sicher endlich. Es gibt also ein $N \in \mathbb{N}$ mit $\mathbf{P}[\tau < N] \geq \frac{1}{2}$. Es folgt

$$\mathbf{P}\left[\bigcup_{n=0}^{N-1} A_n^0\right] = \mathbf{P}[\tau < N] \geq \frac{1}{2}.$$

Wegen $A_n^\varepsilon \uparrow A_n^0$ für $\varepsilon \downarrow 0$ gibt es ein $\varepsilon > 0$ mit $p := \mathbf{P}[A_0^\varepsilon] \geq \frac{1}{4N} > 0$. Da $(X_n)_{n \in \mathbb{N}}$ ergodisch ist, ist auch $(\mathbb{1}_{A_n^\varepsilon})_{n \in \mathbb{N}_0}$ ergodisch. Nach dem individuellen Ergodensatz gilt daher $\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{A_i^\varepsilon} \xrightarrow{n \rightarrow \infty} p$ fast sicher. Also existiert ein $n_0 = n_0(\omega)$ mit $\sum_{i=0}^{n-1} \mathbb{1}_{A_i^\varepsilon} \geq \frac{pn}{2}$ für alle $n \geq n_0$. Es folgt $S_n \geq \frac{pn\varepsilon}{2}$ für $n \geq n_0$, also $\liminf_{n \rightarrow \infty} S_n/n \geq \frac{pn\varepsilon}{2} > 0$.

Der Zusatz folgt, weil $\liminf S_n$ und $\limsup S_n$ keinen endlichen Wert annehmen können und damit terminal messbar sind, also fast sicher konstant gleich $-\infty$ oder $+\infty$. Nach dem schon Gezeigten ist aber $S_n \xrightarrow{n \rightarrow \infty} \infty$ ausgeschlossen, also gilt $\liminf_{n \rightarrow \infty} S_n = -\infty$. Analog folgt $\limsup_{n \rightarrow \infty} S_n = \infty$. \square

Bemerkung 20.22. Satz 20.21 gilt auch ohne die Integrierbarkeitsbedingung für die X_n . Siehe [101]. \diamond

20.5 Mischung

Ergodizität stellt einen relativ schwachen Begriff für „Unabhängigkeit“ oder „Durchmischung“ dar. Auf dem anderen Ende der Skala steht als stärkster Begriff „u.i.v.“. Hier wollen wir dazwischen liegende Mischungsbegriffe betrachten.

Sei im Folgenden stets $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ein maßerhaltendes dynamisches System und $X_n := X_0 \circ \tau^n$. Wir beginnen mit einer einfachen Betrachtung.

Satz 20.23. $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ist genau dann ergodisch, wenn für alle $A, B \in \mathcal{A}$ gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{P}[A \cap \tau^{-k}(B)] = \mathbf{P}[A] \mathbf{P}[B]. \quad (20.7)$$

Beweis. „ \implies “ Sei $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ergodisch. Setze

$$Y_n := \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\tau^{-k}(B)} = \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_B \circ \tau^k.$$

Nach dem Birkhoff'schen Ergodensatz gilt $Y_n \xrightarrow{n \rightarrow \infty} \mathbf{P}[B]$ fast sicher. Also gilt $Y_n \mathbb{1}_A \xrightarrow{n \rightarrow \infty} \mathbb{1}_A \mathbf{P}[B]$ fast sicher. Majorisierte Konvergenz liefert

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{P}[A \cap \tau^{-k}(B)] = \mathbf{E}[Y_n \mathbb{1}_A] \xrightarrow{n \rightarrow \infty} \mathbf{E}[\mathbb{1}_A \mathbf{P}[B]] = \mathbf{P}[A] \mathbf{P}[B].$$

„ \Leftarrow “ Gelte nun (20.7). Sei $A \in \mathcal{I}$ (invariante σ -Algebra) und $B = A$. Offenbar ist $A \cap \tau^{-k}(A) = A$ für jedes $k \in \mathbb{N}_0$. Also ist nach (20.7)

$$\mathbf{P}[A] = \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{P}[A \cap \tau^{-k}(A)] \xrightarrow{n \rightarrow \infty} \mathbf{P}[A]^2.$$

Mithin ist $\mathbf{P}[A] \in \{0, 1\}$, also \mathcal{I} trivial und damit τ ergodisch. \square

Wir betrachten jetzt folgende Verschärfung von (20.7).

Definition 20.24. Ein maßerhaltendes dynamisches System $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ heißt **mischend**, falls

$$\lim_{n \rightarrow \infty} \mathbf{P}[A \cap \tau^{-n}(B)] = \mathbf{P}[A] \mathbf{P}[B] \quad \text{für alle } A, B \in \mathcal{A}. \quad (20.8)$$

Bemerkung 20.25. Gelegentlich wird die Mischungseigenschaft (20.8) auch als **stark mischend** bezeichnet. Im Gegensatz dazu heißt $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ **schwach mischend**, falls

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mathbf{P}[A \cap \tau^{-i}(B)] - \mathbf{P}[A] \mathbf{P}[B]| = 0 \quad \text{für alle } A, B \in \mathcal{A}. \quad \diamond$$

„Stark mischend“ impliziert „schwach mischend“ (siehe Übung 20.5.1). Andererseits gibt es schwach mischende Systeme, die nicht stark mischend sind (siehe [89]).

Beispiel 20.26. Sei $I = \mathbb{N}_0$ oder $I = \mathbb{Z}$ und $(X_n)_{n \in I}$ eine u.i.v. Folge mit Werten im Messraum (E, \mathcal{E}) , also τ der Shift auf dem Produktraum $\Omega = E^I, \mathbf{P} = (\mathbf{P}_{X_0})^{\otimes I}$. Seien $A, B \in \mathcal{E}^{\otimes I}$. Zu jedem $\varepsilon > 0$ gibt es Ereignisse A^ε und B^ε , die nur von endlich vielen Koordinaten abhängen und mit $\mathbf{P}[A \Delta A^\varepsilon] < \varepsilon$ und $\mathbf{P}[B \Delta B^\varepsilon] < \varepsilon$. Offenbar gilt dann auch $\mathbf{P}[\tau^{-n}(A \Delta A^\varepsilon)] < \varepsilon$ und $\mathbf{P}[\tau^{-n}(B \Delta B^\varepsilon)] < \varepsilon$ für jedes $n \in \mathbb{Z}$. Für hinreichend großes $|n|$ hängen A^ε und $\tau^{-n}(B^\varepsilon)$ von unterschiedlichen Koordinaten ab, sind also unabhängig. Es folgt

$$\begin{aligned} \limsup_{|n| \rightarrow \infty} |\mathbf{P}[A \cap \tau^{-n}(B)] - \mathbf{P}[A] \mathbf{P}[B]| \\ \leq \limsup_{|n| \rightarrow \infty} |\mathbf{P}[A^\varepsilon \cap \tau^{-n}(B^\varepsilon)] - \mathbf{P}[A^\varepsilon] \mathbf{P}[B^\varepsilon]| + 4\varepsilon = 4\varepsilon. \end{aligned}$$

Also ist τ mischend. Ist speziell $A = B \in \mathcal{I}$, so erhalten wir das 0-1 Gesetz für invariante Ereignisse: $\mathbf{P}[A] \in \{0, 1\}$. \diamond

Bemerkung 20.27. Offenbar folgt (20.7) aus (20.8), also impliziert „mischend“ schon „ergodisch“. Die umgekehrte Implikation gilt nicht. \diamond

Beispiel 20.28. Sei $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$ und $\mathbf{P} = \lambda$ das Lebesgue-Maß auf $([0, 1], \mathcal{B}([0, 1]))$ sowie für $r \in [0, 1)$

$$\tau_r(x) = x + r - \lfloor x + r \rfloor = x + r \pmod{1}.$$

Ist r irrational, so ist τ_r ergodisch (Beispiel 20.9). Jedoch ist τ_r nicht mischend:

Da r irrational ist, existiert eine Folge $k_n \uparrow \infty$, sodass

$$\tau_r^{k_n}(0) \in \left(\frac{1}{4}, \frac{3}{4} \right) \quad \text{für } n \in \mathbb{N}.$$

Für $A = [0, \frac{1}{4}]$ gilt dann $A \cap \tau_r^{-k_n}(A) = \emptyset$. Also ist

$$\liminf_{n \rightarrow \infty} \mathbf{P}[A \cap \tau_r^{-n}(A)] = 0 \neq \frac{1}{16} = \mathbf{P}[A]^2. \quad \diamond$$

Satz 20.29. Sei X eine irreduzible, positiv rekurrente Markovkette auf dem abzählbaren Raum E mit invariante Verteilung π sowie $\mathbf{P}_\pi = \sum_{x \in E} \pi(x) \mathbf{P}_x$. Dann gelten

- (i) X ist ergodisch (auf $(\Omega, \mathcal{A}, \mathbf{P}_\pi)$).
- (ii) X ist genau dann mischend, wenn X aperiodisch ist.

Beweis. (i) Dies haben wir schon in Beispiel 20.17 gezeigt.

(ii) Da X irreduzibel ist, ist $\pi(\{x\}) > 0$ für jedes $x \in E$ nach Satz 17.51.

„ \Rightarrow “ Sei X periodisch mit Periode $d \geq 2$. Ist $n \in \mathbb{N}$ kein Vielfaches von d , so ist $p^n(x, x) = 0$. Mithin gilt für $A = B = \{X_0 = x\}$

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathbf{P}_\pi[X_0 = x, X_n = x] &= \liminf_{n \rightarrow \infty} \pi(\{x\}) p^n(x, x) \\ &= 0 \neq \pi(\{x\})^2 = \mathbf{P}_\pi[X_0 = x]^2. \end{aligned}$$

Also ist X nicht mischend.

„ \Leftarrow “ Sei X aperiodisch. Zur Vereinfachung der Notation können wir annehmen, dass X der kanonische Prozess auf $E^{\mathbb{N}_0}$ ist. Seien $A, B \subset \Omega = E^{\mathbb{N}_0}$ messbar. Zu jedem $\varepsilon > 0$ existiert ein $N \in \mathbb{N}$ und $\tilde{A}^\varepsilon \in E^{\{0, \dots, N\}}$, sodass, mit $A^\varepsilon = \tilde{A}^\varepsilon \times E^{\{N+1, N+2, \dots\}}$ gilt, dass $\mathbf{P}[A \Delta A^\varepsilon] < \varepsilon$. Nach der Markoveigenschaft ist für $n \geq N$

$$\begin{aligned} \mathbf{P}_\pi[A^\varepsilon \cap \tau^{-n}(B)] &= \mathbf{P}_\pi[(X_0, \dots, X_N) \in \tilde{A}^\varepsilon, (X_n, X_{n+1}, \dots) \in B] \\ &= \sum_{x, y \in E} \mathbf{E}_\pi[\mathbb{1}_{A^\varepsilon} \mathbb{1}_{\{X_N=x\}} \mathbb{1}_{\{X_n=y\}} (X_n, X_{n+1}, \dots) \in B] \\ &= \sum_{x, y \in E} \mathbf{E}_\pi[\mathbb{1}_{A^\varepsilon} \mathbb{1}_{\{X_N=x\}}] p^{n-N}(x, y) \mathbf{P}_y[B]. \end{aligned}$$

Nach Satz 18.13 gilt $p^{n-N}(x, y) \xrightarrow{n \rightarrow \infty} \pi(\{y\})$ für alle $x, y \in E$. (Für periodisches X ist dies falsch.) Majorisierte Konvergenz ergibt daher

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbf{P}_\pi [A^\varepsilon \cap \tau^{-n}(B)] &= \sum_{x, y \in E} \mathbf{E}_\pi [\mathbb{1}_{A^\varepsilon} \mathbb{1}_{\{X_N=x\}}] \pi(\{y\}) \mathbf{P}_y[B] \\ &= \mathbf{P}_\pi [A^\varepsilon] \mathbf{P}_\pi[B].\end{aligned}$$

Wegen $|\mathbf{P}_\pi [A^\varepsilon \cap \tau^{-n}(B)] - \mathbf{P}[A \cap \tau^{-n}(B)]| < \varepsilon$ folgt die Aussage mit $\varepsilon \rightarrow 0$. \square

Übung 20.5.1. Man zeige: „stark mischend“ impliziert „schwach mischend“ und dies impliziert wiederum „ergodisch“. Man gebe ein Beispiel an für ein maßerhaltendes dynamisches System, das ergodisch aber nicht schwach mischend ist. ♣

20.6 Entropie

Die Entropie einer Wahrscheinlichkeitsverteilung (siehe Definition 5.25) misst, wie viel Zufall in der Verteilung steckt. So ist die Entropie einer Deltamasse Null und die Entropie der Gleichverteilung auf n Punkten ist maximal und hat den Wert $\log(n)$ (siehe Übung 5.3.3). Es liegt nahe, den Zufall eines dynamischen Systems mit Hilfe der Entropie zu quantifizieren.

Wir betrachten zunächst nur den Fall eines einfachen Shifts: Sei $\Omega = E^{\mathbb{N}_0}$, wobei E eine endliche Menge ist, mit der Produkt- σ -Algebra $\mathcal{A} = (2^E)^{\otimes \mathbb{N}_0}$. Sei τ der Shift auf Ω und \mathbf{P} ein invariantes Wahrscheinlichkeitsmaß. Für $n \in \mathbb{N}$ sei P_n die Projektion von \mathbf{P} auf $E^n = E^{\{0, \dots, n-1\}}$, d.h.

$$P_n(\{(e_0, \dots, e_{n-1})\}) = \mathbf{P}[\{e_0\} \times \dots \times \{e_{n-1}\} \times E^{\{n, n+1, \dots\}}].$$

Wir bezeichnen mit h_n die Entropie von P_n . Nach Übung 5.3.4 ist die Entropie subadditiv, das heißt, es gilt

$$h_{m+n} \leq h_m + h_n \quad \text{für } m, n \in \mathbb{N}.$$

Folglich existiert der Grenzwert (siehe Übung 20.6.2)

$$h := h(\mathbf{P}, \tau) := \lim_{n \rightarrow \infty} \frac{1}{n} h_n = \inf_{n \in \mathbb{N}} \frac{1}{n} h_n.$$

Definition 20.30 (Entropie des einfachen Shifts). $h(\mathbf{P}, \tau)$ heißt **Entropie des dynamischen Systems** $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$.

Die Entropie, aufgefasst als Maß für die Information, misst also den asymptotischen Informationsgewinn durch Beobachtung von X_n , wenn X_0, \dots, X_{n-1} bereits bekannt sind. Am meisten lernen wir durch die Beobachtung von X_n , wenn

die Zufallsvariablen unabhängig sind, am wenigsten, wenn X_n nach Kenntnis von X_0, \dots, X_{n-1} bereits deterministisch vorhergesagt werden kann. Zwischen diesen Extremen liegt der Fall, wo (X_n) eine Markovkette ist. Wir berechnen für jedes dieser drei Beispiele die Entropie.

Beispiel 20.31. Für das dynamische System, das durch u.i.v. Zufallsvariablen $(X_n)_{n \in \mathbb{N}_0}$ mit Verteilung π gegeben ist, gilt

$$h = H(\pi) = - \sum_{e \in E} \pi(\{e\}) \log(\pi(\{e\})). \quad \diamond$$

Beispiel 20.32 (Markovkette). Sei $(X_n)_{n \in \mathbb{N}_0}$ eine Markovkette mit Übergangsmatrix P und stationärer Verteilung π , und sei $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ das entsprechende dynamische System. Für $x = (x_0, \dots, x_{n-1})$ und $0 \leq k < n-1$ sei

$$p(k, x) = \pi(\{x_k\}) P(x_k, x_{k+1}) \cdots P(x_{n-2}, x_{n-1}).$$

Die Entropie von P_n berechnen wir unter Verwendung der Stationarität von π in der dritten Zeile:

$$\begin{aligned} H(P_n) &= - \sum_{x_0, \dots, x_{n-1} \in E} p(0, x) \log(p(0, x)) \\ &= - \sum_{x_0, \dots, x_{n-1} \in E} p(0, x) \left[\log(\pi(\{x_0\})) + \sum_{k=0}^{n-2} \log(P(x_k, x_{k+1})) \right] \\ &= H(\pi) - \sum_{k=0}^{n-2} \sum_{x_k, \dots, x_{n-1}} p(k, x) \log(P(x_k, x_{k+1})) \\ &= H(\pi) - (n-1) \sum_{x_0, x_1 \in E} \pi(\{x_0\}) P(x_0, x_1) \log(P(x_0, x_1)). \end{aligned}$$

Es folgt, dass die Entropie des dynamischen Systems gegeben ist durch

$$h(\mathbf{P}, \tau) = - \sum_{x, y \in E} \pi(\{x\}) P(x, y) \log(P(x, y)). \quad (20.9)$$

◊

Beispiel 20.33 (Ganzzahlige Rotation). Wir betrachten das Beispiel der Rotation aus Beispiel 20.8. Sei also $n \in \mathbb{N} \setminus \{1\}$, $E = \mathbb{Z}/(n)$ und \mathbf{P} die Gleichverteilung auf Ω . Sei $r \in \{1, \dots, n\}$ und

$$\tau : \Omega \rightarrow \Omega, \quad x \mapsto x + r \pmod{n}.$$

Offenbar ist $\tau^{(n)}$ die Identität, also ist $h_n = h_{2n} = h_{3n} = \dots$ und damit $h(\mathbf{P}, \tau) = 0$.

◊

Wir kommen nun zur Situation eines allgemeinen dynamischen Systems. Sei \mathcal{P} eine endliche messbare Partition von Ω , also $\mathcal{P} = \{A_1, \dots, A_k\}$ für gewisse paarweise disjunkte nichtleere $A_1, \dots, A_k \in \mathcal{A}$ mit $\Omega = A_1 \cup \dots \cup A_k$. Für $n \in \mathbb{N}$ sei \mathcal{P}_n die Partition von Ω , die von den Mengen $\bigcap_{l=0}^{n-1} \tau^{-l}(A_{i_l})$, $i_1, \dots, i_n \in \{1, \dots, k\}$, gebildet wird. Wir definieren

$$h_n(\mathbf{P}, \tau; \mathcal{P}) = - \sum_{A \in \mathcal{P}_n} \mathbf{P}[A] \log(\mathbf{P}[A]).$$

Ähnlich wie für den endlichen Fall erhalten wir Subadditivität der (h_n) und damit die Existenz von

$$h(\mathbf{P}, \tau; \mathcal{P}) := \lim_{n \rightarrow \infty} \frac{1}{n} h_n(\mathbf{P}, \tau; \mathcal{P}) = \inf_{n \in \mathbb{N}} \frac{1}{n} h_n(\mathbf{P}, \tau; \mathcal{P}).$$

Definition 20.34 (Kolmogorov-Sinai Entropie). Die **Entropie** eines allgemeinen maßerhaltenden dynamischen Systems $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ ist

$$h(\mathbf{P}, \tau) = \sup_{\mathcal{P}} h(\mathbf{P}, \tau; \mathcal{P}),$$

wobei das Supremum über alle endlichen messbaren Partitionen von Ω gebildet wird.

Satz 20.35 (Kolmogorov-Sinai). Sei \mathcal{P} ein Erzeuger von \mathcal{A} , das heißt $\mathcal{A} = \sigma(\bigcup_{n \in \mathbb{N}_0} \tau^{-n}(\mathcal{P}))$. Dann gilt

$$h(\mathbf{P}, \tau) = h(\mathbf{P}, \tau; \mathcal{P}).$$

Beweis. Siehe z.B. [95, Theorem 3.2.18], [177, Theorem 4.17] oder [165]. □

Dieser Satz zeigt, dass die in Definition 20.30 eingeführte Entropie für einfache Shifts mit der Entropie aus Definition 20.34 übereinstimmt - man muss lediglich $\mathcal{P} = \{\{e\} \times E^{\mathbb{N}}, e \in E\}$ wählen. Dies ist offenbar ein Erzeuger der Produkt- σ -Algebra auf $\Omega = E^{\mathbb{N}_0}$.

Beispiel 20.36 (Rotation). Wir kommen nun zu der Rotation aus Beispiel 20.9 zurück. Sei $\Omega = [0, 1)$, $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathbf{P} = \lambda$ das Lebesgue-Maß, $r \in (0, 1)$ und $\tau_r(x) = x + r \pmod{1}$.

Sei zunächst r rational. Wir wählen dann eine beliebige endliche messbare Partition \mathcal{P} von Ω . Sei $n \in \mathbb{N}$ so gewählt, dass $rn \in \mathbb{N}_0$ ist. Wie in Beispiel 20.33 erhalten wir $h_n(\mathbf{P}, \tau_r; \mathcal{P}) = h_{kn}(\mathbf{P}, \tau_r; \mathcal{P})$ für jedes $k \in \mathbb{N}$, also $h(\mathbf{P}, \tau_r, \mathcal{P}) = 0$. Damit ist dann auch $h(\mathbf{P}, \tau_r) = 0$.

Sei nun r irrational. Wir wählen $\mathcal{P} = \{[0, 1/2), [1/2, 1]\}$. Da r irrational ist, überlegt man sich leicht, dass \mathcal{A} von $\bigcup_{n \in \mathbb{N}_0} \tau_r^{-n}(\mathcal{P})$ erzeugt wird. Es gilt also

$h(\mathbf{P}, \tau_r) = h(\mathbf{P}, \tau_r, \mathcal{P})$. Um dies auszurechnen, bestimmen wir die zunächst Anzahl der Partitionselemente $\#\mathcal{P}_n$. Betrachte dazu die Abbildung

$$\begin{aligned}\phi_n : [0, 1] &\rightarrow \{0, 1\}^n \\ x &\mapsto (\mathbb{1}_{[1/2,1]}(x), \mathbb{1}_{[1/2,1]}(\tau_r(x)), \dots, \mathbb{1}_{[1/2,1]}(\tau_r^{n-1}(x)))\end{aligned}$$

Offenbar gilt $\#\phi_n([0, 1]) = \#\mathcal{P}_n$. Mit wachsendem $x \in [0, 1]$ wechselt jede Koordinate $\mathbb{1}_{[1/2,1]}(\tau_r^k(x))$, $k = 1, \dots, n - 1$ genau zweimal den Wert. Lediglich $\mathbb{1}_{[1/2,1]}(x)$ wechselt nur einmal den Wert. Insgesamt gilt also $\#\phi_n([0, 1]) \leq 2n$. Die Entropie einer Verteilung auf N Punkten wird durch die Gleichverteilung maximiert und ist $\log(N)$. Folglich $h_n(\mathbf{P}, \tau_r; \mathcal{P}) \leq \log(2n)$. Es folgt

$$h(\mathbf{P}, \tau_r) = h(\mathbf{P}, \tau_r; \mathcal{P}) = 0.$$

◇

Übung 20.6.1. Sei $\Omega = [0, 1]$ und $\tau : x \mapsto 2x \pmod{1}$ sowie \mathbf{P} das Lebesguemaß auf Ω . Man bestimme $h(\mathbf{P}, \tau)$.

Übung 20.6.2. Sei $(a_n)_{n \in \mathbb{N}}$ eine Folge nichtnegativer Zahlen. Die Folge heißt **subadditiv**, falls $a_{m+n} \leq a_m + a_n$ für alle $m, n \in \mathbb{N}$ gilt. Man zeige, dass der Grenzwert $\lim_{n \rightarrow \infty} a_n/n$ existiert, und dass

$$\lim_{n \rightarrow \infty} \frac{1}{n} a_n = \inf_{n \in \mathbb{N}} \frac{1}{n} a_n.$$

♣

21 Die Brown'sche Bewegung

In Beispiel 14.45 hatten wir einen (kanonischen) Prozess $(X_t)_{t \in [0, \infty)}$ hergestellt mit unabhängigen, stationären, normalverteilten Zuwächsen. Ein solcher Prozess kann beispielsweise als Modell eines Flimmerteilchens in einer Suspension dienen oder als Grundlage für Aktienkursmodelle. Jetzt sind wir nicht nur an den Eigenschaften von X zu einem oder mehreren festen Zeitpunkten interessiert, sondern auch an Eigenschaften, die den ganzen Pfad $t \mapsto X_t$ betreffen, beispielsweise am Funktional $F(X) := \sup_{t \in [0,1]} X_t$. Ist aber $F(X)$ überhaupt eine Zufallsvariable?

Wir werden in diesem Kapitel Stetigkeitseigenschaften von Pfaden stochastischer Prozesse untersuchen, die die Messbarkeit von interessanten Funktionalen sichern. Danach konstruieren wir eine Version von X , die stetige Pfade hat, die so genannte *Brown'sche Bewegung*. Ohne Übertreibung kann man sagen, dass dies das zentrale Objekt der Wahrscheinlichkeitstheorie ist.

Zur weiteren Lektüre eignen sich beispielsweise [128, 161, 94, 154].

21.1 Stetige Modifikationen

Die Pfade eines kanonischen Prozesses sind natürlich nicht *per se* stetig, da ja jede Abbildung als Pfad auftaucht. Es wird also wichtig sein zu entscheiden, welche Pfade zumindest \mathbf{P} -fast sicher keine Rolle spielen.

Definition 21.1 (Modifikation / ununterscheidbare Prozesse). Seien X und Y stochastische Prozesse auf $(\Omega, \mathcal{A}, \mathbf{P})$ mit Zeitbereich I und Zustandsraum E .

(i) X und Y heißen **Modifikationen** oder **Versionen** voneinander, falls für jedes $t \in I$ gilt

$$X_t = Y_t \quad \mathbf{P}\text{-fast sicher}.$$

(ii) X und Y heißen **ununterscheidbar**, falls es ein $N \in \mathcal{A}$ gibt mit $\mathbf{P}[N] = 0$ und

$$\{X_t \neq Y_t\} \subset N \quad \text{für jedes } t \in I.$$

Offenbar ist „ununterscheidbar“ stärker als „Modifikation“. Unter gewissen Stetigkeitsannahmen an die Prozesse fallen die Begriffe allerdings zusammen.

Definition 21.2. Seien (E, d) und (E', d') metrische Räume und $\gamma \in (0, 1]$. Eine Abbildung $\varphi : E \rightarrow E'$ heißt im Punkte $r \in E$ Hölder-stetig der Ordnung γ (kurz: Hölder- γ -stetig), falls es ein $\varepsilon > 0$ und ein $C < \infty$ gibt, sodass für jedes $s \in E$ mit $d(s, r) < \varepsilon$ gilt

$$d'(\varphi(r), \varphi(s)) \leq C d(r, s)^\gamma. \quad (21.1)$$

φ heißt lokal Hölder-stetig der Ordnung γ , wenn es zu jedem $t \in E$ ein $\varepsilon > 0$ und ein $C = C(t, \varepsilon) > 0$ gibt, sodass für alle $s, r \in E$ mit $d(s, t) < \varepsilon$ und $d(r, t) < \varepsilon$ die Ungleichung (21.1) gilt. Schließlich heißt φ Hölder-stetig der Ordnung γ , falls es ein C gibt, sodass (21.1) für alle $s, r \in E$ gilt.

Im Falle $\gamma = 1$ ist die Hölder-Stetigkeit natürlich gerade die Lipschitz-Stetigkeit (siehe Definition 13.8). Zudem ist für $E = \mathbb{R}$ und $\gamma > 1$ jede lokal Hölder- γ -stetige Funktion konstant. Offenbar ist eine lokal Hölder- γ -stetige Abbildung in jedem Punkte Hölder- γ -stetig. Andererseits braucht für eine in einem Punkte t Hölder- γ -stetige Abbildung φ keine Umgebung von t zu existieren, in der φ stetig oder gar Hölder- γ -stetig wäre.

Wir sammeln ein paar einfache Eigenschaften der Hölder-Stetigkeit.

Lemma 21.3. Seien $I \subset \mathbb{R}$ und $f : I \rightarrow \mathbb{R}$ lokal Hölder-stetig von der Ordnung $\gamma \in (0, 1]$. Dann gelten:

- (i) f ist lokal Hölder-stetig von der Ordnung γ' für jedes $\gamma' \in (0, \gamma)$.
- (ii) Ist I kompakt, dann ist f auch Hölder-stetig.
- (iii) Ist I ein beschränktes Intervall mit Länge $T > 0$, und gilt für ein $\varepsilon > 0$ und ein $C(\varepsilon) < \infty$ sowie für alle $s, t \in I$ mit $|t - s| \leq \varepsilon$

$$|f(t) - f(s)| \leq C(\varepsilon) |t - s|^\gamma,$$

so ist f Hölder-stetig der Ordnung γ mit Konstante $C := C(\varepsilon) [T/\varepsilon]^{1-\gamma}$.

Beweis. (i) Klar, weil $|t - s|^\gamma \leq |t - s|^{\gamma'}$ für alle $s, t \in I$ mit $|t - s| \leq 1$.

(ii) Für $t \in I$ und $\varepsilon > 0$ sei $U_\varepsilon(t) := \{s \in I : |s - t| < \varepsilon\}$. Für jedes $t \in I$ seien $\varepsilon(t) > 0$ und $C(t) < \infty$ so gewählt, dass

$$|f(r) - f(s)| \leq C(t) \cdot |r - s|^\gamma \quad \text{für alle } r, s \in U_t := U_{\varepsilon(t)}(t).$$

Zu der offenen Überdeckung $\mathfrak{U} := \{U_t, t \in I\}$ von I gibt es eine endliche Teilüberdeckung $\mathfrak{U}' = \{U_{t_1}, \dots, U_{t_n}\}$. Sei $\varrho > 0$ eine Lebesgue'sche Zahl der Überdeckung \mathfrak{U}' , das heißt, $\varrho > 0$ ist so gewählt, dass für jedes $t \in I$ ein $U \in \mathfrak{U}$ existiert mit $U_\varrho(t) \subset U$. Setze

$$\overline{C} := \max \{C(t_1), \dots, C(t_n), 2\|f\|_\infty \varrho^{-\gamma}\}.$$

Für $s, t \in I$ mit $|t - s| < \varrho$ gibt es ein $i \in \{1, \dots, n\}$ mit $s, t \in U_{t_i}$. Nach Voraussetzung ist $|f(t) - f(s)| \leq C(t_i) |t - s|^\gamma \leq \bar{C} |t - s|^\gamma$. Seien nun $s, t \in I$ mit $|s - t| \geq \varrho$. Dann ist

$$|f(t) - f(s)| \leq 2 \|f\|_\infty \left(\frac{|t - s|}{\varrho} \right)^\gamma \leq \bar{C} |t - s|^\gamma.$$

Also ist f Hölder-stetig von der Ordnung γ mit Konstante \bar{C} .

(iii) Sei $n = \lceil \frac{T}{\varepsilon} \rceil$. Für $s, t \in I$ gilt nach Voraussetzung $\frac{|t-s|}{n} \leq \varepsilon$ und daher

$$\begin{aligned} |f(t) - f(s)| &\leq \sum_{k=1}^n \left| f\left(s + (t-s)\frac{k}{n}\right) - f\left(s + (t-s)\frac{k-1}{n}\right) \right| \\ &\leq C(\varepsilon) n^{1-\gamma} |t - s|^\gamma = C |t - s|^\gamma. \end{aligned} \quad \square$$

Definition 21.4 (Pfadeigenschaften). Sei $I \subset \mathbb{R}$ und $X = (X_t, t \in I)$ ein stochastischer Prozess auf einem W-Raum $(\Omega, \mathcal{A}, \mathbf{P})$ mit Werten in einem metrischen Raum (E, d) sowie $\gamma \in (0, 1]$. Für jedes $\omega \in \Omega$ nennen wir die Abbildung $I \rightarrow E$, $t \mapsto X_t(\omega)$ einen **Pfad** von X .

Wir sagen, dass X fast sicher stetige Pfade hat, oder kurz, dass X f.s. stetig ist, falls für fast jedes $\omega \in \Omega$ der Pfad $t \mapsto X_t(\omega)$ stetig ist. Analog definieren wir lokal Hölder- γ -stetige Pfade und so weiter.

Lemma 21.5. Seien X und Y Modifikationen voneinander. Es gelte eine der Bedingungen

- (i) I ist abzählbar.
- (ii) $I \subset \mathbb{R}$ ist ein (möglicherweise unbeschränktes) Intervall und X und Y sind fast sicher rechtsstetig.

Dann sind X und Y ununterscheidbar.

Beweis. Setze $N_t := \{X_t \neq Y_t\}$ für $t \in I$ und $\bar{N} = \bigcup_{t \in I} N_t$. Nach Voraussetzung gilt $\mathbf{P}[N_t] = 0$ für jedes $t \in I$. Zu zeigen ist jeweils: Es existiert $N \in \mathcal{A}$ mit $\bar{N} \subset N$ und $\mathbf{P}[N] = 0$.

- (i) Ist I abzählbar, so ist $N := \bar{N}$ messbar und $\mathbf{P}[N] \leq \sum_{t \in I} \mathbf{P}[N_t] = 0$.
- (ii) Sei nun $I \subset \mathbb{R}$ ein Intervall, und seien X und Y fast sicher rechtsstetig. Setze

$$\bar{R} := \{X \text{ und } Y \text{ sind rechtsstetig}\}$$

und wähle $R \in \mathcal{A}$ mit $R \subset \bar{R}$ und $\mathbf{P}[R] = 1$. Setze

$$\tilde{I} := \begin{cases} \mathbb{Q} \cap I, & \text{falls } I \text{ rechtsseitig offen ist,} \\ (\mathbb{Q} \cap I) \cup \max I, & \text{falls } I \text{ rechtsseitig abgeschlossen ist,} \end{cases}$$

und $\tilde{N} := \bigcup_{r \in \tilde{I}} N_r$. Nach (i) gilt $\mathbf{P}[\tilde{N}] = 0$. Weiter gilt für jedes $t \in I$

$$N_t \cap R \subset \bigcup_{r \geq t, r \in \tilde{I}} (N_r \cap R) \subset \tilde{N}.$$

Also gilt

$$\bar{N} \subset R^c \cup \bigcup_{t \in I} N_t \subset R^c \cup \tilde{N} =: N,$$

und damit $\mathbf{P}[N] \leq \mathbf{P}[R^c] + \mathbf{P}[\tilde{N}] = 0$. \square

Wir kommen zum Hauptsatz dieses Abschnitts.

Satz 21.6 (Kolmogorov-Chentsov). Sei $X = (X_t, t \in [0, \infty))$ ein reellwertiger Prozess. Für jedes $T > 0$ gebe es Zahlen $\alpha, \beta, C > 0$ mit

$$\mathbf{E}[|X_t - X_s|^\alpha] \leq C|t - s|^{1+\beta} \quad \text{für alle } s, t \in [0, T]. \quad (21.2)$$

Dann gelten:

- (i) Es existiert eine Modifikation $\tilde{X} = (\tilde{X}_t, t \in [0, \infty))$ von X , die lokal Hölder-stetige Pfade hat von jeder Ordnung $\gamma \in (0, \frac{\beta}{\alpha})$.
- (ii) Sei $\gamma \in (0, \frac{\beta}{\alpha})$. Zu jedem $\varepsilon > 0$ und $T < \infty$ existiert eine Zahl $K < \infty$, die nur von $\varepsilon, T, \alpha, \beta, C, \gamma$ abhängt, mit

$$\mathbf{P}\left[|\tilde{X}_t - \tilde{X}_s| \leq K|t - s|^\gamma, s, t \in [0, T]\right] \geq 1 - \varepsilon. \quad (21.3)$$

Beweis. (i) Es reicht, für jedes $T > 0$ zu zeigen, dass X auf $[0, T]$ Modifikation X^T besitzt, die Hölder-stetig der Ordnung γ ist für jedes $\gamma \in (0, \beta/\alpha)$. Für $S, T > 0$ sind dann nach Lemma 21.5 die Prozesse X^S und X^T ununterscheidbar auf $[0, S \wedge T]$, also ist

$$\Omega_{S,T} := \left\{ \text{es gibt ein } t \in [0, S \wedge T] \text{ mit } X_t^T \neq X_t^S \right\}$$

eine Nullmenge, und damit ist auch $\Omega_\infty := \bigcup_{S,T \in \mathbb{N}} \Omega_{S,T}$ eine Nullmenge. Definieren

wir $\tilde{X}_t(\omega) := X_t^t(\omega)$, $t \geq 0$, für $\omega \in \Omega \setminus \Omega_\infty$, so ist \tilde{X} eine lokal Hölder-stetige Modifikation von X auf $[0, \infty)$.

Ohne Beschränkung der Allgemeinheit sei $T = 1$. Wir zeigen, dass X eine auf $[0, 1]$ stetige Modifikation besitzt. Die Chebyshev'sche Ungleichung liefert für $\varepsilon > 0$

$$\mathbf{P}[|X_t - X_s| \geq \varepsilon] \leq C\varepsilon^{-\alpha} |t - s|^{1+\beta}, \quad (21.4)$$

also

$$X_s \xrightarrow{s \rightarrow t} X_t \quad \text{stochastisch.} \quad (21.5)$$

Die Idee ist, zunächst \tilde{X} auf den binär rationalen Zahlen zu konstruieren und dann stetig auf $[0, 1]$ fortzusetzen. Dafür wird (21.5) gebraucht. Speziell ist für $\gamma > 0$ sowie $n \in \mathbb{N}$ und $k \in \{1, \dots, 2^n\}$

$$\mathbf{P} [|X_{k2^{-n}} - X_{(k-1)2^{-n}}| \geq 2^{-\gamma n}] \leq C 2^{-n(1+\beta-\alpha\gamma)}.$$

Wir setzen

$$A_n = A_n(\gamma) := \{ \max \{|X_{k2^{-n}} - X_{(k-1)2^{-n}}|, k \in \{1, \dots, 2^n\}\} \geq 2^{-\gamma n} \},$$

sowie

$$B_n := \bigcup_{m=n}^{\infty} A_m \quad \text{und} \quad N := \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} B_n.$$

Es folgt dann für jedes $n \in \mathbb{N}$

$$\mathbf{P}[A_n] \leq \sum_{k=1}^{2^n} \mathbf{P} [|X_{k2^{-n}} - X_{(k-1)2^{-n}}| \geq 2^{-\gamma n}] \leq C 2^{-n(\beta-\alpha\gamma)}.$$

Wir wählen jetzt ein $\gamma \in (0, \beta/\alpha)$ und erhalten

$$\mathbf{P}[B_n] \leq \sum_{m=n}^{\infty} \mathbf{P}[A_m] \leq C \frac{2^{-(\beta-\alpha\gamma)n}}{1 - 2^{\alpha\gamma-\beta}} \xrightarrow{n \rightarrow \infty} 0, \quad (21.6)$$

also $\mathbf{P}[N] = 0$. Sei nun $\omega \in \Omega \setminus N$ fest und $n_0 = n_0(\omega)$ so, dass $\omega \notin \bigcup_{n=n_0}^{\infty} A_n$.

Also gilt

$$|X_{k2^{-n}}(\omega) - X_{(k-1)2^{-n}}(\omega)| < 2^{-\gamma n} \quad \text{für } k \in \{1, \dots, 2^n\}, n \geq n_0. \quad (21.7)$$

Wir definieren die Mengen endlicher dyadischer Zahlen

$$D_m = \{k2^{-m}, k = 0, \dots, 2^m\} \quad \text{und} \quad D = \bigcup_{m \in \mathbb{N}} D_m.$$

Jedes $t \in D_m$ besitzt eine eindeutige Binärdarstellung

$$t = \sum_{i=0}^m b_i(t) 2^{-i} \quad \text{für gewisse } b_i(t) \in \{0, 1\}, i = 0, \dots, m.$$

Seien $m \geq n \geq n_0$ sowie $s, t \in D_m, s \leq t$ mit $|s - t| \leq 2^{-n}$. Setze $u := \max(D_n \cap [0, s])$. Dann ist

$$u \leq s < u + 2^{-n} \quad \text{und} \quad u \leq t < u + 2^{1-n}.$$

Es gilt also $b_i(t - u) = b_i(s - u) = 0$ für $i < n$. Setze

$$t_l = u + \sum_{i=n}^l b_i(t - u) 2^{-i} \quad \text{für } l = n - 1, \dots, m.$$

Dann ist $t_{n-1} = u$, $t_m = t$, $t_l \in D_l$ für jedes l und

$$t_l - t_{l-1} \leq 2^{-l} \quad \text{für } l = n, \dots, m.$$

Also ist nach (21.7)

$$|X_t(\omega) - X_u(\omega)| \leq \sum_{l=n}^m |X_{t_l}(\omega) - X_{t_{l-1}}(\omega)| \leq \sum_{l=n}^m 2^{-\gamma l} \leq \frac{2^{-\gamma n}}{1 - 2^{-\gamma}}.$$

Analog folgt $|X_s(\omega) - X_u(\omega)| \leq 2^{-\gamma n}(1 - 2^{-\gamma})^{-1}$, also nach der Dreiecksungleichung

$$|X_t(\omega) - X_s(\omega)| \leq 2 \frac{2^{-\gamma n}}{1 - 2^{-\gamma}}. \quad (21.8)$$

Setze nun $C_0 = 2^{1+\gamma}(1 - 2^{-\gamma})^{-1} < \infty$. Seien $s, t \in D$ mit $|s - t| \leq 2^{-n_0}$. Indem wir $n \geq n_0$ minimal wählen mit $|t - s| \geq 2^{-n}$, erhalten wir aus (21.8)

$$|X_t(\omega) - X_s(\omega)| \leq C_0 |t - s|^\gamma. \quad (21.9)$$

Wie im Beweis von Lemma 21.3(iii) folgt hieraus, dass (mit $K := C_0 2^{(1-\gamma)n_0}$)

$$|X_t(\omega) - X_s(\omega)| \leq K |t - s|^\gamma \quad \text{für alle } s, t \in D. \quad (21.10)$$

Mit anderen Worten: Auf den binärrationalen Zahlen D ist $X(\omega)$ (global) Hölder- γ -stetig. Speziell ist X auf D gleichmäßig stetig, lässt sich also eindeutig stetig auf $[0, 1]$ fortsetzen: Für $t \in D$ setze $\tilde{X}_t := X_t$. Für $t \in [0, 1] \setminus D$ und $\{s_n, n \in \mathbb{N}\} \subset D$ mit $s_n \rightarrow t$ ist $(X_{s_n}(\omega))_{n \in \mathbb{N}}$ eine Cauchy-Folge. Also existiert der Limes

$$\tilde{X}_t(\omega) := \lim_{D \ni s \rightarrow t} X_s(\omega), \quad (21.11)$$

und es gilt dann die zu (21.10) analoge Aussage auch für nicht nur für $s, t \in D$:

$$|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)| \leq K |t - s|^\gamma \quad \text{für alle } s, t \in [0, 1]. \quad (21.12)$$

Also ist \tilde{X} lokal Hölder-stetig von der Ordnung γ . Nach (21.5) und (21.11) gilt $\mathbf{P}[X_t \neq \tilde{X}_t] = 0$ für jedes $t \in [0, 1]$. Also ist \tilde{X} eine Modifikation von X .

(ii) Sei $\varepsilon > 0$, und sei $n \in \mathbb{N}$ so groß gewählt, dass (siehe (21.6))

$$\mathbf{P}[B_n] \leq C \frac{2^{-(\beta-\alpha\gamma)n}}{1 - 2^{\alpha\gamma-\beta}} < \varepsilon.$$

Für $\omega \notin B_n$ gilt nach dem oben Gezeigten (21.10). Dies ist aber gerade (21.3) mit $T = 1$. Für allgemeine T folgt die Aussage durch lineare Transformationen. \square

Bemerkung 21.7. Die Aussage von Satz 21.6 gilt unverändert, wenn X Werte in einem polnischen Raum (E, ϱ) annimmt, denn im Beweis wurde nirgends ausgenutzt, dass der Wertebereich \mathbb{R} ist. Wenn wir jedoch den Zeitbereich verändern, müssen wir die Voraussetzungen verschärfen: Ist $(X_t)_{t \in \mathbb{R}^d}$ ein Prozess mit Werten in E , und gilt für gewisse $\alpha, \beta > 0$ sowie jedes $T > 0$ und gewisses $C < \infty$

$$\mathbf{E}[\varrho(X_t, X_s)^\alpha] \leq C \|t - s\|_2^{d+\beta} \quad \text{für alle } s, t \in [-T, T]^d, \quad (21.13)$$

so existiert für jedes $\gamma \in (0, \beta/\alpha)$ eine lokal Hölder- γ -stetige Version von X . \diamond

Übung 21.1.1. Man zeige die Aussage von Bemerkung 21.7. \clubsuit

Übung 21.1.2. Sei $X = (X_t)_{t \geq 0}$ ein reellwertiger Prozess mit stetigen Pfaden. Man zeige, dass für alle $0 \leq a < b$ die Abbildung $\omega \mapsto \int_a^b X_t(\omega) dt$ messbar ist. \clubsuit

Übung 21.1.3 (Optional Sampling/Stopping). Sei \mathbb{F} eine Filtration und $(X_t)_{t \geq 0}$ ein \mathbb{F} -Supermartingal mit rechtsstetigen Pfaden. Seien σ und τ beschränkte Stopzeiten mit $\sigma \leq \tau$. Setze $\sigma^n := 2^{-n} \lceil 2^n \sigma \rceil$ und $\tau^n := 2^{-n} \lceil 2^n \tau \rceil$.

- (i) Zeige, dass $\mathbf{E}[X_{\tau^n} | \mathcal{F}_{\sigma^n}] \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_{\tau^n} | \mathcal{F}_\sigma]$ fast sicher und in L^1 sowie $X_{\sigma_n} \xrightarrow{n \rightarrow \infty} X_\sigma$ fast sicher und in L^1 .
- (ii) Folgere das Optional Sampling Theorem für rechtsstetige Supermartingale mit Hilfe der analogen Aussage für diskrete Zeit (Satz 10.11): $X_\sigma \geq \mathbf{E}[X_\tau | \mathcal{F}_\sigma]$.
- (iii) Zeige: Ist Y adaptiert, integrierbar und rechtsstetig, so ist Y genau dann ein Martingal, wenn $\mathbf{E}[Y_\tau] = \mathbf{E}[Y_0]$ für jede beschränkte Stopzeit τ gilt.
- (iv) Zeige, dass $X_\sigma \geq \mathbf{E}[X_\tau | \mathcal{F}_\sigma]$ für lediglich endliche Stopzeiten $\sigma \leq \tau$ gilt, falls X gleichgradig integrierbar ist.
- (v) Sei nun τ eine beliebige Stopzeit. Man folgere das Optional Stopping Theorem für rechtsstetige Supermartingale: $(X_{\tau \wedge t})_{t \geq 0}$ ist ein rechtsstetiges Supermartingal. \clubsuit

Übung 21.1.4. Sei $X = (X_t)_{t \geq 0}$ ein stochastischer Prozess auf $(\Omega, \mathcal{F}, \mathbf{P})$ mit Werten im polnischen Raum E und rechtsstetigen Pfaden. Man zeige:

- (i) Die Abbildung $(\omega, t) \mapsto X_t(\omega)$ ist messbar bezüglich $\mathcal{F} \otimes \mathcal{B}([0, \infty))$ - $\mathcal{B}(E)$.
- (ii) Ist X zudem an die Filtration \mathbb{F} adaptiert, so ist für jedes $t \geq 0$ die Abbildung $\Omega \times [0, t] \rightarrow E, (\omega, s) \mapsto X_s(\omega)$ messbar bezüglich $\mathcal{F}_t \otimes \mathcal{B}([0, t])$ - $\mathcal{B}(E)$.
- (iii) Ist τ eine \mathbb{F} -Stopzeit und X adaptiert, so ist X_τ eine \mathcal{F}_τ -messbare Zufallsvariable. \clubsuit

21.2 Konstruktion und Pfadeigenschaften

Definition 21.8. Ein reellwertiger stochastischer Prozess $B = (B_t, t \in [0, \infty))$ heißt **Brown'sche Bewegung**, falls

- (i) $B_0 = 0$,
- (ii) B hat unabhängige, stationäre Zuwächse (vergleiche Definition 9.7),
- (iii) $B_t \sim \mathcal{N}_{0,t}$ für $t > 0$,
- (iv) \mathbf{P} -fast sicher gilt: $t \mapsto B_t$ ist stetig.

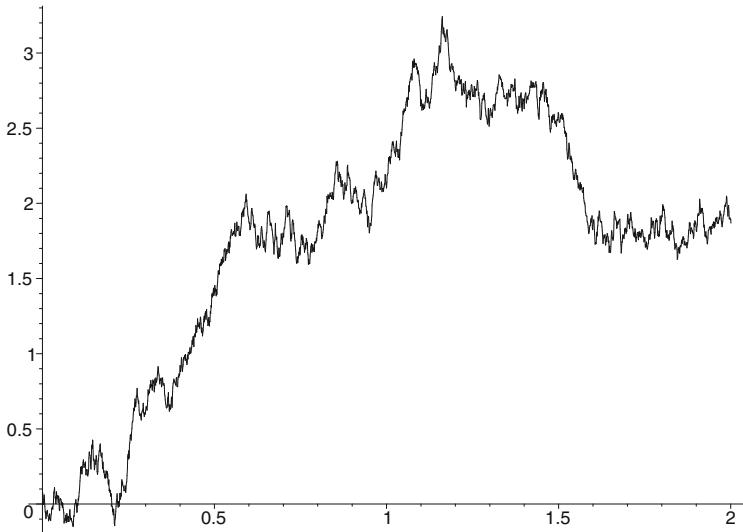


Abb. 21.1. Computersimulation einer Brown'schen Bewegung.

Satz 21.9. Es existiert ein W -Raum $(\Omega, \mathcal{A}, \mathbf{P})$ und eine Brown'sche Bewegung B auf $(\Omega, \mathcal{A}, \mathbf{P})$. Die Pfade von B sind f.s. lokal Hölder- γ -stetig für jedes $\gamma < \frac{1}{2}$.

Beweis. Wie in Beispiel 14.45 oder Korollar 16.10 gibt es einen stochastischen Prozess X , der (i), (ii) und (iii) erfüllt. Offenbar ist $X_t - X_s \stackrel{\mathcal{D}}{=} \sqrt{t-s} X_1 \sim \mathcal{N}_{0,t-s}$ für alle $t > s \geq 0$. Es gilt daher für jedes $n \in \mathbb{N}$ und $C_n := \mathbf{E}[X_1^{2n}] = \frac{(2n)!}{2^n n!} < \infty$

$$\mathbf{E}[(X_t - X_s)^{2n}] = \mathbf{E}\left[\left(\sqrt{t-s} X_1\right)^{2n}\right] = C_n |t-s|^n.$$

Sei nun $n \geq 2$ und $\gamma \in (0, \frac{n-1}{2n})$. Satz 21.6 liefert die Existenz einer Version B von X mit Hölder- γ -stetigen Pfaden. Da alle stetigen Versionen eines Prozesses

äquivalent sind, ist B lokal Hölder- γ -stetig für jedes $\gamma \in (0, \frac{n-1}{2n})$ und jedes $n \geq 2$, also für jedes $\gamma \in (0, \frac{1}{2})$. \square

Erinnerung: Ein stochastischer Prozess $(X_t)_{t \in I}$ heißt **Gauß'scher Prozess**, falls für jedes $n \in \mathbb{N}$ und alle $t_1, \dots, t_n \in I$ gilt

$$(X_{t_1}, \dots, X_{t_n}) \text{ ist } n\text{-dimensional normalverteilt.}$$

Wir nennen X **zentriert**, falls $\mathbf{E}[X_t] = 0$ für jedes $t \in I$. Die Funktion

$$\Gamma(s, t) := \mathbf{Cov}[X_s, X_t] \quad \text{für } s, t \in I,$$

heißt **Kovarianzfunktion** von X .

Bemerkung 21.10. Durch die Kovarianzfunktion sind die endlichdimensionalen Verteilungen eines zentrierten, Gauß'schen Prozesses eindeutig festgelegt, denn eine mehrdimensionale Normalverteilung ist durch den Erwartungswertvektor und Kovarianzmatrix vollständig beschrieben. \diamond

Satz 21.11. Für einen stochastischen Prozess $X = (X_t)_{t \in [0, \infty)}$ sind äquivalent:

- (i) X ist eine Brown'sche Bewegung.
- (ii) X ist ein stetiger, zentrierter, Gauß'scher Prozess mit $\mathbf{Cov}[X_s, X_t] = s \wedge t$ für alle $s, t \geq 0$.

Beweis. Nach Bemerkung 21.10 ist X durch (ii) eindeutig bestimmt. Es reicht also zu zeigen, dass $\mathbf{Cov}[X_s, X_t] = \min(s, t)$ für die Brown'sche Bewegung X gilt. Dies ist aber richtig, denn für $t > s$ sind X_s und $X_t - X_s$ unabhängig, also ist

$$\mathbf{Cov}[X_s, X_t] = \mathbf{Cov}[X_s, X_t - X_s] + \mathbf{Cov}[X_s, X_s] = \mathbf{Var}[X_s] = s. \quad \square$$

Korollar 21.12 (Skalierungseigenschaft der Brown'schen Bewegung). Ist B eine Brown'sche Bewegung und $K \neq 0$, dann ist auch $(K^{-1}B_{K^2 t})_{t \geq 0}$ eine Brown'sche Bewegung.

Beispiel 21.13. Ein weiteres Beispiel für einen stetigen, Gauß'schen Prozess ist die **Brown'sche Brücke** $X = (X_t)_{t \in [0, 1]}$, die die Kovarianzfunktion $\Gamma(s, t) = s \wedge t - st$ hat. Wir konstruieren die Brown'sche Brücke wie folgt: Sei $B = (B_t, t \in [0, 1])$ eine Brown'sche Bewegung und

$$X_t := B_t - tB_1.$$

Offenbar ist X ein zentrierter, Gauß'scher Prozess mit stetigen Pfaden. Die Kovarianzfunktion Γ von X errechnet sich zu

$$\begin{aligned} \Gamma(s, t) &= \mathbf{Cov}[X_s, X_t] = \mathbf{Cov}[B_s - sB_1, B_t - tB_1] \\ &= \mathbf{Cov}[B_s, B_t] - s \mathbf{Cov}[B_1, B_t] - t \mathbf{Cov}[B_s, B_1] + st \mathbf{Cov}[B_1, B_1] \\ &= \min(s, t) - st - st + st = \min(s, t) - st. \end{aligned} \quad \diamond$$

Satz 21.14. Sei $(B_t)_{t \geq 0}$ eine Brown'sche Bewegung und

$$X_t = \begin{cases} tB_{1/t}, & \text{falls } t > 0, \\ 0, & \text{falls } t = 0. \end{cases}$$

Dann ist X eine Brown'sche Bewegung.

Beweis. Offenbar ist X ein Gauß'scher Prozess. Für $s, t > 0$ ist

$$\mathbf{Cov}[X_s, X_t] = ts \cdot \mathbf{Cov}[B_{1/s}, B_{1/t}] = ts \min(s^{-1}, t^{-1}) = \min(s, t).$$

Offenbar ist $t \mapsto X_t$ stetig in allen $t > 0$. Für die Stetigkeit in $t = 0$ betrachte

$$\begin{aligned} \limsup_{t \downarrow 0} X_t &= \limsup_{t \rightarrow \infty} \frac{1}{t} B_t \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} B_n + \limsup_{n \rightarrow \infty} \frac{1}{n} \sup \{B_t - B_n, t \in [n, n+1]\}. \end{aligned}$$

Nach dem Starken Gesetz der großen Zahl ist $\lim_{n \rightarrow \infty} \frac{1}{n} B_n = 0$ f.s. Nach einer Verallgemeinerung des Spiegelprinzips (Satz 17.15, siehe auch Satz 21.19) ist für $x > 0$ (mit der Abkürzung $B_{[a,b]} := \{B_t : t \in [a,b]\}$)

$$\begin{aligned} \mathbf{P}[\sup B_{[n,n+1]} - B_n > x] &= \mathbf{P}[\sup B_{[0,1]} > x] = 2 \mathbf{P}[B_1 > x] \\ &= \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du \leq \frac{1}{x} e^{-x^2/2}. \end{aligned}$$

Speziell ist $\sum_{n=1}^{\infty} \mathbf{P}[\sup B_{[n,n+1]} - B_n > n^\varepsilon] < \infty$ für jedes $\varepsilon > 0$. Nach dem Lemma von Borel-Cantelli (Satz 2.7) ist daher

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sup \{B_t - B_n, t \in [n, n+1]\} = 0 \quad \text{fast sicher.}$$

Mithin ist X auch in 0 stetig. □

Satz 21.15 (Blumenthal'sches 0-1 Gesetz). Sei B eine Brown'sche Bewegung und $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0} = \sigma(B)$ die erzeugte Filtration sowie $\mathcal{F}_0^+ = \bigcap_{t>0} \mathcal{F}_t$. Dann ist \mathcal{F}_0^+ eine \mathbf{P} -triviale σ -Algebra.

Beweis. Setze $Y^n = (B_{2^{-n}+t} - B_{2^{-n}})_{t \in [0, 2^{-n}]}$, $n \in \mathbb{N}$. Dann ist $(Y^n)_{n \in \mathbb{N}}$ eine unabhängige Familie von Zufallsvariablen (mit Werten in $C([0, 2^{-n}])$). Die terminale σ -Algebra $\mathcal{T} = \bigcap_{n \in \mathbb{N}} \sigma(Y^m, m \geq n)$ ist nach dem Kolmogorov'schen 0-1 Gesetz (Satz 2.37) \mathbf{P} -trivial. Andererseits ist $\sigma(Y^m, m \geq n) = \mathcal{F}_{2^{-n+1}}$, also ist

$$\mathcal{F}_0^+ = \bigcap_{t>0} \mathcal{F}_t = \bigcap_{n \in \mathbb{N}} \mathcal{F}_{2^{-n+1}} = \mathcal{T}$$

\mathbf{P} -trivial. □

Beispiel 21.16. Sei B eine Brown'sche Bewegung. Für jedes $K > 0$ ist

$$\mathbf{P}[\inf\{t > 0 : B_t \geq K\sqrt{t}\} = 0] = 1. \quad (21.14)$$

Um dies einzusehen, setze $A_s := \{\inf\{t > 0 : B_t \geq K\sqrt{t}\} \leq s\}$ und

$$A := \left\{ \inf\{t > 0 : B_t \geq K\sqrt{t}\} = 0 \right\} = \bigcap_{s > 0} A_s \in \mathcal{F}_0^+.$$

Dann ist $\mathbf{P}[A] \in \{0, 1\}$. Wegen der Skalierungseigenschaft der Brown'schen Bewegung ist $\mathbf{P}[A] = \inf_{s > 0} \mathbf{P}[A_s] \geq \mathbf{P}[B_1 \geq K] > 0$ und deshalb $\mathbf{P}[A] = 1$. \diamond

Das eben untersuchte Beispiel zeigt insbesondere für jedes $t \geq 0$, dass B in t fast sicher nicht Hölder- $\frac{1}{2}$ -stetig ist. Hier ist Vorsicht mit der Reihenfolge der Quantoren angebracht: Wir haben *nicht* gezeigt, dass B fast sicher in keinem $t \geq 0$ Hölder- $\frac{1}{2}$ -stetig wäre (siehe aber Bemerkung 22.4). Wir können allerdings ohne großen Aufwand den folgenden Satz zeigen, der für den Fall $\gamma = 1$ auf Paley, Wiener und Zygmund [135] zurückgeht. Der hier vorgestellte Beweis beruht auf einer Idee von Dvoretzky, Erdös und Kakutani (siehe [45]).

Satz 21.17 (Paley-Wiener-Zygmund (1933)). Für jedes $\gamma > \frac{1}{2}$ sind die Pfade der Brown'schen Bewegung $(B_t)_{t \geq 0}$ fast sicher in keinem Punkte Hölder-stetig der Ordnung γ . Insbesondere sind die Pfade fast sicher nirgends differenzierbar.

Beweis. Sei $\gamma > \frac{1}{2}$. Es reicht, $B = (B_t)_{t \in [0,1]}$ zu betrachten. Wir bezeichnen mit $H_{\gamma,t}$ die Menge der in t Hölder- γ -stetigen Abbildungen $[0, 1] \rightarrow \mathbb{R}$ und setzen $H_\gamma := \bigcup_{t \in [0,1]} H_{\gamma,t}$. Das Ziel ist zu zeigen, dass fast sicher $B \notin H_\gamma$ gilt.

Ist $t \in [0, 1]$ und $w \in H_{\gamma,t}$, so existiert zu jedem $\delta > 0$ ein $c = c(\delta, w)$ mit der Eigenschaft, dass $|w_s - w_t| \leq c|s - t|^\gamma$ ist für jedes $s \in [0, 1]$ mit $|s - t| < \delta$. Wählen wir $k \in \mathbb{N}$ mit $k > \frac{2}{2\gamma-1}$, so ist für $n \in \mathbb{N}$ mit $n \geq n_0 := \lceil (k+1)/\delta \rceil$, $i = \lfloor tn \rfloor + 1$ und $l \in \{0, \dots, k-1\}$ speziell

$$|w_{(i+l+1)/n} - w_{(i+l)/n}| \leq |w_{(i+l+1)/n} - w_t| + |w_{(i+l)/n} - w_t| \leq 2c(k+1)^\gamma n^{-\gamma}.$$

Für $N \geq 2c(k+1)^\gamma$ ist also $w \in A_{N,n,i}$, wobei

$$A_{N,n,i} := \bigcap_{l=0}^{k-1} \left\{ w : |w_{(i+l+1)/n} - w_{(i+l)/n}| \leq N n^{-\gamma} \right\}.$$

Setzen wir $A_{N,n} = \bigcup_{i=1}^n A_{N,n,i}$, $A_N = \liminf_{n \rightarrow \infty} A_{N,n}$ und $A = \bigcup_{N=1}^{\infty} A_N$, so ist offenbar $H_\gamma \subset A$. Nun ist wegen der Unabhängigkeit der Zuwächse, und weil die Dichte der Standardnormalverteilung nirgends größer als 1 ist

$$\begin{aligned} \mathbf{P}[B \in A_{N,n,i}] &= \mathbf{P}[|B_{1/n}| \leq N n^{-\gamma}]^k = \mathbf{P}[|B_1| \leq N n^{-\gamma+1/2}]^k \\ &\leq N^k n^{k(-\gamma+1/2)}. \end{aligned}$$

Nach Wahl von k und wegen der Stationarität der Zuwächse von B gilt

$$\begin{aligned}\mathbf{P}[B \in A_N] &= \lim_{n \rightarrow \infty} \mathbf{P}\left[\bigcap_{m \geq n} A_{N,m}\right] \leq \limsup_{n \rightarrow \infty} \mathbf{P}[A_{N,n}] \leq \limsup_{n \rightarrow \infty} \sum_{i=1}^n \mathbf{P}[A_{N,n,i}] \\ &\leq \limsup_{n \rightarrow \infty} n \mathbf{P}[B \in A_{N,n,1}] \leq N^k \limsup_{n \rightarrow \infty} n^{1+k(-\gamma+1/2)} = 0\end{aligned}$$

und damit $\mathbf{P}[B \in A] = 0$. Mithin ist fast sicher $B \notin H_\gamma$. \square

Übung 21.2.1. Sei B eine Brown'sche Bewegung und λ das Lebesgue-Maß auf $[0, \infty)$.

- (i) Bestimme Erwartungswert und Varianz von $\int_0^1 B_s ds$. (Für die Messbarkeit des Integrals siehe Übung 21.1.2.)
- (ii) Zeige, dass $\lambda(\{t : B_t = 0\}) = 0$ fast sicher gilt.
- (iii) Bestimme Erwartungswert und Varianz von

$$\int_0^1 \left(B_t - \int_0^1 B_s ds \right)^2 dt. \quad \clubsuit$$

Übung 21.2.2. Sei B eine Brown'sche Bewegung. Zeige, dass auch $(B_t^2 - t)_{t \geq 0}$ ein Martingal ist. \clubsuit

Übung 21.2.3. Sei B eine Brown'sche Bewegung und $\sigma > 0$. Zeige, dass auch $(\exp(\sigma B_t - \frac{\sigma^2}{2}t))_{t \geq 0}$ ein Martingal ist. \clubsuit

Übung 21.2.4. Sei B eine Brown'sche Bewegung. Für $a < 0 < b$ definiere die Stopzeit $\tau_{a,b} = \inf\{t \geq 0 : B_t \in \{a, b\}\}$.

Man zeige, dass $\tau_{a,b} < \infty$ fast sicher gilt und $\mathbf{P}[B_{\tau_{a,b}} = b] = -\frac{a}{b-a}$. Man zeige ferner (mit Übung 21.2.2), dass $\mathbf{E}[\tau_{a,b}] = -ab$. \clubsuit

Übung 21.2.5. Sei B eine Brown'sche Bewegung. Für $b > 0$ sowie definiere die Stopzeit $\tau_b = \inf\{t \geq 0 : B_t = b\}$. Man zeige:

- (i) $\mathbf{E}[e^{-\lambda \tau_b}] = e^{-b\sqrt{2\lambda}}$ für $\lambda \geq 0$. (Hinweis: Verwende Übung 21.2.3 und den Optional Sampling Satz.)

- (ii) τ_b hat eine $\frac{1}{2}$ -stabile Verteilung mit Lévy-Maß

$$\nu(dx) = (b/(\sqrt{2\pi})) x^{-3/2} \mathbb{1}_{\{x>0\}} dx.$$

- (iii) Die Verteilung von τ_b hat die Dichte $f_b(x) = \frac{b}{\sqrt{2\pi}} e^{-b^2/(2x)} x^{-3/2}$. \clubsuit

Übung 21.2.6. Sei B eine Brown'sche Bewegung, $a \in \mathbb{R}$ und $b > 0$ sowie $\tau = \inf\{t \geq 0 : B_t = at + b\}$. Man zeige für $\lambda \geq 0$

$$\mathbf{E}[e^{-\lambda\tau}] = \exp\left(-ba - b\sqrt{a^2 + 2\lambda}\right)$$

und folgere $\mathbf{P}[\tau < \infty] = 1 \wedge e^{-2ba}$. ♣

21.3 Starke Markoveigenschaft

Wir bezeichnen mit \mathbf{P}_x das W-Maß, unter dem $B = (B_t)_{t \geq 0}$ eine in $x \in \mathbb{R}$ gestartete Brown'sche Bewegung ist. Anders gesagt ist unter \mathbf{P}_x der um $-x$ verschobene Prozess $(B_t - x)_{t \geq 0}$ eine (in Null gestartete) Brown'sche Bewegung. Während die (elementare) Markoveigenschaft von $(B, (\mathbf{P}_x)_{x \in \mathbb{R}})$ evident ist, brauchen wir für den Nachweis der starken Markoveigenschaft ein wenig Arbeit.

Satz 21.18 (Starke Markoveigenschaft). Die Brown'sche Bewegung B mit Verteilungen $(\mathbf{P}_x)_{x \in \mathbb{R}}$ hat die starke Markoveigenschaft.

Beweis. Sei $\mathbb{F} = \sigma(B)$ die von B erzeugte Filtration und $\tau < \infty$ eine \mathbb{F} -Stoppzeit. Zu zeigen ist, dass für jedes beschränkte, messbare $F : \mathbb{R}^{[0, \infty)} \rightarrow \mathbb{R}$ gilt:

$$\mathbf{E}_x[F((B_{t+\tau})_{t \geq 0}) | \mathcal{F}_\tau] = \mathbf{E}_{B_\tau}[F(B)]. \quad (21.15)$$

Es reicht, stetige, beschränkte Funktionen F zu betrachten, die nur von endlich vielen Koordinaten t_1, \dots, t_N abhängen, da diese die Verteilung von $(B_{t+\tau})_{t \geq 0}$ eindeutig bestimmen. Sei also $f : \mathbb{R}^N \rightarrow \mathbb{R}$ stetig und beschränkt und $F(B) = f(B_{t_1}, \dots, B_{t_N})$. Offenbar ist die Abbildung $x \mapsto \mathbf{E}_x[F(B)] = \mathbf{E}_0[f(B_{t_1} + x, \dots, B_{t_N} + x)]$ stetig und beschränkt. Sei nun $\tau^n := 2^{-n} \lfloor 2^n \tau + 1 \rfloor$ für $n \in \mathbb{N}$. Dann ist τ^n eine Stoppzeit und $\tau^n \downarrow \tau$, also $B_{\tau^n} \xrightarrow{n \rightarrow \infty} B_\tau$ fast sicher. Nun ist jeder Markovprozess mit abzählbarer Zeitmenge (hier: alle positiven rationalen Linearkombinationen von $1, t_1, \dots, t_N$) ein starker Markovprozess (nach Satz 17.14), also gilt

$$\begin{aligned} \mathbf{E}_x[F((B_{\tau^n+t})_{t \geq 0}) | \mathcal{F}_{\tau^n}] &= \mathbf{E}_x[f(B_{\tau^n+t_1}, \dots, B_{\tau^n+t_N}) | \mathcal{F}_{\tau^n}] \\ &= \mathbf{E}_{B_{\tau^n}}[f(B_{t_1}, \dots, B_{t_N})] \\ &\xrightarrow{n \rightarrow \infty} \mathbf{E}_{B_\tau}[f(B_{t_1}, \dots, B_{t_N})] = \mathbf{E}_{B_\tau}[F(B)]. \end{aligned} \quad (21.16)$$

Aufgrund der Rechtsstetigkeit von B gilt $F((B_{\tau^n+t})_{t \geq 0}) \xrightarrow{n \rightarrow \infty} F((B_{\tau+t})_{t \geq 0})$ fast sicher und in L^1 und damit

$$\begin{aligned} \mathbf{E}\left[\left|\mathbf{E}_x[F((B_{\tau^n+t})_{t \geq 0}) | \mathcal{F}_{\tau^n}] - \mathbf{E}_x[F((B_{\tau+t})_{t \geq 0}) | \mathcal{F}_{\tau^n}]\right|\right] \\ \leq \mathbf{E}_x\left[\left|F((B_{\tau^n+t})_{t \geq 0}) - F((B_{\tau+t})_{t \geq 0})\right|\right] \xrightarrow{n \rightarrow \infty} 0. \end{aligned} \quad (21.17)$$

Weiter gilt

$$\mathcal{F}_{\tau^n} \downarrow \mathcal{F}_{\tau+} := \bigcap_{\sigma > \tau \text{ ist Stoppzeit}} \mathcal{F}_\sigma \supset \mathcal{F}_\tau.$$

Nach (21.16) und (21.17) sowie dem Konvergenzsatz für Rückwärtsmartingale (Satz 12.14) gilt also im Sinne von L^1 -Limiten

$$\begin{aligned} \mathbf{E}_{B_\tau}[F(B)] &= \lim_{n \rightarrow \infty} \mathbf{E}_x[F((B_{\tau^n+t})_{t \geq 0}) | \mathcal{F}_{\tau^n}] \\ &= \lim_{n \rightarrow \infty} \mathbf{E}_x[F((B_{\tau+t})_{t \geq 0}) | \mathcal{F}_{\tau^n}] = \mathbf{E}_x[F((B_{\tau+t})_{t \geq 0}) | \mathcal{F}_{\tau+}]. \end{aligned}$$

Die linke Seite ist \mathcal{F}_τ -messbar. Die Turmeigenschaft der bedingten Erwartung liefert also (21.15). \square

Mit Hilfe der starken Markoveigenschaft zeigen wir das Reflexionsprinzip für die Brown'sche Bewegung.

Satz 21.19 (Reflexionsprinzip für die Brown'sche Bewegung). Für jedes $a > 0$ und $T > 0$ gilt

$$\mathbf{P}\left[\sup\{B_t : t \in [0, T]\} > a\right] = 2 \mathbf{P}[B_T > a] \leq \frac{2\sqrt{T}}{\sqrt{2\pi}} \frac{1}{a} e^{-a^2/2T}.$$

Beweis. Wegen der Skalierungseigenschaft der Brown'schen Bewegung (Korollar 21.12) können wir ohne Einschränkung $T = 1$ annehmen. Sei

$$\tau := \inf\{t \geq 0 : B_t \geq a\} \wedge 1.$$

Aus Symmetriegründen ist $\mathbf{P}_a[B_{1-\tau} > a] = \frac{1}{2}$, falls $\tau < 1$, also

$$\begin{aligned} \mathbf{P}[B_1 > a] &= \mathbf{P}[B_1 > a | \tau < 1] \mathbf{P}[\tau < 1] \\ &= \mathbf{P}_a[B_{1-\tau} > a] \mathbf{P}[\tau < 1] = \frac{1}{2} \mathbf{P}[\tau < 1]. \end{aligned}$$

Für die Ungleichung berechnen wir

$$\begin{aligned} \mathbf{P}[B_1 > a] &= \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-x^2/2} dx \\ &\leq \frac{1}{\sqrt{2\pi}} \frac{1}{a} \int_a^\infty x e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \frac{1}{a} e^{-a^2/2}. \end{aligned} \quad \square$$

Eine Anwendung des Reflexionsprinzips ist das Arkussinus-Gesetzes von Paul Lévy [116, Seite 216] für den letzten Besuch der Brown'schen Bewegung in der Null.

Satz 21.20 (Lévy'sches Arkussinus-Gesetz). Sei $T > 0$ und $\zeta_T := \sup\{t \leq T : B_t = 0\}$. Dann gilt für $t \in [0, T]$

$$\mathbf{P}[\zeta_T \leq t] = \frac{2}{\pi} \arcsin\left(\sqrt{t/T}\right).$$

Beweis. Ohne Einschränkung sei $T = 1$ und $\zeta = \zeta_1$. Sei \tilde{B} eine weitere, unabhängige Brown'sche Bewegung. Nach dem Reflexionsprinzip gilt

$$\begin{aligned}\mathbf{P}[\zeta \leq t] &= \mathbf{P}[B_s \neq 0 \text{ für jedes } s \in [t, 1]] \\ &= \int_{-\infty}^{\infty} \mathbf{P}[B_s \neq 0 \text{ für jedes } s \in [t, 1] \mid B_t = a] \mathbf{P}[B_t \in da] \\ &= \int_{-\infty}^{\infty} \mathbf{P}_{|a|}[\tilde{B}_s > 0 \text{ für jedes } s \in [0, 1-t]] \mathbf{P}[B_t \in da] \\ &= \int_{-\infty}^{\infty} \mathbf{P}_0[|\tilde{B}_{1-t}| \leq |a|] \mathbf{P}[B_t \in da] = \mathbf{P}[|\tilde{B}_{1-t}| \leq |B_t|].\end{aligned}$$

Sind X, Y unabhängig und $\mathcal{N}_{0,1}$ -verteilt, so ist $(B_t, \tilde{B}_{1-t}) \stackrel{\mathcal{D}}{=} (\sqrt{t} X, \sqrt{1-t} Y)$. Es folgt

$$\begin{aligned}\mathbf{P}[\zeta \leq t] &= \mathbf{P}[\sqrt{1-t} |Y| \leq \sqrt{t} |X|] \\ &= \mathbf{P}[Y^2 \leq t(X^2 + Y^2)] \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy e^{-(x^2+y^2)/2} \mathbb{1}_{\{y^2 \leq t(x^2+y^2)\}}.\end{aligned}$$

Durch Polarkoordinatentransformation erhalten wir

$$\mathbf{P}[\zeta \leq t] = \frac{1}{2\pi} \int_0^{\infty} r dr e^{-r^2/2} \int_0^{2\pi} d\varphi \mathbb{1}_{\{\sin(\varphi)^2 \leq t\}} = \frac{2}{\pi} \arcsin(\sqrt{t}). \quad \square$$

Übung 21.3.1. (Schwierig!) Sei \mathbf{P}_x die Verteilung der Brown'schen Bewegung mit Start in $x \in \mathbb{R}$. Sei $a > 0$ und $\tau = \inf \{t \geq 0 : B_t \in \{0, a\}\}$. Man zeige mit Hilfe des Spiegelungsprinzips, dass für jedes $x \in (0, a)$ gilt

$$\mathbf{P}_x[\tau > T] = \sum_{n=-\infty}^{\infty} (-1)^n \mathbf{P}_x[B_T \in [na, (n+1)a]]. \quad (21.18)$$

Ist f die Dichte einer Wahrscheinlichkeitsverteilung auf \mathbb{R} mit charakteristischer Funktion φ und $\sup_{x \in \mathbb{R}} x^2 f(x) < \infty$, so gilt die Poisson'sche Summationsformel (siehe etwa [26, Satz 2.2.2])

$$\sum_{n=-\infty}^{\infty} f(s+n) = \sum_{k=-\infty}^{\infty} \varphi(k) e^{2\pi i s} \quad \text{für jedes } s \in \mathbb{R}. \quad (21.19)$$

Man folgere aus (21.18) und (21.19) (vergleiche auch (21.37))

$$\mathbf{P}_x[\tau > T] = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{1}{2k+1} \exp\left(-\frac{(2k+1)^2 \pi^2 T}{2a^2}\right) \sin\left(\frac{(2k+1)\pi x}{a}\right). \quad (21.20)$$



21.4 Ergänzung: Feller Prozesse

In vielen Situationen kann man keine stetige Version eines Prozesses erwarten, etwa beim Poissonprozess, der ja gewissermaßen von seinen Sprüngen lebt. Oft kann jedoch eine Version mit rechtsstetigen Pfaden, die einen endlichen linksseitigen Grenzwert besitzen, etabliert werden. Wir wollen hier knapp den Existenzsatz für solche Prozesse für Feller'sche Halbgruppen plausibel machen.

Definition 21.21. Sei E ein polnischer Raum. Eine Abbildung $f : [0, \infty) \rightarrow E$ heißt **RCLL** (right continuous with left limits) oder **càdlàg** (continue à droit, limites à gauche), falls $f(t) = f(t+) := \lim_{s \downarrow t} f(s)$ für jedes $t \geq 0$ und falls der linksseitige Grenzwert $f(t-) := \lim_{s \uparrow t} f(s)$ für jedes $t > 0$ existiert und endlich ist.

Definition 21.22. Eine Filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ heißt **rechtsstetig**, falls $\mathbb{F} = \mathbb{F}^+$, wo $\mathcal{F}_t^+ = \bigcap_{s > t} \mathcal{F}_s$. Wir sagen, dass eine Filtration \mathbb{F} die **üblichen Bedingungen** erfüllt, falls \mathbb{F} rechtsstetig ist und \mathcal{F}_0 \mathbf{P} -vollständig ist.

Bemerkung 21.23. Ist \mathbb{F} eine beliebige Filtration und $\mathcal{F}_t^{+,*}$ die Vervollständigung von \mathcal{F}_t^+ , so erfüllt $\mathbb{F}^{+,*}$ die üblichen Bedingungen. ◇

Satz 21.24 (Doob'sche Regularisierung). Sei \mathbb{F} eine Filtration, die die üblichen Bedingungen erfüllt, und $X = (X_t)_{t \geq 0}$ ein \mathbb{F} -Supermartingal mit der Eigenschaft, dass $t \mapsto \mathbf{E}[X_t]$ rechtsstetig ist. Dann gibt es eine Modifikation \tilde{X} von X mit RCLL-Pfaden.

Beweis. Für $a, b \in \mathbb{Q}^+$, $a < b$ und $I \subset [0, \infty)$ sei $U_I^{a,b}$ die Anzahl der Aufkreuzungen von $(X_t)_{t \in I}$ über $[a, b]$. Nach der Aufkreuzungsungleichung (Lemma 11.3) folgt für jedes $N > 0$ und jede endliche Menge $I \subset [0, N]$, dass $\mathbf{E}[U_I^{a,b}] \leq (\mathbf{E}[|X_N|] + |a|)/(b - a)$. Setzen wir $U_N^{a,b} = U_{\mathbb{Q}^+ \cap [0, N]}^{a,b}$, so folgt $\mathbf{E}[U_N^{a,b}] \leq (\mathbf{E}[|X_N|] + |a|)/(b - a)$. Für $\lambda > 0$ ist nach Übung 11.1.1

$$\begin{aligned} \lambda \mathbf{P}[\sup\{|X_t| : t \in \mathbb{Q}^+ \cap [0, N]\} > \lambda] \\ &= \lambda \sup \left\{ \mathbf{P}[\sup\{|X_t| : t \in I\} > \lambda] : I \subset \mathbb{Q}^+ \cap [0, N] \text{ endlich} \right\} \\ &\leq 12 \mathbf{E}[|X_0|] + 9 \mathbf{E}[|X_N|]. \end{aligned}$$

Betrachte das Ereignis

$$A := \bigcap_{N \in \mathbb{N}} \left(\bigcap_{\substack{a, b \in \mathbb{Q}^+ \\ 0 \leq a < b \leq N}} \{U_N^{a,b} < \infty\} \cap \{ \sup\{|X_t| : t \in \mathbb{Q}^+ \cap [0, N]\} < \infty \} \right).$$

Nach dem Gezeigten ist $\mathbf{P}[A] = 1$, also $A \in \mathcal{F}_t$ für jedes $t \geq 0$, da \mathbb{F} die üblichen Bedingungen erfüllt. Für $\omega \in A$ existiert für jedes $t \geq 0$ der Limes

$$\tilde{X}_t(\omega) := \lim_{\mathbb{Q}^+ \ni s \downarrow t, s > t} X_s(\omega)$$

und ist RCLL. Für $\omega \in A^c$ setzen wir $\tilde{X}_t(\omega) = 0$. Da \mathbb{F} die üblichen Bedingungen erfüllt, ist \tilde{X} an \mathbb{F} adaptiert. Da X ein Supermartingal ist, ist $(X_s)_{s \leq N}$ für jedes N gleichgradig integrierbar. Also gilt (nach Voraussetzung), dass

$$\mathbf{E}[\tilde{X}_t] = \lim_{\mathbb{Q}^+ \ni s \downarrow t, s > t} \mathbf{E}[X_s] = \mathbf{E}[X_t].$$

Da X ein Supermartingal ist, ist aber für $s > t$

$$X_t \geq \mathbf{E}[X_s | \mathcal{F}_t] \xrightarrow{\mathbb{Q}^+ \ni s \downarrow t, s > t} \mathbf{E}[\tilde{X}_t | \mathcal{F}_t] = \tilde{X}_t \quad \text{in } L^1.$$

Folglich ist $X_t = \tilde{X}_t$ fast sicher, also \tilde{X} eine Modifikation von X . \square

Korollar 21.25. Sei $(\nu_t)_{t \geq 0}$ eine stetige Faltungshalbgruppe mit $\int |x| \nu_1(dx) < \infty$. Dann existiert ein Markov-Prozess X mit unabhängigen, stationären Zuwächsen $\mathbf{P}_{X_t - X_s} = \nu_{t-s}$ für alle $t > s$ und mit RCLL Pfaden.

Sei E ein lokalkompakter, polnischer Raum und $C_0(E)$ die Menge der (beschränkten) stetigen Funktionen, die im Unendlichen verschwinden. Ist κ ein stochastischer Kern von E nach E und ist f messbar und beschränkt, so schreiben wir $\kappa f(x) = \int \kappa(x, dy) f(y)$.

Definition 21.26. Eine Markov'sche Halbgruppe $(\kappa_t)_{t \geq 0}$ auf E heißt **Feller'sche Halbgruppe**, falls

$$f(x) = \lim_{t \rightarrow 0} \kappa_t f(x) \quad \text{für jedes } x \in E, f \in C_0(E)$$

und $\kappa_t f \in C_0(E)$ für jedes $f \in C_0(E)$.

Sei X ein zu $(\kappa_t)_{t \geq 0}$ gehöriger Markovprozess bezüglich einer Filtration \mathbb{F} , die die üblichen Bedingungen erfüllt.

Sei $g \in C_0(E)$, $g \geq 0$. Setze $h = \int_0^\infty e^{-t} \kappa_t g dt$. Dann ist

$$e^{-s} \kappa_s h = e^{-s} \int_0^\infty e^{-t} \kappa_s \kappa_t g dt = \int_s^\infty e^{-t} \kappa_t g dt \leq h.$$

Also ist $X^g := (e^{-t} h(X_t))_{t \geq 0}$ ein \mathbb{F} -Supermartingal.

Die Fellereigenschaft und Satz 21.24 sichern nun die Existenz einer RCLL Version \tilde{X}^g von X^g . Mit etwas mehr Arbeit kann man zeigen, dass mit einer abzählbaren Menge $G \subset C_0(E)$ ein Prozess \tilde{X} durch alle \tilde{X}^g , $g \in G$, eindeutig festgelegt ist und eine RCLL Version von X ist. Siehe etwa [156, Kapitel III.7ff].

Wir wollen nun rückblicken, wie wir die starke Markoveigenschaft der Brown'schen Bewegung in Abschnitt 21.3 hergeleitet hatten. Tatsächlich wurde dort lediglich die Rechtsstetigkeit der Pfade sowie eine Stetigkeit im Anfangspunkt benötigt, die genau die Fellereigenschaft ist. Mit etwas Arbeit kann man daher den folgenden Satz zeigen (siehe etwa [156, Kapitel III.8ff] oder [154, Kapitel III, Theorem 2.7]).

Satz 21.27. Sei $(\kappa_t)_{t \geq 0}$ eine Feller'sche Halbgruppe auf dem lokalkompakten, polnischen Raum E . Dann existiert ein starker Markovprozess $(X_t)_{t \geq 0}$ mit RCLL Pfaden und Übergangskernen $(\kappa_t)_{t \geq 0}$.

Einen solchen Prozess X nennen wir auch einen **Feller-Prozess**.

Übung 21.4.1 (Doob'sche Ungleichung). Sei $X = (X_t)_{t \geq 0}$ ein Martingal oder nichtnegatives Submartingal mit RCLL Pfaden. Für $T \geq 0$ sei $|X|_T^* = \sup_{t \in [0, T]} |X_t|$.

Man zeige die Doob'schen Ungleichungen:

- (i) Für jedes $p \geq 1$ und $\lambda > 0$ gilt $\lambda^p \mathbf{P}[|X|_T^* \geq \lambda] \leq \mathbf{E}[|X_T|^p]$.
- (ii) Für jedes $p > 1$ gilt $\mathbf{E}[|X_T|^p] \leq \mathbf{E}[(|X|_T^*)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbf{E}[|X_T|^p]$.

Man zeige durch ein Gegenbeispiel, dass auf die Rechtsstetigkeit von X nicht ohne Weiteres verzichtet werden kann. ♣

Übung 21.4.2 (Martingalkonvergenzsätze). Sei X ein stochastischer Prozess mit RCLL Pfaden. Man zeige mit Hilfe der Doob'schen Ungleichung (Übung 21.4.1), dass die Martingalkonvergenzsätze (f.s. Konvergenz (Satz 11.4), f.s. und L^1 -Konvergenz für gleichgradig integrierbare Martingale (Satz 11.7) und der L^p -Martingalkonvergenzsatz (Satz 11.10)) sinngemäß für X gelten. ♣

Übung 21.4.3. Sei $p \geq 1$ und X^1, X^2, X^3, \dots p -fach integrierbare Martingale. Für jedes $t \geq 0$ gebe es ein $\tilde{X}_t \in \mathcal{L}^p(\mathbf{P})$ mit $X_t^n \xrightarrow{n \rightarrow \infty} \tilde{X}_t$ in L^p .

- (i) Zeige: $(\tilde{X}_t)_{t \geq 0}$ ist ein Martingal.
- (ii) Zeige mit Hilfe der Doob'schen Ungleichung: Ist $p > 1$ und sind X^1, X^2, \dots f.s. stetig, so gibt es ein stetiges Martingal X mit den Eigenschaften: X ist eine Modifikation von \tilde{X} und $X_t^n \xrightarrow{n \rightarrow \infty} X_t$ in L^p für jedes $t \geq 0$. ♣

Übung 21.4.4. Sei X ein stochastischer Prozess mit Werten in einem polnischen Raum E mit RCLL Pfaden, und sei $\mathbb{F} = \sigma(X)$ die von X erzeugte Filtration sowie $\mathbb{F}^+ := (\mathcal{F}_t^+)_{t \geq 0}$ definiert durch $\mathcal{F}_t^+ = \bigcap_{s > t} \mathcal{F}_s$. Sei $U \subset E$ offen und $C \subset E$ abgeschlossen. Für jede Menge $A \subset E$ sei $\tau_A := \inf\{t > 0 : X_t \in A\}$. Man zeige:

- (i) τ_C ist eine \mathbb{F} -Stoppzeit (und eine \mathbb{F}^+ -Stoppzeit).
- (ii) τ_U ist eine \mathbb{F}^+ -Stoppzeit, jedoch im Allgemeinen (selbst für stetiges X) keine \mathbb{F} -Stoppzeit. ♣

Übung 21.4.5. Man zeige die Aussage von Bemerkung 21.23 und folgere: Ist \mathbb{F} eine Filtration und B eine Brown'sche Bewegung, die ein \mathbb{F} -Martingal ist. Dann ist B auch ein $\mathbb{F}^{+,*}$ -Martingal. ♣

21.5 Konstruktion durch L^2 -Approximation

Wir geben eine funktionalanalytische Konstruktion der Brown'schen Bewegung durch eine L^2 -Approximation an. Der Einfachheit halber betrachten wir als Zeitintervall $[0, 1]$ statt $[0, \infty)$.

Es sei also $H = L^2([0, 1])$ der Hilbertraum der quadratintegrierbaren (bezüglich des Lebesgue-Maßes λ) Funktionen $[0, 1] \rightarrow \mathbb{R}$ mit Skalarprodukt

$$\langle f, g \rangle = \int_{[0, 1]} f(x)g(x) \lambda(dx)$$

und Norm $\|f\| = \sqrt{\langle f, f \rangle}$ (vergleiche Kapitel 7.3). Zwei Funktionen $f, g \in H$ werden als gleich angesehen, wenn $f = g$ λ -f.ü. Sei $(b_n)_{n \in \mathbb{N}}$ eine Orthonormalbasis (ONB) von H , also $\langle b_m, b_n \rangle = \mathbb{1}_{\{m=n\}}$ und

$$\lim_{n \rightarrow \infty} \left\| f - \sum_{m=1}^n \langle f, b_m \rangle b_m \right\| = 0 \quad \text{für jedes } f \in H.$$

Speziell gilt für jedes $f \in H$ die **Parseval'sche Gleichung**

$$\|f\|^2 = \sum_{m=1}^{\infty} \langle f, b_m \rangle^2 \tag{21.21}$$

und für $f, g \in H$

$$\langle f, g \rangle = \sum_{m=1}^{\infty} \langle f, b_m \rangle \langle g, b_m \rangle. \tag{21.22}$$

Betrachte jetzt eine u.i.v. Folge $(\xi_n)_{n \in \mathbb{N}}$ von $\mathcal{N}_{0,1}$ -Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$. Für $n \in \mathbb{N}$ und $t \in [0, 1]$ setze

$$X_t^n = \int \mathbb{1}_{[0,t]}(s) \left(\sum_{m=1}^n \xi_m b_m(s) \right) \lambda(ds) = \sum_{m=1}^n \xi_m \langle \mathbb{1}_{[0,t]}, b_m \rangle.$$

Offenbar ist für $n \geq m$

$$\begin{aligned} \mathbf{E}[(X_t^m - X_t^n)^2] &= \mathbf{E} \left[\left(\sum_{k=m+1}^n \xi_k \langle \mathbb{1}_{[0,t]}, b_k \rangle \right) \left(\sum_{l=m+1}^n \xi_l \langle \mathbb{1}_{[0,t]}, b_l \rangle \right) \right] \\ &= \sum_{k=m+1}^n \langle \mathbb{1}_{[0,t]}, b_k \rangle^2 \leq \sum_{k=m+1}^{\infty} \langle \mathbb{1}_{[0,t]}, b_k \rangle^2. \end{aligned}$$

Wegen $\sum_{k=1}^{\infty} \langle \mathbb{1}_{[0,t]}, b_k \rangle^2 = \|\mathbb{1}_{[0,t]}\|^2 = t < \infty$, gilt $X_t^n \in L^2(\mathbf{P})$ und

$$\lim_{m \rightarrow \infty} \sup_{n \geq m} \mathbf{E}[(X_t^m - X_t^n)^2] = 0.$$

Also ist $(X_t^n)_{n \in \mathbb{N}}$ eine Cauchy-Folge in $L^2(\mathbf{P})$ und hat wegen der Vollständigkeit von $L^2(\mathbf{P})$ (siehe Satz 7.3) einen L^2 -Grenzwert X_t . Offenbar gilt dann auch für $N \in \mathbb{N}$ und $0 \leq t_1, \dots, t_N \leq 1$

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[\sum_{i=1}^N (X_{t_i}^n - X_{t_i})^2 \right] = 0.$$

Speziell gilt also $(X_{t_1}^n, \dots, X_{t_N}^n) \xrightarrow{n \rightarrow \infty} (X_{t_1}, \dots, X_{t_N})$ \mathbf{P} -stochastisch.

Offenbar ist $(X_{t_1}^n, \dots, X_{t_N}^n)$ Gauß-verteilt und zentriert. Für $s, t \in [0, 1]$ gilt

$$\begin{aligned} \mathbf{Cov}[X_s^n, X_t^n] &= \mathbf{E} \left[\left(\sum_{k=1}^n \xi_k \langle \mathbb{1}_{[0,s]}, b_k \rangle \right) \left(\sum_{l=1}^n \xi_l \langle \mathbb{1}_{[0,t]}, b_l \rangle \right) \right] \\ &= \sum_{k,l=1}^n \mathbf{E}[\xi_k \xi_l] \langle \mathbb{1}_{[0,s]}, b_k \rangle \langle \mathbb{1}_{[0,t]}, b_l \rangle \\ &= \sum_{k=1}^n \langle \mathbb{1}_{[0,s]}, b_k \rangle \langle \mathbb{1}_{[0,t]}, b_k \rangle \\ &\xrightarrow{n \rightarrow \infty} \langle \mathbb{1}_{[0,s]}, \mathbb{1}_{[0,t]} \rangle = \min(s, t). \end{aligned}$$

Also ist $(X_t)_{t \in [0,1]}$ ein zentraler, Gauß'scher Prozess mit

$$\mathbf{Cov}[X_s, X_t] = \min(s, t). \quad (21.23)$$

Lévy-Konstruktion der Brown'schen Bewegung

Bis auf die Stetigkeit der Pfade ist X also eine Brown'sche Bewegung. Eine stetige Version von X liefert jetzt der Satz von Kolmogorov-Chentsov (Satz 21.6). Wir können X aber auch direkt als stetigen Prozess konstruieren, indem wir die ONB $(b_n)_{n \in \mathbb{N}}$ geschickt wählen, beispielsweise die **Haar-Funktionen** $b_{n,k}$: Sei $b_{0,1} \equiv 1$ und für $n \in \mathbb{N}$ und $k = 1, \dots, 2^{n-1}$ sei

$$b_{n,k}(t) = \begin{cases} 2^{(n-1)/2}, & \text{falls } \frac{2k-2}{2^n} \leq t < \frac{2k-1}{2^n}, \\ -2^{(n-1)/2}, & \text{falls } \frac{2k-1}{2^n} \leq t < \frac{2k}{2^n}, \\ 0, & \text{sonst.} \end{cases}$$

Dann ist $(b_{n,k})$ ein Orthonormalsystem: $\langle b_{m,k}, b_{n,l} \rangle = \mathbb{1}_{\{(m,k)=(n,l)\}}$. Man prüft ebenfalls leicht nach, dass $(b_{n,k})$ eine Basis ist (Übung!). Wir definieren die **Schau-funktionen** durch

$$B_{n,k}(t) = \int_{[0,t]} b_{n,k}(s) \lambda(ds) = \langle \mathbb{1}_{[0,t]}, b_{n,k} \rangle.$$

Seien $\xi_{0,1}, (\xi_{n,k})_{n \in \mathbb{N}, k=1, \dots, 2^{n-1}}$ unabhängig und $\mathcal{N}_{0,1}$ -verteilt. Wir setzen

$$X^n := \xi_{0,1} B_{0,1} + \sum_{m=1}^n \sum_{k=1}^{2^{m-1}} \xi_{m,k} B_{m,k}$$

und definieren \tilde{X}_t als den $L^2(\mathbf{P})$ -Limes $\tilde{X}_t = L^2 - \lim_{n \rightarrow \infty} X_t^n$.

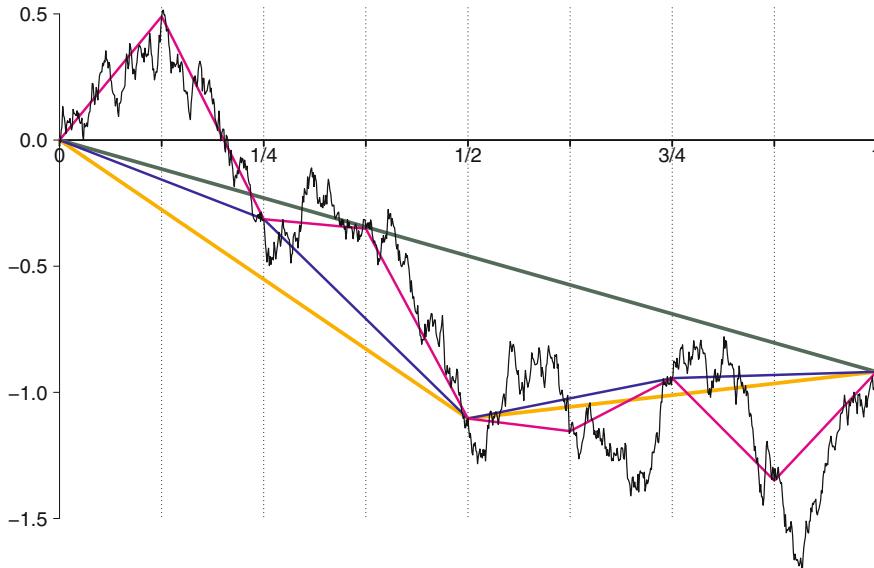


Abb. 21.2. Die Prozesse X^n , $n = 0, 1, 2, 3, 10$ der Lévy-Konstruktion der Brown'schen Bewegung.

Satz 21.28 (Brown'sche Bewegung, L^2 -Approximation). Es existiert eine stetige Modifikation X von \tilde{X} . X ist eine Brown'sche Bewegung, und es gilt

$$\lim_{n \rightarrow \infty} \|X^n - X\|_\infty = 0 \quad \mathbf{P}\text{-fast sicher.} \quad (21.24)$$

Beweis. Für $t \in [0, 1]$ folgt $X_t = \tilde{X}_t$ f.s. aus (21.24). Da gleichmäßige Limiten stetiger Funktionen wieder stetig sind, folgt aus (21.24) die Stetigkeit von X und aus (21.23) (zusammen mit Satz 21.11), dass X eine Brown'sche Bewegung ist. Es reicht also zu zeigen, dass es ein X gibt, sodass (21.24) gilt.

Da $(C([0, 1]), \|\cdot\|_\infty)$ vollständig ist, reicht es zu zeigen, dass \mathbf{P} -fast sicher X^n eine Cauchy-Folge in $(C([0, 1]), \|\cdot\|_\infty)$ ist. Man beachte, dass $\|B_{n,k}\|_\infty \leq 2^{-(n+1)/2}$, falls $n \in \mathbb{N}$, und $B_{n,k} B_{n,l} = 0$, falls $k \neq l$. Also ist

$$\|X^n - X^{n-1}\|_\infty \leq 2^{-(n+1)/2} \max\{|\xi_{n,k}|, k = 1, \dots, 2^{n-1}\}.$$

Mithin ist

$$\begin{aligned} \mathbf{P} \left[\|X^n - X^{n-1}\|_\infty > 2^{-n/4} \right] &\leq \sum_{k=1}^{2^{n-1}} \mathbf{P} \left[|\xi_{n,k}| > 2^{(n+2)/4} \right] \\ &= 2^{n-1} \frac{2}{\sqrt{2\pi}} \int_{2^{(n+2)/4}}^{\infty} e^{-x^2/2} dx \\ &\leq 2^n \exp(-2^{n/2}). \end{aligned}$$

Offenbar ist $\sum_{n=1}^{\infty} \mathbf{P}[\|X^n - X^{n-1}\|_\infty > 2^{-n/4}] < \infty$, also nach dem Lemma von Borel-Cantelli

$$\mathbf{P} \left[\|X^n - X^{n-1}\|_\infty > 2^{-n/4} \text{ höchstens endlich oft} \right] = 1.$$

Es folgt $\lim_{n \rightarrow \infty} \sup_{m \geq n} \|X^m - X^n\|_\infty = 0$ \mathbf{P} -fast sicher. \square

Brown'sche Bewegung und weißes Rauschen

Die Konstruktion der Brown'schen Bewegung über die Haar-Funktionen hatte den Vorteil, dass sie die Stetigkeit der Pfade ohne weiteren Aufwand lieferte. Für manche Anwendungen ist jedoch auch eine Frequenzzerlegung der Brown'schen Bewegung nützlich. Hier wird als Orthonormalbasis von $L^2([0, 1])$ die Folge der Cosinusfunktionen verwendet: $b_0 = 1$ und

$$b_n(x) = \sqrt{2} \cos(n\pi x) \quad \text{für } n \in \mathbb{N}.$$

Diese Funktionen bilden eine Orthonormalbasis von $L^2([0, 1])$.

Für $t \in [0, 1]$ und $n \in \mathbb{N}_0$ definieren wir

$$B_n(t) = \int_0^t b_n(s) \lambda(ds),$$

also $B_0(t) = t$ und

$$B_n(t) = \frac{\sqrt{2}}{n\pi} \sin(n\pi t) \quad \text{für } n \in \mathbb{N}.$$

Seien nun $\xi_n, n \in \mathbb{N}_0$, unabhängig und standardnormalverteilt. Wir definieren $A_0 = \xi_0$ und

$$A_n := \frac{\sqrt{2}}{\pi n} \xi_n \quad \text{für } n \in \mathbb{N}.$$

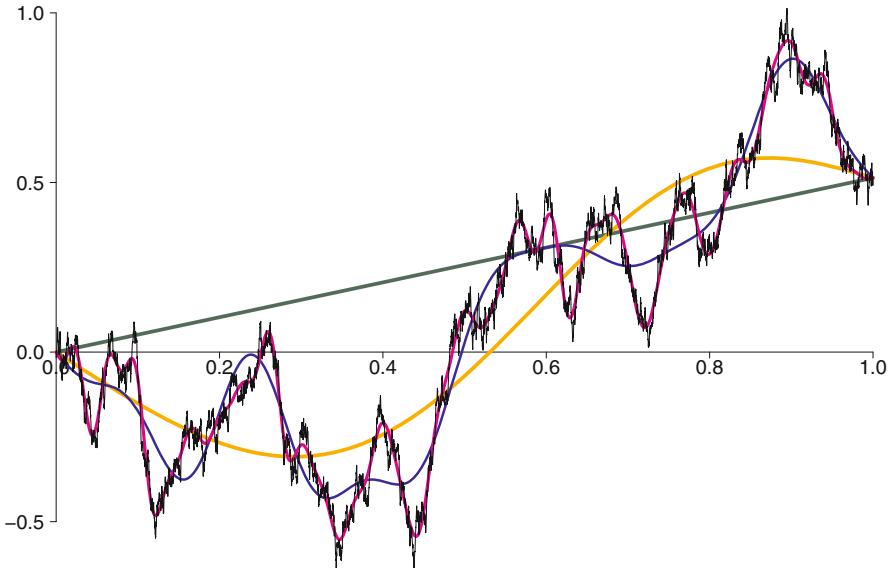


Abb. 21.3. Die Prozesse X^n , $n = 0, 1, 4, 64, 8192$ der Fourier-Konstruktion der Brown'schen Bewegung.

Schließlich setzen wir

$$X^n := \sum_{k=0}^n \xi_k B_k,$$

also

$$X^n(t) = \xi_0 t + \sum_{k=1}^n A_k \sin(k\pi t).$$

Wie oben gezeigt wurde, konvergiert (X^n) in $L^2([0, 1])$ gegen einen Prozess X , der (bis auf die Stetigkeit der Pfade) alle Eigenschaften der Brown'schen Bewegung hat:

$$X_t = \xi_0 t + \sum_{n=1}^{\infty} \frac{\sqrt{2}}{n\pi} \xi_n \sin(n\pi t).$$

Diese Darstellung geht auf Paley und Wiener zurück, die auch zeigen, dass die Reihe entlang einer Teilfolge fast sicher gleichmäßig konvergiert und daher der Limes X stetig ist, siehe [134, Theorem XLIII, Seite 148]. Manchmal wird sie auch als Karhunen-Loève Darstellung der Brown'schen Bewegung bezeichnet. Genauer gesagt, ist es aber (bis auf den ersten Summanden) die Karhunen-Loève Darstellung der Brown'schen Brücke $(X_t - tX_1)_{t \in [0, 1]}$ (vergleiche etwa [1, Kapitel 3.3]).

Wenn wir formal die Ableitung

$$\dot{X}_t := \frac{d}{dt} X_t = \xi_0 + \sqrt{2} \sum_{n=1}^{\infty} \xi_n \cos(n\pi t)$$

bilden, so sehen wir, dass die Fourierkoeffizienten aller Frequenzen unabhängig und identisch verteilt sind. Aus diesem Grunde wird das formale Objekt \dot{X} oft auch als **weißes Rauschen** bezeichnet (im Gegensatz zum farbigen Rauschen, bei dem die Koeffizienten der unterschiedlichen Frequenzen unterschiedliche Verteilungen haben).

Mit der Fourier-Basis ist die Stetigkeit des Grenzprozesses X sehr viel schwieriger nachzuweisen als mit der Haar-Basis. Das hinreichende Kriterium der Absolutsummierbarkeit der Koeffizienten (A_n) etwa schlägt fehl (siehe Übung 21.5.5).

Beispiel 21.29 (Stochastisches Integral à la Paley-Wiener). Wir nehmen an, dass $(\xi_n)_{n \in \mathbb{N}}$ eine u.i.v. Folge von $\mathcal{N}_{0,1}$ verteilten Zufallsvariablen ist sowie $(b_n)_{n \in \mathbb{N}}$ eine Orthonormalbasis von $L^2([0, 1])$, sodass $W_t := \lim_{n \rightarrow \infty} \sum_{k=1}^n \xi_k \langle \mathbf{1}_{[0,t]}, b_k \rangle$, $t \in [0, 1]$, eine Brown'sche Bewegung ist. Für $f \in L^2([0, 1])$ definieren wir

$$I(f) := \sum_{n=1}^{\infty} \xi_n \langle f, b_n \rangle.$$

Nach der Parseval'schen Gleichung und der Bienaymé Formel ist

$$\|f\|_2^2 = \sum_{n=1}^{\infty} \langle f, b_n \rangle^2 = \mathbf{Var}[I(f)] = \mathbf{E}[I(f)^2],$$

also gilt:

$$I : L^2([0, 1]) \rightarrow L^2(\mathbf{P}), \quad f \mapsto I(f) \quad \text{ist eine Isometrie.} \quad (21.25)$$

Wir nennen

$$\int_0^t f(s) dW_s := I(f \mathbf{1}_{[0,t]}), \quad t \in [0, 1], f \in L^2([0, 1]),$$

das **stochastische Integral** von f bezüglich W . Für den Fall der Fourier-Basis $b_0(x) = 1$ und $b_n(x) = \sqrt{2} \cos(n\pi x)$, $n \in \mathbb{N}$, geht diese Konstruktion auf Paley und Wiener [134, Theorem XLV, Seite 154] zurück.

Durch $X_t := \int_0^t f(s) dW_s$ wird ein zentrierter, Gauß'scher Prozess definiert mit Kovarianzfunktion

$$\mathbf{Cov}[X_s, X_t] = \int_0^{s \wedge t} f^2(u) du.$$

In der Tat ist klar, dass X zentriert und Gauß'sch ist (als Limes von Gauß'schen Partialsummenprozessen) mit der angegebenen Kovarianzfunktion. Ferner wie für die Brown'sche Bewegung mit den vierten Momenten der Zuwächse, die wir bei normalverteilten Zufallsvariablen aus den Varianzen berechnen können (vergleiche Satz 21.9), dass es eine stetige Modifikation von X gibt. Im Folgenden nehmen wir stets an, dass für das stochastische Integral diese stetige Modifikation gewählt wurde.

In dem Spezialfall $f = \sum_{i=1}^n \alpha_i \mathbb{1}_{(t_{i-1}, t_i]}$ für gewisses $n \in \mathbb{N}$ und $0 = t_0 < t_1 < \dots < t_n$ und $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ erhalten wir

$$\int_0^1 f(s) dW_s = \sum_{i=1}^n \alpha_i (W_{t_i} - W_{t_{i-1}}). \quad \diamond$$

Übung 21.5.1. Man zeige mit Hilfe der Darstellung der Brown'schen Bewegung $(W_t)_{t \in [0,1]}$ als zufällige Linearkombination der Schauder-Funktionen $(B_{n,k})$, dass die Brown'sche Brücke $Y = (Y_t)_{t \in [0,1]} = (W_t - tW_1)_{t \in [0,1]}$ ein stetiger, Gauß'scher Prozess mit Kovarianzfunktion $\text{Cov}[Y_t, Y_s] = (s \wedge t) - st$ ist. Man zeige ferner

$$\mathbf{P}_Y = \lim_{\varepsilon \downarrow 0} \mathbf{P}[W \in \cdot | W_1 \in (-\varepsilon, \varepsilon)]. \quad \clubsuit$$

Übung 21.5.2. (Vergleiche Beispiel 8.32.) Sei $T \in (0, 1)$ fest gewählt. Man zeige mit Hilfe einer Orthonormalbasis $b_{0,1}, (c_{n,k}), (d_{n,k})$ von geeignet modifizierten Haar-Funktionen (so, dass die $c_{n,k}$ von $[0, T]$ getragen werden und die $d_{n,k}$ von $[T, 1]$): Eine reguläre Version der bedingten Verteilung von W_T gegeben W_1 wird beschrieben durch

$$\mathbf{P}[W_T \in \cdot | W_1 = x] = \mathcal{N}_{Tx, T}. \quad \clubsuit$$

Übung 21.5.3. Definiere $Y := (Y_t)_{t \in [0,1]}$ durch $Y_1 = 0$ und

$$Y_t = (1-t) \int_0^t (1-s)^{-1} dW_s \quad \text{für } t \in [0, 1].$$

Man zeige, dass Y eine Brown'sche Brücke ist.

Hinweis: Zu zeigen ist, dass Y ein stetiger, Gauß'scher Prozess mit der richtigen Kovarianzfunktion ist. Speziell ist zu zeigen, dass $\lim_{t \uparrow 1} Y_t = 0$ fast sicher. \clubsuit

Übung 21.5.4. Sei $d \in \mathbb{N}$. Man zeige mit Hilfe einer geeigneten Orthonormalbasis auf $[0, 1]^d$:

(i) Es gibt einen Gauß'schen Prozess $(W_t)_{t \in [0,1]^d}$ mit Kovarianzfunktion

$$\text{Cov}[W_t, W_s] = \prod_{i=1}^d (t_i \wedge s_i).$$

(ii) Es existiert eine Modifikation von W , sodass $t \mapsto W_t$ fast sicher stetig ist (siehe Bemerkung 21.7).

Ein Prozess W mit den Eigenschaften (i) und (ii) heißt **Brown'sches Blatt**. \clubsuit

Übung 21.5.5. Man betrachte die Koeffizienten $(A_n)_{n \in \mathbb{N}_0}$ der Fourier-Basis bei der Konstruktion der Brown'schen Bewegung und zeige:

- (i) $\sum_{n=0}^{\infty} A_n^2 < \infty$ fast sicher.
- (ii) $\sum_{n=0}^{\infty} |A_n| = \infty$ fast sicher.
- (iii) $\sum_{k=0}^n A_k, n \in \mathbb{N}$ ist fast sicher konvergent.

Hinweis: Kolmogorov'scher Dreireihensatz (Satz 15.50). ♣

Übung 21.5.6. Sei $t \in (0, 1)$ und $f_0(x) := t$ sowie

$$f_n(x) := \frac{2 \sin(n\pi t)}{n\pi} \cos(n\pi x) \quad \text{für } n \in \mathbb{N}, x \in [0, 1].$$

Man zeige: $\sum_{n=0}^{\infty} f_n(x) = \mathbb{1}_{[0,t]}(x)$ für $x \in (0, 1) \setminus \{t\}$. ♣

21.6 Der Raum $C([0, \infty))$

Sind Funktionale, die vom ganzen Pfad der Brown'schen Bewegung abhängen, messbar? Ist beispielsweise $\sup\{X_t, t \in [0, 1]\}$ messbar? Für allgemeine stochastische Prozesse ist dies sicherlich falsch, weil das Supremum von mehr als abzählbar vielen Koordinaten abhängt. Für Prozesse mit stetigen Pfaden ist dies jedoch richtig, wie wir in diesem Abschnitt in allgemeinem Rahmen zeigen werden.

Es liegt nahe, dass man die Brown'sche Bewegung als kanonischen Prozess auf dem Raum $\Omega := C([0, \infty))$ der stetigen Pfade begreift.

Wir sammeln zunächst ein paar Eigenschaften von $\Omega = C([0, \infty)) \subset \mathbb{R}^{[0, \infty)}$. Wir definieren die **Auswertungsabbildung**

$$X_t : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto \omega(t), \tag{21.26}$$

also die Einschränkung der kanonischen Projektion $\mathbb{R}^{[0, \infty)} \rightarrow \mathbb{R}$ auf Ω .

Für $f, g \in C([0, \infty))$ und $n \in \mathbb{N}$ sei $d_n(f, g) := \left\| (f - g)|_{[0, n]} \right\|_{\infty} \wedge 1$ und

$$d(f, g) = \sum_{n=1}^{\infty} 2^{-n} d_n(f, g). \tag{21.27}$$

Satz 21.30. d ist eine vollständige Metrik auf $\Omega := C([0, \infty))$, die die Topologie der gleichmäßigen Konvergenz auf kompakten Mengen erzeugt. Der Raum (Ω, d) ist separabel, also polnisch.

Beweis. Offenbar ist jedes d_n eine vollständige Metrik auf $(C([0, n]), \| \cdot \|_\infty)$. Zu jeder Cauchy-Folge (f_N) in (Ω, d) und jedem $n \in \mathbb{N}$ existiert daher ein $g_n \in \Omega$ mit $d_n(f_N, g_n) \xrightarrow{N \rightarrow \infty} 0$. Offenbar ist $g_n(x) = g_m(x)$ für jedes $x \leq m \wedge n$, also existiert ein $g \in \Omega$ mit $g(x) = g_n(x)$ für jedes $x \leq n$ für jedes $n \in \mathbb{N}$. Offenbar gilt dann $d(f_N, g) \xrightarrow{N \rightarrow \infty} 0$, also ist d vollständig.

Die Menge der Polynome mit rationalen Koeffizienten ist abzählbar und nach dem Satz von Weierstraß dicht in jedem $(C([0, n]), \| \cdot \|_\infty)$ also dicht in (Ω, d) . \square

Satz 21.31. Beziiglich der Borel'schen σ -Algebra $\mathcal{B}(\Omega, d)$ sind die kanonischen Projektionen $X_t, t \in [0, \infty)$ messbar. Andererseits erzeugen die X_t schon $\mathcal{B}(\Omega, d)$. Es gilt also

$$(\mathcal{B}(\mathbb{R}))^{\otimes [0, \infty)}|_{\Omega} = \sigma(X_t, t \in [0, \infty)) = \mathcal{B}(\Omega, d).$$

Beweis. Die erste Gleichung gilt per definitionem. Für die zweite betrachten wir die gegenseitigen Inklusionen.

„ \subseteq “ Offenbar ist jedes $X_t : \Omega \rightarrow \mathbb{R}$ stetig, also $(\mathcal{B}(\Omega, d) - \mathcal{B}(\mathbb{R}))$ messbar. Mithin ist $\sigma(X_t, t \in [0, \infty)) \subset \mathcal{B}(\Omega, d)$.

„ \supseteq “ Wir müssen zeigen, dass offene Mengen aus (Ω, d) in $\mathcal{A} := (\mathcal{B}(\mathbb{R}))^{\otimes [0, \infty)}$ liegen. Da (Ω, d) separabel ist (Satz 21.30), ist jede offene Menge eine abzählbare Vereinigung von ε -Kugeln. Es reicht also zu zeigen, dass für festes $\omega_0 \in \Omega$ die Abbildung $\omega \mapsto Y(\omega) := d(\omega_0, \omega)$ messbar ist bezüglich \mathcal{A} .

Für $t > 0$ ist

$$\omega \mapsto Z_t := |X_t(\omega) - X_t(\omega_0)| \wedge 1$$

\mathcal{A} -messbar. Da ω stetig ist, ist für jedes $n \in \mathbb{N}$ (siehe (21.27))

$$\omega \mapsto Y_n(\omega) := d_n(\omega_0, \omega) = \sup_{t \in [0, n] \cap \mathbb{Q}} Z_t(\omega)$$

\mathcal{A} -messbar, und damit ist auch $Y = \sum_{n=1}^{\infty} 2^{-n} Y_n$ bezüglich \mathcal{A} . \square

Im Folgenden sei stets $\mathcal{A} := \sigma(X_t, t \in [0, \infty))$.

Korollar 21.32. $F_1 : \Omega \rightarrow [0, \infty), \omega \mapsto \sup\{\omega(t) : t \in [0, 1]\}$ ist \mathcal{A} -messbar.

Beweis. F_1 ist stetig bezüglich d , also $\mathcal{B}(\Omega, d)$ -messbar. \square

Ist B eine Brown'sche Bewegung (auf einem W-Raum $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbf{P}})$), so existiert $\overline{\Omega} \in \tilde{\mathcal{A}}$ mit $\tilde{\mathbf{P}}[\overline{\Omega}] = 1$ und $B(\omega) \in C([0, \infty))$ für jedes $\omega \in \overline{\Omega}$. Sei $\overline{\mathcal{A}} = \tilde{\mathcal{A}}|_{\overline{\Omega}}$ und $\overline{\mathbf{P}} = \tilde{\mathbf{P}}|_{\overline{\mathcal{A}}}$. Dann ist $B : \overline{\Omega} \rightarrow C([0, \infty))$ messbar bezüglich $(\overline{\mathcal{A}}, \mathcal{A})$. Beziiglich

des Bildmaßes $\mathbf{P} = \overline{\mathbf{P}} \circ B^{-1}$ auf $\Omega = C([0, \infty))$ ist der kanonische Prozess $X = (X_t, t \in [0, \infty))$ auf $C([0, \infty))$ eine Brown'sche Bewegung.

Definition 21.33. Sei \mathbf{P} das W-Maß auf $\Omega = C([0, \infty))$, bezüglich dessen der kanonische Prozess X eine Brown'sche Bewegung ist. Dann heißt \mathbf{P} **Wiener-Maß**. Das Tripel $(\Omega, \mathcal{A}, \mathbf{P})$ heißt **Wiener-Raum**, und X heißt **kanonische Brown'sche Bewegung** oder **Wiener-Prozess**.

Bemerkung 21.34. Manchmal soll eine Brown'sche Bewegung nicht in $X_0 = 0$ starten, sondern in einem beliebigen Punkt x . Mit \mathbf{P}_x bezeichnen wir dann dasjenige Maß auf $C([0, \infty))$, für das $\tilde{X} = (X_t - x, t \in [0, \infty))$ eine Brown'sche Bewegung (mit $\tilde{X}_0 = 0$) ist. \diamond

Übung 21.6.1. Man zeige: Die Abbildung $F_\infty : \Omega \rightarrow [0, \infty]$, die durch $\omega \mapsto \sup\{\omega(t) : t \in [0, \infty)\}$ definiert wird, ist \mathcal{A} -messbar. \clubsuit

21.7 Konvergenz von W-Maßen auf $C([0, \infty))$

Seien X und $(X^n)_{n \in \mathbb{N}}$ Zufallsvariablen mit Werten in $C([0, \infty))$, also stetige stochastische Prozesse, mit Verteilungen \mathbf{P}_X und $(\mathbf{P}_{X^n})_{n \in \mathbb{N}}$.

Definition 21.35. Wir sagen, dass die endlichdimensionalen Verteilungen (finite dimensional distributions) von (X^n) gegen die von X konvergieren, falls für jedes $k \in \mathbb{N}$ und $t_1, \dots, t_k \in [0, \infty)$ gilt

$$(X_{t_1}^n, \dots, X_{t_k}^n) \xrightarrow{n \rightarrow \infty} (X_{t_1}, \dots, X_{t_k}).$$

Wir schreiben dann $X^n \xrightarrow[\text{fdd}]{n \rightarrow \infty} X$ oder $\mathbf{P}_{X^n} \xrightarrow[\text{fdd}]{n \rightarrow \infty} \mathbf{P}_X$.

Lemma 21.36. Aus $P_n \xrightarrow[\text{fdd}]{n \rightarrow \infty} P$ und $P_n \xrightarrow[\text{fdd}]{n \rightarrow \infty} Q$ folgt $P = Q$.

Beweis. Nach Satz 14.12(iii) legen die endlichdimensionalen Verteilungen P eindeutig fest. \square

Satz 21.37. Schwache Konvergenz in $\mathcal{M}_1(\Omega, d)$ impliziert fdd-Konvergenz:

$$P_n \xrightarrow{n \rightarrow \infty} P \implies P_n \xrightarrow[\text{fdd}]{n \rightarrow \infty} P.$$

Beweis. Sei $k \in \mathbb{N}$ und $t_1, \dots, t_k \in [0, \infty)$. Die Abbildung

$$\varphi : C([0, \infty)) \rightarrow \mathbb{R}^k, \quad \omega \mapsto (\omega(t_1), \dots, \omega(t_k))$$

ist stetig. Nach dem Continuous Mapping Theorem (Satz 13.25 auf Seite 262) gilt $P_n \circ \varphi^{-1} \xrightarrow{n \rightarrow \infty} P \circ \varphi^{-1}$, also $P_n \xrightarrow[\text{fdd}]{n \rightarrow \infty} P$. \square

Die Umkehrung des Satzes ist nicht richtig. Es gilt aber Folgendes.

Satz 21.38. Seien $(P_n)_{n \in \mathbb{N}}$ und P W-Maße auf $C([0, \infty))$. Dann sind äquivalent:

- (i) $P_n \xrightarrow[n \rightarrow \infty]{\text{fdd}} P$ und $(P_n)_{n \in \mathbb{N}}$ ist straff.
- (ii) $P_n \xrightarrow[n \rightarrow \infty]{\text{fdd}} P$ schwach.

Beweis. „(ii) \implies (i)“ Dies folgt direkt aus dem Satz von Prohorov (Satz 13.29 mit $E = C([0, \infty))$).

„(i) \implies (ii)“ Nach dem Satz von Prohorov ist $(P_n)_{n \in \mathbb{N}}$ relativ folgenkompakt. Sei Q ein Limespunkt von $(P_{n_k})_{k \in \mathbb{N}}$ entlang einer Teilfolge (n_k) . Dann gilt $P_{n_k} \xrightarrow{\text{fdd}} Q, k \rightarrow \infty$. Nach Lemma 21.36 ist $P = Q$. \square

Als nächstes wollen wir uns ein nützliches Kriterium für Straffheit von Mengen $\{P_n\} \subset \mathcal{M}_1(C([0, \infty)))$ verschaffen. Wir beginnen mit einer Wiederholung der Charakterisierung von Relativkompaktheit in $C([0, \infty))$ von Arzelà und Ascoli (siehe etwa [179, Satz II.3.4]).

Für $N, \delta > 0$ und $\omega \in C([0, \infty))$ setze

$$V^N(\omega, \delta) := \sup \{ |\omega(t) - \omega(s)| : |t - s| \leq \delta, s, t \leq N \}.$$

Satz 21.39 (Arzelà-Ascoli). Eine Menge $A \subset C([0, \infty))$ ist genau dann relativ kompakt, wenn die beiden folgenden Bedingungen gelten.

- (i) $\{\omega(0), \omega \in A\} \subset \mathbb{R}$ ist beschränkt.
- (ii) Für jedes N gilt $\lim_{\delta \downarrow 0} \sup_{\omega \in A} V^N(\omega, \delta) = 0$.

Satz 21.40. Eine Familie $(P_i, i \in I)$ von W-Maßen auf $C([0, \infty))$ ist genau dann schwach relativkompakt, wenn die beiden folgenden Bedingungen gelten.

- (i) $(P_i \circ X_0^{-1}, i \in I)$ ist straff, das heißt, für jedes $\varepsilon > 0$ gibt es ein $K > 0$, sodass

$$P_i(\{\omega : |\omega(0)| > K\}) \leq \varepsilon \quad \text{für jedes } i \in I. \quad (21.28)$$

- (ii) Für alle $\eta, \varepsilon > 0$ und $N \in \mathbb{N}$ gibt es ein $\delta > 0$, sodass

$$P_i(\{\omega : V^N(\omega, \delta) > \eta\}) \leq \varepsilon \quad \text{für jedes } i \in I. \quad (21.29)$$

Beweis. „ \implies “ Nach dem Satz von Prohorov (Satz 13.29) folgt aus der schwachen Relativkompaktheit von $(P_i, i \in I)$ die Straffheit dieser Familie. Zu jedem $\varepsilon > 0$ gibt es daher eine kompakte Menge $A \subset C([0, \infty))$ mit $P_i(A) > 1 - \varepsilon$

für jedes $i \in I$. Aus der Charakterisierung der Kompaktheit von A im Satz von Arzelà-Ascoli folgen nun (i) und (ii).

„ \Leftarrow “ Wir nehmen jetzt an, dass (i) und (ii) gelten. Seien also für $\varepsilon > 0$ und $k, N \in \mathbb{N}$ die Zahlen K_ε und $\delta_{N,k,\varepsilon}$ so gewählt, dass

$$\sup_{i \in I} P_i(\{\omega : |\omega(0)| > K_\varepsilon\}) \leq \frac{\varepsilon}{2}$$

und

$$\sup_{i \in I} P_i\left(\left\{\omega : V^N(\omega, \delta_{N,k,\varepsilon}) > \frac{1}{k}\right\}\right) \leq 2^{-N-k-1} \varepsilon.$$

Setze

$$C_{N,\varepsilon} = \left\{ \omega : |\omega(0)| \leq K_\varepsilon, V^N(\omega, \delta_{N,k,\varepsilon}) \leq \frac{1}{k} \text{ für jedes } k \in \mathbb{N} \right\}.$$

Nach dem Satz von Arzelà-Ascoli ist $C_\varepsilon := \bigcap_{N \in \mathbb{N}} C_{N,\varepsilon}$ in $C([0, \infty))$ relativ kompakt, und wir haben

$$P_i(C_\varepsilon^c) \leq \frac{\varepsilon}{2} + \sum_{k,N=1}^{\infty} P_i(\{\omega : V^N(\omega, \delta_{N,k,\varepsilon}) > 1/k\}) \leq \varepsilon \quad \text{für jedes } i \in I.$$

Es folgt die Aussage. \square

Korollar 21.41. Sind $(X_i, i \in I)$ und $(Y_i, i \in I)$ Familien von Zufallsvariablen in $C([0, \infty))$, und sind $(\mathbf{P}_{X_i}, i \in I)$ und $(\mathbf{P}_{Y_i}, i \in I)$ straff, dann ist auch $(\mathbf{P}_{X_i+Y_i}, i \in I)$ straff.

Beweis. Wende die Dreiecksungleichung an, um im vorigen Satz (i) und (ii) nachzuweisen. \square

Ein wichtiges Hilfsmittel, um schwache Relativkompaktheit nachzuweisen, ist das folgende.

Satz 21.42 (Kolmogorov'sches Kriterium für schwache Relativkompaktheit).
Sei $(X^i, i \in I)$ eine Folge von stetigen stochastischen Prozessen. Es gelte:

(i) Die Familie $(\mathbf{P}[X_0^i \in \cdot], i \in I)$ der Startverteilungen ist straff.

(ii) Für jedes $N > 0$ gibt es $C, \alpha, \beta > 0$, sodass für alle $s, t \in [0, N]$ und jedes $i \in I$ gilt

$$\mathbf{E}[|X_s^i - X_t^i|^\alpha] \leq C |s - t|^{\beta+1}.$$

Dann ist die Familie $(\mathbf{P}_{X^i}, i \in I) = (\mathcal{L}[X^i], i \in I)$ von Verteilungen der X^i schwach relativkompakt in $\mathcal{M}_1(C([0, \infty)))$.

Beweis. Wir prüfen die Bedingungen von Satz 21.40. Die erste Bedingung aus Satz 21.40 ist genau (i).

Sei $N > 0$. Nach dem Satz von Kolmogorov-Chentsov (Satz 21.6(ii)) gibt es zu $\gamma \in (0, \beta/\alpha)$ und $\varepsilon > 0$ eine Konstante K , sodass für jedes $i \in I$ gilt

$$\mathbf{P} [|X_t^i - X_s^i| \leq K |t - s|^\gamma \text{ für alle } s, t \in [0, N]] \geq 1 - \varepsilon.$$

Offenbar impliziert dies (21.29) mit $\delta = (\eta/K)^{1/\gamma}$. \square

21.8 Satz von Donsker

Seien Y_1, Y_2, \dots u.i.v. Zufallsvariablen mit $\mathbf{E}[Y_1] = 0$ und $\mathbf{Var}[Y_1] = \sigma^2 > 0$. Für $t > 0$ sei $S_t^n = \sum_{i=1}^{\lfloor nt \rfloor} Y_i$ und $\tilde{S}_t^n = \frac{1}{\sqrt{\sigma^2 n}} S_t^n$. Nach dem zentralen Grenzwertsatz gilt $\mathcal{L}[\tilde{S}_t^n - \tilde{S}_s^n] \xrightarrow{n \rightarrow \infty} \mathcal{N}_{0, t-s}$ für $t > s \geq 0$. Bezeichnet $B = (B_t, t \geq 0)$ eine Brown'sche Bewegung, so gilt also

$$\mathcal{L}[\tilde{S}_t^n - \tilde{S}_s^n] \xrightarrow{n \rightarrow \infty} \mathcal{L}[B_t - B_s] \quad \text{für jedes } t > s \geq 0.$$

Für $N \in \mathbb{N}$ und $0 = t_0 < t_1 < \dots < t_N$ sind die Zufallsvariablen $\tilde{S}_{t_i}^n - \tilde{S}_{t_{i-1}}^n$, $i = 1, \dots, N$, unabhängig, also gilt auch

$$\mathcal{L}[(\tilde{S}_{t_1}^n - \tilde{S}_{t_0}^n, \dots, \tilde{S}_{t_N}^n - \tilde{S}_{t_{N-1}}^n)] \xrightarrow{n \rightarrow \infty} \mathcal{L}[(B_{t_1} - B_{t_0}, \dots, B_{t_N} - B_{t_{N-1}})]$$

und damit

$$\mathcal{L}[(\tilde{S}_{t_1}^n, \dots, \tilde{S}_{t_N}^n)] \xrightarrow{n \rightarrow \infty} \mathcal{L}[(B_{t_1}, \dots, B_{t_N})]. \quad (21.30)$$

Wir definieren jetzt \bar{S}^n wie \tilde{S}^n , aber linear interpoliert

$$\bar{S}_t^n = \frac{1}{\sqrt{\sigma^2 n}} \sum_{i=1}^{\lfloor nt \rfloor} Y_i + \frac{tn - \lfloor tn \rfloor}{\sqrt{\sigma^2 n}} Y_{\lfloor nt \rfloor + 1}. \quad (21.31)$$

Dann gilt für $\varepsilon > 0$

$$\mathbf{P} [|\tilde{S}_t^n - \bar{S}_t^n| > \varepsilon] \leq \varepsilon^{-2} \mathbf{E}[(\tilde{S}_t^n - \bar{S}_t^n)^2] \leq \frac{1}{\varepsilon^2 n} \frac{1}{\sigma^2} \mathbf{E}[Y_1^2] = \frac{1}{\varepsilon^2 n} \xrightarrow{n \rightarrow \infty} 0.$$

Nach dem Satz von Slutsky (Satz 13.18) gilt daher die Konvergenz der endlichdimensionalen Verteilungen gegen das Wiener-Maß \mathbf{P}_W :

$$\mathbf{P}_{\bar{S}^n} \xrightarrow[n \rightarrow \infty]{\text{fdd}} \mathbf{P}_W. \quad (21.32)$$

Wir wollen diese Konvergenzaussage verstärken zur schwachen Konvergenz der W-Maße auf $C([0, \infty))$. Dazu formulieren wir als Hauptsatz dieses Abschnitts den

Funktionalen Zentralen Grenzwertsatz, der in dieser Allgemeinheit auf Donsker [40] zurückgeht. Sätze von diesem Typ werden auch **Invarianzprinzipien** genannt, weil die Grenzverteilung die selbe ist für alle Verteilungen von Y_i mit Erwartungswert 0 und selber Varianz.

Satz 21.43 (Donsker'sches Invarianzprinzip). *Im Sinne der schwachen Konvergenz auf $C([0, \infty))$ konvergieren die Verteilungen von \bar{S}^n gegen das Wiener-Maß*

$$\mathcal{L}[\bar{S}^n] \xrightarrow{n \rightarrow \infty} \mathbf{P}_W. \quad (21.33)$$

Beweis. Wegen (21.32) und Satz 21.38 reicht es zu zeigen, dass $(\mathcal{L}[\bar{S}^n], n \in \mathbb{N})$ straff ist. Dafür möchten wir das Kolmogorov'sche Momentenkriterium anwenden. Wie wir schon beim Beweis der Existenz der Brown'schen Bewegung gesehen haben, reichen hierfür aber zweite Momente nicht aus, sondern wir benötigen vierte Momente, damit wir $\beta > 0$ wählen können. Die Strategie ist also, zunächst die Y_i abzuschneiden, um vierte Momente zu erhalten, und dann für den abgeschnittenen Teil und den Hauptteil separat Straffheit zu zeigen.

Für $K > 0$ definieren wir

$$Y_i^K := Y_i \mathbb{1}_{\{|Y_i| \leq K/2\}} - \mathbf{E}[Y_i \mathbb{1}_{\{|Y_i| \leq K/2\}}] \quad \text{und} \quad Z_i^K := Y_i - Y_i^K \quad \text{für } i \in \mathbb{N}.$$

Dann gilt $\mathbf{E}[Y_i^K] = \mathbf{E}[Z_i^K] = 0$ sowie $\mathbf{Var}[Z_i^K] \xrightarrow{K \rightarrow \infty} 0$ und $\mathbf{Var}[Y_i^K] \leq \sigma^2$, $i \in \mathbb{N}$. Außerdem ist offenbar $|Y_i^K| \leq K$ für jedes i . Setze

$$T_n^K := \sum_{i=1}^n Y_i^K \quad \text{und} \quad U_n^K := \sum_{i=1}^n Z_i^K \quad \text{für } n \in \mathbb{N}.$$

Es seien $\bar{T}_t^{K,n}$ und $\bar{U}_t^{K,n}$ die linearen Interpolationen von

$$\tilde{T}_t^{K,n} := \frac{1}{\sqrt{\sigma^2 n}} T_{\lfloor nt \rfloor}^K \quad \text{und} \quad \tilde{U}_t^{K,n} := \frac{1}{\sqrt{\sigma^2 n}} U_{\lfloor nt \rfloor}^K \quad \text{für } t \geq 0.$$

Offenbar ist $\bar{S}^n = \bar{T}^{K,n} + \bar{U}^{K,n}$. Nach Korollar 21.41 reicht es zu zeigen, dass für eine noch zu wählende Folge $(K_n)_{n \in \mathbb{N}}$ gilt: $(\mathcal{L}[\bar{U}^{K_n,n}], n \in \mathbb{N})$ und $(\mathcal{L}[\bar{T}^{K_n,n}], n \in \mathbb{N})$ sind straff.

Wir betrachten zunächst den Restterm. U^K ist ein Martingal. Die Doob'sche Ungleichung (Satz 11.2) liefert

$$\mathbf{P}\left[\sup_{l=1,\dots,n} |U_l^K| > \varepsilon \sqrt{n}\right] \leq \varepsilon^{-2} \mathbf{Var}[Z_1^K] \quad \text{für jedes } \varepsilon > 0.$$

Gilt jetzt $K_n \uparrow \infty$, $n \rightarrow \infty$, so haben wir für jedes $N > 0$

$$\mathbf{P}\left[\sup_{t \in [0,N]} |\bar{U}_t^{K_n,n}| > \varepsilon\right] \leq \frac{N}{\varepsilon^2 \sigma^2} \mathbf{Var}[Z_1^{K_n}] \xrightarrow{n \rightarrow \infty} 0,$$

also $\bar{U}^{K_n, n} \xrightarrow{n \rightarrow \infty} 0$ in $C([0, \infty))$. Speziell ist $(\mathcal{L}[\bar{U}^{K_n, n}], n \in \mathbb{N})$ straff.

Wir berechnen nun für $N > 0$ und $s, t \in [0, N]$ die vierten Momente der Differenzen $\bar{T}_{t+s}^{K_n, n} - \bar{T}_s^{K_n, n}$ des Hauptteils. Im Folgenden setzen wir $K_n = n^{1/4}$. Sei nun $n \in \mathbb{N}$ fest gewählt. Wir unterscheiden zwei Fälle.

Fall 1: $t < n^{-1}$. Sei $k := \lfloor (t+s)n \rfloor$. Ist $sn \geq k$, so ist

$$\bar{T}_{t+s}^{K_n, n} - \bar{T}_s^{K_n, n} = \frac{tn}{\sqrt{n\sigma^2}} Y_{k+1}^{K_n}.$$

Ist $sn < k$, so ist

$$\bar{T}_{t+s}^{K_n, n} - \bar{T}_s^{K_n, n} = \frac{1}{\sqrt{n\sigma^2}} \left(((t+s)n - k)Y_{k+1}^{K_n} + (k - sn)Y_k^{K_n} \right).$$

In beiden Fällen ist

$$|\bar{T}_{t+s}^{K_n, n} - \bar{T}_s^{K_n, n}| \leq \frac{t\sqrt{n}}{\sigma} (|Y_k^{K_n}| + |Y_{k+1}^{K_n}|),$$

also

$$\begin{aligned} \mathbf{E}\left[(\bar{T}_{t+s}^{K_n, n} - \bar{T}_s^{K_n, n})^4\right] &\leq \frac{n^2 t^4}{\sigma^4} (2K_n)^2 \mathbf{E}\left[\left(|Y_1^{K_n}| + |Y_2^{K_n}|\right)^2\right] \\ &\leq \frac{16n^{5/2} t^4}{\sigma^4} \mathbf{Var}[Y_1^{K_n}] \leq \frac{16}{\sigma^2} t^{3/2}. \end{aligned} \quad (21.34)$$

Fall 2: $t \geq n^{-1}$. Mit dem binomischen Lehrsatz erhalten wir (beachte, dass die gemischten Terme mit ungeraden Momenten wegfallen, wegen $\mathbf{E}[Y_1^{K_n}] = 0$)

$$\begin{aligned} \mathbf{E}\left[(T_n^{K_n})^4\right] &= n \mathbf{E}\left[(Y_1^{K_n})^4\right] + 3n(n-1) \mathbf{E}\left[(Y_1^{K_n})^2\right]^2 \\ &\leq n K_n^2 \sigma^2 + 3n(n-1) \sigma^4. \end{aligned} \quad (21.35)$$

Wir bemerken, dass für unabhängige, reelle Zufallsvariablen X, Y mit $\mathbf{E}[X] = \mathbf{E}[Y] = 0$ und $\mathbf{E}[X^4], \mathbf{E}[Y^4] < \infty$ sowie für $a \in [-1, 1]$ gilt:

$$\begin{aligned} \mathbf{E}\left[(aX + Y)^4\right] &= a^4 \mathbf{E}[X^4] + 6a^2 \mathbf{E}[X^2] \mathbf{E}[Y^2] + \mathbf{E}[Y^4] \\ &\leq \mathbf{E}[X^4] + 6 \mathbf{E}[X^2] \mathbf{E}[Y^2] + \mathbf{E}[Y^4] = \mathbf{E}[(X + Y)^4]. \end{aligned}$$

Wir wenden dies zweifach an (mit $a = \lceil (t+s)n \rceil - (t+s)n$ und $a = sn - \lfloor sn \rfloor$ und erhalten (mit der groben Abschätzung $\lceil (t+s)n \rceil - \lfloor sn \rfloor \leq tn + 2 \leq 3tn$) aus (21.35) (wegen $t \leq N$)

$$\begin{aligned}
\mathbf{E}[(\bar{T}_{t+s}^{K_n,n} - \bar{T}_s^{K_n,n})^4] &\leq n^{-2} \sigma^{-4} \mathbf{E}[(T_{\lceil(t+s)n\rceil}^{K_n} - T_{\lfloor sn \rfloor}^{K_n})^4] \\
&= n^{-2} \sigma^{-4} \mathbf{E}[(T_{\lceil(t+s)n\rceil}^{K_n} - T_{\lfloor sn \rfloor}^{K_n})^4] \\
&\leq \frac{3tnK_n^2}{n^2 \sigma^2} + 18t^2 = \frac{3}{\sigma^2} tn^{-1/2} + 18t^2 \\
&\leq \frac{3}{\sigma^2} t^{3/2} + 18t^2 \leq \left(\frac{3}{\sigma^2} + 18\sqrt{N}\right) t^{3/2}.
\end{aligned} \tag{21.36}$$

Nach (21.34) und (21.36) gibt es also zu jedem $N > 0$ eine Konstante $C = C(N, \sigma^2)$, sodass für jedes $n \in \mathbb{N}$ und alle $s, t \in [0, N]$ gilt

$$\mathbf{E}[(\bar{T}_{t+s}^{K_n,n} - \bar{T}_s^{K_n,n})^4] \leq C t^{3/2}.$$

Nach dem Kolmogorov'schen Momentenkriterium (Satz 21.42 mit $\alpha = 4$ und $\beta = 1/2$) ist also $(\mathcal{L}[\bar{T}^{K_n,n}], n \in \mathbb{N})$ straff in $\mathcal{M}_1(C([0, \infty)))$. \square

Übung 21.8.1. Seien X_1, X_2, \dots u.i.v. Zufallsvariablen mit stetiger Verteilungsfunktion F . Es sei $G_n : [0, 1] \rightarrow \mathbb{R}$, $t \mapsto n^{-1/2} \sum_{i=1}^n (\mathbb{1}_{[0,t]}(F(X_i)) - t)$ und $M_n := \|G_n\|_\infty$. Ferner sei $M = \sup_{t \in [0,1]} |B_t|$, wo B eine Brown'sche Brücke ist.

- (i) Man zeige $\mathbf{E}[G_n(t)] = 0$ und $\mathbf{Cov}[G_n(s), G_n(t)] = s \wedge t - st$ für $s, t \in [0, 1]$.
- (ii) Man zeige $\mathbf{E}[(G_n(t) - G_n(s))^4] \leq C((t-s)^2 + |t-s|/n)$ für ein $C > 0$.
- (iii) Man folgere, dass eine geeignete stetige Version von G_n schwach gegen B konvergiert. Beispielsweise kann

$$H_n(t) = n^{-1/2} \sum_{i=1}^n (h_n(F(X_i) - t) - g_n(t))$$

genommen werden, wo $g_n(t) := \int_0^t h_n(u-t) du$ für eine geeignete Glättung h_n von $\mathbb{1}_{(-\infty, 0]}$ ist, etwa $h_n(s) = 1 - (s/\varepsilon_n \vee 0) \wedge 1$ für eine geeignete Folge $\varepsilon_n \downarrow 0$.

- (iv) Man zeige schließlich $M_n \xrightarrow{n \rightarrow \infty} M$.

Bemerkung: Die Verteilung von M lässt sich durch die Formel von Kolmogorov-Smirnov ([108] und [166]) ausdrücken (siehe etwa [142]):

$$\mathbf{P}[M > x] = 2 \sum_{n=1}^{\infty} (-1)^{n-1} e^{-2n^2 x^2}. \tag{21.37}$$

Vergleiche hierzu auch (21.20). Mit Hilfe der Statistik M_n können Zufallsvariablen bei bekannter Verteilung auf Unabhängigkeit getestet werden. Seien X_1, X_2, \dots und $\tilde{X}_1, \tilde{X}_2, \dots$ unabhängige Zufallsvariablen mit unbekannten, stetigen Verteilungsfunktionen F und \tilde{F} und empirischen Verteilungsfunktionen F_n und \tilde{F}_n . Ferner sei

$$D_n := \sup_{t \in \mathbb{R}} |F_n(t) - \tilde{F}_n(t)|.$$

Unter der Annahme, dass $F = \tilde{F}$ gilt, konvergiert $\sqrt{n/2} D_n$ in Verteilung gegen M . Diese Tatsache ist Grundlage von nichtparametrischen Tests auf Verteilungsgleichheit. ♣

21.9 Pfadweise Konvergenz von Verzweigungsprozessen*

In diesem Abschnitt untersuchen wir die Konvergenz reskalierten Galton-Watson-Prozesse (Verzweigungsprozesse). Ähnlich wie für Summen unabhängiger Zufallsvariablen zeigen wir zunächst die Konvergenz zu einem festen Zeitpunkt gegen die Verteilungen eines Grenzprozesses. Hernach zeigen wir Konvergenz der endlichdimensionalen Verteilungen und schließlich mit Hilfe des Kolmogorov'schen Straffheitskriteriums die Konvergenz im Pfadraum $C([0, \infty))$.

Wir betrachten einen Galton-Watson-Prozess $(Z_n)_{n \in \mathbb{N}_0}$ mit geometrischer Nachkommenverteilung

$$p(k) = 2^{-k-1} \quad \text{für } k \in \mathbb{N}_0.$$

Das heißt, wir betrachten u.i.v. Zufallsvariablen $X_{n,i}$, $n, i \in \mathbb{N}_0$ auf \mathbb{N}_0 mit $\mathbf{P}[X_{n,i} = k] = p(k)$, $k \in \mathbb{N}_0$ und definieren, ausgehend vom Startzustand Z_0 , rekursiv

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}.$$

Z ist also eine Markovkette mit Übergangswahrscheinlichkeiten $p(i, j) = p^{*i}(j)$, wobei p^{*i} die i -te Faltungspotenz von p ist. Mit anderen Worten: Sind Z, Z^1, \dots, Z^i unabhängige Kopien des Galton-Watson-Prozesses, mit $Z_0 = i$ und $Z_0^1 = \dots = Z_0^i = 1$, so ist

$$Z \stackrel{\mathcal{D}}{=} Z^1 + \dots + Z^i. \quad (21.38)$$

Wir betrachten nun die Erzeugendenfunktion $\psi^{(1)}(s) := \psi(s) := \mathbf{E}[s^{X_{1,1}}]$ von $X_{1,1}$, $s \in [0, 1]$, und deren Iterierte $\psi^{(n)} := \psi^{(n-1)} \circ \psi$ für $n \in \mathbb{N}$. Dann ist nach Lemma 3.10 $\mathbf{E}_i[s^{Z_n}] = \mathbf{E}_1[s^{Z_n}]^i = (\psi^{(n)}(s))^i$. Für die geometrische Verteilung können wir $\psi^{(n)}$ leicht ausrechnen.

Lemma 21.44. *Für den Verzweigungsprozess mit kritischer, geometrischer Nachkommenverteilung ist die n -te Iterierte der Erzeugendenfunktion*

$$\psi^{(n)}(s) = \frac{n - (n-1)s}{n + 1 - ns}.$$

Beweis. Wir berechnen

$$\psi(s) = \sum_{k=0}^{\infty} 2^{-k-1} s^k = \frac{1}{-s+2}.$$

Um die Iterierten auszurechnen, betrachten wir zunächst allgemeine linear rationale Funktionen der Form $f(x) = \frac{ax+b}{cx+d}$. Für f von dieser Form definieren wir die Matrix $M_f = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Für zwei linear rationale Funktionen f und g ist $M_{f \circ g} = M_f \cdot M_g$. Wir berechnen leicht

$$M_\psi = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}, \quad M_\psi^2 = \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix}, \quad M_\psi^3 = \begin{pmatrix} -2 & 3 \\ -3 & 4 \end{pmatrix}$$

und induktiv

$$M_\psi^n = \begin{pmatrix} -(n-1) & n \\ -n & n+1 \end{pmatrix}. \quad \square$$

Setzen wir $s = e^{-\lambda}$, so erhalten wir die Laplace-Transformierten von Z_n

$$\mathbf{E}_i[e^{-\lambda Z_n}] = \psi^{(n)}(e^{-\lambda})^i.$$

Nach Beispiel 6.29 ergeben sich die Momente von Z_n durch Differenzieren. Es gilt also:

Lemma 21.45. *Die Momente von Z_n sind*

$$\mathbf{E}_i[Z_n^k] = (-1)^k \frac{d^k}{d\lambda^k} \left(\psi^{(n)}(e^{-\lambda})^i \right) \Big|_{\lambda=0}. \quad (21.39)$$

Speziell sind die ersten sechs Momente

$$\begin{aligned} \mathbf{E}_i[Z_n] &= i \\ \mathbf{E}_i[Z_n^2] &= 2i n + i^2, \\ \mathbf{E}_i[Z_n^3] &= 6i n^2 + 6i^2 n + i^3, \\ \mathbf{E}_i[Z_n^4] &= 24i n^3 + 36i^2 n^2 + (12i^3 + 2i) n + i^4, \\ \mathbf{E}_i[Z_n^5] &= 120i n^4 + 240i^2 n^3 + (120i^3 + 30i) n^2 + (20i^4 + 10i^2) n + i^5, \\ \mathbf{E}_i[Z_n^6] &= 720i n^5 + 1800i^2 n^4 + (1200i^3 + 360i) n^3, \\ &\quad + (300i^4 + 240i^2)n^2 + (30i^5 + 30i^3 + 2i)n + i^6. \end{aligned} \quad (21.40)$$

Insbesondere ist Z ein Martingal, und die ersten sechs zentrierten Momente sind

$$\begin{aligned} \mathbf{E}_i[(Z_n - i)^2] &= 2i n, \\ \mathbf{E}_i[(Z_n - i)^3] &= 6i n^2, \\ \mathbf{E}_i[(Z_n - i)^4] &= 24i n^3 + 12i^2 n^2 + 2i n, \\ \mathbf{E}_i[(Z_n - i)^5] &= 120i n^4 + 120i^2 n^3 + 30i n^2, \\ \mathbf{E}_i[(Z_n - i)^6] &= 720i n^5 + 1080i^2 n^4 + (120i^3 + 360i) n^3 + 60i^2 n^2 + 2i n. \end{aligned} \quad (21.41)$$

Beweis. Die genauen Formeln für die ersten sechs Momente erhält man durch stures Ausrechnen von (21.39). \square

Wir betrachten jetzt die folgende Reskalierung: Wir fixieren $x \geq 0$ und starten mit $Z_0 = \lfloor nx \rfloor$ Individuen und betrachten $\tilde{Z}_t^n := \frac{Z_{\lfloor nt \rfloor}}{n}$ für $t \geq 0$. Wir schreiben kurz

$$\mathcal{L}_x[\tilde{Z}^n] := \mathcal{L}_{\lfloor nx \rfloor}[(n^{-1} Z_{\lfloor nt \rfloor})_{t \geq 0}]. \quad (21.42)$$

Offenbar ist $\mathbf{E}_x[\tilde{Z}_t^n] = \frac{\lfloor nx \rfloor}{n} \leq x$ für jedes n , also ist $(\mathcal{L}_x[\tilde{Z}_t^n], n \in \mathbb{N})$ straff. Indem wir Laplace-Transformierte betrachten, sehen wir sogar, dass für jedes $\lambda \geq 0$ die Folge der Verteilungen konvergiert:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{E}_x[e^{-\lambda \tilde{Z}_t^n}] &= \lim_{n \rightarrow \infty} \left(\psi^{(\lfloor tn \rfloor)}(e^{-\lambda/n}) \right)^{nx} \\ &= \lim_{n \rightarrow \infty} \left(\frac{nt - (nt - 1)e^{-\lambda/n}}{nt + 1 - nt e^{-\lambda/n}} \right)^{nx} \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{1 - e^{-\lambda/n}}{n(1 - e^{-\lambda/n})t + 1} \right)^{nx} \\ &= \exp \left(- \lim_{n \rightarrow \infty} \frac{x n (1 - e^{-\lambda/n})}{n(1 - e^{-\lambda/n})t + 1} \right) \\ &= \exp \left(- \frac{\lambda}{\lambda + 1/t} (x/t) \right) := \psi_t(\lambda)^x. \end{aligned} \quad (21.43)$$

Die Funktion ψ_t^x ist aber die Laplace-Transformierte der zusammengesetzten Poisson-Verteilung $\text{CPoi}_{(x/t) \exp_{1/t}}$ (siehe Definition 16.3).

Wir betrachten jetzt den stochastischen Kern $\kappa_t(x, dy) := \text{CPoi}_{(x/t) \exp_{1/t}}(dy)$. Dies ist genau derjenige Kern auf $[0, \infty)$, dessen Laplace-Transformierte gegeben ist durch

$$\int_0^\infty \kappa_t(x, dy) e^{-\lambda y} = \psi_t(\lambda)^x. \quad (21.44)$$

Lemma 21.46. $(\kappa_t)_{t \geq 0}$ ist eine Markov'sche Halbgruppe, und es existiert ein Markovprozess $(Y_t)_{t \geq 0}$ mit Übergangskernen $\mathbf{P}_x[Y_t \in dy] = \kappa_t(x, dy)$.

Beweis. Es reicht, die Chapman-Kolmogorov Gleichung $\kappa_t \cdot \kappa_s = \kappa_{s+t}$ zu zeigen. Wir berechnen die Laplace-Transformierten dieser Kerne: Für $\lambda \geq 0$ erhalten wir durch zweimaliges Anwenden von (21.44)

$$\begin{aligned}
\int \int \kappa_t(x, dy) \kappa_s(y, dz) e^{-\lambda z} &= \int \kappa_t(x, dy) \exp \left(-\frac{\lambda y}{\lambda s + 1} \right) \\
&= \exp \left(-\frac{\frac{\lambda}{\lambda s + 1}}{\frac{\lambda}{\lambda s + 1} t + 1} x \right) \\
&= \exp \left(-\frac{\lambda x}{\lambda(t + s) + 1} \right) \\
&= \int \kappa_{t+s}(x, dz) e^{-\lambda z}.
\end{aligned}
\quad \square$$

Als nächstes zeigen wir, dass Y eine stetige Version besitzt. Dafür berechnen wir Momente und ziehen den Satz von Kolmogorov-Chentsov (Satz 21.6) heran.

Lemma 21.47. *Wir erhalten die k -ten Momente von Y_t durch Ableiten der Laplace-Transformierten*

$$\mathbf{E}_x[Y_t^k] = (-1)^k \frac{d^k}{d\lambda^k} (\psi(\lambda)^x) \Big|_{\lambda=0},$$

wobei $\psi_t(\lambda) = \exp \left(-\frac{\lambda}{\lambda t + 1} \right)$. Speziell sind die ersten Momente

$$\begin{aligned}
\mathbf{E}_x[Y_t] &= x, \\
\mathbf{E}_x[Y_t^2] &= 2x t + x^2, \\
\mathbf{E}_x[Y_t^3] &= 6x t^2 + 6x^2 t + x^3, \\
\mathbf{E}_x[Y_t^4] &= 24x t^3 + 36x^2 t^2 + 12x^3 t + x^4, \\
\mathbf{E}_x[Y_t^5] &= 120x t^4 + 240x^2 t^3 + 120x^3 t^2 + 20x^4 t + x^5, \\
\mathbf{E}_x[Y_t^6] &= 720x t^5 + 1800x^2 t^4 + 1200x^3 t^3 + 300x^4 t^2 + 30x^5 t + x^6.
\end{aligned} \tag{21.45}$$

Es ist also Y ein Martingal, und die ersten zentrierten Momente sind

$$\begin{aligned}
\mathbf{E}_x[(Y_t - x)^2] &= 2x t, \\
\mathbf{E}_x[(Y_t - x)^3] &= 6x t^2, \\
\mathbf{E}_x[(Y_t - x)^4] &= 24x t^3 + 12x^2 t^2, \\
\mathbf{E}_x[(Y_t - x)^5] &= 120x t^4 + 120x^2 t^3, \\
\mathbf{E}_x[(Y_t - x)^6] &= 720x t^5 + 1080x^2 t^4 + 120x^3 t^3.
\end{aligned} \tag{21.46}$$

Satz 21.48. *Es existiert eine stetige Version des Markovprozesses Y mit Übergangskernen $(\kappa_t)_{t \geq 0}$ gegeben durch (21.44). Diese Version nennen wir **Feller'sche Verzweigungsdiffusion** oder den **Feller'schen stetigen Verzweigungsprozess**.*

Beweis. Für festes $N > 0$ und $s, t \in [0, N]$ gilt

$$\begin{aligned}\mathbf{E}_x[(Y_{t+s} - Y_s)^4] &= \mathbf{E}_x[\mathbf{E}_{Y_s}[(Y_t - Y_0)^4]] = \mathbf{E}_x[24Y_s t^3 + 12Y_s^2 t^2] \\ &= 24x t^3 + 12(2sx + x^2) t^2 \leq (48Nx + 12x^2) t^2.\end{aligned}$$

Mithin erfüllt Y die Bedingung aus Satz 21.6 (Kolmogorov-Chentsov) mit $\alpha = 4$ und $\beta = 1$. \square

Bemerkung 21.49. (i) Indem man alle höheren Momente heranzieht, kann man zeigen, dass die Pfade von Y Hölder-stetig sind von jeder Ordnung $\gamma \in (0, \frac{1}{2})$.

(ii) Man kann zeigen, dass Y die (eindeutige, starke) Lösung der stochastischen (Itô'schen) Differentialgleichung (siehe Beispiele 26.11 und 26.31)

$$dY_t = \sqrt{2Y_t} dW_t \quad (21.47)$$

ist, wobei W eine Brown'sche Bewegung ist. \diamond

Satz 21.50. Es gilt $\mathcal{L}_x[\tilde{Z}^n] \xrightarrow[n \rightarrow \infty]{\text{fdd}} \mathcal{L}_x[Y]$.

Beweis. Wie in (21.43) erhalten wir für $0 \leq t_1 \leq t_2$ und $\lambda_1, \lambda_2 \geq 0$ sowie $x \geq 0$

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbf{E}_x[e^{-(\lambda_1 \tilde{Z}_{t_1}^n + \lambda_2 \tilde{Z}_{t_2}^n)}] &= \lim_{n \rightarrow \infty} \mathbf{E}_x\left[\mathbf{E}_x\left[e^{-\lambda_2 \tilde{Z}_{t_2}^n} \middle| \tilde{Z}_{t_1}^n\right] e^{-\lambda_1 \tilde{Z}_{t_1}^n}\right] \\ &= \lim_{n \rightarrow \infty} \mathbf{E}_x\left[\exp\left(-\frac{\lambda_2}{\lambda_2(t_2 - t_1) + 1} \tilde{Z}_{t_1}^n\right) e^{-\lambda_1 \tilde{Z}_{t_1}^n}\right] \\ &= \exp\left(-\frac{\left(\frac{\lambda_2}{\lambda_2(t_2 - t_1) + 1} + \lambda_1\right)x}{\left(\frac{\lambda_2}{\lambda_2(t_2 - t_1) + 1} + \lambda_1\right)t_1 + 1}\right) \\ &= \mathbf{E}_x[\exp(-(\lambda_1 Y_{t_1} + \lambda_2 Y_{t_2}))].\end{aligned}$$

Wir erhalten also

$$\mathcal{L}_x[\lambda_1 \tilde{Z}_{t_1}^n + \lambda_2 \tilde{Z}_{t_2}^n] \xrightarrow{n \rightarrow \infty} \mathcal{L}_x[\lambda_1 Y_{t_1} + \lambda_2 Y_{t_2}].$$

Nach der Cramér-Wold Device (Satz 15.56) folgt hieraus

$$\mathcal{L}_x[(\tilde{Z}_{t_1}^n, \tilde{Z}_{t_2}^n)] \xrightarrow{n \rightarrow \infty} \mathcal{L}_x[(Y_{t_1}, Y_{t_2})].$$

Wir können dieses Vorgehen jetzt iterieren und erhalten so für jedes $k \in \mathbb{N}$ und $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$

$$\mathcal{L}_x[(\tilde{Z}_{t_i}^n)_{i=1,\dots,k}] \xrightarrow{n \rightarrow \infty} \mathcal{L}_x[(Y_{t_i})_{i=1,\dots,k}].$$

Dies ist aber die Behauptung. \square

Wir zeigen nun, dass die Konvergenz sogar im Pfadraum gilt. Hierzu müssen wir den reskalierten Prozess noch stetig machen. Wir nehmen an, dass $(Z_i^n)_{i \in \mathbb{N}_0}, n \in \mathbb{N}$

eine Folge von Galton-Watson-Prozessen ist, mit $Z_0^n = \lfloor nx \rfloor$. Wir definieren die linearen Interpolationen

$$\bar{Z}_t^n := (t - n^{-1} \lfloor tn \rfloor) \left(Z_{\lfloor tn \rfloor + 1}^n - Z_{\lfloor tn \rfloor}^n \right) + \frac{1}{n} Z_{\lfloor tn \rfloor}^n.$$

Satz 21.51 (Lindvall (1972)). *Die reskalierten Galton-Watson-Prozesse \bar{Z}^n konvergieren für $n \rightarrow \infty$ gegen die Feller'sche Diffusion Y im Sinne der schwachen Konvergenz in $\mathcal{M}_1(C([0, \infty)))$:*

$$\mathcal{L}_x[\bar{Z}^n] \xrightarrow{n \rightarrow \infty} \mathcal{L}_x[Y].$$

Beweis. Die Konvergenz der endlichdimensionalen Verteilungen ist schon gezeigt. Nach Satz 21.38 reicht es, die Straffheit von $(\mathcal{L}_x[\bar{Z}^n], n \in \mathbb{N})$ in $\mathcal{M}_1(C([0, \infty)))$ zu zeigen. Hierzu verwenden wir das Kriterium von Kolmogorov (Satz 21.42 mit $\alpha = 4$ und $\beta = 1$). Wir berechnen also die vierten Momente $\mathbf{E}_x[(\bar{Z}_{t+s}^n - \bar{Z}_s^n)^4]$ für $s, t \in [0, N]$ und für festes $N > 0$. Wir unterscheiden zwei Fälle.

Fall 1: $t < \frac{1}{n}$. Sei $k = \lfloor (t+s)n \rfloor$. Wir nehmen zunächst an, dass $\lfloor sn \rfloor = k$. Dann ist (nach Lemma 21.45)

$$\begin{aligned} \mathbf{E}_x[(\bar{Z}_{t+s}^n - \bar{Z}_s^n)^4] &= n^{-4}(tn)^4 \mathbf{E}_{\lfloor nx \rfloor}[(Z_{k+1}^n - Z_k^n)^4] \\ &= t^4 \mathbf{E}_{\lfloor nx \rfloor}[24Z_k^n + 12(Z_k^n)^2 + 2Z_k^n] \\ &= t^4 (26\lfloor nx \rfloor + 24\lfloor nx \rfloor k + \lfloor nx \rfloor^2) \\ &\leq 26xt^3 + 24xst^2 + x^2t^2 \\ &\leq (50Nx + x^2)t^2. \end{aligned}$$

Der Fall $\lfloor sn \rfloor = k - 1$ liefert eine ähnliche Abschätzung. Insgesamt erhalten wir eine Konstante $C = C(N, x)$ mit

$$\mathbf{E}_x[(\bar{Z}_{s+t}^n - \bar{Z}_s^n)^4] \leq C t^2 \quad \text{für alle } s, t \in [0, N] \text{ mit } t < \frac{1}{n}. \quad (21.48)$$

Fall 2: $t \geq \frac{1}{n}$. Wir setzen jetzt $k := \lceil (t+s)n \rceil - \lfloor sn \rfloor \leq tn + 1 \leq 2tn$. Dann ist (nach Lemma 21.45)

$$\begin{aligned} &\mathbf{E}_x[(\bar{Z}_{t+s}^n - \bar{Z}_s^n)^4] \\ &\leq n^{-4} \mathbf{E}_{\lfloor nx \rfloor}[(Z_{\lceil (t+s)n \rceil}^n - Z_{\lfloor sn \rfloor}^n)^4] \\ &= n^{-4} \mathbf{E}_{\lfloor nx \rfloor}[\mathbf{E}_{Z_{\lfloor sn \rfloor}^n}[(Z_k^n - Z_0^n)^4]] \\ &= n^{-4} \mathbf{E}_{\lfloor nx \rfloor} \left[24Z_{\lfloor sn \rfloor}^n k^3 + 12(Z_{\lfloor sn \rfloor}^n)^2 k^2 + 2Z_{\lfloor sn \rfloor}^n k \right] \\ &\leq n^{-4} (24xn(2tn)^3 + (24xn sn + 12x^2n^2)(2tn)^2 + 4xtn^2) \\ &\leq 192xt^3 + (96xs + 48x^2)t^2 + 4xn^{-1}t^2 \\ &\leq (292Nx + 48x^2)t^2. \end{aligned} \quad (21.49)$$

Die Abschätzungen aus (21.48) und (21.49) ergeben zusammen, dass die Voraussetzungen des Kolmogorov'schen Straffheitskriteriums (Satz 21.42) erfüllt sind mit $\alpha = 4$ und $\beta = 1$. Also ist die Folge $(\mathcal{L}_x[\bar{Z}^n], n \in \mathbb{N})$ straff. \square

21.10 Quadratische Variation und lokale Martingale

Nach dem Satz von Paley-Wiener-Zygmund (Satz 21.17) sind die Pfade $t \mapsto W_t$ der Brown'schen Bewegung fast sicher nirgends differenzierbar, sind also von lokal unendlicher Variation. Insbesondere lässt sich das in Beispiel 21.29 betrachtete stochastische Integral $\int_0^1 f(s) dW_s$ nicht als Lebesgue-Stieltjes Integral verstehen. Um jedoch Integrale von diesem Typ auch für eine größere Klasse von Integranden und Integratoren zu definieren, wollen wir hier vorbereitend die Pfadeigenschaften der Brown'schen Bewegung, und allgemeiner von stetigen lokalen Martingalen, genauer untersuchen.

Definition 21.52. Sei $G : [0, \infty) \rightarrow \mathbb{R}$ stetig. Wir definieren für jedes $t \geq 0$ die **Variation bis t** durch

$$V_t^1(G) := \sup \left\{ \sum_{i=0}^{n-1} |G_{t_{i+1}} - G_{t_i}| : 0 = t_0 \leq t_1 \leq \dots \leq t_n = t, n \in \mathbb{N} \right\}.$$

Wir sagen, dass G von **lokal endlicher Variation** ist, falls $V_t^1(G) < \infty$ für alle $t \geq 0$ und schreiben \mathcal{C}_V für den Vektorraum der stetigen Funktionen G mit stetiger Variation $t \mapsto V_t^1(G)$.

Bemerkung 21.53. Offenbar gilt $V^1(F + G) \leq V^1(F) + V^1(G)$ und $V^1(\alpha G) = |\alpha| V^1(G)$ für alle stetigen $F, G : [0, \infty) \rightarrow \mathbb{R}$ und für alle $\alpha \in \mathbb{R}$. Also ist \mathcal{C}_V tatsächlich ein Vektorraum. \diamond

Bemerkung 21.54. (i) Ist G von der Form $G_t = \int_0^t f(s) ds$ für eine lokal integrierbare Funktion f , so ist $G \in \mathcal{C}_V$ mit $V_t^1(G) = \int_0^t |f(s)| ds$.

(ii) Ist $G = G^+ - G^-$ die Differenz zweier stetiger, monoton wachsender Funktionen G^+ und G^- , so ist

$$V_t^1(G) - V_s^1(G) \leq (G_t^+ - G_s^+) + (G_t^- - G_s^-) \quad \text{für alle } t > s, \quad (21.50)$$

also ist $G \in \mathcal{C}_V$. Gleichheit gilt in (21.50), wenn G^- und G^+ „nicht auf den selben Mengen wachsen“, also formal gesprochen die Verteilungsfunktionen gegenseitig singulärer Maße μ^- und μ^+ sind. Diese Maße μ^- und μ^+ sind dann die Jordan-Zerlegung des signierten Maßes $\mu = \mu^+ - \mu^-$, dessen Verteilungsfunktion G ist. Das **Lebesgue-Stieltjes Integral** wird dann definiert durch

$$\int_0^t F(s) dG_s := \int_{[0,t]} F d\mu^+ - \int_{[0,t]} F d\mu^- . \quad (21.51)$$

(iii) Ist $G \in \mathcal{C}_V$, so sind offenbar

$$G_t^+ := \frac{1}{2}(V_t^1(G) + G_t) \quad \text{und} \quad G_t^- := \frac{1}{2}(V_t^1(G) - G_t)$$

eine Zerlegung von G wie in (ii) beschrieben. \diamond

Dass die Pfade der Brown'schen Bewegung unendliche Variation haben, folgt schon aus ihrer Nichtdifferenzierbarkeit. Wir können dies aber auch leicht direkt einsehen.

Satz 21.55. *Sei W eine Brown'sche Bewegung. Dann gilt $V_t^1(W) = \infty$ fast sicher für jedes $t > 0$.*

Beweis. Es reicht, $t = 1$ zu betrachten und zu zeigen, dass

$$Y_n := \sum_{i=1}^{2^n} |W_{i2^{-n}} - W_{(i-1)2^{-n}}| \xrightarrow{n \rightarrow \infty} \infty \quad \text{f.s.} \quad (21.52)$$

Es ist $\mathbf{E}[Y_n] = 2^{n/2} \mathbf{E}[|W_1|] = 2^{n/2} \sqrt{2/\pi}$ und $\mathbf{Var}[Y_n] = 1 - 2/\pi$. Nach der Chebyshev'schen Ungleichung ist

$$\sum_{n=1}^{\infty} \mathbf{P}\left[Y_n \leq \frac{1}{2} 2^{n/2} \sqrt{2/\pi}\right] \leq \sum_{n=1}^{\infty} \frac{2\pi - 4}{2^n} = 2\pi - 4 < \infty.$$

Nach dem Lemma von Borel-Cantelli folgt (21.52). \square

Offenbar ist die Variation ein zu grobes Maß, um wesentliche Pfadeigenschaften der Brown'schen Bewegung zu messen. Wir wollen daher statt der Zuwächse (in der Definition der Variation) die (kleineren) quadratischen Zuwächse summieren. Für die Definition dieser *quadratischen* Variation ist etwas mehr Vorsicht nötig als in Definition 21.52 für die Variation.

Definition 21.56. Eine Folge $\mathcal{P} = (\mathcal{P}^n)_{n \in \mathbb{N}}$ abzählbarer Teilmengen von $[0, \infty)$

$$\mathcal{P}^n := \{t_0, t_1, t_2, \dots\} \quad \text{mit } 0 = t_0 < t_1 < t_2 < \dots$$

heißt **zulässige Zerlegungsfolge**, falls gilt:

- (i) $\mathcal{P}^1 \subset \mathcal{P}^2 \subset \dots$,
- (ii) $\sup \mathcal{P}^n = \infty$ für jedes $n \in \mathbb{N}$,

(iii) die **Feinheit**

$$|\mathcal{P}^n| := \sup_{t \in \mathcal{P}^n} \min_{s \in \mathcal{P}^n, s \neq t} |s - t|$$

geht für $n \rightarrow \infty$ gegen 0.

Sind $0 \leq S < T$, so schreiben wir

$$\mathcal{P}_{S,T}^n := \mathcal{P}^n \cap [S, T] \quad \text{und} \quad \mathcal{P}_T^n := \mathcal{P}^n \cap [0, T].$$

Ist $t = t_k \in \mathcal{P}_T^n$, so schreiben wir $t' := t_{k+1} \wedge T = \min \{s \in \mathcal{P}_T^n \cup \{T\} : s > t\}$.

Beispiel 21.57. $\mathcal{P}^n = \{k2^{-n} : k = 0, 1, 2, \dots\}$. ◇

Definition 21.58. Für stetige $F, G : [0, \infty) \rightarrow \mathbb{R}$ und $p \geq 1$ definieren wir die p -Variation von G (entlang \mathcal{P}) durch

$$V_T^p(G) := V_T^{\mathcal{P}, p}(G) := \lim_{n \rightarrow \infty} \sum_{t \in \mathcal{P}_T^n} |G_{t'} - G_t|^p \quad \text{für } T \geq 0,$$

falls der Grenzwert existiert. Speziell heißt $\langle G \rangle := V^2(G)$ die **quadratische Variation** von G . Ist $T \mapsto V_T^2(G)$ stetig, so schreiben wir $G \in \mathcal{C}_{\text{qV}} := \mathcal{C}_{\text{qV}}^{\mathcal{P}}$.

Existiert für jedes $T \geq 0$ der Grenzwert

$$V_T^{\mathcal{P}, 2}(F, G) := \lim_{n \rightarrow \infty} \sum_{t \in \mathcal{P}_T^n} (F_{t'} - F_t)(G_{t'} - G_t),$$

so nennen wir $\langle F, G \rangle := V^2(F, G) := V^{\mathcal{P}, 2}(F, G)$ die **quadratische Kovariation** von F und G (entlang \mathcal{P}).

Bemerkung 21.59. Ist $p' > p$ und $V_T^p(G) < \infty$, so ist $V_T^{p'}(G) = 0$. Speziell ist $\langle G \rangle \equiv 0$, falls G von lokal endlicher Variation ist. ◇

Bemerkung 21.60. Aufgrund der Dreiecksungleichung ist

$$\sum_{t \in \mathcal{P}_T^{n+1}} |G_{t'} - G_t| \geq \sum_{t \in \mathcal{P}_T^n} |G_{t'} - G_t| \quad \text{für alle } n \in \mathbb{N}, T \geq 0.$$

Daher existiert der Limes im Fall $p = 1$ stets und stimmt, unabhängig von der Zerlegungsfolge \mathcal{P} , mit $V^1(G)$ aus Definition 21.52 überein. Ähnliche Ungleichungen gelten für V^2 nicht, daher braucht der Limes nicht zu existieren oder kann von der Wahl von \mathcal{P} abhängen. Wir werden im Folgenden jedoch für die Pfade einer großen Klasse von stetigen stochastischen Prozessen zeigen, dass V^2 zumindest für eine geeignete Zerlegungsfolge fast sicher existiert und (unabhängig von der gewählten Zerlegungsfolge) fast sicher eindeutig ist. ◇

Bemerkung 21.61. (i) Existieren $\langle F + G \rangle_T$ und $\langle F - G \rangle_T$, so existiert die Kovarianz $\langle F, G \rangle_T$, und es gilt die **Polarisationsformel**

$$\langle F, G \rangle_T = \frac{1}{4} (\langle F + G \rangle_T - \langle F - G \rangle_T).$$

(ii) Existieren $\langle F \rangle_T$, $\langle G \rangle_T$ und $\langle F, G \rangle_T$, so folgt aus der Cauchy-Schwarz'schen Ungleichung für die approximierenden Summen

$$V_T^1(\langle F, G \rangle) \leq \sqrt{\langle F \rangle_T \langle G \rangle_T}. \quad \diamond$$

Bemerkung 21.62. Ist $f \in C^1(\mathbb{R})$ und $G \in \mathcal{C}_{\text{qV}}$, so ist (Übung!) im Sinne des Lebesgue-Stieltjes Integrals

$$\langle f(G) \rangle_T = \int_0^T (f'(G_s))^2 d\langle G \rangle_s. \quad \diamond$$

Korollar 21.63. Ist F von lokal endlicher quadratischer Variation und gilt $\langle G \rangle \equiv 0$ (speziell also, falls G von lokal endlicher Variation ist), so ist $\langle F, G \rangle \equiv 0$ und $\langle F + G \rangle = \langle F \rangle$.

Satz 21.64. Für die Brown'sche Bewegung W und jede zulässige Zerlegungsfolge gilt

$$\langle W \rangle_T = T \quad \text{für alle } T \geq 0 \quad \text{f.s.}$$

Beweis. Wir beweisen dies nur für den Fall, wo

$$\sum_{n=1}^{\infty} |\mathcal{P}^n| < \infty \quad (21.53)$$

gilt. Für den allgemeinen Fall skizzieren wir das Vorgehen.

Gelebt also (21.53). Falls $\langle W \rangle$ existiert, ist $T \mapsto \langle W \rangle_T$ monoton wachsend. Daher reicht es zu zeigen, dass $\langle W \rangle_T$ für jedes $T \in \mathbb{Q}^+ = \mathbb{Q} \cap [0, \infty)$ existiert und $\langle W \rangle_T = T$ fast sicher gilt. Da $(\tilde{W}_t)_{t \geq 0} = (T^{-1/2} W_{tT})_{t \geq 0}$ eine Brown'sche Bewegung ist und $\langle \tilde{W} \rangle_1 = T^{-1} \langle W \rangle_T$ gilt, reicht es, den Fall $T = 1$ zu betrachten.

Setze

$$Y_n := \sum_{t \in \mathcal{P}_1^n} (W_{t'} - W_t)^2 \quad \text{für alle } n \in \mathbb{N}.$$

Dann ist $\mathbf{E}[Y_n] = \sum_{t \in \mathcal{P}_1^n} (t' - t) = 1$ und

$$\mathbf{Var}[Y_n] = \sum_{t \in \mathcal{P}_1^n} \mathbf{Var}[(W_{t'} - W_t)^2] = 2 \sum_{t \in \mathcal{P}_1^n} (t' - t)^2 \leq 2 |\mathcal{P}^n|.$$

Nach Voraussetzung (21.53) gilt also $\sum_{n=1}^{\infty} \mathbf{Var}[Y_n] \leq 2 \sum_{n=1}^{\infty} |\mathcal{P}^n| < \infty$, also gilt $Y_n \xrightarrow{n \rightarrow \infty} 1$ fast sicher.

Verzichten wir auf die Bedingung (21.53), so gilt immer noch $\mathbf{Var}[Y_n] \xrightarrow{n \rightarrow \infty} 0$, also $Y_n \xrightarrow{n \rightarrow \infty} 1$ stochastisch. Es ist allerdings nicht zu schwer zu zeigen, dass $(Y_n)_{n \in \mathbb{N}}$ ein Rückwärtsmartingal ist (siehe etwa [149, Theorem I.28]) und daher fast sicher gegen 1 konvergiert. \square

Korollar 21.65. Sind W und \widetilde{W} unabhängige Brown'sche Bewegungen, so gilt $\langle W, \widetilde{W} \rangle_T = 0$.

Beweis. Die stetigen Prozesse $((W + \widetilde{W})/\sqrt{2})$ und $((W - \widetilde{W})/\sqrt{2})$ haben unabhängige, normalverteilte Zuwächse, sind also Brown'sche Bewegungen. Nach Bemerkung 21.61(i) gilt

$$\begin{aligned} 4\langle W, \widetilde{W} \rangle_T &= \langle W + \widetilde{W} \rangle_T - \langle W - \widetilde{W} \rangle_T \\ &= 2\langle (W + \widetilde{W})/\sqrt{2} \rangle_T - 2\langle (W - \widetilde{W})/\sqrt{2} \rangle_T = 2T - 2T = 0. \end{aligned} \quad \square$$

Nach Übung 21.4.2 ist $(W_t^2 - t)_{t \geq 0}$ ein stetiges Martingal. Offenbar ist auch $(W_t \widetilde{W}_t)_{t \geq 0}$ ein stetiges Martingal. Nach dem Gezeigten sind also die Prozesse $W^2 - \langle W \rangle$ und $W\widetilde{W} - \langle W, \widetilde{W} \rangle$ Martingale. Wir werden sehen (Satz 21.70), dass die quadratische Variation $\langle M(\omega) \rangle$ eines quadratintegrierbaren, stetigen Martingals M stets existiert (für fast alle ω), und dass der Prozess $\langle M \rangle$ eindeutig charakterisiert ist durch die Eigenschaft, dass $M^2 - \langle M \rangle$ ein Martingal ist.

Um eine ähnliche Aussage auch für stetige Martingale zu erhalten, die nicht quadratisch integrierbar sind, treffen wir die folgende Definition.

Definition 21.66 (Lokales Martingal). Sei \mathbb{F} eine Filtration auf $(\Omega, \mathcal{F}, \mathbf{P})$ und τ ein \mathbb{F} -Stoppzeit. Ein adaptierter, reeller stochastischer Prozess $M = (M_t)_{t \geq 0}$ heißt **lokales Martingal bis τ** , falls es eine Folge $(\tau_n)_{n \in \mathbb{N}}$ von Stoppzeiten gibt mit $\tau_n \uparrow \tau$ fast sicher, und so, dass für jedes $n \in \mathbb{N}$ der gestoppte Prozess $M^{\tau_n} = (M_{\tau_n \wedge t})_{t \geq 0}$ ein gleichgradig integrierbares Martingal ist. Eine solche Folge $(\tau_n)_{n \in \mathbb{N}}$ heißt **lokalisierende Folge** für M . M heißt **lokales Martingal schlechthin**, falls M ein lokales Martingal bis $\tau \equiv \infty$ ist. Mit $\mathcal{M}_{loc,c}$ bezeichnen wir den Raum der stetigen lokalen Martingale.

Bemerkung 21.67. Sei M ein stetiger, adaptierter Prozess und τ eine Stoppzeit. Dann sind äquivalent:

- (i) M ist ein lokales Martingal bis τ .
- (ii) Es gibt eine Folge $(\tau_n)_{n \in \mathbb{N}}$ von Stoppzeiten mit $\tau_n \uparrow \tau$ fast sicher, sodass jedes M^{τ_n} ein Martingal ist.

- (iii) Es gibt eine Folge $(\tau_n)_{n \in \mathbb{N}}$ von Stoppzeiten mit $\tau_n \uparrow \tau$ fast sicher, sodass jedes M^{τ_n} ein beschränktes Martingal ist.

In der Tat: (iii) \implies (i) \implies (ii) ist trivial. Gelte also (ii), und sei τ'_n definiert durch

$$\tau'_n := \inf\{t \geq 0 : |M_t| \geq n\} \quad \text{für jedes } n \in \mathbb{N}.$$

Da M stetig ist, gilt $\tau'_n \uparrow \infty$. Also ist $(\sigma_n)_{n \in \mathbb{N}} := (\tau_n \wedge \tau'_n)_{n \in \mathbb{N}}$ eine lokalisierende Folge für M , sodass jedes M^{σ_n} ein beschränktes Martingal ist. \diamond

Bemerkung 21.68. Ein beschränktes lokales Martingal M ist stets auch ein Martingal. In der Tat: Sei $|M_t| \leq C < \infty$ fast sicher für alle $t \geq 0$ und $(\tau_n)_{n \in \mathbb{N}}$ eine lokalisierende Folge für M . Seien $t > s \geq 0$ und $A \in \mathcal{F}_s$. Dann ist $A \cap \{\tau_n \leq s\} \in \mathcal{F}_{\tau_n \wedge s}$, also

$$\mathbf{E}[M_{\tau_n \wedge t} \mathbb{1}_{A \cap \{\tau_n \leq s\}}] = \mathbf{E}[M_{\tau_n \wedge s} \mathbb{1}_{A \cap \{\tau_n \leq s\}}].$$

Wegen $\tau_n \uparrow \infty$, liefert der Satz von der majorisierten Konvergenz (Korollar 6.26)

$$\mathbf{E}[M_t \mathbb{1}_A] = \mathbf{E}[M_s \mathbb{1}_A].$$

Also ist $\mathbf{E}[M_t | \mathcal{F}_s] = M_s$ und damit M ein Martingal. \diamond

Beispiel 21.69. (i) Ein Martingal ist offenbar stets ein lokales Martingal.

(ii) In Bemerkung 21.68 hatten wir gesehen, dass beschränkte lokale Martingale auch Martingale sind. Andererseits ist selbst ein gleichgradig integrierbares lokales Martingal nicht notwendigerweise ein Martingal: Sei $W = (W^1, W^2, W^3)$ eine dreidimensionale Brown'sche Bewegung (das heißt, W^1, W^2 und W^3 sind unabhängige Brown'sche Bewegungen) mit Start in $W_0 = x \in \mathbb{R}^3 \setminus \{0\}$. Sei

$$u(y) = \|y\|^{-1} \quad \text{für } y \in \mathbb{R}^3 \setminus \{0\}.$$

Man prüft leicht nach, dass u harmonisch ist, dass also $\Delta u(y) = 0$ ist für alle $y \neq 0$. Wir werden später sehen (Korollar 25.34 zur Itô-Formel), dass hieraus folgt, dass $M := (u(W_t))_{t \geq 0}$ ein lokales Martingal ist. Durch

$$\tau_n := \inf\{t > 0 : M_t \geq n\} = \inf\{t > 0 : \|W_t\| \leq 1/n\}, \quad n \in \mathbb{N},$$

wird eine lokalisierende Folge für M definiert. Andererseits liefert eine explizite Rechnung mit der dreidimensionalen Normalverteilung $\mathbf{E}[M_t] \leq t^{-1/2} \xrightarrow{t \rightarrow \infty} 0$, also ist M integrierbar aber kein Martingal. Wegen $M_t \xrightarrow{t \rightarrow \infty} 0$ in L^1 ist M sogar gleichgradig integrierbar. \diamond

Satz 21.70. Sei M ein stetiges lokales Martingal.

(i) Es existiert ein eindeutig bestimmter, stetiger, monoton wachsender, adaptierter Prozess $\langle M \rangle = (\langle M \rangle_t)_{t \geq 0}$ mit $\langle M \rangle_0 = 0$, sodass gilt:

$$(M_t^2 - \langle M \rangle_t)_{t \geq 0} \quad \text{ist ein stetiges lokales Martingal.}$$

(ii) Ist M ein stetiges, quadratisch integrierbares Martingal, so ist $M^2 - \langle M \rangle$ ein Martingal.

(iii) Für jede zulässige Zerlegungsfolge $\mathcal{P} = (\mathcal{P}^n)_{n \in \mathbb{N}}$ gilt

$$U_T^n := \sum_{t \in \mathcal{P}_T^n} (M_{t'} - M_t)^2 \xrightarrow{n \rightarrow \infty} \langle M \rangle_T \quad \text{stochastisch} \quad \text{für alle } T \geq 0.$$

Der Prozess $\langle M \rangle$ heißt **quadratischer Variationsprozess** von M .

Bemerkung 21.71. Indem wir in (iii) gegebenenfalls zu einer (von T abhängigen) Teilfolge \mathcal{P}' übergehen, können wir annehmen, dass $U_T^n \xrightarrow{n \rightarrow \infty} \langle M \rangle_T$ fast sicher gilt. Durch ein Diagonalfolgenargument erhalten wir (wie im Beweis des Satzes von Helly) eine Zerlegungsfolge, sodass $U_T^n \xrightarrow{n \rightarrow \infty} \langle M \rangle_T$ fast sicher für alle $T \in \mathbb{Q}^+$ gilt. Aufgrund der Monotonie und der Stetigkeit von $T \mapsto U_T^n$ und $T \mapsto \langle M \rangle_T$ folgt $U_T^n \xrightarrow{n \rightarrow \infty} \langle M \rangle_T$ für alle $T \geq 0$ fast sicher. Also ist für diese Zerlegungsfolge die pfadweise definierte quadratische Variation fast sicher gleich dem quadratischen Variationsprozess:

$$\langle M(\omega) \rangle = V^2(M(\omega)) = \langle M \rangle(\omega). \quad \diamond$$

Beweis (von Satz 21.70). **Schritt 1.** Sei zunächst $|M_t| \leq C$ fast sicher für alle $t \geq 0$ für ein $C < \infty$. Insbesondere ist M dann ein Martingal (nach Bemerkung 21.68). Es gilt $U_T^n = M_T^2 - M_0^2 - N_T^n$, wobei

$$N_T^n = 2 \sum_{t \in \mathcal{P}_T^n} M_t (M_{t'} - M_t), \quad T \geq 0,$$

ein stetiges Martingal ist. Wenn wir zeigen können, dass $(U_T^n)_{n \in \mathbb{N}}$ für jedes $T \geq 0$ eine Cauchy-Folge in $L^2(\mathbf{P})$ ist, so ist auch $(N_T^n)_{n \in \mathbb{N}}$ eine Cauchy-Folge, und wir können den L^2 -Limes \tilde{N}_T von $(N_T^n)_{n \in \mathbb{N}}$ definieren. Nach Übung 21.4.3 hat \tilde{N} eine stetige Modifikation N , und es gilt $N_T^n \xrightarrow{n \rightarrow \infty} N_T$ in L^2 für alle $T \geq 0$. Mithin existiert ein stetiger Prozess $\langle M \rangle$ mit

$$U_T^n \xrightarrow{n \rightarrow \infty} \langle M \rangle_T \quad \text{in } L^2 \quad \text{für alle } T \geq 0, \quad (21.54)$$

und $N = M^2 - M_0^2 - \langle M \rangle$ ist ein stetiges Martingal.

Wir zeigen nun also für $T \geq 0$

$$(U_T^n)_{n \in \mathbb{N}} \text{ ist eine Cauchy-Folge in } L^2. \quad (21.55)$$

Für $m \in \mathbb{N}$ sei

$$Z_m := \max \left\{ (M_t - M_s)^2 : s \in \mathcal{P}_T^m, t \in \mathcal{P}_{s,s'}^n, n \geq m \right\}.$$

Da M fast sicher gleichmäßig stetig auf $[0, T]$ ist, gilt $Z_m \xrightarrow{m \rightarrow \infty} 0$ fast sicher.
Wegen $Z_m \leq 4C^2$ folgt

$$\mathbf{E}[Z_m^2] \xrightarrow{m \rightarrow \infty} 0. \quad (21.56)$$

Für $n \in \mathbb{N}$ und Zahlen a_0, \dots, a_n gilt

$$(a_n - a_0)^2 - \sum_{k=0}^{n-1} (a_{k+1} - a_k)^2 = 2 \sum_{k=0}^{n-1} (a_k - a_0)(a_{k+1} - a_k).$$

Wenden wir dies in der folgenden Gleichung auf jeden einzelnen Summanden der äußeren Summe an, so erhalten wir für $m \in \mathbb{N}$ und $n \geq m$

$$\begin{aligned} U_T^m - U_T^n &= \sum_{s \in \mathcal{P}_T^m} \left((M_{s'} - M_s)^2 - \sum_{t \in \mathcal{P}_{s,s'}^n} (M_{t'} - M_t)^2 \right) \\ &= 2 \sum_{s \in \mathcal{P}_T^m} \sum_{t \in \mathcal{P}_{s,s'}^n} (M_t - M_s)(M_{t'} - M_t). \end{aligned} \quad (21.57)$$

Da M ein Martingal ist, ist für $s_1, s_2 \in \mathcal{P}_T^m$ und $t_1 \in \mathcal{P}_{s_1,s'_1}^n$, $t_2 \in \mathcal{P}_{s_2,s'_2}^n$ mit $t_1 < t_2$

$$\begin{aligned} &\mathbf{E}\left[(M_{t_1} - M_{s_1})(M_{t'_1} - M_{t_1})(M_{t_2} - M_{s_2})(M_{t'_2} - M_{t_2})\right] \\ &= \mathbf{E}\left[(M_{t_1} - M_{s_1})(M_{t'_1} - M_{t_1})(M_{t_2} - M_{s_2}) \mathbf{E}[M_{t'_2} - M_{t_2} | \mathcal{F}_{t_2}]\right] = 0. \end{aligned}$$

Wenn wir mit Hilfe von (21.57) den Erwartungswert von $(U_T^m - U_T^n)^2$ berechnen, fallen also die gemischten Terme weg, und wir erhalten (unter Benutzung der Cauchy-Schwarz'schen Ungleichung in der dritten Zeile)

$$\begin{aligned} \mathbf{E}[(U_T^n - U_T^m)^2] &= 4 \mathbf{E}\left[\sum_{s \in \mathcal{P}_T^m} \sum_{t \in \mathcal{P}_{s,s'}^n} (M_t - M_s)^2 (M_{t'} - M_t)^2\right] \\ &\leq 4 \mathbf{E}\left[Z_m \sum_{t \in \mathcal{P}_T^n} (M_{t'} - M_t)^2\right] \\ &\leq 4 \mathbf{E}[Z_m^2]^{1/2} \mathbf{E}\left[\left(\sum_{t \in \mathcal{P}_T^n} (M_{t'} - M_t)^2\right)^2\right]^{1/2}. \end{aligned} \quad (21.58)$$

Wir schreiben den zweiten Faktor als

$$\begin{aligned} \mathbf{E} \left[\left(\sum_{t \in \mathcal{P}_T^n} (M_{t'} - M_t)^2 \right)^2 \right] &= \mathbf{E} \left[\sum_{t \in \mathcal{P}_T^n} (M_{t'} - M_t)^4 \right] \\ &\quad + 2 \mathbf{E} \left[\sum_{s \in \mathcal{P}_T^n} (M_{s'} - M_s)^2 \sum_{t \in \mathcal{P}_{s', T}^n} (M_{t'} - M_t)^2 \right]. \end{aligned} \tag{21.59}$$

Der erste Summand in (21.59) ist beschränkt durch

$$4C^2 \mathbf{E} \left[\sum_{t \in \mathcal{P}_T^n} (M_{t'} - M_t)^2 \right] = 4C^2 \mathbf{E}[(M_T - M_0)^2] \leq 16C^4.$$

Der zweite Summand in (21.59) ist gleich

$$\begin{aligned} 2 \mathbf{E} \left[\sum_{s \in \mathcal{P}_T^n} (M_{s'} - M_s)^2 \mathbf{E} \left[\sum_{t \in \mathcal{P}_{s', T}^n} (M_{t'} - M_t)^2 \mid \mathcal{F}_{s'} \right] \right] \\ = 2 \mathbf{E} \left[\sum_{s \in \mathcal{P}_T^n} (M_{s'} - M_s)^2 \mathbf{E}[(M_T - M_{s'})^2 \mid \mathcal{F}_{s'}] \right] \\ \leq 8C^2 \mathbf{E}[(M_T - M_0)^2] \leq 32C^4. \end{aligned}$$

Es gilt also nach (21.58) und (21.56)

$$\sup_{n \geq m} \mathbf{E}[(U_T^n - U_T^m)^2] \leq 16\sqrt{3} C^2 \mathbf{E}[Z_m^2]^{1/2} \xrightarrow{m \rightarrow \infty} 0.$$

Damit ist (21.55) gezeigt.

Schritt 2. Sei nun $M \in \mathcal{M}_{loc, c}$ und $(\tau_N)_{N \in \mathbb{N}}$ eine lokalisierende Folge, sodass jedes M^{τ_N} ein beschränktes Martingal ist (siehe Bemerkung 21.67). Nach Schritt 1 gilt für $T \geq 0$ und $N \in \mathbb{N}$

$$U_T^{N,n} := \sum_{t \in \mathcal{P}_T^n} (M_{t'}^{\tau_N} - M_t^{\tau_N})^2 \xrightarrow{n \rightarrow \infty} \langle M^{\tau_N} \rangle_T \text{ in } L^2.$$

Wegen $U_T^{N,n} = U_T^{N+1,n}$, falls $T \leq \tau_N$, gibt es einen stetigen Prozess U mit $U_T^{N,n} \xrightarrow{n \rightarrow \infty} U_T$ stochastisch, falls $T \leq \tau_N$. Also gilt $\langle M^{\tau_N} \rangle_T = \langle M \rangle_T := U_T$, falls $T \leq \tau_N$. Wegen $\tau_N \uparrow \infty$ fast sicher, gilt für alle $T \geq 0$

$$U_T^n \xrightarrow{n \rightarrow \infty} \langle M \rangle_T \text{ stochastisch.}$$

Da $((M_T^{\tau_N})^2 - \langle M^{\tau_N} \rangle_T)_{T \geq 0}$ ein stetiges Martingal ist und $\langle M^{\tau_N} \rangle = \langle M \rangle^{\tau_N}$ gilt, folgt $M^2 - \langle M \rangle \in \mathcal{M}_{loc, c}$.

Schritt 3. Wir müssen noch (ii) zeigen. Sei also M ein stetiges, quadratintegrierbares Martingal und $(\tau_n)_{n \in \mathbb{N}}$ eine lokalisierende Folge für das lokale Martingal

$M^2 - \langle M \rangle$. Sei $T > 0$ und $\tau \leq T$ eine Stoppzeit. Da M^2 ein nichtnegatives Submartingal ist, ist $M_{\tau_n \wedge \tau}^2 \leq \mathbf{E}[M_T^2 | \mathcal{F}_{\tau_n \wedge \tau}]$, also ist $(M_{\tau_n \wedge \tau}^2)_{n \in \mathbb{N}}$ gleichgradig integrierbar und damit

$$\mathbf{E}[M_\tau^2] = \lim_{n \rightarrow \infty} \mathbf{E}[M_{\tau_n \wedge \tau}^2] = \lim_{n \rightarrow \infty} \mathbf{E}[\langle M \rangle_{\tau_n \wedge \tau}] + \mathbf{E}[M_0^2] = \mathbf{E}[\langle M \rangle_\tau] + \mathbf{E}[M_0^2],$$

wobei wir im letzten Schritt den Satz über monotone Konvergenz ausgenutzt haben. Nach dem Optional Sampling Theorem ist also $M^2 - \langle M \rangle$ ein Martingal.

Schritt 4 (Eindeutigkeit). Seien A und A' stetige, monoton wachsende, adaptierte Prozesse mit $A_0 = A'_0$, sodass $M^2 - A$ und $M^2 - A'$ lokale Martingale sind. Dann ist auch $N = A - A'$ ein lokales Martingal, und für fast alle ω hat der Pfad $N(\omega)$ endliche Variation. Daher ist $\langle N \rangle \equiv 0$ und damit $N^2 - \langle N \rangle = N^2$ ein stetiges lokales Martingal mit $N_0 = 0$. Sei $(\tau_n)_{n \in \mathbb{N}}$ eine lokalisierte Folge für N^2 . Dann ist $\mathbf{E}[N_{\tau_n \wedge t}^2] = 0$ für jedes $n \in \mathbb{N}$ und $t \geq 0$, also ist $N_{\tau_n \wedge t}^2 = 0$ fast sicher und damit $N_t^2 = \lim_{n \rightarrow \infty} N_{\tau_n \wedge t}^2 = 0$ fast sicher. Es folgt $A = A'$. \square

Korollar 21.72. Sei M ein stetiges lokales Martingal mit $\langle M \rangle \equiv 0$. Dann ist $M_t = M_0$ für alle $t \geq 0$ fast sicher. Speziell gilt dies, falls die Pfade von M von lokal endlicher Variation sind.

Korollar 21.73. Seien $M, N \in \mathcal{M}_{loc,c}$. Dann existiert ein eindeutig bestimmter stetiger, adaptierter Prozess $\langle M, N \rangle$ von fast sicher lokal endlicher Variation mit $\langle M, N \rangle_0 = 0$, sodass gilt:

$MN - \langle M, N \rangle$ ist ein stetiges lokales Martingal.

$\langle M, N \rangle$ heißt **quadratischer Kovariationsprozess** von M und N . Es gilt für jede zulässige Zerlegungsfolge \mathcal{P} und jedes $T \geq 0$

$$\langle M, N \rangle_T = \lim_{n \rightarrow \infty} \sum_{t \in \mathcal{P}_T^n} (M_{t'} - M_t)(N_{t'} - N_t) \text{ stochastisch.} \quad (21.60)$$

Beweis. Existenz. Offenbar gilt $M + N, M - N \in \mathcal{M}_{loc,c}$. Wir definieren

$$\langle M, N \rangle := \frac{1}{4} (\langle M + N \rangle - \langle M - N \rangle).$$

Als Differenz monoton wachsender Funktionen ist $\langle M, N \rangle$ von lokal endlicher Variation. Wegen Satz 21.70(iii) folgt (21.60). Weiter ist

$$MN - \langle M, N \rangle = \frac{1}{4} ((M + N)^2 - \langle M + N \rangle) - \frac{1}{4} ((M - N)^2 - \langle M - N \rangle)$$

ein lokales Martingal.

Eindeutigkeit. Seien A und A' mit $A_0 = A'_0 = 0$ stetig, adaptiert und von lokal endlicher Variation, sodass $MN - A$ und $MN - A'$ in $\mathcal{M}_{loc,c}$ sind. Dann ist $A - A' \in \mathcal{M}_{loc,c}$ von lokal endlicher Variation, also $A - A' = 0$. \square

Korollar 21.74. Ist $M \in \mathcal{M}_{loc,c}$ und A stetig und adaptiert mit $\langle A \rangle \equiv 0$, so ist $\langle M + A \rangle = \langle M \rangle$.

Ist M ein stetiges lokales Martingal bis zur Stoppzeit τ , so ist $M^\tau \in \mathcal{M}_{loc,c}$, und wir schreiben $\langle M \rangle_t := \langle M^\tau \rangle_t$ für $t < \tau$.

Satz 21.75. Sei τ eine Stoppzeit und M ein stetiges lokales Martingal bis τ sowie $\tau_0 < \tau$ eine Stoppzeit mit $\mathbf{E}[\langle M \rangle_{\tau_0}] < \infty$. Dann ist $\mathbf{E}[M_{\tau_0}] = \mathbf{E}[M_0]$, und M^{τ_0} ist ein L^2 -beschränktes Martingal, falls $\mathbf{E}[M_0^2] < \infty$.

Beweis. Sei $\tau_n \uparrow \tau$ eine lokalisierte Folge von Stoppzeiten für M , sodass jedes M^{τ_n} sogar ein beschränktes Martingal ist (siehe Bemerkung 21.67). Dann ist auch $M^{\tau_0 \wedge \tau_n}$ ein beschränktes Martingal, und es gilt für jedes $t \geq 0$

$$\mathbf{E}[M_{\tau_0 \wedge \tau_n \wedge t}^2] = \mathbf{E}[M_0^2] + \mathbf{E}[\langle M \rangle_{\tau_0 \wedge \tau_n \wedge t}] \leq \mathbf{E}[M_0^2] + \mathbf{E}[\langle M \rangle_{\tau_0}] < \infty. \quad (21.61)$$

Da also $((M_{\tau_0 \wedge \tau_n \wedge t}), n \in \mathbb{N}, t \geq 0)$ beschränkt in L^2 und damit gleichgradig integrierbar ist, gilt (nach dem Optional Sampling Theorem für gleichgradig integrierbare Martingale)

$$\mathbf{E}[M_{\tau_0}] = \lim_{n \rightarrow \infty} \mathbf{E}[M_{\tau_0 \wedge \tau_n}] = \mathbf{E}[M_0],$$

und für $t > s$

$$\begin{aligned} \mathbf{E}[M_t^{\tau_0} \mid \mathcal{F}_s] &= \mathbf{E}\left[\lim_{n \rightarrow \infty} M_s^{\tau_0 \wedge \tau_n} \mid \mathcal{F}_s\right] \\ &= \lim_{n \rightarrow \infty} \mathbf{E}[M_s^{\tau_0 \wedge \tau_n} \mid \mathcal{F}_s] \\ &= \lim_{n \rightarrow \infty} M_s^{\tau_0 \wedge \tau_n} = M_s^{\tau_0}. \end{aligned}$$

Also ist M^{τ_0} ein Martingal. □

Korollar 21.76. Ist $M \in \mathcal{M}_{loc,c}$ mit $\mathbf{E}[M_0^2] < \infty$ und $\mathbf{E}[\langle M \rangle_t] < \infty$ für jedes $t \geq 0$, so ist M ein quadratintegrierbares Martingal.

Übung 21.10.1. Zeige, dass die Zufallsvariablen $(Y_n)_{n \in \mathbb{N}}$ aus dem Beweis von Satz 21.64 ein Rückwärtsmartingal bilden. ♣

Übung 21.10.2. Sei $f : [0, \infty) \rightarrow \mathbb{R}$ stetig und $X \in \mathcal{C}_{qV}^{\mathcal{P}}$ für die zulässige Zerlegungsfolge \mathcal{P} . Man zeige:

$$\int_0^T f(s) d\langle X \rangle_s = \lim_{n \rightarrow \infty} \sum_{t \in \mathcal{P}_T^n} f(t)(X_{t'} - X_t)^2 \quad \text{für alle } T \geq 0. \quad \clubsuit$$

Übung 21.10.3. Man zeige durch ein Gegenbeispiel: Ist M ein stetiges lokales Martingal mit $M_0 = 0$ und τ eine Stoppzeit mit $\mathbf{E}[\langle M \rangle_\tau] = \infty$, so folgt hieraus nicht notwendigerweise $\mathbf{E}[M_\tau^2] = \infty$. ♦

22 Gesetz vom iterierten Logarithmus

Für Summen unabhängiger Zufallsvariablen kennen wir bislang zwei Grenzwertsätze: das Gesetz der großen Zahl und den Zentralen Grenzwertsatz. Das Gesetz der großen Zahl beschreibt für großes $n \in \mathbb{N}$ das typische oder Mittelwertverhalten von Summen von n Zufallsvariablen, während der Zentrale Grenzwertsatz die typischen Fluktuationen um diesen Mittelwert quantitativ erfasst.

In Kapitel 23 werden wir die *untypisch* großen Fluktuationen (große Abweichungen) quantitativ erfassen. Dagegen ist das Thema dieses Kapitels die genauere quantitative Erfassung der *typischen* Fluktuationen, aber nun im gesamten zeitlichen Verlauf $n \rightarrow \infty$. Die Botschaft lautet in etwa: Während zu fester Zeit die Partialsumme S_n um etwa \sqrt{n} von ihrem Erwartungswert abweicht (Zentraler Grenzwertsatz), ist die *maximale* Fluktuation von der Ordnung $\sqrt{n} \log \log n$ (Satz von Hartman und Wintner, Satz 22.11).

Wir beginnen mit der etwas leichteren Aufgabe, diese Fluktuationen zunächst für die Brown'sche Bewegung auszurechnen (Satz 22.1). Danach werden wir sehen, wie man Summen unabhängiger Zufallsvariablen (mit endlicher Varianz) in eine Brown'sche Bewegung einbetten kann (Satz von Skorohod, Satz 22.5), um damit die Aussage des Satzes von Hartman und Wintner zu zeigen.

Wir folgen in diesem Kapitel in Teilen der Darstellung in [44, Kapitel 8.8].

22.1 Iterierter Logarithmus für die Brown'sche Bewegung

Sei $(B_t)_{t \geq 0}$ eine Brown'sche Bewegung. In Beispiel 21.16 haben wir als Anwendung des Blumenthal'schen 0-1 Gesetzes gesehen, dass $\limsup_{t \downarrow 0} B_t / \sqrt{t} = \infty$ f.s. Da nach Satz 21.14 auch $(tB_{1/t})_{t \geq 0}$ eine Brown'sche Bewegung ist, folgt

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{t}} = \infty \quad \text{f.s.}$$

Unser Ziel in diesem Abschnitt ist es, \sqrt{t} durch eine Funktion zu ersetzen, sodass der Limes superior endlich und nichttrivial wird.

Satz 22.1 (Gesetz vom iterierten Logarithmus für die Brown'sche Bewegung).
Es gilt

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log(t)}} = 1 \quad f.s. \quad (22.1)$$

Bevor wir den Satz beweisen, bringen wir ein elementares Lemma.

Lemma 22.2. Sei $X \sim \mathcal{N}_{0,1}$ standardnormalverteilt. Dann ist für jedes $x > 0$

$$\frac{1}{\sqrt{2\pi}} \frac{1}{x + \frac{1}{x}} e^{-x^2/2} \leq \mathbf{P}[X \geq x] \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}. \quad (22.2)$$

Beweis. Sei $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ die Dichte der Standardnormalverteilung. Partielle Integration liefert die zweite Ungleichung in (22.2):

$$\mathbf{P}[X \geq x] = \int_x^\infty \frac{1}{t} (t\varphi(t)) dt = -\frac{1}{t} \varphi(t) \Big|_x^\infty - \int_x^\infty \frac{1}{t^2} \varphi(t) dt \leq \frac{1}{x} \varphi(x).$$

Analog ist

$$\mathbf{P}[X \geq x] \geq \frac{1}{x} \varphi(x) - \frac{1}{x^2} \int_x^\infty \varphi(t) dt = \frac{1}{x} \varphi(x) - \frac{1}{x^2} \mathbf{P}[X \geq x].$$

Hieraus folgt die erste Ungleichung in (22.2). \square

Beweis von Satz 22.1.

1. Schritt: „ \leq “. Betrachte zunächst die Folge $t_n = \alpha^n$ für ein $\alpha > 1$. Später wollen wir $\alpha \downarrow 1$ gehen lassen. Setze $f(t) = 2\alpha^2 \log \log t$. Dann ist nach dem Spiegelungsprinzip (Satz 21.19) und mit der Abkürzung $B_{[a,b]} := \{B_t : t \in [a,b]\}$

$$\begin{aligned} \mathbf{P} \left[\sup B_{[t_n, t_{n+1}]} > \sqrt{t_n f(t_n)} \right] &\leq \mathbf{P} \left[t_{n+1}^{-1/2} \sup B_{[0, t_{n+1}]} > \sqrt{f(t_n)/\alpha} \right] \\ &= \mathbf{P} \left[\sup B_{[0,1]} > \sqrt{f(t_n)/\alpha} \right] \\ &\leq \sqrt{\frac{\alpha}{f(t_n)}} e^{-f(t_n)/2\alpha} \\ &= (\log \alpha)^{-\alpha} \sqrt{\frac{\alpha}{f(t_n)}} n^{-\alpha} \\ &\leq n^{-\alpha} \quad \text{für hinreichend großes } n, \end{aligned} \quad (22.3)$$

wobei wir im vorletzten Schritt benutzt haben, dass

$$\frac{f(t_n)}{2\alpha} = \alpha(\log(n \log \alpha)) = \alpha \log n + \alpha \log \log \alpha.$$

Wegen $\alpha > 1$ ist die rechte Seite von (22.3) summierbar in n

$$\sum_{n=1}^{\infty} \mathbf{P} \left[\sup B_{[t_n, t_{n+1}]} > \sqrt{t_n f(t_n)} \right] < \infty.$$

Das Lemma von Borel-Cantelli (Satz 2.7) liefert jetzt (merke: $t \mapsto \sqrt{tf(t)}$ ist monoton wachsend)

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{tf(t)}} \leq 1 \quad \text{f.s.}$$

Wir lassen $\alpha \downarrow 1$ gehen und erhalten

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log t}} \leq 1 \quad \text{f.s.} \quad (22.4)$$

2. Schritt: „ \geq “. Wir zeigen nun die andere Ungleichung in (22.1). Hierfür lassen wir $\alpha \rightarrow \infty$ gehen. Setze $\beta := \frac{\alpha}{\alpha-1} > 1$ und $g(t) = \frac{2}{\beta^2} \log \log t$. Wähle n_0 so groß, dass $\beta g(t_n) \geq 1$ ist für $n \geq n_0$. Dann ist nach der Brown'schen Skalierung (merke: $t_n - t_{n-1} = \frac{1}{\beta} t_n$ und (22.2) (wegen $(x + \frac{1}{x})^{-1} \geq \frac{1}{2} \frac{1}{x}$ für $x = (\beta g(t_n))^{1/2} \geq 1$)

$$\begin{aligned} \mathbf{P} \left[B_{t_n} - B_{t_{n-1}} > \sqrt{t_n g(t_n)} \right] &= \mathbf{P} \left[B_1 > \sqrt{\beta g(t_n)} \right] \\ &\geq \frac{1}{\sqrt{2\pi}} \frac{1}{2} \frac{1}{\sqrt{\beta g(t_n)}} e^{-\beta g(t_n)/2} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{2} (\log \alpha)^{-1/\beta} \frac{1}{\sqrt{\beta g(t_n)}} n^{-1/\beta}. \end{aligned}$$

Ist $\varepsilon \in (0, 1 - 1/\beta)$, so ist für hinreichend großes $n \in \mathbb{N}$ die rechte Seite der vorangehenden Gleichung $\geq n^{-\varepsilon} n^{-1/\beta} \geq n^{-1}$. Also ist

$$\sum_{n=2}^{\infty} \mathbf{P} \left[B_{t_n} - B_{t_{n-1}} > \sqrt{t_n g(t_n)} \right] = \infty.$$

Die Ereignisse sind unabhängig, daher liefert das Lemma von Borel-Cantelli

$$\mathbf{P} \left[B_{t_n} - B_{t_{n-1}} > \sqrt{t_n g(t_n)} \quad \text{für unendlich viele } n \right] = 1. \quad (22.5)$$

Wegen $\frac{t_n \log \log t_n}{t_{n-1} \log \log t_{n-1}} \xrightarrow{n \rightarrow \infty} \alpha$ folgt aus (22.4) zusammen mit der Symmetrie der Brown'schen Bewegung für $\varepsilon > 0$

$$B_{t_{n-1}} > -(1 + \varepsilon) \alpha^{-1/2} \sqrt{2t_n \log \log t_n} \quad \text{für fast jedes } n \in \mathbb{N} \quad \text{f.s.} \quad (22.6)$$

Aus (22.5) und (22.6) folgt

$$\limsup_{n \rightarrow \infty} \frac{B_{t_n}}{\sqrt{2t_n \log \log t_n}} \geq \frac{1}{\beta} - (1 + \varepsilon) \alpha^{-1/2} = \frac{\alpha - 1}{\alpha} - (1 + \varepsilon) \alpha^{-1/2} \quad \text{f.s.}$$

Lassen wir nun $\alpha \rightarrow \infty$, so erhalten wir $\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log t}} \geq 1$ f.s. Zusammen mit (22.4) folgt die Aussage des Satzes. \square

Korollar 22.3. Es gilt fast sicher $\limsup_{t \downarrow 0} \frac{B_t}{\sqrt{2t \log \log(1/t)}} = 1$.

Beweis. Nach Satz 21.14 ist $(tB_{1/t})$ eine Brown'sche Bewegung. Wende hierauf Satz 22.1 an. \square

Bemerkung 22.4. Die Aussage von Korollar 22.3 betrifft die *typischen Punkte* der Brown'schen Bewegung B . Wie sieht es aber aus, wenn wir nach der Existenz von Punkten t fragen, in denen sich B schneller als $\sqrt{2t \log \log(1/t)}$ bewegt? Auskunft gibt hier ein Satz von Paul Lévy [115]: Bezeichnen wir mit $h(\delta) := \sqrt{2\delta \log(1/\delta)}$ den **Lévy'schen Stetigkeitsmodul**, so ist

$$\mathbf{P} \left[\lim_{\delta \downarrow 0} \sup_{\substack{s, t \in [0, 1] \\ 0 \leq t - s \leq \delta}} |B_t - B_s| / h(\delta) = 1 \right] = 1. \quad (22.7)$$

(Siehe etwa [154, Theorem I.2.5] für einen Beweis.) Hieraus folgt insbesondere, dass B fast sicher nicht lokal Hölder- $\frac{1}{2}$ -stetig ist. \diamond

22.2 Skorohod'scher Einbettungssatz

Um das Ergebnis des vorigen Abschnitts auf Summen von quadratintegrierbaren, zentrierten Zufallsvariablen zu übertragen, brauchen wir eine Einbettung von solchen Zufallsvariablen in eine Brown'sche Bewegung. Die gewünschte Darstellung liefert der Satz von Skorohod. Mit dieser Technik lässt sich auch ein alternativer Beweis des Satzes von Donsker (Invarianzprinzip, Satz 21.43) angeben.

Satz 22.5 (Skorohod'scher Einbettungssatz). Sei X eine reelle Zufallsvariable mit $\mathbf{E}[X] = 0$ und $\mathbf{Var}[X] < \infty$. Auf einem geeigneten Wahrscheinlichkeitsraum lassen sich folgende Objekte konstruieren: Eine Zufallsvariable Ξ , eine von Ξ unabhängige Brown'sche Bewegung B und eine \mathbb{F} -Stopzeit τ mit

$$B_\tau \stackrel{\mathcal{D}}{=} X \quad \text{und} \quad \mathbf{E}[\tau] = \mathbf{Var}[X].$$

Dabei ist $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ und $\mathcal{F}_t = \sigma(\Xi, (B_s)_{s \leq t})$ für $t \geq 0$.

Bemerkung 22.6. In der Tat kann man auf die Zufallsvariable Ξ verzichten, also für \mathbb{F} die natürliche Filtration $\mathbb{F} = \sigma(B)$ wählen. Der Beweis hierfür ist allerdings aufwändiger (siehe Seite 524). \diamond

Korollar 22.7. Seien X_1, X_2, \dots u.i.v. reelle Zufallsvariablen mit $\mathbf{E}[X_1] = 0$ und $\mathbf{Var}[X_1] < \infty$. Ferner sei $S_n = X_1 + \dots + X_n$, $n \in \mathbb{N}$. Dann gibt es auf einem geeigneten Wahrscheinlichkeitsraum eine Filtration \mathbb{F} und eine Brown'sche Bewegung B , sowie \mathbb{F} -Stopzeiten $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \dots$ mit: $(\tau_n - \tau_{n-1})_{n \in \mathbb{N}}$ ist u.i.v., $\mathbf{E}[\tau_1] = \mathbf{Var}[X_1]$ und $(B_{\tau_n})_{n \in \mathbb{N}} \stackrel{\mathcal{D}}{=} (S_n)_{n \in \mathbb{N}}$.

Beweis (von Korollar 22.7). Wir skizzieren den Beweis nur. Die Details bleiben zur Übung.

Seien $(B^{(n)}, \Xi^{(n)}, \tau^{(n)})$, $n \in \mathbb{N}$, unabhängig und jeweils wie in Satz 22.5 gewählt. Sei $\tau_n = \tau^{(1)} + \dots + \tau^{(n)}$. Definiere $B_t := B_t^{(1)}$, falls $t \leq \tau_1$ ist, und iterativ

$$B_t = B_{\tau_n} + B_{t-\tau_n}^{(n+1)}, \quad \text{falls } \tau_n < t \leq \tau_{n+1}.$$

Durch wiederholtes Ausnutzen der starken Markoveigenschaft der Brown'schen Bewegung sieht man, dass B eine Brown'sche Bewegung ist. Setze nun $\mathcal{F}_t = \sigma((\Xi_n)_{n \in \mathbb{N}}, (B_s)_{s \leq t})$. \square

Zur Vorbereitung des Beweises von Satz 22.5 bringen wir ein Lemma. Wir wollen dabei als Integranden Maße zulassen. Wir verwenden deshalb folgende Notation: Ist $\mu \in \mathcal{M}(E)$ ein Maß und $f \in \mathcal{L}^1(\mu)$ nichtnegativ, so wird $\int \mu(dx)f(x)\delta_x := f\mu$ definiert, wobei $f\mu$ das Maß mit Dichte f bezüglich μ ist. Dies ist konsistent, denn für messbares $A \subset E$ ist dann

$$\left(\int \mu(dx)f(x)\delta_x \right)(A) = \int \mu(dx)f(x)\delta_x(A) = \int \mu(dx)f(x)\mathbb{1}_A(x) = f\mu(A).$$

Lemma 22.8. Sei $\mu \in \mathcal{M}_1(\mathbb{R})$ mit $\int x \mu(dx) = 0$ und $\sigma^2 := \int x^2 \mu(dx) < \infty$. Dann existiert ein W-Maß $\theta \in \mathcal{M}_1((-\infty, 0) \times [0, \infty))$ mit

$$\mu = \int \theta(d(u, v)) \left(\frac{v}{v-u} \delta_u + \frac{-u}{v-u} \delta_v \right). \quad (22.8)$$

Es ist $\sigma^2 = - \int uv \theta(d(u, v))$.

Beweis. Wir setzen $m := \int_{[0, \infty)} v \mu(dv) = - \int_{(-\infty, 0)} u \mu(du)$. Ist $m = 0$, so ist $\theta = \delta_{(-1, 0)}$ eine geeignete Wahl. Sei im Folgenden $m > 0$. Wir setzen

$$\theta(d(u, v)) := m^{-1}(v - u) \mu(du) \mu(dv) \quad \text{für } u < 0 \text{ und } v \geq 0.$$

Dann ist

$$\begin{aligned} \int \theta(d(u, v)) &= m^{-1} \int_{(-\infty, 0)} \mu(du) \int_{[0, \infty)} \mu(dv) (v - u) \\ &= m^{-1} \int_{(-\infty, 0)} \mu(du) [m - u\mu([0, \infty))] \\ &= m^{-1} (m\mu((-\infty, 0)) + m\mu([0, \infty))) = 1. \end{aligned}$$

Es ist also tatsächlich θ ein W-Maß. Weiter gilt

$$\begin{aligned} \int \theta(d(u, v)) \left(\frac{v}{v-u} \delta_u + \frac{-u}{v-u} \delta_v \right) &= m^{-1} \int_{(-\infty, 0)} \mu(du) \int_{[0, \infty)} \mu(dv) (v\delta_u - u\delta_v) \\ &= \int_{(-\infty, 0)} \mu(du) \delta_u + \int_{[0, \infty)} \mu(dv) \delta_v = \mu. \end{aligned}$$

Aus (22.8) folgt direkt

$$\sigma^2 = \int \mu(dx) x^2 = \int \theta(d(u, v)) \left(\frac{v}{v-u} u^2 + \frac{-u}{v-u} v^2 \right) = - \int \theta(d(u, v)) uv. \quad \square$$

Beweis (Satz 22.5). Wir nehmen zunächst an, dass X nur die zwei Werte $u < 0$ und $v \geq 0$ annimmt: $\mathbf{P}[X = u] = \frac{v}{v-u} = 1 - \mathbf{P}[X = v]$. Sei

$$\tau_{u,v} = \inf \{t > 0 : B_t \in \{u, v\}\}.$$

Nach Übung 21.2.4 ist $\mathbf{E}[B_{\tau_{u,v}}] = 0$ also $B_{\tau_{u,v}} \stackrel{\mathcal{D}}{=} X$ sowie $\mathbf{E}[\tau_{u,v}] = -uv$.

Sei nun X beliebig mit $\mathbf{E}[X] = 0$ und $\sigma^2 := \mathbf{E}[X^2] < \infty$. Setze $\mu = \mathbf{P}_X$ und $\theta = \theta_\mu$ wie in Lemma 22.8. Ferner sei $\Xi = (\Xi_u, \Xi_v)$ eine Zufallsvariable mit Werten in $(-\infty, 0) \times [0, \infty)$ und Verteilung θ .

Sei $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$, wobei $\mathcal{F}_t := \sigma(\Xi, B_s : s \in [0, t])$ ist. Setze $\tau := \tau_{\Xi_u, \Xi_v}$. Auf Grund der Stetigkeit von B und wegen $\tau \leq \tau_{u,v}$, falls $u < \Xi_u$ und $v > \Xi_v$, ist für jedes $t \geq 0$

$$\{\tau \leq t\} = \bigcap_{\substack{u, v \in \mathbb{Q} \\ u < 0 \leq v}} \left(\{\Xi \in (-\infty, u] \times [v, \infty)\} \cap \{\tau_{u,v} \leq t\} \right) \in \mathcal{F}_t.$$

Also ist τ eine \mathbb{F} -Stoppzeit (aber keine Stoppzeit bezüglich $\sigma(B)$). Für $x < 0$ ist

$$\begin{aligned} \mathbf{P}[X \leq x] &= \int_{(-\infty, x] \times [0, \infty)} \theta(d(u, v)) \frac{v}{v-u} \\ &= \int_{(-\infty, x] \times [0, \infty)} \theta(d(u, v)) \mathbf{P}[B_{\tau_{u,v}} = u] = \mathbf{P}[B_\tau \leq x]. \end{aligned}$$

Für $x \geq 0$ ist analog $\mathbf{P}[X > x] = \mathbf{P}[B_\tau > x]$, also insgesamt $B_\tau \stackrel{\mathcal{D}}{=} X$. Weiter ist

$$\mathbf{E}[\tau] = -\mathbf{E}[\Xi_u \Xi_v] = - \int \theta(d(u, v)) uv = \sigma^2. \quad \square$$

Ergänzung: Beweis von Bemerkung 22.6

Wir beweisen hier, dass man im Skorohod'schen Einbettungssatz tatsächlich ohne *randomisierte* Stoppzeiten auskommt, also eine Stoppzeit bezüglich der von der Brown'schen Bewegung B erzeugten Filtration wählen kann. Mit anderen Worten, man kann die Stoppzeit τ wählen, ohne dass man auf zusätzliche Zufallsvariablen, wie Ξ im obigen Beweis, zurückgreifen muss.

Ein eleganter Beweis, der allerdings auf Methoden der stochastischen Analysis basiert, findet sich bei Azéma und Yor, siehe [7] oder [128, Chapter 5.3]. Wir folgen hier einer elementaren Route, die auf Ideen von Dubins zurückgreift.

Für eine Zufallsvariable X , die nur zwei Werte $u < 0 < v$ annimmt und für die $\mathbf{E}[X] = 0$ gilt, ist die Aussage einfach gezeigt. Sei $\tau_{u,v} = \inf\{t > 0 : B_t \in \{u, v\}\}$. Nach Übung 21.2.4 ist $\mathbf{E}[B_{\tau_{u,v}}] = 0$ und $\mathbf{E}[\tau_{u,v}] = -uv = \mathbf{Var}[X]$.

In einem ersten Schritt verallgemeinern wir diese Aussage nun auf binär aufspaltende Martingale (vergleiche Definition 9.42), also solche Martingale, die je nach bisherigem Verlauf in jedem Zeitschritt immer nur zwischen zwei möglichen Werten wählen können. Im zweiten Schritt zeigen wir dann, dass jede quadratisch integrierbare Zufallsvariable als Grenzwert solcher Martingale auftritt.

Satz 22.9. *Sei $(X_n)_{n \in \mathbb{N}_0}$ ein binär aufspaltendes quadratisch integrierbares Martingal mit $X_0 = 0$, und sei B eine Brown'sche Bewegung. Dann existieren Stoppzeiten $0 = \tau_0 \leq \tau_1 \leq \dots$ für B , sodass*

$$(X_n)_{n \in \mathbb{N}_0} \stackrel{\mathcal{D}}{=} (B_{\tau_n})_{n \in \mathbb{N}_0},$$

und so, dass $\mathbf{E}[\tau_n] = \mathbf{E}[X_n^2]$ für jedes $n \in \mathbb{N}_0$ gilt. Ist $(X_n)_{n \in \mathbb{N}_0}$ in L^2 beschränkt und damit fast sicher und in L^2 konvergent gegen ein X_∞ , so ist $\tau := \sup_{n \in \mathbb{N}} \tau_n < \infty$ fast sicher, $\mathbf{E}[\tau] = \mathbf{Var}[X_\infty]$ und $X_\infty \stackrel{\mathcal{D}}{=} B_\tau$.

Beweis. Für $n \in \mathbb{N}$ seien $f_n : \mathbb{R}^{n-1} \times \{-1, +1\} \rightarrow \mathbb{R}$ und eine $\{-1, +1\}$ -wertige Zufallsvariable D_n so gewählt, dass $X_n = f_n(X_1, \dots, X_{n-1}, D_n)$ gilt (vergleiche Definition 9.42). Ohne Einschränkung sei f_n monoton wachsend in D_n . Setze $\tau_0 := 0$ und sukzessive

$$\tau_n := \inf \left\{ t > \tau_{n-1} : B_t \in \{f_n(B_{\tau_1}, \dots, B_{\tau_{n-1}}, -1), f_n(B_{\tau_1}, \dots, B_{\tau_{n-1}}, +1)\} \right\}.$$

Setze $\tilde{X}_n := B_{\tau_n}$ und

$$\tilde{D}_n := \begin{cases} 1, & \text{falls } \tilde{X}_n \geq \tilde{X}_{n-1}, \\ -1, & \text{sonst.} \end{cases}$$

Dann ist nach der starken Markoveigenschaft (angewandt auf τ_{n-1}) und nach Übung 21.2.4

$$\begin{aligned} \mathbf{P}[\tilde{D}_n = 1 | \tilde{X}_1, \dots, \tilde{X}_{n-1}] \\ = \frac{\tilde{X}_{n-1} - f_n(\tilde{X}_1, \dots, \tilde{X}_{n-1}, -1)}{f_n(\tilde{X}_1, \dots, \tilde{X}_{n-1}, +1) - f_n(\tilde{X}_1, \dots, \tilde{X}_{n-1}, -1)} \end{aligned}$$

und $\mathbf{E}[\tau_n - \tau_{n-1}] = \mathbf{E}[(\tilde{X}_n - \tilde{X}_{n-1})^2]$. Da $(X_n)_{n \in \mathbb{N}_0}$ ein Martingal ist, gilt andererseits,

$$\begin{aligned} X_{n-1} &= \mathbf{E}[X_n | X_0, \dots, X_{n-1}] \\ &= \sum_{i=-1,+1} \mathbf{P}[D_n = i | X_0, \dots, X_{n-1}] f_n(X_1, \dots, X_{n-1}, i). \end{aligned}$$

Also ist

$$\begin{aligned} \mathbf{P}[D_n = 1 | X_1, \dots, X_{n-1}] \\ = \frac{X_{n-1} - f_n(X_1, \dots, X_{n-1}, -1)}{f_n(X_1, \dots, X_{n-1}, +1) - f_n(X_1, \dots, X_{n-1}, -1)}. \end{aligned}$$

Es folgt $(X_n)_{n \in \mathbb{N}_0} \stackrel{\mathcal{D}}{=} (\tilde{X}_n)_{n \in \mathbb{N}_0}$. Wegen $\mathbf{E}[\tau_n - \tau_{n-1}] = \mathbf{E}[(X_n - X_{n-1})^2]$ und weil die Martingaldifferenzen $(X_i - X_{i-1})$ unkorreliert sind, gilt $\mathbf{E}[\tau_n] = \mathbf{E}[X_n^2]$.

Ist schließlich (X_n) in L^2 beschränkt, so existiert nach dem Martingalkonvergenzsatz eine quadratisch integrierbare Zufallsvariable X_∞ mit $X_n \xrightarrow{n \rightarrow \infty} X_\infty$ fast sicher und in L^2 . Speziell gilt $\mathbf{E}[X_n^2] \xrightarrow{n \rightarrow \infty} \mathbf{E}[X_\infty^2]$. Nun ist offenbar $(\tau_n)_{n \in \mathbb{N}}$ monoton wachsend und damit konvergent gegen eine Stoppzeit τ . Nach dem Satz über monotone Konvergenz ist $\mathbf{E}[\tau] = \lim_{n \rightarrow \infty} \mathbf{E}[\tau_n] = \lim_{n \rightarrow \infty} \mathbf{E}[X_n^2] = \mathbf{E}[X_\infty^2] < \infty$. Insbesondere ist also $\tau < \infty$ fast sicher. Da die Brown'sche Bewegung stetig ist, gilt fast sicher

$$B_\tau = \lim_{n \rightarrow \infty} B_{\tau_n} = \lim_{n \rightarrow \infty} \tilde{X}_n \stackrel{\mathcal{D}}{=} X_\infty. \quad \square$$

Wir haben nun also die Aussage von Bemerkung 22.6 für den Fall gezeigt, dass die Zufallsvariable X sich als Grenzwert eines binär aufspaltenden Martingals darstellen lässt. Den allgemeine Fall erhalten wir daher durch den folgenden Satz.

Satz 22.10. *Sei X eine quadratisch integrierbare zentrierte Zufallsvariable. Dann existiert ein binär aufspaltendes Martingal $(X_n)_{n \in \mathbb{N}_0}$ mit $X_0 = 0$ und so, dass $X_n \xrightarrow{n \rightarrow \infty} X$ fast sicher und in L^2 .*

Beweis. Wir folgen der Beweisidee von [128]. Sei $X_0 := \mathbf{E}[X] = 0$. Definiere induktiv für $n \in \mathbb{N}$

$$D_n := \begin{cases} 1, & \text{falls } X \geq X_{n-1}, \\ -1, & \text{falls } X < X_{n-1}, \end{cases}$$

$$\mathcal{F}_n := \sigma(D_1, \dots, D_n)$$

und

$$X_n := \mathbf{E}[X | \mathcal{F}_n].$$

Es gibt also eine Abbildung $g_n : \{-1, +1\}^n \rightarrow \mathbb{R}$ mit $g_n(D_1, \dots, D_n) = X_n$. Offenbar ist $\mathbb{1}_{D_k=1} = \mathbb{1}_{X_k \geq X_{k-1}}$ fast sicher für jedes $k \in \mathbb{N}$. Also lassen sich die D_1, \dots, D_k aus X_1, \dots, X_k berechnen. Daher existiert eine Abbildung $f_n : \mathbb{R}^{n-1} \times \{-1, +1\} \rightarrow \mathbb{R}$ mit $f_n(X_1, \dots, X_{n-1}, D_n) = X_n$. Also ist (X_n) binär aufspaltend.

Offenbar ist $(X_n)_{n \in \mathbb{N}_0}$ ein Martingal. Nach der Jensen'schen Ungleichung ist $\mathbf{E}[X_n^2] \leq \mathbf{E}[X^2] < \infty$, also ist $(X_n)_{n \in \mathbb{N}_0}$ in L^2 beschränkt und damit fast sicher und in L^2 konvergent gegen ein quadratisch integrierbares X_∞ . Es bleibt also zu zeigen, dass $X_\infty = X$ fast sicher gilt. Hierzu zeigen wir zunächst

$$\lim_{n \rightarrow \infty} D_n(\omega) (X(\omega) - X_n(\omega)) = |X(\omega) - X_\infty(\omega)| \quad \text{für fast alle } \omega. \quad (22.9)$$

Gilt $X(\omega) = X_\infty(\omega)$, so gilt (22.9) trivialerweise. Gilt $X(\omega) > X_\infty(\omega)$, so ist $X(\omega) > X_n(\omega)$ und damit $D_n(\omega) = 1$ für alle hinreichend großen n , also gilt (22.9). Ähnlich erhalten wir (22.9) falls $X(\omega) < X_\infty(\omega)$.

Offenbar ist

$$\mathbf{E}[D_n(X - X_n)] = \mathbf{E}[D_n \mathbf{E}[X - X_n | \mathcal{F}_n]] = 0.$$

Da $(D_n(X - X_n))_{n \in \mathbb{N}}$ in L^2 beschränkt ist (und damit gleichgradig integrierbar ist), folgt $\mathbf{E}[|X - X_\infty|] = \lim_{n \rightarrow \infty} \mathbf{E}[D_n(X - X_n)] = 0$, also $X = X_\infty$ f.s. \square

22.3 Satz von Hartman-Wintner

Ziel dieses Abschnitts ist der Beweis des Gesetzes vom iterierten Logarithmus für u.i.v. Zufallsvariablen X_n , $n \in \mathbb{N}$ mit zweiten Momenten, der auf Hartman und Wintner (1941) (siehe [75]) zurückgeht. (In der einfacheren Situation, wo die X_n Bernoulli Zufallsvariablen sind, hat bereits Khinchin (1923) die obere Abschätzung im Gesetz vom iterierten Logarithmus gefunden.)

Satz 22.11 (Hartman-Wintner, Gesetz vom iterierten Logarithmus).

Seien X_1, X_2, \dots u.i.v. reelle Zufallsvariablen mit $\mathbf{E}[X_1] = 0$ und $\mathbf{Var}[X_1] = 1$. Sei $S_n = X_1 + \dots + X_n$, $n \in \mathbb{N}$. Dann gilt

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = 1 \quad \text{f.s.} \quad (22.10)$$

Wir beweisen den Satz, indem wir ihn auf das Gesetz vom iterierten Logarithmus für die Brown'sche Bewegung zurückführen. Zu diesem Zweck fassen wir die Partialsummen S_n als Werte der Brown'schen Bewegung B zu gewissen Stoppzeiten $\tau_1 \leq \tau_2 \leq \dots$ auf. Dass dies funktioniert, sichert der Skorohod'sche Einbettungssatz. Wir folgen der Darstellung in [44, Kapitel 8.8].

Beweis. Nach Korollar 22.7 gibt es auf einem geeigneten Wahrscheinlichkeitsraum eine Filtration \mathbb{F} und eine Brown'sche Bewegung B , die ein \mathbb{F} -Martingal ist, sowie Stoppzeiten $\tau_1 \leq \tau_2 \leq \dots$, sodass $(S_n)_{n \in \mathbb{N}} \stackrel{\mathcal{D}}{=} (B_{\tau_n})_{n \in \mathbb{N}}$. Ferner sind $(\tau_n - \tau_{n-1})_{n \in \mathbb{N}}$ u.i.v. mit $\mathbf{E}[\tau_n - \tau_{n-1}] = \mathbf{Var}[X_1] = 1$.

Nach dem Gesetz vom iterierten Logarithmus für die Brown'sche Bewegung (siehe Satz 22.1) ist

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log t}} = 1 \quad \text{f.s.}$$

Es reicht also zu zeigen, dass

$$\limsup_{t \rightarrow \infty} \frac{|B_t - B_{\tau_{[t]}}|}{\sqrt{2t \log \log t}} = 0 \quad \text{f.s.}$$

Nach dem starken Gesetz der großen Zahl (Satz 5.17) gilt $\frac{1}{n} \tau_n \xrightarrow{n \rightarrow \infty} 1$ f.s. Sei also $\varepsilon > 0$ und $t_0 = t_0(\omega)$ so groß, dass

$$\frac{1}{1 + \varepsilon} \leq \frac{\tau_{[t]}}{t} \leq 1 + \varepsilon \quad \text{für jedes } t \geq t_0.$$

Setze

$$M_t := \sup_{s \in [t/(1+\varepsilon), t(1+\varepsilon)]} |B_s - B_t|.$$

Es reicht zu zeigen, dass $\limsup_{t \rightarrow \infty} \frac{M_t}{\sqrt{2t \log \log t}} = 0$ gilt. Betrachte hierzu die Folge $t_n = (1 + \varepsilon)^n$, $n \in \mathbb{N}$, und setze

$$M'_n := \sup_{s \in [t_{n-1}, t_{n+2}]} |B_s - B_{t_{n-1}}|.$$

Dann ist (nach der Dreiecksungleichung) für $t \in [t_n, t_{n+1}]$

$$M_t \leq 2M'_n.$$

Setze $\delta := (1 + \varepsilon)^3 - 1$. Dann ist $t_{n+2} - t_{n-1} = \delta t_{n-1}$. Brown'sche Skalierung und das Spiegelungsprinzip (Satz 21.19) ergeben nun

$$\begin{aligned} \mathbf{P}\left[M'_n > \sqrt{3\delta t_{n-1} \log \log t_{n-1}}\right] &= \mathbf{P}\left[\sup_{s \in [0, 1]} |B_s| > \sqrt{3 \log \log t_{n-1}}\right] \\ &\leq 2 \mathbf{P}\left[\sup_{s \in [0, 1]} B_s > \sqrt{3 \log \log t_{n-1}}\right] \\ &= 4 \mathbf{P}\left[B_1 > \sqrt{3 \log \log t_{n-1}}\right] \\ &\leq \frac{2}{\sqrt{3 \log \log t_{n-1}}} \exp\left(-\frac{3}{2} \log \log t_{n-1}\right) \quad (\text{Lemma 22.2}) \\ &\leq n^{-3/2} \quad \text{für } n \text{ hinreichend groß.} \end{aligned}$$

Die Wahrscheinlichkeiten lassen sich also über n summieren, und das Lemma von Borel-Cantelli liefert

$$\limsup_{t \rightarrow \infty} \frac{M_t}{\sqrt{t \log \log t}} \leq \limsup_{n \rightarrow \infty} \frac{2M'_n}{\sqrt{t_{n-1} \log \log t_{n-1}}} \leq 2\sqrt{3\delta}.$$

Lassen wir nun $\varepsilon \rightarrow 0$ gehen, so geht $\delta = (1 + \varepsilon)^3 - 1 \rightarrow 0$, und der Beweis ist vollständig. \square

23 Große Abweichungen

Wir haben (bis auf das Gesetz vom iterierten Logarithmus) bislang zwei Typen von Grenzwertsätzen für Partialsummen $S_n = X_1 + \dots + X_n$, $n \in \mathbb{N}$, von identisch verteilten, reellen Zufallsvariablen $(X_i)_{i \in \mathbb{N}}$ mit Verteilungsfunktion F gesehen:

- (1) (Schwache) Gesetze der großen Zahl besagen (unter gewissen Annahmen an die Familie $(X_i)_{i \in \mathbb{N}}$), dass für jedes $x > 0$

$$\mathbf{P} [|S_n - n \mathbf{E}[X_1]| \geq xn] \xrightarrow{n \rightarrow \infty} 0. \quad (23.1)$$

Hieraus abgeleitet erhalten wir für die empirischen Verteilungsfunktionen

$$F_n : x \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, x]}(X_i)$$

die stochastische Konvergenz $\|F_n - F\|_\infty \xrightarrow{n \rightarrow \infty} 0$. Wir wollen dies umformulieren zu: Für jede Verteilungsfunktion $G \neq F$ und jedes $\varepsilon > 0$ mit $\varepsilon < \|F - G\|_\infty$ gilt

$$\mathbf{P} [\|F_n - G\|_\infty < \varepsilon] \xrightarrow{n \rightarrow \infty} 0. \quad (23.2)$$

- (2) Zentrale Grenzwertsätze besagen (unter wieder anderen Annahmen an die Familie $(X_i)_{i \in \mathbb{N}}$), dass für $x \in \mathbb{R}$

$$\mathbf{P} [S_n - n \mathbf{E}[X_1] \geq x\sqrt{n}] \xrightarrow{n \rightarrow \infty} 1 - \Phi \left(\frac{x}{\sqrt{\text{Var}[X_1]}} \right), \quad (23.3)$$

wobei $\Phi : t \mapsto \mathcal{N}_{0,1}((-\infty, t])$ die Verteilungsfunktion der Standardnormalverteilung ist.

Der *typische Wert* von S_n ist jeweils $n \mathbf{E}[X_1]$. Gleichung (23.3) macht präzise Aussagen über mittelgroße Abweichungen (der Ordnung \sqrt{n}) vom typischen Wert. Hieraus folgt natürlich, dass die Wahrscheinlichkeit für *große Abweichungen* (der Ordnung n) vom typischen Wert gegen 0 geht, also (23.1).

In diesem Kapitel wollen wir die *Konvergenzgeschwindigkeit* in (23.1) bestimmen (Satz von Cramér), beziehungsweise in (23.2) (Satz von Sanov).

Wir folgen in Teilen den Darstellungen [36] und [81].

23.1 Satz von Cramér

Seien X_1, X_2, \dots u.i.v. mit $\mathbf{P}_{X_i} = \mathcal{N}_{0,1}$. Dann ist für jedes $x > 0$

$$\mathbf{P}[S_n \geq xn] = \mathbf{P}[X_1 \geq x\sqrt{n}] = 1 - \Phi(x\sqrt{n}) = (1 + \varepsilon_n) \frac{1}{x\sqrt{2\pi n}} e^{-n x^2/2},$$

wobei (nach Lemma 22.2) $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$ gilt. Es gilt also

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}[S_n \geq xn] = -\frac{x^2}{2} \quad \text{für jedes } x > 0. \quad (23.4)$$

Man könnte versucht sein zu glauben, dass ein Zentraler Grenzwertsatz die Aussage (23.4) auch für alle zentrierten u.i.v. Folgen (X_i) mit endlicher Varianz liefert. Dies ist allerdings falsch, wie der folgende Satz zeigt. Die großen Abweichungen werden eben stärker durch die Schwänze der Verteilung von X_i beeinflusst, als dies bei den mittleren Fluktuationen der Fall ist, die durch die Varianz komplett determiniert werden. Der folgende Satz zeigt dies exemplarisch anhand der Bernoulli-Verteilung.

Satz 23.1. Seien X_1, X_2, \dots u.i.v. mit $\mathbf{P}[X_1 = -1] = \mathbf{P}[X_1 = 1] = \frac{1}{2}$. Dann gilt für jedes $x \geq 0$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}[S_n \geq xn] = -I(x), \quad (23.5)$$

wobei die **Ratenfunktion** I gegeben ist durch

$$I(z) = \begin{cases} \frac{1+z}{2} \log(1+z) + \frac{1-z}{2} \log(1-z), & \text{falls } z \in [-1, 1], \\ \infty, & \text{falls } |z| > 1. \end{cases} \quad (23.6)$$

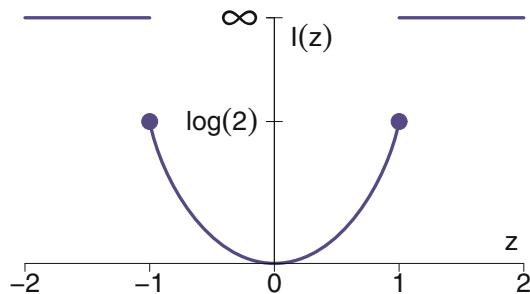


Abb. 23.1. Ratenfunktion $I(z)$ aus (23.6).

Bemerkung 23.2. Wir verstehen hierbei $0 \log 0 = 0$, wodurch die Einschränkung von I auf $[-1, 1]$ stetig wird mit $I(-1) = I(1) = \log 2$. Man bemerke: I ist strikt konkav auf $[-1, 1]$ mit $I(0) = 0$; I ist monoton wachsend auf $[0, 1]$ und monoton fallend auf $[-1, 0]$. \diamond

Beweis. Die Fälle $x = 0$ und $x > 1$ sind trivial. Für $x = 1$ ist $\mathbf{P}[S_n \geq n] = 2^{-n}$, daher gilt auch hier (23.5) trivialerweise. Es reicht also, $x \in (0, 1)$ zu betrachten. Es ist $\frac{S_n+n}{2} \sim b_{n,1/2}$ binomialverteilt, also

$$\mathbf{P}[S_n \geq xn] = 2^{-n} \sum_{k \geq (1+x)n/2} \binom{n}{k}.$$

Wir setzen $a_n(x) = \lceil n(1+x)/2 \rceil$ für $n \in \mathbb{N}$ und erhalten, weil $k \mapsto \binom{n}{k}$ monoton fallend ist für $k \geq \frac{n}{2}$:

$$Q_n(x) := \max \left\{ \binom{n}{k} : a_n(x) \leq k \leq n \right\} = \binom{n}{a_n(x)}. \quad (23.7)$$

Wir machen die Abschätzung

$$2^{-n} Q_n(x) \leq \mathbf{P}[S_n \geq xn] \leq (n+1) 2^{-n} Q_n(x). \quad (23.8)$$

Die Stirling'sche Formel

$$\lim_{n \rightarrow \infty} \frac{1}{n!} n^n e^{-n} \sqrt{2\pi n} = 1$$

liefert nun

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log Q_n(x) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{n!}{a_n(x)! \cdot (n - a_n(x))!} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{n^n}{a_n(x)^{a_n(x)} \cdot (n - a_n(x))^{n - a_n(x)}} \\ &= \lim_{n \rightarrow \infty} \left(\log(n) - \frac{a_n(x)}{n} \log(a_n(x)) - \frac{n - a_n(x)}{n} \log(n - a_n(x)) \right) \\ &= \lim_{n \rightarrow \infty} \left(\log(n) - \frac{1+x}{2} \left(\log \frac{1+x}{2} + \log(n) \right) \right. \\ &\quad \left. - \frac{1-x}{2} \left(\log \frac{1-x}{2} + \log(n) \right) \right) \\ &= -\frac{1+x}{2} \log \frac{1+x}{2} - \frac{1-x}{2} \log \frac{1-x}{2} = -I(x) + \log 2. \end{aligned}$$

Wegen (23.8) folgt hieraus (23.5). \square

Ein allgemeines Verfahren zur Bestimmung der Ratenfunktion I (unter gewissen restriktiven Annahmen an die Verteilung von (X_i)) liefert der Satz von Cramér [34].

Satz 23.3 (Cramér (1938)). Seien X_1, X_2, \dots u.i.v. reelle Zufallsvariablen mit logarithmischer momentenerzeugender Funktion

$$\Lambda(t) := \log \mathbf{E}[e^{tX_1}] < \infty \quad \text{für jedes } t \in \mathbb{R}. \quad (23.9)$$

Sei

$$\Lambda^*(x) := \sup_{t \in \mathbb{R}} (tx - \Lambda(t)) \quad \text{für } x \in \mathbb{R},$$

die Legendre-Transformierte von Λ . Dann gilt für jedes $x > \mathbf{E}[X_1]$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}[S_n \geq xn] = -I(x) := -\Lambda^*(x). \quad (23.10)$$

Beweis. Indem wir gegebenenfalls $X_i - x$ betrachten, können wir $\mathbf{E}[X_i] < 0$ und $x = 0$ annehmen. (Ist nämlich $\tilde{X}_i := X_i - x$ und sind $\tilde{\Lambda}$ und $\tilde{\Lambda}^*$ wie Λ und Λ^* definiert, jedoch mit \tilde{X}_i statt X_i , so ist $\tilde{\Lambda}(t) = \Lambda(t) - t \cdot x$ und daher gilt $\tilde{\Lambda}^*(0) = \sup_{t \in \mathbb{R}} (-\tilde{\Lambda}(t)) = \Lambda^*(x)$.)

Setze $\varphi(t) := e^{\Lambda(t)}$ und

$$\varrho := e^{-\Lambda^*(0)} = \inf_{t \in \mathbb{R}} \varphi(t).$$

Nach (23.9) und dem Differentiationslemma (Satz 6.28) ist φ unendlich oft differenzierbar, und die ersten beiden Ableitungen sind

$$\varphi'(t) = \mathbf{E}[X_1 e^{tX_1}] \quad \text{und} \quad \varphi''(t) = \mathbf{E}[X_1^2 e^{tX_1}].$$

Also ist φ strikt konvex und $\varphi'(0) = \mathbf{E}[X_1] < 0$.

Sei zunächst der Fall $\mathbf{P}[X_1 \leq 0] = 1$ betrachtet. Dann ist $\varphi'(t) < 0$ für jedes $t \in \mathbb{R}$ und $\varrho = \lim_{t \rightarrow \infty} \varphi(t) = \mathbf{P}[X_1 = 0]$. Die Behauptung folgt jetzt wegen

$$\mathbf{P}[S_n \geq 0] = \mathbf{P}[X_1 = \dots = X_n = 0] = \varrho^n.$$

Sei nun $\mathbf{P}[X_1 < 0] > 0$ und $\mathbf{P}[X_1 > 0] > 0$. Dann ist $\lim_{t \rightarrow \infty} \varphi(t) = \infty = \lim_{t \rightarrow -\infty} \varphi(t)$. Da φ strikt konvex ist, besitzt φ eine eindeutige Minimalstelle $\tau \in \mathbb{R}$, also

$$\varphi(\tau) = \varrho \quad \text{und} \quad \varphi'(\tau) = 0.$$

Wegen $\varphi'(0) < 0$ ist $\tau > 0$. Mit Hilfe der Markov'schen Ungleichung (Satz 5.11) erhalten wir die Abschätzung

$$\mathbf{P}[S_n \geq 0] = \mathbf{P}[e^{\tau S_n} \geq 1] \leq \mathbf{E}[e^{\tau S_n}] = \varphi(\tau)^n = \varrho^n.$$

Wir erhalten so die obere Schranke:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}[S_n \geq 0] \leq \log \varrho = -\Lambda^*(0).$$

Im Rest des Beweises müssen wir also die umgekehrte Ungleichung zeigen:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}[S_n \geq 0] \geq \log \varrho. \quad (23.11)$$

Wir verwenden eine Methode der exponentiellen Größenverzerrung der Verteilung $\mu := \mathbf{P}_{X_1}$ von X_1 , die untypische Werte typisch macht, damit man sie besser untersuchen kann. Wir definieren also die **Cramér-Transformierte** $\hat{\mu} \in \mathcal{M}_1(\mathbb{R})$ von μ durch

$$\hat{\mu}(dx) = \varrho^{-1} e^{\tau x} \mu(dx) \quad \text{für } x \in \mathbb{R}.$$

Seien $\hat{X}_1, \hat{X}_2, \dots$ unabhängig und identisch verteilt mit $\mathbf{P}_{\hat{X}_i} = \hat{\mu}$. Dann ist

$$\hat{\varphi}(t) := \mathbf{E}[e^{t\hat{X}_1}] = \frac{1}{\varrho} \int_{\mathbb{R}} e^{tx} e^{\tau x} \mu(dx) = \frac{1}{\varrho} \varphi(t + \tau).$$

Also ist

$$\begin{aligned} \mathbf{E}[\hat{X}_1] &= \hat{\varphi}'(0) = \frac{1}{\varrho} \varphi'(\tau) = 0, \\ \mathbf{Var}[\hat{X}_1] &= \hat{\varphi}''(0) = \frac{1}{\varrho} \varphi''(\tau) \in (0, \infty). \end{aligned}$$

Setzen wir $\hat{S}_n = \hat{X}_1 + \dots + \hat{X}_n$, so ist

$$\begin{aligned} \mathbf{P}[S_n \geq 0] &= \int_{\{x_1 + \dots + x_n \geq 0\}} \mu(dx_1) \cdots \mu(dx_n) \\ &= \int_{\{x_1 + \dots + x_n \geq 0\}} (\varrho e^{-\tau x_1}) \hat{\mu}(dx_1) \cdots (\varrho e^{-\tau x_n}) \hat{\mu}(dx_n) \\ &= \varrho^n \mathbf{E}\left[e^{-\tau \hat{S}_n} \mathbb{1}_{\{\hat{S}_n \geq 0\}}\right]. \end{aligned}$$

Wir erhalten also (23.11), wenn wir zeigen können, dass

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{E}\left[e^{-\tau \hat{S}_n} \mathbb{1}_{\{\hat{S}_n \geq 0\}}\right] \geq 0. \quad (23.12)$$

Nach dem Zentralen Grenzwertsatz (Satz 15.37) ist für $c > 0$

$$\begin{aligned} \frac{1}{n} \log \mathbf{E}\left[e^{-\tau \hat{S}_n} \mathbb{1}_{\{\hat{S}_n \geq 0\}}\right] &\geq \frac{1}{n} \log \mathbf{E}\left[e^{-\tau \hat{S}_n} \mathbb{1}_{\{0 \leq \hat{S}_n \leq c\sqrt{n}\}}\right] \\ &\geq \frac{1}{n} \log \left(e^{-\tau c\sqrt{n}} \mathbf{P}\left[\frac{\hat{S}_n}{\sqrt{n}} \in [0, c]\right] \right) \\ &\stackrel{n \rightarrow \infty}{\longrightarrow} \lim_{n \rightarrow \infty} \frac{-\tau c\sqrt{n}}{n} + \lim_{n \rightarrow \infty} \frac{1}{n} \log (\mathcal{N}_{0, \mathbf{Var}[\hat{X}_1]}([0, c])) \\ &= 0. \end{aligned}$$

□

Beispiel 23.4. Ist $\mathbf{P}_{X_1} = \mathcal{N}_{0,1}$, so ist

$$\Lambda(t) = \log(\mathbf{E}[e^{tX_1}]) = \log\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx\right) = \frac{t^2}{2}.$$

Weiter ist

$$\Lambda^*(z) = \sup_{t \in \mathbb{R}} (tz - \Lambda(t)) = \sup_{t \in \mathbb{R}} \left(tz - \frac{t^2}{2} \right) = \frac{z^2}{2}.$$

Die Ratenfunktion stimmt also mit der aus (23.4) überein. \diamond

Beispiel 23.5. Ist $\mathbf{P}_{X_1} = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, so ist $\Lambda(t) = \log \cosh(t)$. Der Maximierer $t^* = t^*(z)$ aus dem Variationsproblem für Λ^* erfüllt die Gleichung $z = \Lambda'(t^*) = \tanh(t^*)$. Also ist

$$\Lambda^*(z) = zt^* - \Lambda(t^*) = z \operatorname{arctanh}(z) - \log(\cosh(\operatorname{arctanh}(z))).$$

Nun ist $\operatorname{arctanh}(z) = \frac{1}{2} \log \frac{1+z}{1-z}$ für $z \in (-1, 1)$ und

$$\cosh(\operatorname{arctanh}(z)) = \frac{1}{\sqrt{1-z^2}} = \frac{1}{\sqrt{(1-z)(1+z)}}.$$

Es folgt

$$\begin{aligned} \Lambda^*(z) &= \frac{z}{2} \log(1+z) - \frac{z}{2} \log(1-z) + \frac{1}{2} \log(1-z) + \frac{1}{2} \log(1+z) \\ &= \frac{1+z}{2} \log(1+z) + \frac{1-z}{2} \log(1-z). \end{aligned}$$

Dies ist aber gerade die Ratenfunktion aus Satz 23.1. \diamond

Übung 23.1.1. Sei X eine reelle Zufallsvariable mit Dichte $f(x) = c^{-1} \frac{e^{-|x|}}{1+|x|^3}$, wobei $c = \int_{-\infty}^{\infty} \frac{e^{-|x|}}{1+|x|^3} dx$. Man untersuche die logarithmische momentenerzeugende Funktion Λ auf Unstetigkeitsstellen und skizziere den Graphen von Λ . \clubsuit

23.2 Prinzip der großen Abweichungen

Wir wollen in diesem Abschnitt die Idee des Satzes von Cramér, die Wahrscheinlichkeiten seltener, oder untypischer, Ereignisse mittels einer exponentiellen Rate und einer Ratenfunktion zu quantifizieren, in einen formalen Rahmen stellen. In diesem Rahmen kann die gesamte Theorie großer Abweichungen entwickelt werden; der Leser sei etwa auf die Bücher [36], [37] oder [81] verwiesen.

Sei E ein polnischer Raum mit vollständiger Metrik d . Die offene Kugel um $x \in E$ mit Radius $\varepsilon > 0$ bezeichnen wir mit $B_\varepsilon(x) = \{y \in E : d(x, y) < \varepsilon\}$.

Eine Abbildung $f : E \rightarrow \overline{\mathbb{R}} = [-\infty, \infty]$ heißt **halbstetig von unten**, falls für jedes $a \in \overline{\mathbb{R}}$ die **Niveaumenge** $f^{-1}([-\infty, a]) \subset E$ abgeschlossen ist. (Speziell sind also stetige Abbildungen stets halbstetig von unten. Allerdings ist $\mathbb{1}_{(0,1)} : \mathbb{R} \rightarrow \mathbb{R}$ halbstetig von unten, jedoch nicht stetig.) Äquivalent hierzu ist die Bedingung, dass $\lim_{\varepsilon \downarrow 0} \inf f(B_\varepsilon(x)) = f(x)$ ist für jedes $x \in E$. (Man beachte, dass $\inf f(A) = \inf\{f(x) : x \in A\}$.) Ist $K \subset E$ kompakt und nichtleer, so nimmt f auf K das Infimum an. In der Tat: Für den Fall, wo $f(x) = \infty$ für jedes $x \in K$ ist, ist die Aussage trivial. Sei nun $\inf f(K) < \infty$. Ist $a_n \downarrow \inf f(K)$ streng monoton fallend, so ist $K \cap f^{-1}([-\infty, a_n]) \neq \emptyset$ kompakt für jedes $n \in \mathbb{N}$, also ist auch der unendliche Schnitt nichtleer

$$f^{-1}(\inf f(K)) = K \cap \bigcap_{n=1}^{\infty} f^{-1}([-\infty, a_n]) \neq \emptyset.$$

Definition 23.6 (Ratenfunktion). Eine von unten halbstetige Funktion $I : E \rightarrow [0, \infty]$ heißt **Ratenfunktion**. Sind alle Niveaumengen $I^{-1}([-\infty, a])$, $a \in [0, \infty)$, kompakt, so nennen wir I eine gute **Ratenfunktion**.

Definition 23.7 (Prinzip großer Abweichungen). Sei I eine Ratenfunktion und $(\mu_\varepsilon)_{\varepsilon > 0}$ eine Familie von W-Maßen auf E . Wir sagen, dass $(\mu_\varepsilon)_{\varepsilon > 0}$ ein **Prinzip großer Abweichungen** (kurz: **LDP** für **Large Deviations Principle**) mit Ratenfunktion I erfüllt, falls

$$(LDP 1) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \log(\mu_\varepsilon(U)) \geq -\inf I(U) \quad \text{für jedes offene } U \subset E,$$

$$(LDP 2) \quad \limsup_{\varepsilon \rightarrow 0} \varepsilon \log(\mu_\varepsilon(C)) \leq -\inf I(C) \quad \text{für jedes abgeschlossene } C \subset E.$$

Wir sagen, dass eine Familie $(P_n)_{n \in \mathbb{N}}$ von W-Maßen auf E ein LDP mit Rate $r_n \uparrow \infty$ und Ratenfunktion I erfüllt, falls (LDP 1) und (LDP 2) für die Folge $\varepsilon_n = 1/r_n$ und für $\mu_{1/r_n} = P_n$ gelten.

Oftmals werden die Bedingungen (LDP 1) und (LDP 2) kurz *untere Schranke* und *obere Schranke* genannt. In vielen Fällen ist die untere Schranke leichter zu zeigen als die obere.

Bevor wir zeigen, dass der Satz von Cramér im Wesentlichen schon ein LDP ist, bringen wir noch zwei mehr technische Aussagen.

Satz 23.8. Die Ratenfunktion in einem LDP ist eindeutig.

Beweis. Es erfülle $(\mu_\varepsilon)_{\varepsilon>0}$ das LDP mit Ratenfunktionen I und J . Dann ist für jedes $x \in E$ und $\delta > 0$

$$\begin{aligned} I(x) &\geq \inf I(B_\delta(x)) \\ &\geq -\liminf_{\varepsilon \rightarrow 0} \varepsilon \log(\mu_\varepsilon(B_\delta(x))) \\ &\geq -\limsup_{\varepsilon \rightarrow 0} \varepsilon \log(\mu_\varepsilon(\overline{B_\delta(x)})) \\ &\geq \inf J(\overline{B_\delta(x)}) \xrightarrow{\delta \rightarrow 0} J(x). \end{aligned}$$

Es folgt $I(x) \geq J(x)$ und analog $J(x) \geq I(x)$. \square

Lemma 23.9. Sei $N \in \mathbb{N}$, und seien a_ε^i , $i = 1, \dots, N$, $\varepsilon > 0$, nichtnegative Zahlen. Dann gilt

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \sum_{i=1}^N a_\varepsilon^i = \max_{i=1, \dots, N} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log(a_\varepsilon^i).$$

Beweis. Summe und Maximum unterscheiden sich höchstens um den Faktor N :

$$\max_{i=1, \dots, N} \varepsilon \log(a_\varepsilon^i) \leq \varepsilon \log \sum_{i=1}^N a_\varepsilon^i \leq \varepsilon \log(N) + \max_{i=1, \dots, N} \varepsilon \log(a_\varepsilon^i).$$

Maximum und Limes (superior) vertauschen, also ist

$$\begin{aligned} \max_{i=1, \dots, N} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log(a_\varepsilon^i) &= \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \left(\max_{i=1, \dots, N} a_\varepsilon^i \right) \\ &\leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \left(\sum_{i=1}^N a_\varepsilon^i \right) \\ &\leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \log(N) + \max_{i=1, \dots, N} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log(a_\varepsilon^i) \\ &= \max_{i=1, \dots, N} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log(a_\varepsilon^i). \end{aligned} \quad \square$$

Beispiel 23.10. Wir nehmen an, dass die Bedingungen aus dem Satz von Cramér (Satz 23.3) gelten. Es seien also X_1, X_2, \dots u.i.v. reelle Zufallsvariablen mit $\Lambda(t) = \log(\mathbf{E}[e^{tX_1}]) < \infty$ für jedes $t \in \mathbb{R}$. Ferner sei $S_n = X_1 + \dots + X_n$ für jedes n . Wir wollen zeigen, dass aus dem Satz von Cramér folgt, dass $P_n := \mathbf{P}_{S_n/n}$ ein LDP mit Rate n und guter Ratenfunktion $I(x) = \Lambda^*(x) := \sup_{t \in \mathbb{R}} (tx - \Lambda(t))$ erfüllt. Ohne Einschränkung können wir annehmen, dass $\mathbf{E}[X_1] = 0$ ist. Die Funktion I ist halbstetig von unten, strikt konvex (wo sie endlich ist) und hat die eindeutige Minimalstelle bei $I(0) = 0$. Für $y > x \geq 0$ (oder $y < x \leq 0$) mit $I(x) < \infty$ gilt auf Grund der Konvexität $I(y) > I(x)$.

Der Satz von Cramér besagt, dass $\lim_{n \rightarrow \infty} \frac{1}{n} \log(P_n([x, \infty))) = -I(x)$ für $x > 0$ und (aus Symmetriegründen) $\lim_{n \rightarrow \infty} \frac{1}{n} \log(P_n((-∞, x])) = -I(x)$ für $x < 0$. Offenbar ist für $x > 0$

$$\begin{aligned} -I(x) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n((x, \infty)) \\ &\geq \sup_{y > x} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n([y, \infty)) = -\inf_{y > x} I(y) \end{aligned}$$

und für $x < 0$ analog $\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n((-\infty, x)) \geq -\inf_{y < x} I(y)$. Ferner gilt nach dem Gesetz der großen Zahl natürlich für jedes $x > 0$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log P_n((-x, \infty)) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log P_n([-x, \infty)) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log P_n((-\infty, x)) = \lim_{n \rightarrow \infty} \frac{1}{n} \log P_n((-\infty, x]) = 0 = -I(0). \end{aligned}$$

Die Familie $(P_n)_{n \in \mathbb{N}}$ erfüllt die Bedingungen (LDP 1) und (LDP 2) also zumindest für unbeschränkte Intervalle. Hiermit ist die Hauptarbeit getan, der Rest besteht darin, mit Standardargumenten (LDP 1) und (LDP 2) für beliebige offene, beziehungsweise abgeschlossene Mengen zu zeigen.

Sei zunächst $C \subset \mathbb{R}$ abgeschlossen. Wir definieren $x_+ := \inf(C \cap [0, \infty))$ und $x_- := \sup(C \cap (-\infty, 0])$. Aufgrund der Monotonie von I auf $(-\infty, 0]$ und $[0, \infty)$ folgt $\inf I(C) = I(x_-) \wedge I(x_+)$ (mit der Konvention $I(-\infty) = I(\infty) = \infty$). Ist $x_- = 0$ oder $x_+ = 0$, so ist $\inf I(C) = 0$, und (LDP 2) gilt trivialerweise. Sei nun also $x_- < 0 < x_+$. Wir erhalten mit Hilfe von Lemma 23.9

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_n(C) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log (P_n((-\infty, x_-]) + P_n([x_+, \infty))) \\ &= \max \left\{ \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_n((-\infty, x_-]), \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_n([x_+, \infty)) \right\} \\ &= \max \{-I(x_-), -I(x_+)\} = -\inf I(C). \end{aligned}$$

Damit ist (LDP 2) gezeigt.

Sei nun $U \subset \mathbb{R}$ offen und $x \in U \cap [0, \infty)$ mit $I(x) < \infty$ (falls es solch ein x gibt). Dann existiert ein $\varepsilon > 0$ mit $(x - \varepsilon, x + \varepsilon) \subset U$. Nun ist

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n((x - \varepsilon, \infty)) &\geq -I(x) > -I(x + \varepsilon) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log P_n([x + \varepsilon, \infty)). \end{aligned}$$

Es folgt

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n(U) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n((x - \varepsilon, x + \varepsilon)) \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \log (P_n((x - \varepsilon, \infty)) - P_n([x + \varepsilon, \infty))) \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \log (P_n((x - \varepsilon, \infty))) \geq -I(x). \end{aligned}$$

Analog folgt dies auch für $x \in U \cap (-\infty, 0)$ mit $I(x) < \infty$, also ist

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n(U) \geq -\inf I(U).$$

Damit ist die untere Schranke (LDP 1) gezeigt. \diamond

Tatsächlich kann man auf die Bedingung, dass $\Lambda(t) < \infty$ für alle $t \in \mathbb{R}$ gilt, verzichten. Da offenbar $\Lambda(0) = 0$ ist, ist $\Lambda^*(x) \geq 0$ für jedes $x \in \mathbb{R}$. Die Abbildung Λ^* ist eine konvexe Ratenfunktion, jedoch im Allgemeinen keine gute Ratenfunktion. Wir zitieren die folgende Verstärkung des Satzes von Cramér (siehe [36, Theorem 2.2.3]).

Satz 23.11 (Cramér). Sind X_1, X_2, \dots u.i.v. reelle Zufallsvariablen, dann erfüllt $(P_{S_n/n})_{n \in \mathbb{N}}$ ein LDP mit Ratenfunktion Λ^* .

Übung 23.2.1. Sei $E = \mathbb{R}$. Man zeige, dass $\mu_\varepsilon := \mathcal{N}_{0,\varepsilon}$ ein LDP mit guter Ratenfunktion $I(x) = x^2/2$ erfüllt. Man zeige ferner, dass in der *oberen Schranke* (LDP 2) strikte Ungleichheit auftreten kann. \clubsuit

Übung 23.2.2. Sei $E = \mathbb{R}$. Man zeige, dass $\mu_\varepsilon := \mathcal{N}_{0,\varepsilon^2}$ ein LDP mit guter Ratenfunktion $I(x) = \infty \cdot \mathbb{1}_{\mathbb{R} \setminus \{0\}}(x)$ erfüllt. Man zeige ferner, dass in der *unteren Schranke* (LDP 1) strikte Ungleichheit auftreten kann. \clubsuit

Übung 23.2.3. Sei $E = \mathbb{R}$. Man zeige, dass $\mu_\varepsilon := \frac{1}{2}\mathcal{N}_{-1,\varepsilon} + \frac{1}{2}\mathcal{N}_{1,\varepsilon}$ ein LDP mit guter Ratenfunktion $I(x) = \min(\frac{1}{2}(x+1)^2, \frac{1}{2}(x-1)^2)$ erfüllt. \clubsuit

Übung 23.2.4. Man berechne Λ und Λ^* für den Fall, wo $X_1 \sim \exp_\theta$ für $\theta > 0$ und interpretiere die Aussage von Satz 23.11 für diesen Fall. Man prüfe, dass Λ^* die eindeutige Nullstelle bei $\mathbf{E}[X_1]$ hat. (Ergebnis: $\Lambda^*(x) = \theta x - \log(\theta x) - 1$ falls $x > 0$ und $= \infty$ sonst.) \clubsuit

Übung 23.2.5. Man berechne Λ und Λ^* für den Fall, wo X_1 Cauchy verteilt ist und interpretiere die Aussage von Satz 23.11 für diesen Fall. \clubsuit

Übung 23.2.6. Sei $X_\lambda \sim \text{Poi}_\lambda$ für jedes $\lambda > 0$. Man zeige, dass $\mu_\varepsilon := \mathbf{P}_{\varepsilon X_\lambda/\varepsilon}$ ein LDP mit guter Ratenfunktion $I(x) = x \log(x/\lambda) + \lambda - x$ für $x \geq 0$ (und $= \infty$ sonst) erfüllt. \clubsuit

Übung 23.2.7. Sei $(X_t)_{t \geq 0}$ die Irrfahrt auf \mathbb{Z} in stetiger Zeit, die mit Rate $\frac{1}{2}$ einen Schritt nach rechts springt und mit Rate $\frac{1}{2}$ einen Schritt nach links springt. Man zeige, dass $(P_{\varepsilon X_{1/\varepsilon}})_{\varepsilon > 0}$ ein LDP erfüllt mit der konvexen guten Ratenfunktion $I(x) = 1 + x \operatorname{arcsinh}(x) - \sqrt{1+x^2}$. \clubsuit

23.3 Satz von Sanov

Dieser Abschnitt ist an die Darstellung in [36] angelehnt.

Wir wollen hier ein Prinzip der großen Abweichungen vorstellen, das nicht auf einem linearen Raum basiert, wie der Satz von Cramér, sondern für empirische Verteilungen unabhängiger Zufallsvariablen mit Werten in einer endlichen Menge Σ , die meist *Alphabet* genannt wird, annehmen.

Sei μ ein W-Maß auf Σ mit $\mu(\{x\}) > 0$ für jedes $x \in \Sigma$. Seien ferner X_1, X_2, \dots u.i.v. Zufallsvariablen mit Werten in Σ und Verteilung $\mathbf{P}_{X_1} = \mu$. Wir wollen ein Prinzip großer Abweichungen für die empirischen Maße

$$\xi_n(X) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

herleiten. Man beachte, dass nach dem Gesetz der großen Zahl \mathbf{P} -fast sicher gilt, dass $\xi_n(X) \xrightarrow{n \rightarrow \infty} \mu$. Als Zustandsraum ergibt sich also $E = \mathcal{M}_1(\Sigma)$, ausgestattet mit der Metrik $d(\mu, \nu) = \|\mu - \nu\|_{TV}$ der Totalvariation. (Da Σ nur endlich viele Punkte enthält, sind in E die vage Konvergenz, die schwache Konvergenz und die Konvergenz in Totalvariation identisch.) Es sei weiterhin

$$E_n := \left\{ \mu \in \mathcal{M}_1(\Sigma) : n\mu(\{x\}) \in \mathbb{N}_0 \text{ für jedes } x \in \Sigma \right\}$$

der mögliche Wertebereich der Zufallsvariablen $\xi_n(X)$.

Wir erinnern an den Begriff der **Entropie** von μ

$$H(\mu) := - \int \log(\mu(\{x\})) \mu(dx).$$

Ist $\nu \in \mathcal{M}_1(\Sigma)$, so definieren wir die **relative Entropie** (oder **Kullback-Leibler Information**) nach [112]) von ν gegeben μ durch

$$H(\nu | \mu) := \int \log\left(\frac{\nu(\{x\})}{\mu(\{x\})}\right) \nu(dx). \quad (23.13)$$

Da $\mu(\{x\}) > 0$ ist für alle $x \in \Sigma$, ist der Integrand ν -f.s. endlich und damit ist auch das Integral endlich. Eine einfache Anwendung der Jensen'schen Ungleichung liefert, dass $H(\mu) \geq 0$ und $H(\nu | \mu) \geq 0$ ist (siehe Lemma 5.26 und Übung 5.3.3) sowie $H(\nu | \mu) = 0$ genau dann, wenn $\nu = \mu$ ist. Außerdem ist offenbar

$$H(\nu | \mu) + H(\nu) = - \int \log(\mu(\{x\})) \nu(dx). \quad (23.14)$$

Da die Abbildung $\nu \mapsto I_\mu(\nu) := H(\nu | \mu)$ stetig ist, ist I_μ eine Ratenfunktion.

Lemma 23.12. Für jedes $n \in \mathbb{N}$ und $\nu \in E_n$ gilt

$$(n+1)^{-\#\Sigma} e^{-n H(\nu | \mu)} \leq \mathbf{P}[\xi_n(X) = \nu] \leq e^{-n H(\nu | \mu)}. \quad (23.15)$$

Beweis. Wir betrachten die Menge möglicher Werte für das n -Tupel (X_1, \dots, X_n) , sodass $\xi_n(X) = \nu$ ist:

$$A_n(\nu) := \left\{ k = (k_1, \dots, k_n) \in \Sigma^n : \frac{1}{n} \sum_{i=1}^n \delta_{k_i} = \nu \right\}.$$

Für jedes $k \in A_n(\nu)$ ist (vergleiche (23.14))

$$\begin{aligned} \mathbf{P}[\xi_n(X) = \nu] &= \#A_n(\nu) \mathbf{P}[X_1 = k_1, \dots, X_n = k_n] \\ &= \#A_n(\nu) \prod_{x \in \Sigma} \mu(\{x\})^{n\nu(\{x\})} \\ &= \#A_n(\nu) \exp \left(n \int \nu(dx) \log \mu(\{x\}) \right) \\ &= \#A_n(\nu) \exp(-n[H(\nu) + H(\nu|\mu)]). \end{aligned}$$

Seien nun Y_1, Y_2, \dots u.i.v. Zufallsvariablen mit Werten in Σ und Verteilung $\mathbf{P}_{Y_1} = \nu$. Dann ist wie in der Rechnung für X (wegen $H(\nu|\nu) = 0$)

$$1 \geq \mathbf{P}[\xi_n(Y) = \nu] = \#A_n(\nu) e^{-nH(\nu)},$$

also $\#A_n(\nu) \leq e^{nH(\nu)}$. Hieraus folgt die zweite Ungleichung in (23.15).

Die Zufallsvariable $n \xi_n(Y)$ ist multinomialverteilt mit Parametern $(n\nu(\{x\}))_{x \in \Sigma}$, also ist die Abbildung $E_n \rightarrow [0, 1]$, $\nu' \mapsto \mathbf{P}[\xi_n(Y) = \nu']$ maximal in $\nu' = \nu$. Es folgt

$$\#A_n(\nu) = e^{nH(\nu)} \mathbf{P}[\xi_n(Y) = \nu] \geq \frac{e^{nH(\nu)}}{\#E_n} \geq (n+1)^{-\#\Sigma} e^{nH(\nu)}.$$

Hieraus folgt die erste Ungleichung in (23.15). \square

Wir kommen jetzt zum Hauptsatz dieses Abschnitts, dem Satz von Sanov (siehe [159] und [160]).

Satz 23.13 (Sanov (1957)). Seien X_1, X_2, \dots u.i.v. Zufallsvariablen mit Werten in der endlichen Menge Σ und mit Verteilung μ . Dann erfüllt die Familie $(\mathbf{P}_{\xi_n(X)})_{n \in \mathbb{N}}$ der Verteilungen der empirischen Maße ein LDP mit Rate n und Ratenfunktion $I_\mu := H(\cdot|\mu)$.

Beweis. Für jedes $A \subset E$ ist nach Lemma 23.12

$$\begin{aligned} \mathbf{P}[\xi_n(X) \in A] &= \sum_{\nu \in A \cap E_n} \mathbf{P}[\xi_n(X) = \nu] \\ &\leq \sum_{\nu \in A \cap E_n} e^{-nH(\nu|\mu)} \\ &\leq \#(A \cap E_n) \exp(-n \inf I_\mu(A \cap E_n)) \\ &\leq (n+1)^{-\#\Sigma} \exp(-n \inf I_\mu(A)). \end{aligned}$$

Es folgt

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}[\xi_n(X) \in A] \leq -\inf I_\mu(A),$$

also die obere Schranke im LDP (sogar für allgemeines A).

Analog erhalten wir mit der ersten Ungleichung aus Lemma 23.12

$$\mathbf{P}[\xi_n(X) \in A] \geq (n+1)^{-\#\Sigma} \exp(-n \inf I_\mu(A \cap E_n))$$

und damit

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}[\xi_n(X) \in A] \geq -\limsup_{n \rightarrow \infty} \inf I_\mu(A \cap E_n). \quad (23.16)$$

Man beachte, dass wir für diese Ungleichung im Infimum nicht einfach $A \cap E_n$ durch A ersetzen können. Wir zeigen vielmehr, dass dies für offenes A zumindest asymptotisch geht. Sei also $A \subset E$ offen. Für $\nu \in A$ gibt es ein $\varepsilon > 0$ mit $B_\varepsilon(\nu) \subset A$. Für $n \geq (2\#\Sigma)/\varepsilon$ ist $E_n \cap B_\varepsilon(\nu) \neq \emptyset$, also existiert eine Folge $\nu_n \xrightarrow{n \rightarrow \infty} \nu$ mit $\nu_n \in E_n \cap A$ für hinreichend großes $n \in \mathbb{N}$. Da I_μ stetig ist, gilt

$$\limsup_{n \rightarrow \infty} \inf I_\mu(A \cap E_n) \leq \lim_{n \rightarrow \infty} I_\mu(\nu_n) = I_\mu(\nu).$$

Da $\nu \in A$ beliebig war, folgt $\limsup_{n \rightarrow \infty} \inf I_\mu(A \cap E_n) = \inf I_\mu(A)$. \square

Beispiel 23.14. Sei $\Sigma = \{-1, 1\}$ und $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ die Gleichverteilung auf Σ . Schreiben wir $m = m(\nu) = \nu(\{1\}) - \nu(\{-1\})$, dann ist die relative Entropie von $\nu \in \mathcal{M}_1(\Sigma)$

$$H(\nu | \mu) = \frac{1+m}{2} \log(1+m) + \frac{1-m}{2} \log(1-m).$$

Dies ist genau die Ratenfunktion, die wir bereits aus Satz 23.1 kennen. \diamond

Wir wollen den Zusammenhang zwischen den LDPS von Sanov und von Cramér, der im letzten Beispiel angedeutet wurde, nun formal herstellen, indem wir eine Variante des Satzes von Cramér für \mathbb{R}^d -wertige Zufallsvariablen, die nur endlich viele Werte annehmen, aus dem Satz von Sanov herleiten.

Beispiel 23.15. Sei $\Sigma \subset \mathbb{R}^d$ endlich und μ ein W-Maß auf Σ . Seien ferner X_1, X_2, \dots u.i.v. Zufallsvariablen mit Werten in Σ und Verteilung $\mathbf{P}_{X_1} = \mu$ sowie $S_n = X_1 + \dots + X_n$ für jedes $n \in \mathbb{N}$. Wir setzen $\Lambda(t) = \log \mathbf{E}[e^{\langle t, X_1 \rangle}]$ für $t \in \mathbb{R}^d$ und $\Lambda^*(x) = \sup_{t \in \mathbb{R}^d} (\langle t, x \rangle - \Lambda(t))$ für $x \in \mathbb{R}^d$.

Wir zeigen, dass $(\mathbf{P}_{S_n/n})_{n \in \mathbb{N}}$ ein LDP mit Rate n und Ratenfunktion Λ^* erfüllt.

Es sei $\xi_n(X)$ das empirische Maß von X_1, \dots, X_n . Sei $E := \mathcal{M}_1(\Sigma)$. Definiere die Abbildung

$$m : E \rightarrow \mathbb{R}^d, \quad \nu \mapsto \int x \nu(dx) = \sum_{x \in \Sigma} x \nu(\{x\}),$$

die ν das erste Moment zuordnet. Offenbar ist dann $\frac{1}{n} S_n = m(\xi_n(X))$. Für $x \in \mathbb{R}^d$ und $A \subset \mathbb{R}^d$ seien $E_x := m^{-1}(\{x\}) = \{\nu \in E : m(\nu) = x\}$ und $E_A = m^{-1}(A) = \{\nu \in E : m(\nu) \in A\}$. Die Abbildung $\nu \mapsto m(\nu)$ ist stetig, also ist E_A offen (beziehungsweise abgeschlossen), falls A offen (beziehungsweise abgeschlossen) ist. Mit $\tilde{I}(x) := \inf I_\mu(E_x)$ (wobei $I_\mu(\nu) = H(\nu|\mu)$ die relative Entropie ist) gilt nach dem Satz von Sanov für offenes $U \subset \mathbb{R}^d$

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_{S_n/n}(U) &= \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_{\xi_n(X)}(m^{-1}(U)) \\ &\geq -\inf I_\mu(m^{-1}(U)) = -\inf \tilde{I}(U). \end{aligned}$$

Analog ist für abgeschlossenes $C \subset \mathbb{R}^d$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_{S_n/n}(C) \leq -\inf \tilde{I}(C).$$

Mit anderen Worten: $(\mathbf{P}_{S_n/n})_{n \in \mathbb{N}}$ erfüllt ein LDP mit Rate n und Ratenfunktion \tilde{I} . Es ist also nur noch zu zeigen, dass $\tilde{I} = \Lambda^*$ gilt.

Man beachte, dass $t \mapsto \Lambda(t)$ differenzierbar (mit Ableitung Λ') und strikt konvex ist. Daher besitzt das Variationsproblem für $\Lambda^*(x)$ einen eindeutigen Maximierer $t^*(x)$. Genauer gilt

$$\Lambda^*(x) = \langle t^*(x), x \rangle - \Lambda(t^*(x))$$

und $\Lambda^*(x) > \langle t, x \rangle - \Lambda(t)$ für alle $t \neq t^*(x)$ sowie $\Lambda'(t^*(x)) = x$.

Nach der Jensen'schen Ungleichung ist für jedes $\nu \in \mathcal{M}_1(\Sigma)$

$$\begin{aligned} \Lambda(t) &= \log \int e^{\langle t, y \rangle} \mu(dy) \\ &= \log \int \left(e^{\langle t, y \rangle} \frac{\mu(\{y\})}{\nu(\{y\})} \right) \nu(dy) \\ &\geq \int \log \left(e^{\langle t, y \rangle} \frac{\mu(\{y\})}{\nu(\{y\})} \right) \nu(dy) \\ &= \langle t, m(\nu) \rangle - H(\nu|\mu) \end{aligned}$$

mit Gleichheit genau dann, wenn $\nu = \nu_t$, wo $\nu_t(\{y\}) = \mu(\{y\})e^{\langle t, y \rangle - \Lambda(t)}$. Also ist

$$\langle t, x \rangle - \Lambda(t) \leq \inf_{\nu \in E_x} H(\nu|\mu)$$

mit Gleichheit, falls $\nu_t \in E_x$. Nun ist aber $m(\nu_t) = \Lambda'(t)$, also ist $\nu_{t^*(x)} \in E_x$ und damit

$$\Lambda^*(x) = \langle t^*(x), x \rangle - \Lambda(t^*(x)) = \inf_{\nu \in E_x} H(\nu|\mu) = \tilde{I}(x). \quad \diamond$$

Das Beweisprinzip, das wir im letzten Beispiel verwandt haben, um das LDP mit Ratenfunktion \tilde{I} herzuleiten, wird **Kontraktionsprinzip** genannt. Wir formulieren es als Satz.

Satz 23.16 (Kontraktionsprinzip). *Die Familie $(\mu_\varepsilon)_{\varepsilon > 0}$ von W-Maßen auf E erfülle ein LDP mit Ratenfunktion I . Ist F ein topologischer Raum und $m : E \rightarrow F$ stetig, so erfüllen die Bildmaße $(\mu_\varepsilon \circ m^{-1})_{\varepsilon > 0}$ ein LDP mit Ratenfunktion $\tilde{I}(x) = \inf I(m^{-1}(\{x\}))$.*

23.4 Varadhan'sches Lemma und Freie Energie

Wir nehmen an, dass $(\mu_\varepsilon)_{\varepsilon > 0}$ eine Familie von W-Maßen ist, die ein LDP mit Ratenfunktion I erfüllt. Wir wissen also, dass die Masse von μ_ε für kleine $\varepsilon > 0$ mehr und mehr um die Nullstellen von I herum konzentriert liegt. In der statistischen Physik ist es oftmals von Interesse, Funktionen bezüglich μ_ε (wobei $1/\varepsilon$ als „Systemgröße“ verstanden wird) zu integrieren, die ihren größten Wert nicht in den Nullstellen von I annehmen, und die zudem noch exponentiell mit $1/\varepsilon$ skalieren. Es soll also die Asymptotik von $Z_\varepsilon^\phi := \int e^{\phi(x)/\varepsilon} \mu_\varepsilon(dx)$ für $\varepsilon \rightarrow 0$ untersucht werden. Unter harmlosen Stetigkeitsannahmen tragen zu dem Integral hauptsächlich diejenigen Punkte x bei, für die $\phi(x)$ groß ist, die aber gleichzeitig nicht zu unwahrscheinlich sind, also diejenigen x , für die $\phi(x) - I(x)$ die größten Werte annimmt. Die Beiträge messen wir durch die *gekippten* W-Maße $\mu_\varepsilon^\phi(dx) = (Z_\varepsilon^\phi)^{-1} e^{\phi(x)/\varepsilon} \mu_\varepsilon(dx)$, für die wir ein LDP herleiten. Als Anwendung folgern wir das Prinzip der Minimierung der freien Energie in der statistischen Physik und analysieren speziell den Weiss'schen Ferromagneten.

Satz 23.17 (Varadhan'sches Lemma (1966)). *Sei I eine gute Ratenfunktion und $(\mu_\varepsilon)_{\varepsilon > 0}$ eine Familie von W-Maßen auf E , die ein LDP mit Ratenfunktion I erfüllt. Sei ferner $\phi : E \rightarrow \mathbb{R}$ stetig und erfülle die Bedingung*

$$\inf_{M > 0} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \int e^{\phi(x)/\varepsilon} \mathbb{1}_{\{\phi(x) \geq M\}} \mu_\varepsilon(dx) = -\infty. \quad (23.17)$$

Dann gilt

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \int e^{\phi(x)/\varepsilon} \mu_\varepsilon(dx) = \sup_{x \in E} (\phi(x) - I(x)). \quad (23.18)$$

Bemerkung 23.18. Die Bedingung (23.17) folgt aus der etwas griffigeren Bedingung, dass es ein $\alpha > 1$ gibt mit

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \int e^{\alpha \phi/\varepsilon} d\mu_\varepsilon < \infty. \quad (23.19)$$

In der Tat: Für jedes $M \in \mathbb{R}$ ist

$$\begin{aligned} \varepsilon \log \int e^{\phi(x)/\varepsilon} \mathbb{1}_{\{\phi(x) \geq M\}} \mu_\varepsilon(dx) &= M + \varepsilon \log \int e^{(\phi(x)-M)/\varepsilon} \mathbb{1}_{\{\phi(x) \geq M\}} \mu_\varepsilon(dx) \\ &\leq M + \varepsilon \log \int e^{\alpha(\phi(x)-M)/\varepsilon} \mu_\varepsilon(dx) \\ &= -(\alpha-1)M + \varepsilon \log \int e^{\alpha\phi(x)/\varepsilon} \mu_\varepsilon(dx). \end{aligned}$$

Hieraus und aus (23.19) folgt sofort (23.17). \diamond

Beweis. Wir zeigen mit unterschiedlichen Argumenten, dass die rechte Seite in (23.18) eine untere Schranke und eine obere Schranke für die linke Seite ist.

Untere Schranke. Für jedes $x \in E$ und $r > 0$ ist

$$\begin{aligned} \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \int e^{\phi/\varepsilon} d\mu_\varepsilon &\geq \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \int_{B_r(x)} e^{\phi/\varepsilon} d\mu_\varepsilon \\ &\geq \inf \phi(B_r(x)) - I(x) \xrightarrow{r \rightarrow 0} \phi(x) - I(x). \end{aligned}$$

Obere Schranke. Für $M > 0$ und $\varepsilon > 0$ definieren wir

$$F_M^\varepsilon := \int_{\{\phi \geq M\}} e^{\phi(x)/\varepsilon} \mu_\varepsilon(dx) \quad \text{und} \quad G_M^\varepsilon := \int_{\{\phi < M\}} e^{\phi(x)/\varepsilon} \mu_\varepsilon(dx).$$

Wir setzen

$$F_M := \limsup_{\varepsilon \rightarrow 0} \varepsilon \log F_M^\varepsilon \quad \text{und} \quad G_M := \limsup_{\varepsilon \rightarrow 0} \varepsilon \log G_M^\varepsilon.$$

Nach Lemma 23.9 gilt für jedes $M > 0$

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \int e^{\phi(x)/\varepsilon} \mu_\varepsilon(dx) = F_M \vee G_M.$$

Da nach Voraussetzung $\inf_{M>0} F_M = -\infty$ gilt, reicht es zu zeigen, dass

$$\sup_{M>0} G_M \leq \sup_{x \in E} (\phi(x) - I(x)). \quad (23.20)$$

Sei $\delta > 0$. Für jedes $x \in E$ gibt es ein $r(x) > 0$ mit

$$\inf I(B_{2r(x)}(x)) \geq I(x) - \delta \quad \text{und} \quad \sup \phi(B_{2r(x)}(x)) \leq \phi(x) + \delta.$$

Sei $a \geq 0$. Da I eine gute Ratenfunktion ist, ist die Niveaumenge $K := I^{-1}([0, a])$ kompakt. Wir finden also endlich viele Punkte $x_1, \dots, x_N \in I^{-1}([0, a])$, sodass $\bigcup_{i=1}^N B_{r(x_i)}(x_i) \supset K$. Es gilt daher

$$\begin{aligned}
G_M^\varepsilon &\leq \int_{\{\phi < M\} \cap K^c} e^{\phi(x)/\varepsilon} \mu_\varepsilon(dx) + \sum_{i=1}^N \int_{\{\phi < M\} \cap B_{r(x_i)}(x_i)} e^{\phi(x)/\varepsilon} \mu_\varepsilon(dx) \\
&\leq e^{M/\varepsilon} \mu_\varepsilon(K^c) + \sum_{i=1}^N e^{(\phi(x_i) \wedge M + \delta)/\varepsilon} \mu_\varepsilon(B_{r(x_i)}(x_i)) \\
&= e^{(M + \varepsilon \log(\mu_\varepsilon(K^c)))/\varepsilon} + \sum_{i=1}^N e^{(\phi(x_i) \wedge M + \delta + \varepsilon \log(\mu_\varepsilon(B_{r(x_i)}(x_i))))/\varepsilon}.
\end{aligned}$$

Es folgt nach Lemma 23.9 und dem LDP

$$\begin{aligned}
G_M &\leq (M - a) \vee \max_{i=1,\dots,N} (\phi(x_i) - I(x_i) + 2\delta) \\
&\leq (M - a) \vee \sup_{x \in E} (\phi(x) - I(x)) + 2\delta.
\end{aligned}$$

Indem wir erst $\delta \downarrow 0$ und dann $a \uparrow \infty$ gehen lassen, folgt (23.20). \square

Satz 23.19 (Gekipptes LDP). Es erfülle $(\mu_\varepsilon)_{\varepsilon>0}$ ein LDP mit der guten Ratenfunktion I . Ferner sei $\phi : E \rightarrow \mathbb{R}$ stetig und erfülle die Bedingung (23.17). Wir definieren $Z_\varepsilon^\phi := \int e^{\phi/\varepsilon} d\mu_\varepsilon$ und $\mu_\varepsilon^\phi \in \mathcal{M}_1(E)$ durch

$$\mu_\varepsilon^\phi(dx) = (Z_\varepsilon^\phi)^{-1} e^{\phi(x)/\varepsilon} \mu_\varepsilon(dx).$$

Ferner definieren wir $I^\phi : E \rightarrow [0, \infty]$ durch

$$I^\phi(x) = \sup_{z \in E} (\phi(z) - I(z)) - (\phi(x) - I(x)). \quad (23.21)$$

Dann erfüllt $(\mu_\varepsilon^\phi)_{\varepsilon>0}$ ein LDP mit der Ratenfunktion I^ϕ .

Beweis. Dies bleibt dem Leser zur Übung überlassen. (Vergleiche [37, Aufgabe 2.1.24], siehe auch [48, Abschnitt II.7].) \square

Wir wollen das Varadhan'sche Lemma im Kontext der statistischen Physik betrachten. Sei hierzu Σ ein polnischer Raum, den wir als Raum der möglichen Zustände eines einzelnen Teilchens auffassen wollen. Ferner sei $\lambda \in \mathcal{M}_1(\Sigma)$ eine Verteilung, die wir als *a priori* Verteilung eines Teilchens ohne Berücksichtigung der Energie auffassen wollen. Ist Σ endlich oder eine beschränkte Menge eines \mathbb{R}^d , so ist λ typischerweise die Gleichverteilung auf Σ . Wenn wir n ununterscheidbare Teilchen unabhängig nach λ auf Positionen $z_1, \dots, z_n \in \Sigma$ setzen, so können wir den Zustand dieses Ensembles als $x := \frac{1}{n} \sum_{i=1}^n \delta_{z_i}$ beschreiben. Mit $\mu_n^0 \in \mathcal{M}_1(\mathcal{M}_1(\Sigma))$ bezeichnen wir die so gewonnene *a priori* Verteilung von x .

Wir machen nun die Annahme, dass sich die Energie $U_n(x)$ eines Zustandes schreiben lässt als $U_n(x) = nU(x)$, wo $U(x)$ als die mittlere Energie eines Teilchens bei Gesamtzustand x interpretiert wird.

Es sei $T > 0$ die Temperatur des Systems und $\beta := 1/T$ die so genannte **inverse Temperatur**. Eine wichtige Rolle in der statistischen Physik spielt die **Zustandssumme** oder **Partitionsfunktion**

$$Z_n^\beta := \int e^{-\beta U_n} d\mu_n^0.$$

Ein Postulat der statistischen Physik besagt, dass der Zustand x nach der **Boltzmann-Verteilung** verteilt ist:

$$\mu_n^\beta(dx) = (Z_n^\beta)^{-1} e^{-\beta U_n(x)} \mu_n^0(dx). \quad (23.22)$$

Das Varadhan'sche Lemma (genauer: das gekippte LDP) und der Satz von Sanov erlauben uns, die Brücke zum Variationsprinzip für die freie Energie zu schlagen. Wir nehmen nun an, dass Σ eine endliche Menge ist und $\lambda = \mathcal{U}_\Sigma$ die Gleichverteilung auf Σ . Nach dem Satz von Sanov erfüllt $(\mu_n^0)_{n \in \mathbb{N}}$ ein LDP mit Rate n und Ratenfunktion $I(x) = H(x|\lambda)$, wo $H(x|\lambda)$ die relative Entropie von x bezüglich λ ist. Nach (23.14) ist $H(x|\lambda) = \log(\#\Sigma) - H(x)$, wo $H(x)$ die Entropie von x ist.

Wir definieren die **freie Energie** (oder das **Helmholtz-Potential**) pro Teilchen als

$$F^\beta(x) := U(x) - \beta^{-1} H(x).$$

Der Satz über das gekippte LDP liefert nun, dass die Folge der Boltzmann-Verteilungen $(\mu_n^\beta)_{n \in \mathbb{N}}$ ein LDP erfüllt mit Rate n und Ratenfunktion

$$I^\beta(x) = F^\beta(x) - \inf_{y \in \mathcal{M}_1(\Sigma)} F^\beta(y).$$

Für großes n ist die Boltzmann Verteilung auf diejenigen x konzentriert, die die freie Energie minimieren. Dies können für unterschiedliche Temperaturen (also Werte von β) sehr unterschiedliche Zustände sein. Daher treten bei kritischen Temperaturen Phasenübergänge auf, und chemische Reaktionen laufen bei unterschiedlichen Temperaturen in unterschiedlichen Richtungen ab.

Beispiel 23.20. Wir betrachten den **Weiss'schen Ferromagneten**. Dies ist ein mikroskopisches Modell für Magnetismus, das davon ausgeht, dass jedes von n magnetischen Teilchen eine von den zwei Ausrichtungen $\sigma_i \in \Sigma = \{-1, +1\}$ hat. Die mittlere Magnetisierung $m = \frac{1}{n} \sum_{i=1}^n \sigma_i$ beschreibt den Zustand des Systems vollständig (da die Teilchen ununterscheidbar sind) und ist die relevante makroskopische Messgröße. Die Grundidee ist, dass es energetisch günstiger ist, wenn Teilchen magnetisch parallel ausgerichtet sind, als wenn sie antiparallel ausgerichtet sind. Wir wollen die räumliche Struktur der Wechselwirkung ignorieren und annehmen, dass jedes Teilchen mit jedem anderen in gleicher Weise wechselwirkt (**mean field** Annahme). Außerdem wollen wir annehmen, dass es ein äußeres Magnetfeld der Stärke h gibt. Bis auf Konstanten ist die Energie pro Teilchen daher

$$U(m) = -\frac{1}{2}m^2 - hm.$$

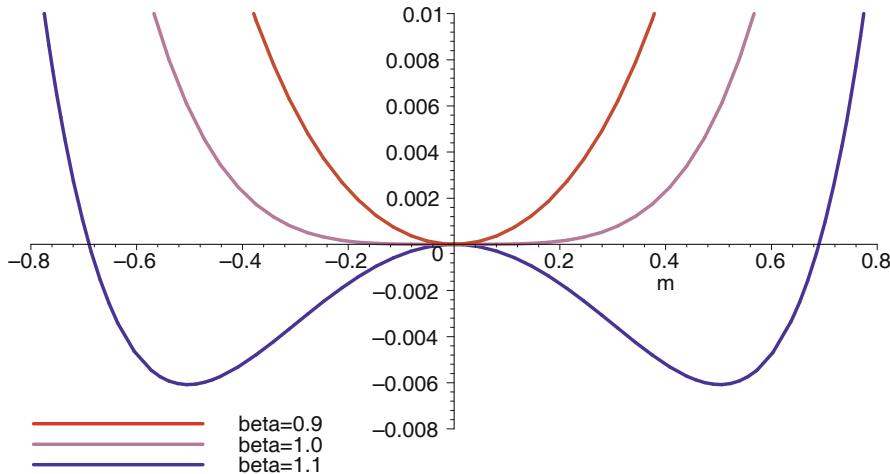


Abb. 23.2. Die verschobene freie Energie $F^\beta(m) - F^\beta(0)$ des Weiss'schen Ferromagneten ohne äußerer Feld ($h = 0$).

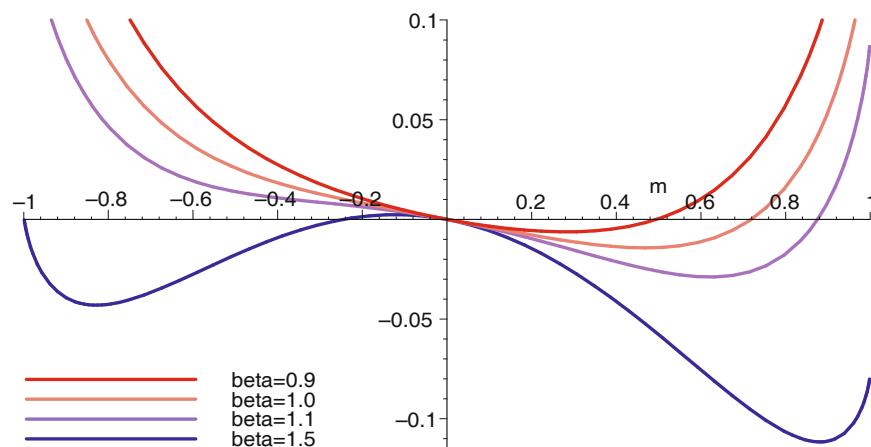


Abb. 23.3. Die verschobene freie Energie $F^\beta(m) - F^\beta(0)$ des Weiss'schen Ferromagneten mit äußerem Feld $h = 0.04$.

Die Entropie des Zustands m ist

$$H(m) = -\frac{1+m}{2} \log \left(\frac{1+m}{2} \right) - \frac{1-m}{2} \log \left(\frac{1-m}{2} \right).$$

Die freie Energie pro Teilchen ist also

$$F^\beta(m) = -\frac{1}{2}m^2 - hm + \beta^{-1} \left[\frac{1+m}{2} \log \left(\frac{1+m}{2} \right) + \frac{1-m}{2} \log \left(\frac{1-m}{2} \right) \right].$$

Um die Minimalstellen von F^β zu bestimmen, berechnen wir die Ableitung

$$0 \stackrel{!}{=} \frac{d}{dm} F^\beta(m) = -m - h + \beta^{-1} \operatorname{arctanh}(m).$$

Wir erhalten so für m die Gleichung

$$m = \tanh(\beta(m+h)). \quad (23.23)$$

Im Fall $h = 0$ hat (23.23) stets die Lösung $m = 0$. Ist $\beta \leq 1$, so ist diese Lösung eindeutig, und F^β hat das globale Minimum in $m = 0$. Ist $\beta > 1$, so besitzt (23.23) zwei weitere Lösungen $m_-^{\beta,0} \in (-1, 0)$ und $m_+^{\beta,0} = -m_-^{\beta,0}$, die nur numerisch bestimmt werden können.

In diesem Fall besitzt F^β in 0 ein lokales Maximum und in $m_\pm^{\beta,0}$ globale Minima. Da für große n nur noch solche Werte angenommen werden, für die F^β minimal ist, liegt die Verteilung konzentriert um 0, falls $\beta \leq 1$ und konzentriert um $m_\pm^{\beta,0}$, falls $\beta > 1$. Im letzteren Fall ist die betragsmäßige Magnetisierung $|m_\pm^{\beta,0}| = m_+^{\beta,0} > 0$. Wir haben also einen **Phasenübergang** zwischen einer Phase bei hoher Temperatur ($\beta \leq 1$), wo keine Magnetisierung auftritt, und niedriger Temperatur ($\beta > 1$), wo so genannte spontane Magnetisierung auftritt (das heißt ohne Einwirkung eines äußeren Feldes).

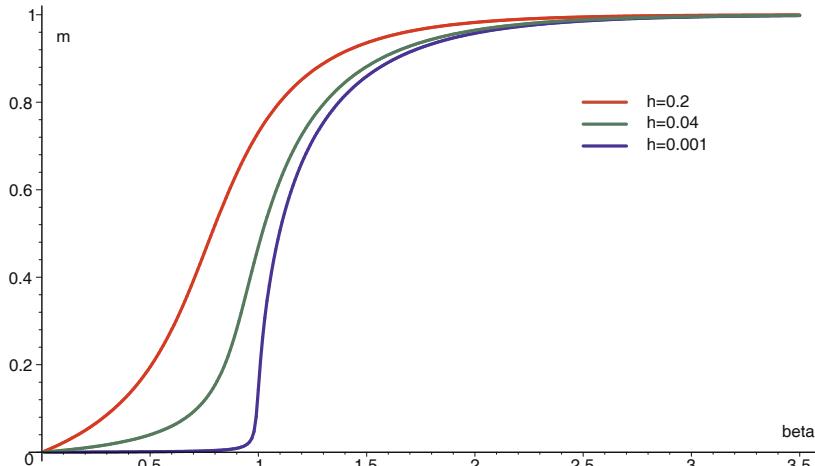


Abb. 23.4. Weiss'scher Ferromagnet: Magnetisierung $m^{\beta,h}$ als Funktion von β .

Ist $h \neq 0$, so besitzt F^β in $m = 0$ keine Minimalstelle. Vielmehr ist F^β asymmetrisch und besitzt ein globales Minimum $m^{\beta,h}$ mit selbem Vorzeichen wie h sowie für großes β noch eine weiteres lokales Minimum mit dem entgegengesetzten Vorzeichen. Die exakten Werte für die Magnetisierung können wieder nur numerisch

bestimmt werden. Wir können $m^{\beta,h}$ jedoch für hohe Temperaturen (β klein) approximativ bestimmen, indem wir die Näherung $\tanh(\beta(m+h)) \approx \beta(m+h)$ verwenden. Wir erhalten so

$$m^{\beta,h} \approx \frac{h}{\beta^{-1} - 1} = \frac{h}{T - T_c} \quad \text{für } T \rightarrow \infty, \quad (23.24)$$

wo die *Curie-Temperatur* $T_c = 1$ die kritische Temperatur für das Auftreten von spontaner Magnetisierung ist. Die Beziehung (23.24) heißt **Curie-Weiss'sches Gesetz**. \diamond

24 Der Poisson'sche Punktprozess

Poisson'sche Punktprozesse können als ein Grundbaustein zur Konstruktion sehr unterschiedlicher stochastischer Objekte verwendet werden, wie etwa unbegrenzt teilbare Verteilungen, Markovprozesse mit komplexer Dynamik, Objekte der stochastischen Geometrie und so fort.

Wir geben in diesem Kapitel kurz den allgemeinen Rahmen zufälliger Maße an, konstruieren den Poisson'schen Punktprozess und charakterisieren ihn durch seine Laplace-Transformierte. Als Anwendung konstruieren wir einen Subordinator und zeigen, dass der Poisson'sche Punktprozess das invariante Maß von Systemen unabhängiger Irrfahrten ist. Über den Zusammenhang zu Subordinatoren schlagen wir im dritten Abschnitt die Brücke zu den in der Populationsgenetik wichtigen Poisson-Dirichlet und GEM Verteilungen.

24.1 Zufällige Maße

Sei E im Folgenden ein lokalkompakter, polnischer Raum (und damit insbesondere σ -kompakt), etwa $E = \mathbb{R}^d$ oder $E = \mathbb{Z}^d$, mit Borel'scher σ -Algebra $\mathcal{B}(E)$. Sei

$$\mathcal{B}_b(E) = \{B \in \mathcal{B}(E) : B \text{ ist relativ kompakt}\}$$

das System der *beschränkten* Borel'schen Mengen und $\mathcal{M}(E)$ der Raum der Radon-Maße auf E (siehe Definition 13.3).

Definition 24.1. Wir bezeichnen mit $\mathbb{M} = \sigma(I_A : A \in \mathcal{B}_b(E))$ die kleinste σ -Algebra auf $\mathcal{M}(E)$, bezüglich der alle Abbildungen

$$I_A : \mu \mapsto \mu(A), \quad A \in \mathcal{B}_b(E),$$

messbar sind.

Wir schreiben $\mathcal{B}^+(E)$ für die Menge der messbaren Abbildungen $E \rightarrow [0, \infty]$ und $\mathcal{B}_b^+(E)$ für die Menge der beschränkten, messbaren Abbildungen $E \rightarrow \mathbb{R}$ mit kompaktem Träger. Das Integral $I_f(\mu) := \int f d\mu$ ist für jedes $f \in \mathcal{B}^+(E)$ wohldefiniert und für jedes $f \in \mathcal{B}_b^+(E)$ wohldefiniert und endlich.

Satz 24.2. Sei τ_v die vage Topologie auf $\mathcal{M}(E)$. Dann ist

$$\mathbb{M} = \mathcal{B}(\tau_v) = \sigma(I_f : f \in C_c(E)) = \sigma(I_f : f \in C_c^+(E)).$$

Beweis. Übung! (Siehe [90, Lemma 4.1].) \square

Sei $\widetilde{\mathcal{M}}(E)$ der Raum aller Maße auf E mit σ -Algebra $\widetilde{\mathbb{M}} = \sigma(I_A : A \in \mathcal{B}_b(E))$. Wählen wir eine abzählbare dichte Menge $F \subset E$ und zu jedem $x \in F$ eine kompakte Umgebung K_x , so erhalten wir (vergleiche Übung 13.1.8)

$$\mathcal{M}(E) = \bigcap_{x \in F} \{\mu \in \widetilde{\mathcal{M}}(E) : \mu(K_x) < \infty\}.$$

Also ist $\mathcal{M}(E) \in \widetilde{\mathbb{M}}$. Offenbar ist $\mathbb{M} = \widetilde{\mathbb{M}}|_{\mathcal{M}(E)}$ die Spur- σ -Algebra von $\widetilde{\mathbb{M}}$ auf $\mathcal{M}(E)$. Wir brauchen diesen etwas größeren Raum, um zufällige Maße so zu definieren, dass fast sicher wohldefinierte Operationen wieder zufällige Maße ergeben.

Definition 24.3. Ein zufälliges Maß auf E ist eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$ mit Werten in $(\widetilde{\mathcal{M}}(E), \widetilde{\mathbb{M}})$ und mit der Eigenschaft $\mathbf{P}[X \in \mathcal{M}(E)] = 1$.

Satz 24.4. Sei X ein zufälliges Maß auf E . Dann ist die Mengenfunktion $\mathbf{E}[X] : \mathcal{B}(E) \rightarrow [0, \infty]$, $A \mapsto \mathbf{E}[X(A)]$ ein Maß. Wir nennen $\mathbf{E}[X]$ das Intensitätsmaß von X . X heißt integrierbar, falls $\mathbf{E}[X] \in \mathcal{M}(E)$.

Beweis. Offenbar ist $\mathbf{E}[X]$ endlich additiv. Seien $A, A_1, A_2, \dots \in \mathcal{B}(E)$ mit $A_n \uparrow A$. Betrachte die Zufallsvariablen $Y_n := X(A_n)$ und $Y = X(A)$. Dann gilt $Y_n \uparrow Y$, also nach dem Satz über monotone Konvergenz $\mathbf{E}[X](A_n) = \mathbf{E}[Y_n] \xrightarrow{n \rightarrow \infty} \mathbf{E}[Y] = \mathbf{E}[X](A)$. Mithin ist $\mathbf{E}[X]$ stetig von unten und damit ein Maß (nach Satz 1.36). \square

Satz 24.5. Die Verteilung \mathbf{P}_X eines zufälligen Maßes X ist eindeutig bestimmt sowohl durch die Verteilungen von

$$((I_{f_1}, \dots, I_{f_n}) : n \in \mathbb{N}; f_1, \dots, f_n \in C_c^+(E)) \quad (24.1)$$

als auch von

$$((I_{A_1}, \dots, I_{A_n}) : n \in \mathbb{N}; A_1, \dots, A_n \in \mathcal{B}_b(E) \text{ paarweise disjunkt}). \quad (24.2)$$

Beweis. Das Mengensystem

$$\mathcal{I} = \{(I_{f_1}, \dots, I_{f_n})^{-1}(A) : n \in \mathbb{N}; f_1, \dots, f_n \in C_c^+(E), A \in \mathcal{B}([0, \infty)^n)\}$$

ist schnittstabil und nach Satz 24.2 ein Erzeuger von \mathbb{M} . Also ist das Maß \mathbf{P}_X eindeutig durch die Werte auf \mathcal{I} festgelegt.

Die Aussage folgt analog für

$$(I_{A_1}, \dots, I_{A_n}) : n \in \mathbb{N}; A_1, \dots, A_n \in \mathcal{B}_b(E).$$

Sind $A_1, \dots, A_n \in \mathcal{B}_b(E)$ beliebig, so existieren $2^n - 1$ paarweise disjunkte Mengen B_1, \dots, B_{2^n-1} mit $A_i = \bigcup_{k: B_k \subset A_i} B_k$ für jedes $i = 1, \dots, n$. Die Verteilung von $(I_{A_1}, \dots, I_{A_n})$ lässt sich aus der von $(I_{B_1}, \dots, I_{B_{2^n-1}})$ berechnen. \square

Im Folgenden sei $i = \sqrt{-1}$.

Definition 24.6. Sei X ein zufälliges Maß auf E . Wir bezeichnen mit

$$\mathcal{L}_X(f) = \mathbf{E} \left[\exp \left(- \int f dX \right) \right], \quad f \in \mathcal{B}^+(E),$$

die **Laplace-Transformierte** von X und mit

$$\varphi_X(f) = \mathbf{E} \left[\exp \left(i \int f dX \right) \right], \quad f \in \mathcal{B}_b^{\mathbb{R}}(E),$$

die **charakteristische Funktion** von X .

Satz 24.7. Die Verteilung \mathbf{P}_X eines zufälligen Maßes X ist eindeutig bestimmt sowohl durch die Werte der Laplace-Transformierten $\mathcal{L}_X(f)$, $f \in C_c^+(E)$, als auch durch die Werte der charakteristischen Funktion $\varphi_X(f)$, $f \in C_c(E)$.

Beweis. Dies folgt aus Satz 24.5 und dem Eindeutigkeitssatz für charakteristische Funktionen (Satz 15.8) beziehungsweise Laplace-Transformierte (Übung 15.1.2) von Zufallsvariablen auf $[0, \infty)^n$. \square

Definition 24.8. Wir sagen, dass ein zufälliges Maß X auf E **unabhängige Zuwächse** hat, falls für je endlich viele paarweise disjunkte messbare Mengen A_1, \dots, A_n die Zufallsvariablen $X(A_1), \dots, X(A_n)$ unabhängig sind.

Korollar 24.9. Die Verteilung eines zufälligen Maßes X auf E mit unabhängigen Zuwächsen ist durch $(\mathbf{P}_{X(A)}, A \in \mathcal{B}_b(E))$ eindeutig bestimmt.

Beweis. Dies folgt direkt aus Satz 24.5. \square

Definition 24.10. Sei $\mu \in \mathcal{M}(E)$. Ein zufälliges Maß X mit unabhängigen Zuwächsen heißt **Poisson'scher Punktprozess (PPP)** mit Intensitätsmaß μ , falls für jedes $A \in \mathcal{B}_b(E)$ gilt, dass $\mathbf{P}_{X(A)} = \text{Poi}_{\mu(A)}$. Wir schreiben dann $\text{PPP}_\mu := \mathbf{P}_X \in \mathcal{M}_1(\mathcal{M}(E))$ und sagen kurz, dass X ein PPP $_\mu$ ist.

Bemerkung 24.11. Die Definition des PPP (und die Konstruktion im folgenden Satz) funktioniert auch, wenn (E, \mathcal{E}, μ) lediglich ein σ -endlicher Maßraum ist. Die Charakterisierung mit Hilfe von Laplace-Transformierten und charakteristischen Funktionen ist allerdings etwas einfacher im hier betrachteten Fall lokalkompakter, polnischer Räume. \diamond

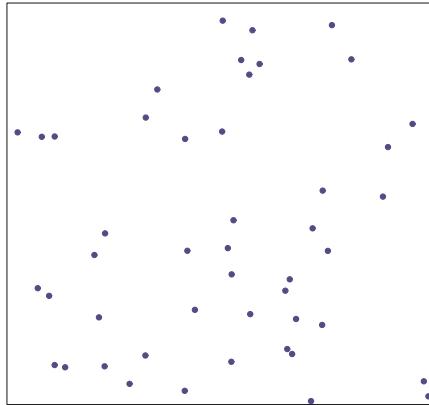


Abb. 24.1. Poisson'scher Punktprozess auf dem Einheitsquadrat mit Intensitätsmaß 50λ .

Satz 24.12. Zu jedem $\mu \in \mathcal{M}(E)$ existiert ein Poisson'scher Punktprozess X mit Intensitätsmaß μ .

Beweis. Da $\mu \in \mathcal{M}(E)$ ist, ist μ σ -endlich. Sei also $E_n \uparrow E$ mit $\mu(E_n) < \infty$ für jedes $n \in \mathbb{N}$. Setze $\mu_1 = \mu(E_1 \cap \cdot)$ und $\mu_n = \mu((E_n \setminus E_{n-1}) \cap \cdot)$ für $n \geq 2$. Sind X_1, X_2, \dots unabhängige Poisson'sche Punktprozesse mit Intensitätsmaßen μ_1, μ_2, \dots , so hat $X = \sum_{n=1}^{\infty} X_n$ das Intensitätsmaß $\mathbf{E}[X] = \mu$, also ist X ein zufälliges Maß (siehe Übung 24.1.1). Außerdem sieht man leicht, dass X unabhängige Zuwächse hat und

$$\mathbf{P}_{X(A)} = \mathbf{P}_{X_1(A)} * \mathbf{P}_{X_2(A)} * \dots = \text{Poi}_{\mu_1(A)} * \text{Poi}_{\mu_2(A)} * \dots = \text{Poi}_{\mu(A)}.$$

Also ist $X \sim \text{PPP}_\mu$.

Es reicht also, den Fall $\mu(E) \in (0, \infty)$ zu betrachten, den wir im Folgenden annehmen wollen. Setze $\nu = \mu(\cdot)/\mu(E) \in \mathcal{M}_1(E)$. Seien N, Y_1, Y_2, \dots unabhängige Zufallsvariablen mit $N \sim \text{Poi}_{\mu(E)}$ und $\mathbf{P}_{Y_i} = \nu$ für jedes $i \in \mathbb{N}$. Wir definieren

$$X(A) = \sum_{n=1}^N \mathbb{1}_A(Y_n) \quad \text{für } A \in \mathcal{B}(E).$$

Die Zufallsvariablen $\mathbb{1}_A(Y_1), \mathbb{1}_A(Y_2), \dots$ sind unabhängig und $\text{Ber}_{\nu(A)}$ -verteilt, also ist $X(A) \sim \text{Poi}_{\mu(A)}$ (siehe Satz 15.14(iii)). Sei $n \in \mathbb{N}$, und seien $A_1, \dots, A_n \in \mathcal{B}(E)$ paarweise disjunkt. Dann ist

$$\psi(t) := \mathbf{E} \left[\exp \left(i \sum_{l=1}^n t_l \mathbb{1}_{A_l}(Y_1) \right) \right] = 1 + \sum_{l=1}^n \nu(A_l) (e^{i t_l} - 1), \quad t \in \mathbb{R}^n,$$

die charakteristische Funktion von $(\mathbb{1}_{A_1}(Y_1), \dots, \mathbb{1}_{A_n}(Y_1))$. Sei ferner φ die charakteristische Funktion von $(X(A_1), \dots, X(A_n))$ und φ_l die von $X(A_l)$ für $l = 1, \dots, n$, also $\varphi_l(t_l) = \exp(\mu(A_l)(e^{i t_l} - 1))$. Nach Satz 15.14(iii) ist

$$\begin{aligned}
\varphi(t) &= \mathbf{E} \left[\exp \left(i \sum_{l=1}^n t_l X(A_l) \right) \right] \\
&= \exp (\mu(E)(\psi(t) - 1)) \\
&= \exp \left(\sum_{l=1}^n \mu(A_l) (e^{i t_l} - 1) \right) = \prod_{l=1}^n \varphi_l(t_l).
\end{aligned}$$

Also sind $X(A_1), \dots, X(A_n)$ unabhängig. Es folgt $X \sim \text{PPP}_\mu$. \square

Übung 24.1.1. Seien X_1, X_2, \dots zufällige Maße und $\lambda_1, \lambda_2, \dots \in [0, \infty)$ sowie $X := \sum_{n=1}^{\infty} \lambda_n X_n$. Man zeige, dass X genau dann ein zufälliges Maß ist, wenn $\mathbf{P}[X(B) < \infty] = 1$ für jedes $B \in \mathcal{B}_b(E)$. Man folgere: Ist X eine Zufallsvariable mit Werten in $(\widetilde{\mathcal{M}}(E), \widetilde{\mathbb{M}}(E))$ und $\mathbf{E}[X] \in \mathcal{M}(E)$, so ist X ein zufälliges Maß. ♣

Übung 24.1.2. Sei τ_w die Topologie der schwachen Konvergenz auf $\mathcal{M}_1(E)$ und $\sigma(\tau_w)$ die Borel'sche σ -Algebra auf $\mathcal{M}_1(E)$. Man zeige: $\mathbb{M}|_{\mathcal{M}_1(E)} = \sigma(\tau_w)$. ♣

24.2 Eigenschaften des Poisson'schen Punktprozesses

Satz 24.13. Sei $\mu \in \mathcal{M}(E)$ atomlos, also $\mu(\{x\}) = 0$ für jedes $x \in E$, und sei X ein zufälliges Maß auf E mit $\mathbf{P}[X(A) \in \mathbb{N}_0 \cup \{\infty\}] = 1$ für jedes $A \in \mathcal{B}(E)$. Dann sind äquivalent:

- (i) $X \sim \text{PPP}_\mu$.
- (ii) X ist fast sicher doppelpunktfrei, also $\mathbf{P}[X(\{x\}) \geq 2 \text{ für ein } x \in E] = 0$, und

$$\mathbf{P}[X(A) = 0] = e^{-\mu(A)} \quad \text{für jedes } A \in \mathcal{B}_b(E). \quad (24.3)$$

Beweis. (i) \Rightarrow (ii) Das ist klar.

(ii) \Rightarrow (i) Sind $A_1, \dots, A_n \in \mathcal{B}_b(E)$ paarweise disjunkt, so ist

$$\begin{aligned}
\mathbf{P}[X(A_1) = 0, \dots, X(A_n) = 0] &= \mathbf{P}[X(A_1 \cup \dots \cup A_n) = 0] \\
&= e^{-\mu(A_1 \cup \dots \cup A_n)} \\
&= \prod_{l=1}^n e^{-\mu(A_l)} = \prod_{l=1}^n \mathbf{P}[X(A_l) = 0].
\end{aligned}$$

Also sind die Zufallsvariablen $\tilde{X}(A) := X(A) \wedge 1$ unabhängig für disjunkte Mengen A . Der Rest des Beweises geht wie im Beweis von Satz 5.34. Sei $A \in \mathcal{B}_b(E)$. Wähle $A_0 \subset A$ mit $\mu(A_0) = \mu(A)/2$ (das geht nach Übung 8.3.1, weil μ atomlos ist) und setze $A_1 = A \setminus A_0$. Wähle nun in gleicher Weise $A_{i,0}, A_{i,1} \subset A_i$ für

$i = 0, 1$ und sukzessive disjunkte Mengen $A_{i,0}, A_{i,1} \subset A_i$ für $i \in \{0, 1\}^{n-1}$ mit $\mu(A_i) = 2^{-n}\mu(A)$ für jedes $i \in \{0, 1\}^n$. Setze

$$N_n(A) := \sum_{i \in \{0,1\}^n} \tilde{X}(A_i).$$

Da X doppelpunktfrei ist, gilt $N_n(A) \uparrow X(A)$ fast sicher. Andererseits ist nach Voraussetzung $N_n(A) \sim b_{2^n, 1 - \exp(-2^{-n}\mu(A))}$ für $n \in \mathbb{N}$, also konvergiert die charakteristische Funktion

$$\begin{aligned} \varphi_{N_n(A)}(t) &= \left(1 + (1 - e^{-2^{-n}\mu(A)})(e^{it} - 1)\right)^{2^n} \\ &\xrightarrow{n \rightarrow \infty} \exp(\mu(A)(e^{it} - 1)) = \varphi_{\text{Poi}_{\mu(A)}}(t). \end{aligned}$$

Mithin gilt $\mathbf{P}_{N_n(A)} \xrightarrow{n \rightarrow \infty} \text{Poi}_{\mu(A)}$, also $X(A) \sim \text{Poi}_{\mu(A)}$.

Sind nun $A_1, \dots, A_k \in \mathcal{B}_b(E)$ paarweise disjunkt, so sind die analog konstruierten $N_n(A_1), \dots, N_n(A_k)$ unabhängig, also sind auch die Limiten $X(A_l) = \lim_{n \rightarrow \infty} N_n(A_l)$, $l = 1, \dots, k$ unabhängig. \square

Satz 24.14. Sei $\mu \in \mathcal{M}(E)$ und X ein Poisson'scher Punktprozess mit Intensitätsmaß μ . Dann hat X die Laplace-Transformierte

$$\mathcal{L}_X(f) = \exp \left(\int \mu(dx) (e^{-f(x)} - 1) \right), \quad f \in \mathcal{B}^+(E),$$

und die charakteristische Funktion

$$\varphi_X(f) = \exp \left(\int \mu(dx) (e^{if(x)} - 1) \right), \quad f \in \mathcal{B}_b^{\mathbb{R}}(E).$$

Beweis. Es reicht, die Aussage für Elementarfunktion $f = \sum_{l=1}^n \alpha_l \mathbb{1}_{A_l}$ mit komplexen Zahlen $\alpha_1, \dots, \alpha_n$ und paarweise disjunkten Mengen $A_1, \dots, A_n \in \mathcal{B}_b(E)$ zu zeigen. (Die Aussagen für allgemeines f folgen dann mit den üblichen Approximationssargumenten.) Für solches f ist aber

$$\begin{aligned} \mathbf{E} \left[\exp(-I_f(X)) \right] &= \mathbf{E} \left[\prod_{l=1}^n e^{-\alpha_l X(A_l)} \right] = \prod_{l=1}^n \mathbf{E} \left[e^{-\alpha_l X(A_l)} \right] \\ &= \prod_{l=1}^n \exp(\mu(A_l)(e^{-\alpha_l} - 1)) \\ &= \exp \left(\sum_{l=1}^n \mu(A_l)(e^{-\alpha_l} - 1) \right) \\ &= \exp \left(\int \mu(dx) (e^{-f(x)} - 1) \right). \end{aligned} \quad \square$$

Korollar 24.15 (Momente des PPP). Sei $\mu \in \mathcal{M}(E)$ und $X \sim \text{PPP}_\mu$.

- (i) Ist $f \in \mathcal{L}^1(\mu)$, so ist $\mathbf{E}[\int f dX] = \int f d\mu$.
- (ii) Ist $f \in \mathcal{L}^2(\mu) \cap \mathcal{L}^1(\mu)$, so ist $\mathbf{Var}[\int f dX] = \int f^2 d\mu$.

Beachte, dass nur für endliches μ die Inklusion $\mathcal{L}^2(\mu) \subset \mathcal{L}^1(\mu)$ gilt.

Beweis. Ist $f = f^+ - f^- \in \mathcal{L}^1(\mu)$, so vertauschen für die charakteristische Funktion Integral und Differentiation $\frac{d}{dt}\varphi_X(tf^+) = i\varphi_X(tf^+) \int f(x) e^{itf^+(x)} \mu(dx)$, also ist (nach Übung 15.4.4(iii))

$$\mathbf{E}[I_{f^+}(X)] = \frac{1}{i} \frac{d}{dt} \varphi_X(tf^+) \Big|_{t=0} = \int f^+ d\mu.$$

Das gleiche Argument für f^- und Aufsummieren liefert (i).

Ist $f \in \mathcal{L}^1(\mu) \cap \mathcal{L}^2(\mu)$, so lässt sich das Argument iterieren (mit Satz 15.34)

$$\frac{d^2}{dt^2}\varphi_X(tf) = -\varphi_X(tf) \left[\int f^2(x) e^{itf(x)} \mu(dx) + \left(\int f(x) e^{itf(x)} \mu(dx) \right)^2 \right],$$

also gilt $\mathbf{E}[I_f(X)^2] = -\frac{d^2}{dt^2}\varphi_X(tf) \Big|_{t=0} = I_{f^2}(\mu) + I_f(\mu)^2$. □

Satz 24.16 (Abbildungssatz). Seien E und F lokalkompakte, polnische Räume und $\phi : E \rightarrow F$ eine messbare Abbildung. Sei $\mu \in \mathcal{M}(E)$ mit $\mu \circ \phi^{-1} \in \mathcal{M}(F)$ und X ein PPP auf E mit Intensitätsmaß μ . Dann ist $X \circ \phi^{-1}$ ein PPP auf F mit Intensitätsmaß $\mu \circ \phi^{-1}$.

Beweis. Für $f \in \mathcal{B}^+(F)$ ist

$$\begin{aligned} \mathcal{L}_{X \circ \phi^{-1}}(f) &= \mathcal{L}_X(f \circ \phi) = \exp \left(\int (e^{-f(\phi(x))} - 1) \mu(dx) \right) \\ &= \exp \left(\int (e^{-f(y)} - 1) (\mu \circ \phi^{-1})(dy) \right). \end{aligned}$$

Die Aussage folgt nun aus Satz 24.14 und Satz 24.7. □

Satz 24.17. Sei $\nu \in \mathcal{M}((0, \infty))$ und $X \sim \text{PPP}_\nu$ auf $(0, \infty)$. Setze $Y := \int x X(dx)$. Dann sind äquivalent

- (i) $\mathbf{P}[Y < \infty] > 0$,
- (ii) $\mathbf{P}[Y < \infty] = 1$,
- (iii) $\int \nu(dx)(1 \wedge x) < \infty$.

Gelten (i)–(iii), so ist Y eine unbegrenzt teilbare, nichtnegative Zufallsvariable mit Lévy-Maß ν .

Beweis. Sei $Y_\infty = \int_{[1,\infty)} x X(dx)$ und $Y_t := \int_{(t,1)} x X(dx)$ für $t \in [0, 1]$. Offenbar ist $Y = Y_0 + Y_\infty$. Außerdem ist offenbar

$$\mathbf{P}[Y_\infty < \infty] > 0 \iff \mathbf{P}[Y_\infty < \infty] = 1 \iff \nu([1, \infty)) < \infty. \quad (24.4)$$

Gilt (iii), so ist $\mathbf{E}[Y_0] = \int_{(0,1)} x \nu(dx) < \infty$, also $Y_0 < \infty$ f.s. (und wegen (24.4) auch $Y < \infty$ f.s.). Gilt andererseits (iii) nicht, so ist $Y_\infty = \infty$ f.s. oder $\mathbf{E}[Y_0] = \infty$. Während für Y_∞ die Erwartung unendlich sein kann, auch wenn Y_∞ f.s. endlich ist, ist dies für Y_0 nicht möglich, denn Y_0 setzt sich im Gegensatz zu Y_∞ nicht aus wenigen großen, sondern aus vielen kleinen Beiträgen zusammen, sodass ein Gesetz der großen Zahl gilt. Konkret ist nach Korollar 24.15

$$\mathbf{Var}[Y_t] = \int_{(t,1)} x^2 \nu(dx) \leq \int_{(t,1)} x \nu(dx) = \mathbf{E}[Y_t] < \infty$$

für jedes $t \in (0, 1)$, also nach der Chebyshev'schen Ungleichung

$$\mathbf{P}\left[Y_t < \frac{\mathbf{E}[Y_t]}{2}\right] \leq \frac{4 \mathbf{Var}[Y_t]}{\mathbf{E}[Y_t]^2} \xrightarrow{t \rightarrow 0} 0.$$

Also ist $Y_0 = \sup_{t \in (0,1)} Y_t \geq \mathbf{E}[Y_0]/2 = \infty$ fast sicher.

Es gelten nun (i) – (iii). Nach Satz 24.14 hat Y die Laplace-Transformierte

$$\mathbf{E}[e^{-tY}] = \exp\left(\int \nu(dx)(e^{-tx} - 1)\right).$$

Nach der Lévy-Khintchin Formel (Satz 16.14) ist Y unbegrenzt teilbar mit Lévy-Maß ν . \square

Korollar 24.18. Seien $\mu_i \in \mathcal{M}_1([0, \infty))$, $i = 1, 2$, unbegrenzt teilbar mit kanonischem Maß $\nu_i \in \mathcal{M}((0, \infty))$ und deterministischem Anteil $\alpha_i \geq 0$ (vergleiche Satz 16.14). Gelten

$$\alpha_1 \leq \alpha_2 \quad \text{und} \quad \nu_1([x, \infty)) \leq \nu_2([x, \infty)) \quad \text{für alle } x > 0, \quad (24.5)$$

so ist μ_1 stochastisch kleiner als μ_2 , d.h. $\mu_1 \leq_{st} \mu_2$.

Beweis. (Der Beweis folgt [107, Beweis von Lemma 6.1].) Die Idee ist, eine Kopp lung zu konstruieren, die auf einem Poisson'schen Punktprozess beruht und zwei Zufallsvariablen Y_1, Y_2 ergibt mit $Y_i \sim \mu_i$, $i = 1, 2$, und $Y_1 \leq Y_2$ fast sicher. Nach Satz 17.58 folgt hieraus die Behauptung.

Sei $G_i(x) := \nu_i([x, \infty))$, $i = 1, 2$, $x > 0$, und

$$\phi_i(y) := G_i^{-1}(y) = \inf \{x \geq 0 : G_i(x) \leq y\} \quad \text{für } y > 0.$$

Ist ν_i endlich, so ist $\phi_i(y) = 0$ für $y \geq \nu_i((0, \infty))$. Bezeichnet λ das Lebesguemaß auf $[0, \infty)$, so gilt offenbar für das auf $(0, \infty)$ eingeschränkte Bildmaß

$$(\lambda \circ \phi_i^{-1})|_{(0, \infty)} = \nu_i, \quad i = 1, 2.$$

Sei nun X ein PPP auf $(0, \infty)$ mit Intensitätsmaß λ . Dann ist

$$X_i := \left(\int \phi_i(x) X(dx) \right)|_{(0, \infty)} = X \circ \phi_i^{-1}$$

nach Satz 24.16 ein PPP mit Intensitätsmaß ν_i , $i = 1, 2$. Nach Satz 24.17 ist daher

$$Y_i := \alpha_i + \int \phi_i(x) X(dx) \sim \mu_i \quad \text{für } i = 1, 2.$$

Nach Voraussetzung ist aber $G_1 \geq G_2$, also $\phi_1 \leq \phi_2$ und damit $Y_1 \leq Y_2$ f.s. \square

Beispiel 24.19. Nach Korollar 16.10 existiert zu jeder nichtnegativen unbegrenzt teilbaren Verteilung μ mit Lévy-Maß ν ein stochastischer Prozess $(Y_t)_{t \geq 0}$ mit unabhängigen stationären Zuwächsen und $Y_t \sim \mu^{*t}$ (also mit Lévy-Maß $t\nu$). Diesen Prozess können wir hier direkt konstruieren: Sei X ein PPP auf $(0, \infty) \times [0, \infty)$ mit Intensitätsmaß $\nu \otimes \lambda$ (wo λ das Lebesgue-Maß ist). Setze $Y_0 = 0$ und

$$Y_t := \int_{(0, \infty) \times (0, t]} x X(d(x, s)).$$

Nach dem Abbildungssatz ist $X(\cdot \times (s, t]) \sim \text{PPP}_{(t-s)\nu}$, also ist $Y_t - Y_s$ unbegrenzt teilbar mit Lévy-Maß $(t-s)\nu$. Die Unabhängigkeit der Zuwächse ist evident. Man beachte, dass $t \mapsto Y_t$ rechtsstetig und monoton wachsend ist.

Der so konstruierte Prozess Y heißt **Subordinator** mit Lévy-Maß ν . \diamond

Wir können das Vorgehen des letzten Beispiels verallgemeinern, indem wir als Zeitmenge allgemeinere Mengen als $[0, \infty)$ zulassen.

Definition 24.20. Ein zufälliges Maß Y heißt unbegrenzt teilbar, wenn für jedes $n \in \mathbb{N}$ u.i.v. zufällige Maße Y_1, \dots, Y_n existieren mit $Y = Y_1 + \dots + Y_n$.

Satz 24.21. Sei $\nu \in \mathcal{M}((0, \infty) \times E)$ mit

$$\int \mathbb{1}_A(t) (1 \wedge x) \nu(d(x, t)) < \infty \quad \text{für jedes } A \in \mathcal{B}_b(E),$$

und sei $\alpha \in \mathcal{M}(E)$. Sei X ein PPP $_\nu$ und

$$Y(A) := \alpha(A) + \int x \mathbb{1}_A(t) X(d(x, t)) \quad \text{für } A \in \mathcal{B}(E).$$

Dann ist Y ein unbegrenzt teilbares zufälliges Maß mit unabhängigen Zuwächsen. Für $A \in \mathcal{B}(E)$ hat $Y(A)$ das Lévy-Maß $\nu(\cdot \times A)$.

Wir nennen ν das kanonische Maß und α den deterministischen Anteil von Y .

Beweis. Das folgt direkt aus Satz 24.16 und Satz 24.17. \square

Bemerkung 24.22. Wir können Y schreiben als $Y = \alpha + \int x \delta_t X(d(x,t))$, wo δ_t die Einheitsmasse in $t \in E$ ist. Wenn wir nun statt $x \delta_t$ allgemeinere Maße $\chi \in \mathcal{M}(E)$ zulassen, so erhalten wir eine Darstellung

$$Y = \alpha + \int_{\mathcal{M}(E)} \chi X(d\chi),$$

wo $X \sim \text{PPP}_\nu$ auf $\mathcal{M}(E)$ und $\nu \in \mathcal{M}(\mathcal{M}(E))$ mit $\int \nu(d\chi)(\chi(A) \wedge 1) < \infty$ für jedes $A \in \mathcal{B}_b(E)$. Man kann zeigen, dass dies die allgemeinste Form eines unbegrenzt teilbaren Maßes auf E ist. Wir nennen ν das kanonische Maß von Y und α den deterministischen Anteil. Y ist charakterisiert durch die Laplace-Transformierte, die der Lévy-Khinchin Formel genügt:

$$\mathcal{L}_Y(f) = \exp \left(- \int f d\alpha + \int \nu(d\chi) (e^{-\int f d\chi} - 1) \right). \quad \diamond$$

Satz 24.23 (Färbungssatz). Sei F ein weiterer lokalkompakter, polnischer Raum und $\mu \in \mathcal{M}(E)$ atomlos sowie $(Y_x)_{x \in E}$ von X unabhängige u.i.v. Zufallsvariablen mit Werten in F und Verteilung $\nu \in \mathcal{M}_1(F)$. Dann ist

$$Z(A) := \int \mathbb{1}_A(x, Y_x) X(dx), \quad A \in \mathcal{B}(E \times F),$$

ein $\text{PPP}_{\mu \otimes \nu}$ auf $E \times F$.

Beweis. Übung! \square

Wir wollen die Aussage des Färbungssatzes in nahe liegender Weise verallgemeinern: Die Annahme, dass das Maß μ atomlos ist, sorgt schließlich nur dafür, dass X keine Doppelpunkte hat, also für jede Einheitsmasse, die X produziert, eine andere Zufallsvariable Y_x zur Verfügung steht. Außerdem wollen wir für jeden Punkt x eine eigene Verteilung von Y_x erlauben.

Seien also E, F lokalkompakte, polnische Räume, $\mu \in \mathcal{M}(E)$ und κ ein stochastischer Kern von E nach F mit $\mu\kappa := \int \mu(dx)\kappa(x, \cdot) \in \mathcal{M}(F)$. Seien $(Y_{x,t})_{x \in E, t \in [0,1]}$ unabhängige Zufallsvariablen mit Verteilungen $\mathbf{P}_{Y_{x,t}} = \kappa(x, \cdot)$ für $x \in E$ und $t \in [0, 1]$.

Wir definieren zu $X \sim \text{PPP}_\mu$ das Lifting \tilde{X} als den PPP auf $E \times [0, 1]$ mit Intensitätsmaß $\mu \otimes \lambda|_{[0,1]}$, wo λ das Lebesgue-Maß ist. Offenbar ist dann $X \stackrel{\mathcal{D}}{=} \tilde{X}(\cdot \times [0, 1])$. Das zufällige Maß \tilde{X} können wir also als Realisierung von X auffassen, wobei wir den einzelnen Punkten von X willkürlich eine Markierung mit Werten aus $[0, 1]$ gegeben haben, um sie zu unterscheiden. Wir setzen nun

$$X^\kappa(A) := \int \tilde{X}(d(x,t)) \mathbb{1}_A(Y_{x,t}) \quad \text{für } A \in \mathcal{B}(F).$$

Satz 24.24. X^κ ist ein zufälliges Maß mit $\mathbf{P}_{X^\kappa} = \text{PPP}_{\mu\kappa}$.

Beweis. Offenbar ist X^κ fast sicher ein Maß. Für $A \in \mathcal{B}_b(F)$ ist

$$\mathbf{E}[X^\kappa(A)] = \mathbf{E}\left[\int \tilde{X}(d(x,t)) \kappa(x, A)\right] = (\mu\kappa)(A) < \infty$$

nach Voraussetzung, also ist $X^\kappa(A) < \infty$ fast sicher, und damit ist X^κ ein zufälliges Maß. Wir berechnen die Laplace-Transformierte von X^κ . Sei $g(x) := -\log \mathbf{E}[e^{-f(Y_{x,t})}]$. Dann ist (weil \tilde{X} doppelpunktfrei ist)

$$\begin{aligned} \mathcal{L}_{X^\kappa}(f) &= \mathbf{E}\left[\exp\left(-\int \tilde{X}(d(x,t)) f(Y_{x,t})\right)\right] \\ &= \mathbf{E}\left[\prod_{(x,t): \tilde{X}(\{(x,t)\})=1} e^{-f(Y_{x,t})}\right] = \mathbf{E}\left[\prod_{(x,t): \tilde{X}(\{(x,t)\})=1} \mathbf{E}[e^{-f(Y_{x,t})}]\right] \\ &= \mathbf{E}\left[\prod_{(x,t): \tilde{X}(\{(x,t)\})=1} e^{-g(x)}\right] = \mathcal{L}_X(g) \\ &= \exp\left(\int \mu(dx)(\mathbf{E}[e^{-f(Y_{x,t})}] - 1)\right) \\ &= \exp\left(\int \mu(dx) \int \kappa(x, dy)(e^{-f(y)} - 1)\right) \\ &= \exp\left(\int \mu\kappa(dy)(e^{-f(y)} - 1)\right). \end{aligned} \quad \square$$

Beispiel 24.25 (PPP als invariante Verteilung). Als Anwendung des letzten Satzes betrachten wir einen stochastischen Prozess auf $E = \mathbb{Z}^d$ oder $E = \mathbb{R}^d$, der aus unabhängigen Irrfahrten besteht. Wir nehmen also an, dass wir u.i.v. Zufallsvariablen Z_n^i , $i, n \in \mathbb{N}$ mit Verteilung $\nu \in \mathcal{M}_1(E)$ haben. Wir nehmen zudem an, dass das i -te Teilchen unseres Irrfahrtenprozesses zur Zeit n die Position $S_n^i := S_0^i + \sum_{l=1}^n Z_l^i$ hat, wobei S_0^i ein willkürlicher, eventuell zufälliger, Startpunkt ist. Wenn wir die Teilchen als ununterscheidbar annehmen, reicht es, die Teilchen an jedem Ort zusammenzuzählen. Wir betrachten also

$$X_n(A) := \sum_{i=1}^{\infty} \mathbb{1}_A(S_n^i) \quad \text{für } A \subset E.$$

Jedes X_n ist ein Maß auf E und, wenn wir die Teilchen anfangs nicht zu sehr konzentrieren, lokal endlich, also ein zufälliges Maß. Nehmen wir an, dass $X_0 \sim \text{PPP}_\mu$ für ein $\mu \in \mathcal{M}(E)$ ist. Wir setzen $\kappa(x, \cdot) = \delta_x * \nu$ und schreiben κ^n für die n -fache

Anwendung von κ , also $\kappa^n(x, \cdot) = \delta_x * \nu^{*n}$. Wir erhalten so $X_0^\kappa \stackrel{\mathcal{D}}{=} X_1$. In der Tat: Das unabhängige Bewegen der einzelnen Teilchen in der Definition von X_0^κ ist genau die Dynamik der unabhängigen Irrfahrten. Da nun auch X_1 ein PPP ist, erhalten wir iterativ $X_n^\kappa \stackrel{\mathcal{D}}{=} X_{n+1}$ und damit $X_n \sim \text{PPP}_{\mu \kappa^n} = \text{PPP}_{\mu * \nu^{*n}}$. Speziell ist $X_0 \stackrel{\mathcal{D}}{=} X_n$ genau dann, wenn $\mu * \nu = \mu$ gilt. Offenbar ist dies richtig, wenn $E = \mathbb{Z}^d$ und μ das Zählmaß ist oder $E = \mathbb{R}^d$ und μ das Lebesgue-Maß. Ist beispielsweise $E = \mathbb{Z}^d$, so kann man unter relativ schwachen Annahmen an ν zeigen, dass das Zählmaß $\mu = \lambda$ die einzige Lösung von $\mu * \nu = \mu$ ist. In dem Fall ist jedes invariante Maß eine Konvexitätskombination von PPPs mit verschiedenen Intensitätsmaßen $\theta \lambda$. \diamond

Übung 24.2.1. Man zeige die Aussage von Korollar 24.15 ohne charakteristische Funktionen direkt über die Approximation mit Elementarfunktionen. ♣

Übung 24.2.2. Man zeige den Färbungssatz (Satz 24.23). ♣

Übung 24.2.3. Seien $p_1, p_2 \in (0, 1]$ und $r_1, r_2 > 0$. Man zeige folgende Aussage zur stochastischen Ordnung der negativen Binomialverteilungen: $b_{r_1, p_1}^- \leq_{\text{st}} b_{r_2, p_2}^-$ genau dann, wenn

$$p_1 \geq p_2 \quad \text{und} \quad p_1^{r_1} \geq p_2^{r_2}.$$



24.3 Die Poisson-Dirichlet-Verteilung*

Ziel dieses Abschnitts ist die Lösung des folgenden Problems: Wir brechen einen Stock der Länge 1 an einer zufälligen (uniform verteilten) Stelle in zwei Stücke und legen das linke Stück (mit der Länge W_1) beiseite. Mit dem restlichen Stock verfahren wir in gleicher Weise und legen das linke Stück der Länge W_2 beiseite. Sukzessive sammeln wir die Bruchstücke mit Längen W_1, W_2, W_3, \dots Wie sieht die gemeinsame Verteilung von (W_1, W_2, \dots) aus? Ferner wollen wir die Zahlen W_1, W_2, \dots der Größe nach umsortieren und $W_{(1)} \geq W_{(2)} \geq \dots$ nennen. Wie sieht die Verteilung von $(W_{(1)}, W_{(2)}, \dots)$ aus? Und schließlich: was hat dies mit Poisson'schen Punktprozessen zu tun?

Zur Beantwortung der Fragen müssen wir etwas weiter ausholen. Wir hatten gesehen, wie die Beta-Verteilung in natürlicher Weise bei dem Pólya'schen Urnenmodell als Grenzverteilung der Frequenzen der beiden Kugelfarben auftritt. Offenbar kann man das Pólya'sche Modell auch mit $n \geq 2$ Farben betrachten. Die Grenzverteilung ist dann die n -dimensionale Verallgemeinerung der Beta-Verteilung, nämlich die so genannte Dirichlet-Verteilung.

Definition 24.26. Sei $n \in \{2, 3, \dots\}$ und $\theta_1, \dots, \theta_n > 0$. Die **Dirichlet-Verteilung** $\text{Dir}_{\theta_1, \dots, \theta_n}$ ist die Verteilung auf dem $(n - 1)$ -dimensionalen Simplex

$$\Delta_n := \{(x_1, \dots, x_n) \in [0, 1]^n : x_1 + \dots + x_n = 1\},$$

die für messbares $A \subset \Delta_n$ definiert ist durch

$$\text{Dir}_{\theta_1, \dots, \theta_n}(A) = \int \mathbb{1}_A(x_1, \dots, x_n) f_{\theta_1, \dots, \theta_n}(x_1, \dots, x_n) dx_1 \cdots dx_{n-1},$$

wobei

$$f_{\theta_1, \dots, \theta_n}(x_1, \dots, x_n) = \frac{\Gamma(\theta_1 + \dots + \theta_n)}{\Gamma(\theta_1) \cdots \Gamma(\theta_n)} x_1^{\theta_1-1} \cdots x_n^{\theta_n-1}.$$

Die Parameter $\theta_1, \dots, \theta_n$ entsprechen (falls ganzzahlig) den Anzahlen der Kugeln der einzelnen Farben, die ursprünglich in der Urne liegen. Wenn wir nun nicht ganz so genau hinschauen und Kugeln zweier Farben, etwa $n-1$ und n zusammenfassen, so sollten wir als Grenzverteilung für die Frequenzen $\text{Dir}_{\theta_1, \dots, \theta_{n-2}, \theta_{n-1}+\theta_n}$ erhalten. Sei $(M_t)_{t \geq 0}$ der **Moran-Gamma-Subordinator**, also ein stochastischer Prozess mit rechtsstetigen, monoton wachsenden Pfaden $t \mapsto M_t$ und unabhängigen, stationären, Gamma-verteilten Zuwächsen: $M_t - M_s \sim \Gamma_{1,t-s}$ für $t > s \geq 0$. Einen wichtigen Zusammenhang zwischen der Dirichlet-Verteilung und M liefern die Korollare des folgenden Satzes und speziell Satz 24.32.

Satz 24.27. Seien $n \in \mathbb{N}$ und $\theta_1, \dots, \theta_n > 0$ sowie $\Theta := \theta_1 + \dots + \theta_n$. Seien $X \sim \text{Dir}_{\theta_1, \dots, \theta_n}$ und $Z \sim \Gamma_{1,\Theta}$ unabhängige Zufallsvariablen. Dann sind die Zufallsvariablen $S_i := Z \cdot X_i$, $i = 1, \dots, n$ unabhängig und $S_i \sim \Gamma_{1,\theta_i}$.

Beweis. Sei im Folgenden stets $x_n := 1 - \sum_{i=1}^{n-1} x_i$ und $s = \sum_{j=1}^n s_j$. Sei

$$\Delta'_n := \left\{ (x_1, \dots, x_{n-1}) \in (0, 1)^{n-1} : \sum_{i=1}^{n-1} x_i < 1 \right\}.$$

Die Verteilung von (X_1, \dots, X_{n-1}, Z) hat (für $x \in \Delta'_n$ und $z \geq 0$) die Dichte

$$f(x_1, \dots, x_{n-1}, z) = \prod_{j=1}^n \left(x_j^{\theta_j-1} / \Gamma(\theta_j) \right) z^{\Theta-1} e^{-z}.$$

Betrachte die Abbildung $F : \Delta'_n \times (0, \infty) \rightarrow (0, \infty)^n$, $(x_1, \dots, x_{n-1}, z) \mapsto (zx_1, \dots, zx_{n-1})$. Die Abbildung ist invertierbar mit Umkehrabbildung

$$F^{-1} : (s_1, \dots, s_n) \mapsto (s_1/s, \dots, s_{n-1}/s, s).$$

Die Ableitung von F hat die Determinante $\det(F'(x_1, \dots, x_{n-1}, z)) = z^{n-1}$. Nach der Transformationsformel für Dichten (Satz 1.101) hat (S_1, \dots, S_n) die Dichte

$$\begin{aligned} g(s_1, \dots, s_n) &= \frac{f(F^{-1}(s_1, \dots, s_n))}{|\det(F'(F^{-1}(s_1, \dots, s_n)))|} \\ &= \frac{s^{\Theta-1} e^{-s}}{s^{n-1}} \prod_{j=1}^n ((s_j/s)^{\theta_j-1} / \Gamma(\theta_j)) \\ &= \prod_{j=1}^n \left(s_j^{\theta_j-1} e^{-s_j} / \Gamma(\theta_j) \right). \end{aligned}$$

Dies ist aber die Dichte von unabhängigen Gamma-Verteilungen. \square

Korollar 24.28. Ist $t_i := \sum_{j=1}^i \theta_j$ für $i = 0, \dots, n$, so sind die Zufallsvariablen $X = ((M_{t_i} - M_{t_{i-1}})/M_{t_n}, i = 1, \dots, n)$ und $S := M_{t_n}$ unabhängig und $X \sim \text{Dir}_{\theta_1, \dots, \theta_n}$ sowie $S \sim \Gamma_{1, t_n}$.

Korollar 24.29. Sei $(X_1, \dots, X_n) \sim \text{Dir}_{\theta_1, \dots, \theta_n}$. Dann sind $X_1 \sim \beta_{\theta_1, \sum_{i=2}^n \theta_i}$ und $(X_2/(1-X_1), \dots, X_n/(1-X_1)) \sim \text{Dir}_{\theta_2, \dots, \theta_n}$ unabhängig.

Beweis. Sei M wie in Korollar 24.28. Dann ist $X_1 = M_{t_1}/M_{t_n} \sim \beta_{\theta_1, t_n - \theta_1}$ und $X_1 = \left(\frac{M_{t_n} - M_{t_1}}{M_{t_1}} + 1 \right)^{-1}$ nur von M_{t_1} und $M_{t_n} - M_{t_1}$ abhängig. Andererseits ist

$$\left(\frac{X_2}{1-X_1}, \dots, \frac{X_n}{1-X_1} \right) = \left(\frac{M_{t_2} - M_{t_1}}{M_{t_n} - M_{t_1}}, \dots, \frac{M_{t_n} - M_{t_{n-1}}}{M_{t_n} - M_{t_1}} \right)$$

unabhängig von M_{t_1} und nach Korollar 24.28 auch unabhängig von $M_{t_n} - M_{t_1}$ sowie $\text{Dir}_{\theta_2, \dots, \theta_n}$ -verteilt. \square

Korollar 24.30. Seien V_1, \dots, V_{n-1} unabhängig und $V_i \sim \beta_{\theta_i, \theta_{i+1} + \dots + \theta_n}$ sowie $V_n = 1$. Dann ist

$$\left(V_1, (1-V_1)V_2, (1-V_1)(1-V_2)V_3, \dots, \left(\prod_{i=1}^{n-1} (1-V_i) \right) V_n \right) \sim \text{Dir}_{\theta_1, \dots, \theta_n}.$$

Beweis. Das folgt durch Iteration der Aussage von Korollar 24.29. \square

Eine natürliche Fragestellung ist, was passiert, wenn wir immer mehr Farben differenzieren (statt zusammenzufassen). Wir wollen der Einfachheit halber eine symmetrische Situation annehmen, bei der $\theta_1 = \dots = \theta_n = \theta/n$ für ein $\theta > 0$ ist. Wir betrachten also

$$\text{Dir}_{\theta; n} := \text{Dir}_{\theta, \dots, \theta} \quad \text{für } \theta > 0.$$

Ist $X^n = (X_1^n, \dots, X_n^n) \sim \text{Dir}_{\theta/n; n}$, so ist aus Symmetriegründen $\mathbf{E}[X_i^n] = 1/n$ für jedes $n \in \mathbb{N}$ und $i = 1, \dots, n$. Offenbar gilt also $(X_1^n, \dots, X_k^n) \xrightarrow{n \rightarrow \infty} 0$ für jedes $k \in \mathbb{N}$. Eine Möglichkeit, einen nicht-trivialen Grenzwert zu erhalten ist, die Werte der Größe nach zu ordnen $X_{(1)}^n \geq X_{(2)}^n \geq \dots$

Definition 24.31. Sei $\theta > 0$ und $(M_t)_{t \in [0, \theta]}$ ein Moran-Gamma-Subordinator. Seien $m_1 \geq m_2 \geq \dots \geq 0$ die der Größe nach sortierten Sprunghöhen von M und $\tilde{m}_i = m_i/M_\theta$, $i = 1, 2, \dots$. Die Verteilung der Zufallsvariablen $(\tilde{m}_1, \tilde{m}_2, \dots)$ auf $S := \{(x_1 \geq x_2 \geq \dots \geq 0) : x_1 + x_2 + \dots = 1\}$ heißt **Poisson-Dirichlet-Verteilung PD θ** mit Parameter $\theta > 0$.

Genau genommen müssen wir noch nachweisen, dass $\sum_{i=1}^{\infty} \tilde{m}_i = 1$ ist. Sei hierzu Y ein PPP auf $(0, \infty) \times (0, \theta]$ mit Intensitätsmaß $\nu \otimes \lambda$, wo λ das Lebesgue-Maß ist und $\nu(dx) = e^{-x}x^{-1} dx$ das Lévy-Maß der $\Gamma_{1,1}$ -Verteilung. Wir können M definieren durch $M_t := \sum_{(x,s): Y(\{x\})=1, s \leq t} x$. Nun ist $m_1 = \sup\{x \in (0, \infty) : Y(\{x\} \times (0, \theta]) = 1\}$ und sukzessive $m_n = \sup\{x < m_{n-1} : Y(\{x\} \times (0, \theta]) = 1\}$ für $n \geq 2$. Durch Vertauschung der Summationsreihenfolge erhalten wir also $M_\theta = \sum_{n=1}^{\infty} m_n$.

Satz 24.32. Ist $X^n \sim \text{Dir}_{\theta/n;n}$ für $n \in \mathbb{N}$, so gilt $\mathbf{P}_{(X_{(1)}^n, X_{(2)}^n, \dots)} \xrightarrow{n \rightarrow \infty} \text{PD}_\theta$.

Beweis. Die Idee ist, die Zufallsvariablen X^n , $n \in \mathbb{N}$, so durch die Zuwächse des Moran-Gamma-Subordinators $(M_t)_{t \in [0, \theta]}$ darzustellen, dass aus der Verteilungskonvergenz eine fast sichere Konvergenz wird. Es sei also $X_i^n = (M_{\theta i/n} - M_{\theta(i-1)/n})/M_\theta$. Nach Korollar 24.28 ist $X^n \sim \text{Dir}_{\theta/n;n}$. Seien $t_1, t_2, \dots \in (0, \theta]$ die Positionen der Sprünge $m_1 \geq m_2 \geq \dots$. Offenbar ist $X_{(1)}^n \geq \tilde{m}_1$ für jedes n . Ist n so groß, dass $|t_1 - t_2| > \theta/n$ ist, so ist $X_{(2)}^n \geq \tilde{m}_2$. Sukzessive erhalten wir $\liminf_{n \rightarrow \infty} X_{(i)}^n \geq \tilde{m}_i$ fast sicher. Nun ist aber (mit der Festsetzung $X_{(i)}^n = 0$ für $i > n$) $\sum_{i=1}^{\infty} X_{(i)}^n = 1$ für jedes $n \in \mathbb{N}$. Nach dem Lemma von Fatou ist daher

$$1 = \sum_{i=1}^{\infty} \tilde{m}_i \leq \sum_{i=1}^{\infty} \liminf_{n \rightarrow \infty} X_{(i)}^n \leq \liminf_{n \rightarrow \infty} \sum_{i=1}^{\infty} X_{(i)}^n = 1.$$

Es folgt $\lim_{n \rightarrow \infty} X_{(i)}^n = \tilde{m}_i$ fast sicher. \square

Anstatt die Werte von X^n strikt der Größe nach zu ordnen, können wir ein anderes Verfahren anwenden, das Konvergenz der Verteilungen sichert. Stellen wir uns vor, dass wir in einer Population ein genetisches Merkmal haben, das wir unterschiedlich fein messen können. Wenn wir n unterschiedliche Werte unterscheiden wollen, so soll X_i^n den Anteil der Bevölkerung mit dem Merkmal i bezeichnen.

Wir greifen nun sukzessive zufällig Individuen aus der Population heraus. Das erste Individuum habe den Typ I_1^n . Mit I_2^n bezeichnen wir den Typ des ersten Individuums, das nicht vom Typ I_1^n ist. Sukzessive sei I_k^n der Typ des ersten Individuums, das nicht von einem der Typen I_1^n, \dots, I_{k-1}^n ist. Wir betrachten nun den Vektor $\hat{X}^n = (\hat{X}_1^n, \dots, \hat{X}_n^n)$, wo $\hat{X}_k^n = X_{I_k^n}^n$. Da die Wahrscheinlichkeit für $I_1^n = i$ proportional zur Größe der Sub-Population mit Merkmal i ist, nennen wir \hat{X}^n den sukzessiven Größenverzerrten gezogenen Vektor.

Die Verteilung von \hat{X}^n ändert sich nicht, wenn wir die Reihenfolge der X_1^n, \dots, X_n^n verändern. Speziell können wir statt X^n die Ordnungsstatistik $(X_{(1)}^n, \dots, X_{(n)}^n)$ wählen und erhalten ebenfalls \hat{X}^n als sukzessive Größenverzerrten gezogenen Vektor. Insbesondere können wir für $X \sim \text{PD}_\theta$ den sukzessiv Größenverzerrten Vektor \hat{X} definieren. Gilt $X^n \sim \text{Dir}_{\theta/n;n}$, so folgt aus Satz 24.32 sofort, dass $\hat{X}^n \xrightarrow{n \rightarrow \infty} \hat{X}$. Hiermit können wir die Verteilung von \hat{X} ausrechnen.

Satz 24.33. Sei $\theta > 0$ und seien $X^n \sim \text{Dir}_{\theta/n;n}$, $n \in \mathbb{N}$, sowie $X \sim \text{PD}_\theta$. Seien ferner V_1, V_2, \dots u.i.v. Zufallsvariablen auf $[0, 1]$ mit Dichte $x \mapsto \theta(1-x)^{\theta-1}$. Wir setzen $Z_1 = V_1$ und $Z_k = (\prod_{i=1}^{k-1} (1-V_i))V_k$ für $k \geq 2$. Dann gilt:

$$(i) \hat{X}^n \xrightarrow{n \rightarrow \infty} \hat{X}.$$

$$(ii) \hat{X} \stackrel{\mathcal{D}}{=} Z.$$

Die Verteilung von Z heißt GEM $_\theta$ -Verteilung (für Griffiths-Engen-McCloskey).

Beweis. Die Aussage (i) wurde bereits in der Diskussion vor dem Satz gezeigt. Um (ii) zu zeigen, berechnen wir die Verteilung von \hat{X}^n und zeigen, dass sie gegen die von Z konvergiert.

Sei $\hat{X}^{n,1}$ der Vektor $X^{n,1} = (X_{I_1^n}^n, X_2, \dots, X_{I_1^n-1}^n, X_{I_1^n+1}^n, \dots, X_n^n)$, bei dem nur die erste Koordinate gröszenverzerrt gezogen wurde. Wir zeigen:

$$\hat{X}^{n,1} \sim \text{Dir}_{(\theta/n)+1, \theta/n, \dots, \theta/n}. \quad (24.6)$$

Sei $f(x) = (\Gamma(\theta)/\Gamma(\theta/n)^n) \cdot \prod_{k=1}^n x_k^{(\theta/n)-1}$ die Dichte von $\text{Dir}_{\theta/n;n}$. Die Dichte $f^{n,1}$ von $X^{n,1}$ berechnen wir durch Zerlegung nach dem Wert i von I_1^n :

$$\begin{aligned} f^{n,1}(x) &= \sum_{i=1}^n x_1 f(x_2, \dots, x_i, x_1, x_{i+1}, \dots, x_n) = n x_1 f(x) \\ &= \frac{n \Gamma(\theta)}{\Gamma(\theta/n)^n} x_1^{\theta/n} \prod_{i=2}^n x_i^{(\theta/n)-1} \\ &= \frac{\Gamma(\theta+1)}{\Gamma((\theta/n)+1) \Gamma(\theta/n)^{n-1}} x_1^{\theta/n} \prod_{i=2}^n x_i^{(\theta/n)-1}. \end{aligned}$$

Dies ist aber die Dichte von $\text{Dir}_{(\theta/n)+1, \theta/n, \dots, \theta/n}$. Nach Korollar 24.29 ist

$$\hat{X}^{n,1} \stackrel{\mathcal{D}}{=} (V_1^n, (1-V_1^n)Y_1, \dots, (1-V_1^n)Y_{n-1}),$$

wobei

$$V_1^n \sim \beta_{(\theta/n)+1, \theta(n-1)/n} \quad \text{und} \quad Y = (Y_1, \dots, Y_{n-1}) \sim \text{Dir}_{\theta/n; n-1}$$

unabhängig sind. Indem wir das Gezeigte nun auf Y anwenden, erhalten wir sukzessive

$$\hat{X}^n \stackrel{\mathcal{D}}{=} Z^n, \quad (24.7)$$

wobei

$$Z_1^n = V_1^n \quad \text{und} \quad Z_k^n = \left(\prod_{i=1}^{k-1} (1-V_i^n) \right) V_k^n \quad \text{für } k \geq 2,$$

und wobei V_1^n, \dots, V_{n-1}^n unabhängig sind und $V_i^n \sim \beta_{(\theta/n)+1, \theta(n-i)/n}$. Nun prüft man aber leicht nach, dass $\beta_{(\theta/n)+1, \theta(n-i)/n} \xrightarrow{n \rightarrow \infty} \beta_{1,\theta}$ für jedes $i \in \mathbb{N}$, und $\beta_{1,\theta}$ hat die Dichte $x \mapsto \theta(1-x)^{\theta-1}$. Es gilt also $V_i^n \xrightarrow{n \rightarrow \infty} V_i$ für jedes i und damit $Z^n \xrightarrow{n \rightarrow \infty} Z$ und $\hat{X}^n \xrightarrow{n \rightarrow \infty} Z$. Zusammen mit (i) folgt hieraus die Aussage (ii). \square

Unsere eingangs gestellte Frage nach den Größen W_1, W_2, \dots der Bruchstücke von sukzessiv uniform verteilt zerbrochenen Stöcken ist damit geklärt: Der Vektor $(W_{(1)}, W_{(2)}, \dots)$ ist PD₁-verteilt, und (W_1, W_2, \dots) ist GEM₁-verteilt.

Der China-Restaurant Prozess

Wir wollen eine weitere Situation kennen lernen, in der die Poisson-Dirichlet-Verteilung in natürlicher Weise auftaucht. Da die technischen Details etwas knifflig werden, begnügen wir uns damit, die Situation zu beschreiben und zwei wichtige Sätze anzugeben. Eine exzellente und vollständige Beschreibung findet sich in [139].

Wir betrachten ein China-Restaurant mit abzählbar vielen (natürlich runden) nummerierten Tischen, an denen jeweils beliebig viele Gäste Platz finden. Anfangs sei das Restaurant leer. Nacheinander treffen (abzählbar viele) Gäste ein. Der erste Gast setzt sich an den (natürlich freien) Tisch mit der Nummer Eins. Sitzen bereits n Gäste an k Tischen, so hat der $(n+1)$ -te Gast die Möglichkeit, sich entweder an einen der k besetzten Tische zu setzen, oder sich an den freien Tisch mit der kleinsten Nummer zu setzen. Wir wollen annehmen, dass die Wahl zufällig erfolgt und dass sich der Gast an den l -ten besetzten Tisch (mit N_l^n Gästen) mit Wahrscheinlichkeit $(N_l^n - \alpha)/(n + \theta)$ setzt, mit Wahrscheinlichkeit $(\theta + k\alpha)/(n + \theta)$ jedoch den ersten noch freien Tisch besetzt. Hierbei sind $\alpha \in [0, 1]$ und $\theta > -\alpha$. Bezeichnet N_l^n die Anzahl der Gäste zur Zeit n am l -ten besetzten Tisch, so nennen wir $(N^n)_{n \in \mathbb{N}} = (N_1^n, N_2^n, \dots)_{n \in \mathbb{N}}$ den **China-Restaurant Prozess** mit Parametern (α, θ) .

Ist speziell $\alpha = 0$, so können wir den China-Restaurant Prozess auch so interpretieren: Die Wahrscheinlichkeit, sich links neben einen der Gäste zu setzen (also an dessen Tisch) beträgt $1/(n + \theta)$, die Wahrscheinlichkeit, einen neuen Tisch zu besetzen dagegen $\theta/(n + \theta)$. Um das asymptotische Verhalten von $N^n/n = (N_1^n/n, N_2^n/n, \dots)$ zu beschreiben, müssen wir die Poisson-Dirichlet-Verteilung und die GEM Verteilung um einen Parameter erweitern.

Definition 24.34. Sei $\alpha \in [0, 1)$ und $\theta > -\alpha$. Seien V_1, V_2, \dots unabhängig und $V_i \sim \beta_{1-\alpha, \theta+i\alpha}$. Wir definieren $Z = (Z_1, Z_2, \dots)$ durch $Z_1 = V_1$ und $Z_k = (\prod_{i=1}^{k-1} (1 - V_i))V_k$ für $k \geq 2$. Dann heißt $\text{GEM}_{\alpha, \theta} := \mathbf{P}_Z$ die **GEM-Verteilung** mit Parametern (α, θ) . Die Verteilung des nach Größe sortierten Vektors $(Z_{(1)}, Z_{(2)}, \dots)$ heißt **Poisson-Dirichlet-Verteilung** mit Parametern (α, θ) , oder kurz $\text{PD}_{\alpha, \theta}$.

Explizite Formeln für die Dichte der endlichdimensionalen Verteilungen von $\text{PD}_{\alpha,\theta}$ finden sich etwa in [141]. Man bemerke, dass wir im Falle $\alpha = 0$ die bisherigen einparametrischen Verteilungen $\text{GEM}_\theta = \text{GEM}_{0,\theta}$ und $\text{PD}_\theta = \text{PD}_{0,\theta}$ zurückgewinnen.

Satz 24.35. Seien $\alpha \in [0, 1)$, $\theta > -\alpha$ und $(N^n)_{n \in \mathbb{N}}$ der China-Restaurant Prozess mit Parametern (α, θ) . Dann gilt $\mathbf{P}_{N^n/n} \xrightarrow{n \rightarrow \infty} \text{PD}_{\alpha,\theta}$.

Beweis. Siehe [138] oder [139, Theorem 25]. □

Ähnlich wie für die einparametrische Poisson-Dirichlet-Verteilung gibt es eine Darstellung von $\text{PD}_{\alpha,\theta}$ durch die nach Größe geordneten Sprünge eines geeigneten Subordinators. Sei im Folgenden $\alpha \in (0, 1)$ und $(M_t)_{t \in [0,1]}$ ein α -stabiler Subordinator, also ein Subordinator mit Lévy-Maß $\nu(dx) = x^{-\alpha-1} dx$. Seien ferner $m_1 \geq m_2 \geq \dots \geq 0$ die Sprünge von M und $\tilde{m}_i = m_i/M_1$ für $i \in \mathbb{N}$, und $\tilde{m} = (\tilde{m}_1, \tilde{m}_2, \dots)$. Wir zitieren den folgenden Satz aus [139, Section 4.2].

Satz 24.36. Sei $\alpha \in (0, 1)$.

- (i) Es gilt $\tilde{m} \sim \text{PD}_{\alpha,0}$.
- (ii) Ist $\theta > -\alpha$, so ist $\text{PD}_{\alpha,\theta} \ll \text{PD}_{\alpha,0} = \mathbf{P}[\tilde{m} \in \cdot]$ mit

$$\text{PD}_{\alpha,\theta}(dx) = \frac{M_1^{-\theta}}{\mathbf{E}[M_1^{-\theta}]} \mathbf{P}[\tilde{m} \in dx].$$

Übung 24.3.1. Sei $(X, 1 - X) \sim \text{Dir}_{\theta_1, \theta_2}$. Man zeige, dass dann $X \sim \beta_{\theta_1, \theta_2}$ Beta-verteilt ist. ♣

Übung 24.3.2. Sei $X = (X_1, \dots, X_n) \sim \text{Dir}_{\theta_1, \dots, \theta_n}$. Man zeige:

- (i) Für jede Permutation σ auf $\{1, \dots, n\}$ ist

$$(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \sim \text{Dir}_{\theta_{\sigma(1)}, \dots, \theta_{\sigma(n)}}.$$

- (ii) Es gilt $(X_1, \dots, X_{n-2}, X_{n-1} + X_n) \sim \text{Dir}_{\theta_1, \dots, \theta_{n-2}, \theta_{n-1} + \theta_n}$. ♣

Übung 24.3.3. Sei $(N^n)_{n \in \mathbb{N}}$ der China-Restaurant Prozess mit Parametern $(0, \theta)$.

- (i) Man zeige für $\theta = 1$:
 - (a) $\mathbf{P}[N_1^n = k] = 1/n$ für jedes $k = 1, \dots, n$.
 - (b) $\mathbf{P}[N_l^n = k_l | N_1^n = k_1, \dots, N_{l-1}^n = k_{l-1}] = 1/(n - (k_1 + \dots + k_{l-1}))$ für $k_l = 1, \dots, n - (k_1 + \dots + k_{l-1})$.
 - (c) Man folgere die Aussage von Satz 24.35 für den Fall $\alpha = 0$ und $\theta = 1$.

(ii) Man zeige für $\theta > 0$:

- (a) $n \mathbf{P}[N_1^n = \lfloor nx \rfloor] \xrightarrow{n \rightarrow \infty} \theta(1-x)^{\theta-1}$ für $x \in (0, 1)$.
- (b) $n \mathbf{P}[N_l^n = \lfloor nx_l \rfloor \mid N_1^n = \lfloor nx_1 \rfloor, \dots, N_{l-1}^n = \lfloor nx_{l-1} \rfloor] \xrightarrow{n \rightarrow \infty} (\theta/y_l)(1-x_l/y_l)^{\theta-1}$ für $x_1, \dots, x_l \in (0, 1)$ mit $y_l = 1 - (x_1 + \dots + x_{l-1}) > x_l$.
- (c) Man folgere wie in (i) die Aussage von Satz 24.35 für $\alpha = 0$ und $\theta > 0$. ♣

25 Das Itô-Integral

Das Itô-Integral erlaubt es, stochastische Prozesse bezüglich der Zuwächse einer Brown'schen Bewegung oder etwas allgemeinerer Prozesse zu integrieren. Wir entwickeln das Itô-Integral zunächst für die Brown'sche Bewegung und dann für verallgemeinerte Diffusionsprozesse (sogenannte Itô-Prozesse). Im dritten Abschnitt leiten wir die Itô-Formel her. Diese Substitutionsformel für das Itô-Integral erlaubt es, in konkreten Fällen, mit dem Itô-Integral wirklich zu rechnen. Wir wenden die Itô-Formel im vierten Abschnitt an, um eine stochastische Lösung des Dirichlet-Problems zu formulieren. Hiermit zeigen wir im fünften Abschnitt, dass die Brown'sche Bewegung (wie die symmetrische einfache Irrfahrt) in niedrigen Dimensionen rekurrent ist, in hohen Dimensionen hingegen transient.

25.1 Das Itô-Integral bezüglich der Brown'schen Bewegung

Sei $W = (W_t)_{t \geq 0}$ eine Brown'sche Bewegung auf dem Raum $(\Omega, \mathcal{F}, \mathbf{P})$ bezüglich der Filtration \mathbb{F} , die die üblichen Bedingungen erfüllt (siehe Definition 21.22). Das heißt, W ist eine Brown'sche Bewegung und ist ein \mathbb{F} -Martingal. Das Ziel dieses Abschnittes ist es, für eine möglichst große Klasse von sinnvollen Integranden $H : \Omega \times [0, \infty) \rightarrow \mathbb{R}$, $(\omega, t) \mapsto H_t(\omega)$ ein Integral

$$I_t^W(H) = \int_0^t H_s dW_s$$

zu definieren, sodass $(I_t^W(H))_{t \geq 0}$ ein stetiges \mathbb{F} -Martingal ist. Da fast alle Pfade $s \mapsto W_s(\omega)$ der Brown'schen Bewegung lokal unendliche Variation haben, ist $W(\omega)$ nicht die Verteilungsfunktion eines signierten Lebesgue-Stieltjes-Maßes auf $[0, \infty)$. Daher können wir $I_t^W(H)$ nicht im klassischen Rahmen der Integrationstheorie definieren. Die grundlegende Idee, um dieses Integral zu konstruieren, besteht darin, es im Sinne eines L^2 -Grenzwertes zu etablieren. Hierzu betrachten wir zunächst ein elementares Beispiel.

Beispiel 25.1. Es seien X_1, X_2, \dots u.i.v. Zufallsvariablen mit $\mathbf{P}[X_n = 1] = \mathbf{P}[X_n = -1] = \frac{1}{2}$. Sei $(h_n)_{n \in \mathbb{N}}$ eine Folge reeller Zahlen. Unter welchen Bedingungen an $(h_n)_{n \in \mathbb{N}}$ ist die Reihe

$$R := \sum_{n \in \mathbb{N}} h_n X_n \tag{25.1}$$

wohldefiniert? Ist $\sum_{n \in \mathbb{N}} |h_n| < \infty$, so konvergiert die Reihe für jedes ω absolut. In diesem Falle tritt kein Problem auf. Wie steht es aber, wenn nur die schwächere Summierbarkeitsbedingung $\sum_{n \in \mathbb{N}} h_n^2 < \infty$ gilt? In diesem Falle konvergiert die Reihe in (25.1) nicht mehr für jedes ω , allerdings gilt $\mathbf{E}[h_n X_n] = 0$ für jedes $n \in \mathbb{N}$ und $\sum_{n=1}^{\infty} \mathbf{Var}[h_n X_n] = \sum_{n=1}^{\infty} h_n^2 < \infty$. Also ist $R_N := \sum_{k=1}^N h_k X_k$, $N \in \mathbb{N}$, konvergent im L^2 -Sinne (für $N \rightarrow \infty$). Wir können daher die Reihe R in (25.1) als L^2 -Limes der Partialsummen R_N definieren. Dabei ist zu beachten, dass (zumindest formal) bei den approximierenden Summen die Reihenfolge der Summanden eine Rolle spielt. Wir haben also gewissermaßen $\sum_{n=1}^{\infty}$ anstatt $\sum_{n \in \mathbb{N}}$ konstruiert.

Eine äquivalente Betrachtung, die allerdings einen leicht anderen Geschmack hat und von der formalen Beschreibung her auf das Kommende hinweist, ist die folgende. Mit ℓ^2 bezeichnen wir den Hilbertraum der quadratsummierbaren Folgen reeller Zahlen mit Skalarprodukt $\langle h, g \rangle = \sum_{n=1}^{\infty} h_n g_n$ und Norm $\|g\| = \langle g, g \rangle^{1/2}$. Sei ℓ^f der Unterraum der Folgen, die nur endlich viele Glieder ungleich Null haben. Dann ist $R(h) = \sum_{n \in \mathbb{N}} h_n X_n$ für $h \in \ell^f$ wohldefiniert (als endliche Summe). Wegen

$$\mathbf{E}[R(h)^2] = \mathbf{Var}[R(h)] = \sum_{n \in \mathbb{N}} \mathbf{Var}[h_n X_n] = \sum_{n \in \mathbb{N}} h_n^2 = \|h\|^2$$

ist die Abbildung $R : \ell^f \rightarrow \mathcal{L}^2(\mathbf{P})$ eine Isometrie. Da $\ell^f \subset \ell^2$ dicht liegt, können wir R stetig auf ℓ^2 fortsetzen. Ist also $h \in \ell^2$ und $(h^N)_{N \in \mathbb{N}}$ eine Folge in ℓ^f mit $\|h^N - h\| \xrightarrow{N \rightarrow \infty} 0$, so ist $R(h^N) \xrightarrow{N \rightarrow \infty} R(h)$ im L^2 -Sinne. Speziell ist $h_n^N := h_n \mathbb{1}_{\{n \leq N\}}$, $n \in \mathbb{N}$, $N \in \mathbb{N}$, eine approximierende Folge für h , und es gilt $R(h^N) = \sum_{n=1}^N h_n X_n$. Daher ist die oben beschriebene Approximation von R mit den Partialsummen R_N als Spezialfall in dieser Konstruktion enthalten. \diamond

Das Programm für die Konstruktion des Itô-Integrals $I_t^W(H)$ sieht nun so aus: Zunächst betrachten wir *elementare* Integranden H , für die Abbildung $t \mapsto H_t(\omega)$ eine Treppenfunktion ist, sodass das Integral als endliche Summe definiert werden kann. Danach erweitern wir das Integral wie in Beispiel 25.1 auf Integranden, die sich in einem gewissen L^2 -Sinne durch elementare Integranden approximieren lassen.

Definition 25.2. Wir bezeichnen mit \mathcal{E} den Vektorraum der Abbildungen $H : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ von der Form

$$H_t(\omega) = \sum_{i=1}^n h_{i-1}(\omega) \mathbb{1}_{(t_{i-1}, t_i]},$$

wobei $n \in \mathbb{N}$, $0 = t_0 < t_1 < \dots < t_n$ und h_{i-1} beschränkt und $\mathcal{F}_{t_{i-1}}$ -messbar ist für jedes $i = 1, \dots, n$.

Wir nennen \mathcal{E} den Vektorraum der elementaren vorhersagbaren Prozesse und versehen \mathcal{E} mit einer (Pseudo-)Norm $\|\cdot\|_{\mathcal{E}}$ durch

$$\|H\|_{\mathcal{E}}^2 = \sum_{i=1}^n \mathbf{E}[h_{i-1}^2] (t_i - t_{i-1}) = \mathbf{E}\left[\int_0^\infty H_s^2 ds\right].$$

Definition 25.3. Für $H \in \mathcal{E}$ und $t \geq 0$ definieren wir

$$I_t^W(H) = \sum_{i=1}^n h_{i-1} (W_{t_i \wedge t} - W_{t_{i-1} \wedge t})$$

und

$$I_\infty^W(H) = \sum_{i=1}^n h_{i-1} (W_{t_i} - W_{t_{i-1}}).$$

Offenbar ist für jede beschränkte Stoppzeit τ

$$\begin{aligned} \mathbf{E}[I_\tau^W(H)] &= \sum_{i=1}^n \mathbf{E}[h_{i-1} (W_{t_i}^\tau - W_{t_{i-1}}^\tau)] \\ &= \sum_{i=1}^n \mathbf{E}[h_{i-1} \mathbf{E}[W_{t_i}^\tau - W_{t_{i-1}}^\tau | \mathcal{F}_{t_{i-1}}]] = 0, \end{aligned}$$

da die gestoppte Brown'sche Bewegung W^τ nach den Optional Stopping Theorem ein \mathbb{F} -Martingal ist. Also ist (wieder nach dem OST) $(I_t^W(H))_{t \geq 0}$ ein \mathbb{F} -Martingal. Speziell ist $\mathbf{E}[(I_{t_{i+1}}^W(H) - I_{t_i}^W(H))(I_{t_{j+1}}^W(H) - I_{t_j}^W(H))] = 0$ für $i \neq j$, also gilt

$$\begin{aligned} \mathbf{E}[I_\infty^W(H)^2] &= \sum_{i=1}^n \mathbf{E}[(I_{t_i}^W(H) - I_{t_{i-1}}^W(H))^2] \\ &= \sum_{i=1}^n \mathbf{E}[h_{i-1}^2 (W_{t_i} - W_{t_{i-1}})^2] \\ &= \sum_{i=1}^n \mathbf{E}[h_{i-1}^2] (t_i - t_{i-1}) = \|H\|_{\mathcal{E}}^2. \end{aligned} \tag{25.2}$$

Aus diesen Betrachtungen folgt sofort:

- Satz 25.4.** (i) Die Abbildung $I_\infty^W : \mathcal{E} \rightarrow \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$ ist eine isometrische lineare Abbildung (bezüglich $\|\cdot\|_{\mathcal{E}}$ und $\|\cdot\|_2$).
(ii) Der Prozess $(I_t^W(H))_{t \geq 0}$ ist ein L^2 -beschränktes, stetiges \mathbb{F} -Martingal.

Beweis. Lediglich die Linearität ist noch zu zeigen. Dies ist aber trivial. \square

Die Idee ist nun, die Abbildung I_∞^W von \mathcal{E} auf einen geeigneten Abschluss $\bar{\mathcal{E}}$ von \mathcal{E} stetig fortzusetzen. Als Unterraum von welchem Raum sollen wir aber \mathcal{E} abschließen? Eine minimale Forderung ist die Messbarkeit von $(\omega, t) \mapsto H_t(\omega)$ (bezüglich $\mathcal{F} \otimes \mathcal{B}([0, \infty))$) sowie die Adaptiertheit von H .

Definition 25.5. Ein stochastischer Prozess $X = (X_t)_{t \geq 0}$ mit Werten in einem polnischen Raum E heißt

- (i) **produktmessbar**, falls $(\omega, t) \mapsto X_t(\omega)$ messbar ist bezüglich $\mathcal{F} \otimes \mathcal{B}([0, \infty))$ - $\mathcal{B}(E)$,
- (ii) **progressiv messbar**, falls für jedes $t \geq 0$ die Abbildung $\Omega \times [0, t] \rightarrow E$, $(\omega, s) \mapsto X_s(\omega)$ messbar ist bezüglich $\mathcal{F}_t \otimes \mathcal{B}([0, t])$ - $\mathcal{B}(E)$,
- (iii) **vorhersagbar** (oder **previsibel**), falls $(\omega, t) \mapsto X_t(\omega)$ messbar ist bezüglich der vorhersagbaren σ -Algebra \mathcal{P} auf $\Omega \times [0, \infty)$:

$$\mathcal{P} := \sigma(X : X \text{ ist linksstetiger, adaptierter Prozess}).$$

Bemerkung 25.6. Jedes $H \in \mathcal{E}$ ist vorhersagbar. Diese Eigenschaft sichert, dass $I^M(H)$ für jedes (auch unstetiges) Martingal M ein Martingal ist. Da wir jedoch hier nicht die Integrationstheorie für unstetige Martingale entwickeln wollen, ist der Begriff der Vorhersagbarkeit für uns im Folgenden nicht so wichtig. \diamond

Bemerkung 25.7. Ist H progressiv messbar, so ist H offenbar auch produktmessbar und adaptiert. Mit etwas mehr Aufwand kann man die partielle Umkehrung zeigen: Ist H adaptiert und produktmessbar, so gibt es eine progressiv messbare Modifikation von H . (Siehe etwa [125, Seite 68ff].) \diamond

Satz 25.8. Ist H adaptiert und rechtsstetig oder linksstetig, so ist H progressiv messbar. Ist H adaptiert und f.s. rechtsstetig oder linksstetig, so existiert ein Version von H , die progressiv messbar ist.

Insbesondere ist jeder vorhersagbare Prozess progressiv messbar.

Beweis. Siehe Übung 21.1.4. \square

Wir betrachten \mathcal{E} als Unterraum von

$$\mathcal{E}_0 := \left\{ H : \text{ produktmessbar, adaptiert und } \|H\|^2 := \mathbf{E} \left[\int_0^\infty H_t^2 dt \right] < \infty \right\}.$$

Sei $\bar{\mathcal{E}}$ der Abschluss von \mathcal{E} in \mathcal{E}_0 .

Satz 25.9. Ist H progressiv messbar (etwa linksstetig oder rechtsstetig und adaptiert) und $\mathbf{E} \left[\int_0^\infty H_t^2 dt \right] < \infty$, so ist $H \in \bar{\mathcal{E}}$.

Beweis. Sei H progressiv messbar und $\mathbf{E}\left[\int_0^\infty H_t^2 dt\right] < \infty$. Es reicht zu zeigen, dass für jedes $T > 0$ eine Folge $(H^n)_{n \in \mathbb{N}}$ in \mathcal{E} existiert mit

$$\mathbf{E}\left[\int_0^T (H_s - H_s^n)^2 ds\right] \xrightarrow{n \rightarrow \infty} 0. \quad (25.3)$$

Schritt 1. Sei zunächst H stetig und beschränkt. Setze $H_0^n = 0$ und

$$H_t^n = H_{i2^{-n}T} \quad \text{falls } i2^{-n}T < t \leq (i+1)2^{-n}T \quad \text{für ein } i = 0, \dots, 2^n - 1$$

und $H_t^n = 0$ für $t > T$. Dann ist $H^n \in \mathcal{E}$, und es gilt $H_t^n(\omega) \xrightarrow{n \rightarrow \infty} H_t(\omega)$ für alle $t > 0$ und $\omega \in \Omega$. Nach dem Satz von der majorisierten Konvergenz gilt (25.3).

Schritt 2. Sei nun H progressiv messbar und beschränkt. Es reicht zu zeigen, dass es stetige, adaptierte Prozesse H^n , $n \in \mathbb{N}$, gibt, für die (25.3) gilt. Sei

$$H_t^n := n \int_{(t-1/n)\vee 0}^{t \wedge T} H_s ds \quad \text{für } t \geq 0, n \in \mathbb{N}.$$

Dann ist H^n stetig und adaptiert und durch $\|H\|_\infty$ beschränkt. Nach dem Hauptatz der Differential- und Integralrechnung (siehe Übung 13.1.7) gilt

$$H_t^n(\omega) \xrightarrow{n \rightarrow \infty} H_t(\omega) \quad \text{für } \lambda\text{-fast alle } t \in [0, T] \text{ und für jedes } \omega \in \Omega. \quad (25.4)$$

Nach dem Satz von Fubini und dem Satz über majorisierte Konvergenz gilt daher

$$\mathbf{E}\left[\int_0^T (H_s - H_s^n)^2 ds\right] = \int_{\Omega \times [0, T]} (H_s(\omega) - H_s^n(\omega))^2 (\mathbf{P} \otimes \lambda)(d(\omega, s)) \xrightarrow{n \rightarrow \infty} 0.$$

Schritt 3. Sei nun H progressiv messbar und $\mathbf{E}\left[\int_0^\infty H_t^2 dt\right] < \infty$. Es reicht zu zeigen, dass es eine Folge $(H^n)_{n \in \mathbb{N}}$ von beschränkten, progressiv messbaren Prozessen gibt, sodass (25.3) gilt. Offenbar kann hierzu aber $H_t^n = H_t \mathbb{1}_{\{|H_t| < n\}}$ gewählt werden. \square

Definition 25.10 (Itô-Integral). Für $H \in \overline{\mathcal{E}}$ definieren wir das **Itô-Integral**

$$\int_0^\infty H_s dW_s := I_\infty^W(H)$$

durch die stetige Fortsetzung der Abbildung $I_\infty^W : \mathcal{E} \rightarrow \mathcal{L}^2(\mathbf{P})$ auf den Abschluss $\overline{\mathcal{E}}$ von \mathcal{E} . Mit anderen Worten: Ist $(H^n)_{n \in \mathbb{N}}$ eine Folge in \mathcal{E} mit $\|H - H^n\| \xrightarrow{n \rightarrow \infty} 0$, so definieren wir $I_\infty^W(H)$ durch

$$I_\infty^W(H) := \lim_{n \rightarrow \infty} I_\infty^W(H^n) \quad \text{in } L^2.$$

Ist τ eine Stoppzeit, so kürzen wir im Folgenden ab:

$$H_t^{(\tau)} := H_t \mathbb{1}_{\{t \leq \tau\}} \quad \text{für } t \geq 0.$$

(Man beachte den Gegensatz zum gestoppten Prozess $H_t^\tau = H_{\tau \wedge t}$.)

Satz 25.11. (i) Die Abbildung $I_\infty^W : \overline{\mathcal{E}} \rightarrow \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$ ist linear und

$$\mathbf{E}[I_\infty^W(H)^2] = \mathbf{E}\left[\int_0^\infty H_s^2 ds\right].$$

(ii) Für jedes $H \in \overline{\mathcal{E}}$ wird durch $\tilde{I}_t^W(H) := I_\infty^W(H^{(t)})$ ein L^2 -beschränktes \mathbb{F} -Martingal $\tilde{I}^W(H)$ definiert, das eine stetige Modifikation $I^W(H)$ besitzt.

Definition 25.12 (Itô-Integral als Prozess). Sei $I^W(H)$ die stetige Modifikation des Martingals $(I_\infty^W(H^{(t)}))_{t \geq 0}$ aus Satz 25.11(ii). Wir bezeichnen mit

$$\int_s^t H_r dW_r := I_t^W(H) - I_s^W(H) \quad \text{für } 0 \leq s \leq t \leq \infty$$

das Itô-Integral von s bis t von H bezüglich der Brown'schen Bewegung W .

Beweis (von Satz 25.11). (i) Dies folgt direkt aus der Definition von $I_\infty^W(H)$.

(ii) Sei $(H^n)_{n \in \mathbb{N}}$ eine Folge in \mathcal{E} mit $\|H^n - H\| \xrightarrow{n \rightarrow \infty} 0$. Nach Satz 25.4(ii) ist

$$I_\infty^W((H^n)^{(t)}) = I_t^W(H^n) = \mathbf{E}[I_\infty^W(H^n) | \mathcal{F}_t] \quad \text{für alle } t \geq 0, n \in \mathbb{N}.$$

Wegen $\|(H^n)^{(t)} - H^{(t)}\| \leq \|H^n - H\| \xrightarrow{n \rightarrow \infty} 0$ folgt (zusammen mit Korollar 8.21)

$$\tilde{I}_t^W(H) = \lim_{n \rightarrow \infty} I_t^W(H^n) = \lim_{n \rightarrow \infty} \mathbf{E}[I_\infty^W(H^n) | \mathcal{F}_t] = \mathbf{E}[I_\infty^W(H) | \mathcal{F}_t].$$

Mithin ist $\tilde{I}^W(H)$ ein L^2 -beschränktes Martingal und $I_t^W(H^n) \xrightarrow{n \rightarrow \infty} \tilde{I}_t^W(H)$ in L^2 für jedes $t \geq 0$. Nach Satz 25.4(ii) ist $I^W(H^n)$ stetig für jedes $n \in \mathbb{N}$, und nach Übung 21.4.3 existiert eine stetige Modifikation $I^W(H)$ von $\tilde{I}^W(H)$. \square

Als letzten Schritt bei der Konstruktion des Itô-Integrals wollen wir uns von der strengen Integrierbarkeitsbedingung $\mathbf{E}[\int_0^\infty H_s^2 ds] < \infty$ lösen. Hierzu machen wir zunächst die eine einfache Feststellung.

Sei τ eine Stoppzeit. Wir erinnern daran, dass mit $\int_0^\tau H_s dW_s$ für jedes ω die Zufallsvariable $\left(\int_0^{\tau(\omega)} H_s dW_s\right)(\omega)$ bezeichnet.

Lemma 25.13. Sei τ eine Stoppzeit und $H \in \overline{\mathcal{E}}$.

(i) Es gilt

$$\int_0^\tau H_s dW_s = \int_0^\infty H_s^{(\tau)} dW_s := \int_0^\infty H_s \mathbb{1}_{\{s \leq \tau\}} dW_s \quad f.s.$$

(ii) Speziell gilt für jedes $t \geq 0$ auf $\{\tau \geq t\}$

$$\int_0^t H_s dW_s = \int_0^t H_s^{(\tau)} dW_s \quad f.s.$$

(iii) Sei nun zudem $G \in \bar{\mathcal{E}}$ mit $H_s = G_s$ für jedes $s \leq \tau$. Dann gilt

$$\int_0^\tau H_s dW_s = \int_0^\tau G_s dW_s \quad f.s.$$

Beweis. (i) Für τ mit Werten in $\{k/2^n : k \in \mathbb{N}_0\} \cup \{\infty\}$ für ein $n \in \mathbb{N}$ ist $\mathbb{1}_{\{k/2^n \leq \tau\}} \mathbb{1}_{\{t \in ((k-1)/2^n, k/2^n]\}} \in \mathcal{E}$ für jedes $k \in \mathbb{N}$. Ist zudem $H \in \mathcal{E}$, so ist also auch $H^{(\tau)} \in \mathcal{E}$ und die Aussage folgt direkt aus der Definition des Itô-Integrals (Definition 25.3). Sei nun $H \in \bar{\mathcal{E}}$ und $(H^k)_{k \in \mathbb{N}}$ eine Folge in \mathcal{E} mit $\|H^k - H\|_{\mathcal{E}} \xrightarrow{k \rightarrow \infty} 0$. Wir schreiben $H_t^{k,(\tau)} := H_t^k \mathbb{1}_{\{t \leq \tau\}}$. Dann gilt $\|H^{k,(\tau)} - H^{(\tau)}\|_{\mathcal{E}} \xrightarrow{k \rightarrow \infty} 0$. Durch Übergang zu einer geeigneten Folge $k_m \uparrow \infty$ erhalten wir

$$\begin{aligned} \int_0^\tau H_s dW_s &= \lim_{m \rightarrow \infty} \int_0^\tau H_s^{k_m} dW_s \\ &= \lim_{m \rightarrow \infty} \int_0^\infty H_s^{k_m,(\tau)} dW_s = \int_0^\infty H_s^{(\tau)} dW_s \quad f.s. \end{aligned}$$

Sei schließlich τ eine beliebige Stoppzeit und $\tau_n := 2^{-n} \lceil 2^n \tau \rceil$ für $n \in \mathbb{N}$. Dann ist (τ_n) eine Folge von Stoppzeiten mit $\tau_n \downarrow \tau$. Da $I^W(H)$ nach Definition stetig ist, und wegen $\|H^{(\tau_n)} - H^{(\tau)}\|_{\mathcal{E}} \xrightarrow{n \rightarrow \infty} 0$ folgt nach Übergang zu einer geeigneten Folge $n(m) \uparrow \infty$

$$\begin{aligned} \int_0^\tau H_s dW_s &= \lim_{m \rightarrow \infty} \int_0^{\tau_{n(m)}} H_s dW_s \\ &= \lim_{m \rightarrow \infty} \int_0^\infty H_s^{(\tau_{n(m)})} dW_s = \int_0^\infty H_s^{(\tau)} dW_s \quad f.s. \end{aligned}$$

(ii), (iii) Diese Aussagen folgen direkt aus (i). □

Definition 25.14. Sei \mathcal{E}_{loc} der Raum der progressiv messbaren stochastischen Prozesse H mit

$$\int_0^T H_s^2 ds < \infty \quad f.s. \quad \text{für jedes } T > 0.$$

Lemma 25.15. Für jedes $H \in \mathcal{E}_{loc}$ existiert eine Folge $(\tau_n)_{n \in \mathbb{N}}$ von Stoppzeiten mit $\tau_n \uparrow \infty$ fast sicher und $\mathbf{E}[\int_0^{\tau_n} H_s^2 ds] < \infty$, also mit $H^{(\tau_n)} \in \overline{\mathcal{E}}$ für jedes $n \in \mathbb{N}$.

Beweis. Setze

$$\tau_n := \inf \left\{ t \geq 0 : \int_0^t H_s^2 ds \geq n \right\}.$$

Nach der Definition von \mathcal{E}_{loc} gilt $\tau_n \uparrow \infty$ fast sicher und nach Konstruktion ist $\|H^{(\tau_n)}\|^2 = \mathbf{E}[\int_0^{\tau_n} H_s^2 ds] \leq n$. \square

Definition 25.16. Sei $H \in \mathcal{E}_{loc}$ und $(\tau_n)_{n \in \mathbb{N}}$ wie in Lemma 25.15. Wir definieren für $t \geq 0$ das Itô-Integral als den fast sicheren Grenzwert

$$\int_0^t H_s dW_s := \lim_{n \rightarrow \infty} \int_0^t H_s^{(\tau_n)} dW_s. \quad (25.5)$$

Satz 25.17. Sei $H \in \mathcal{E}_{loc}$.

- (i) Der Grenzwert in (25.5) ist wohldefiniert, stetig in t und (f.s.) unabhängig von der Wahl der Folge $(\tau_n)_{n \in \mathbb{N}}$.
- (ii) Ist τ eine Stoppzeit mit $\mathbf{E}[\int_0^\tau H_s^2 ds] < \infty$, so ist das gestoppte Itô-Integral $\left(\int_0^{\tau \wedge t} H_s dW_s \right)_{t \geq 0}$ ein L^2 -beschränktes, stetiges Martingal.
- (iii) Ist speziell $\mathbf{E}[\int_0^T H_s^2 ds] < \infty$ für jedes $T > 0$, so ist $\left(\int_0^t H_s dW_s \right)_{t \geq 0}$ ein quadratintegrierbares, stetiges Martingal.

Beweis. (i) Nach Lemma 25.13(ii) ist auf dem Ereignis $\{\tau_n \geq t\}$

$$\int_0^t H_s dW_s = \int_0^t H_s^{(\tau_n)} dW_s.$$

Also existiert der Limes, ist stetig und unabhängig von der Wahl der Folge $(\tau_n)_{n \in \mathbb{N}}$.

(ii) Dies folgt direkt aus Satz 25.11.

(iii) Da wir $\tau_n = n$ wählen können, folgt dies aus (ii). \square

Satz 25.18. Sei H progressiv messbar und $\mathbf{E}[\int_0^T H_s^2 ds] < \infty$ für alle $T > 0$. Dann definiert

$$M_t := \int_0^t H_s dW_s, \quad t \geq 0,$$

ein quadratintegrierbares, stetiges Martingal, und

$$(N_t)_{t \geq 0} := \left(M_t^2 - \int_0^t H_s^2 ds \right)_{t \geq 0}$$

ist ein stetiges Martingal mit $N_0 = 0$.

Beweis. Es reicht zu zeigen, dass N ein Martingal ist. Offenbar ist N adaptiert. Sei τ ein beschränkte Stoppzeit. Dann ist

$$\begin{aligned} \mathbf{E}[N_\tau] &= \mathbf{E}\left[M_\tau^2 - \int_0^\tau H_s^2 ds\right] \\ &= \mathbf{E}\left[\left(\int_0^\infty H_s^{(\tau)} dW_s\right)^2\right] - \mathbf{E}\left[\int_0^\infty (H_s^{(\tau)})^2 ds\right] = 0. \end{aligned}$$

Nach dem Optional Stopping Theorem (siehe Übung 21.1.3(iii)) ist N damit als Martingal erkannt. \square

Wir erinnern an den Begriff des lokalen Martingals und der quadratischen Variation aus Kapitel 21.10.

Korollar 25.19. Ist $H \in \mathcal{E}_{loc}$, so ist das Itô-Integral $M_t = \int_0^t H_s dW_s$ ein stetiges lokales Martingal mit quadratischem Variationsprozess $\langle M \rangle_t = \int_0^t H_s^2 ds$.

Beispiel 25.20. (i) $W_t = \int_0^t 1 dW_s$ ist ein quadratintegrierbares Martingal, und $(W_t^2 - t)_{t \geq 0}$ ist ein stetiges Martingal.

(ii) Wegen $\mathbf{E}\left[\int_0^T W_s^2 ds\right] = \frac{T^2}{2} < \infty$ für alle $T \geq 0$ ist $M_t := \int_0^t W_s dW_s$ ein stetiges, quadratintegrierbares Martingal, und $(M_t^2 - \int_0^t W_s^2 ds)_{t \geq 0}$ ist ein stetiges Martingal.

(iii) Sei H progressiv messbar und beschränkt sowie $M_t := \int_0^t H_s dW_s$. Dann ist M progressiv messbar (weil stetig und adaptiert) und

$$\mathbf{E}\left[\int_0^T M_s^2 ds\right] = \int_0^T \left(\int_0^s \mathbf{E}[H_r^2] dr\right)^2 ds \leq \frac{T^2 \|H\|_\infty^2}{2}.$$

Also ist $\widetilde{M}_t := \int_0^t M_s dW_s$ ein quadratisch integrierbares, stetiges Martingal und $(\widetilde{M}_t^2 - \int_0^t M_s^2 ds)_{t \geq 0}$ ist ein stetiges Martingal. \diamond

25.2 Itô-Integral bezüglich Diffusionen

Ist

$$H = \sum_{i=1}^n h_{i-1} \mathbb{1}_{(t_{i-1}, t_i]} \in \mathcal{E}, \quad (25.6)$$

so ist das elementare Integral

$$I_t^M(H) = \sum_{i=1}^n h_{i-1} (M_{t_i \wedge t} - M_{t_{i-1} \wedge t})$$

ein Martingal (beziehungsweise lokales Martingal), wenn M ein Martingal (beziehungsweise lokales Martingal) ist, und es gilt

$$\begin{aligned} \mathbf{E}[(I_\infty^M(H))^2] &= \sum_{i=1}^n \mathbf{E}[h_{i-1}^2 (M_{t_i} - M_{t_{i-1}})^2] = \sum_{i=1}^n \mathbf{E}[h_{i-1}^2 (\langle M \rangle_{t_i} - \langle M \rangle_{t_{i-1}})] \\ &= \mathbf{E}\left[\int_0^\infty H_t^2 d\langle M \rangle_t\right], \end{aligned}$$

falls der Ausdruck auf der rechten Seite endlich ist. Grob gesprochen können wir die Prozedur, mit der wir das Itô-Integral für die Brown'sche Bewegung in Abschnitt 25.1 für Integranden $H \in \bar{\mathcal{E}}$ definiert hatten, wiederholen, um ein Integral bezüglich M für eine große Klasse von Integranden zu definieren. Für die Definition der Norm auf \mathcal{E} müssen wir im Prinzip nur dt (die quadratische Variation der Brown'schen Bewegung) durch $d\langle M \rangle_t$ ersetzen:

$$\|H\|_M^2 := \mathbf{E}\left[\int_0^\infty H_t^2 d\langle M \rangle_t\right].$$

Das Problem besteht nicht darin, das elementare Integral auf $\bar{\mathcal{E}}$ fortzusetzen, sondern darin zu prüfen, welche Prozesse in $\bar{\mathcal{E}}$ liegen. Für unstetige Martingale etwa müssen die Integranden vorhersagbar sein, damit das Integral ein Martingal wird (abgesehen von der Schwierigkeit, dass wir die Existenz einer quadratischen Variation für solche Martingale nicht etabliert haben und dies in diesem Rahmen auch nicht tun werden). Dies hatten wir in Kapitel 9.3 schon für den Fall diskreter Zeit gesehen. Haben wir nun ein stetiges Martingal M mit stetiger quadratischer Variation $\langle M \rangle$ vorliegen, so tritt immer noch folgendes Problem auf: Im Beweis von Satz 25.9 wurde in Schritt 2 benutzt, dass $H_t^n(\omega) \xrightarrow{n \rightarrow \infty} H_t(\omega)$ für Lebesgue-fast alle t und alle ω gilt, um zu zeigen, dass progressiv messbare H in $\bar{\mathcal{E}}$ liegen. Ist $d\langle M \rangle_t$ nun nicht absolutstetig bezüglich des Lebesgue-Maßes, so reicht dies aber nicht aus, um die Konvergenz der Integrale bezüglich $d\langle M \rangle_t$ zu folgern. Im Fall absolutstetiger quadratischer Variation hingegen geht der Beweis glatt durch. Wie in Abschnitt 25.1 erhalten wir:

Satz 25.21. Sei M ein stetiges lokales Martingal mit absolutstetiger quadratischer Variation $\langle M \rangle$ und H progressiv messbar mit $\int_0^T H_s^2 d\langle M \rangle_s < \infty$ f.s. für jedes $T \geq 0$. Dann ist das Itô-Integral $N_t := \int_0^t H_s dM_s$ wohldefiniert und ist ein stetiges lokales Martingal mit quadratischer Variation $\langle N \rangle_t = \int_0^t H_s^2 d\langle M \rangle_s$. Für jede Folge $(\tau_n)_{n \in \mathbb{N}}$ mit $\tau_n \uparrow \infty$ und $\|H^{(\tau_n)}\|_M < \infty$ und jede Familie $(H^{n,m}, n, m \in \mathbb{N}) \subset \mathcal{E}$ mit $\|H^{n,m} - H^{(\tau_n)}\|_M \xrightarrow{m \rightarrow \infty} 0$ gilt

$$\int_0^t H_s dM_s = \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} I_t^M(H^{m,n}) \quad \text{für alle } t \geq 0 \text{ stochastisch.}$$

Als gewisse Verallgemeinerung erhalten wir den folgenden Satz.

Satz 25.22. Seien M^1 und M^2 stetige lokale Martingale mit absolutstetiger quadratischer Variation. Sei H^i progressiv messbar mit $\int_0^T (H_s^i)^2 d\langle M^i \rangle_s < \infty$ für alle $i = 1, 2$ und $T < \infty$. Sei $N_t^i := \int_0^t H_s^i dM_s^i$ für $i = 1, 2$. Dann sind N^1 und N^2 stetige lokale Martingale mit quadratischer Kovariation $\langle N^i, N^j \rangle_t = \int_0^t H_s^i H_s^j d\langle M^i, M^j \rangle_s$. Sind M^1 und M^2 unabhängig, so ist $\langle N^1, N^2 \rangle \equiv 0$.

Beweis. Seien zunächst $H^1, H^2 \in \mathcal{E}$. Dann gibt es Zahlen $0 = t_0 < t_1 < \dots < t_n$ und \mathcal{F}_{t_k} -messbare beschränkte Abbildungen h_k^i , $i = 1, 2$, $k = 0, \dots, n-1$ mit

$$H_t^i(\omega) = \sum_{k=1}^n h_{k-1}^i(\omega) \mathbb{1}_{(t_{k-1}, t_k]}(t).$$

Es folgt

$$N_t^i N_t^j = \sum_{k,l=1}^n h_{k-1}^i h_{l-1}^j (M_{t_k \wedge t}^i - M_{t_{k-1} \wedge t}^i) (M_{t_l \wedge t}^j - M_{t_{l-1} \wedge t}^j).$$

Die Summanden mit $k \neq l$ sind jeweils lokale Martingale. Für jeden einzelnen Summanden mit $k = l$ ist

$$\begin{aligned} & \left(h_{k-1}^i h_{k-1}^j \left((M_{t_k \wedge t}^i - M_{t_{k-1} \wedge t}^i) (M_{t_k \wedge t}^j - M_{t_{k-1} \wedge t}^j) \right. \right. \\ & \quad \left. \left. - (\langle M^i, M^j \rangle_{t_k \wedge t} - \langle M^i, M^j \rangle_{t_{k-1} \wedge t}) \right) \right)_{t \geq 0} \end{aligned}$$

ein lokales Martingal. Wegen

$$\sum_{k=1}^n h_{k-1}^i h_{k-1}^j (\langle M^i, M^j \rangle_{t_k \wedge t} - \langle M^i, M^j \rangle_{t_{k-1} \wedge t}) = \int_0^t H_s^i H_s^j d\langle M^i, M^j \rangle_s$$

ist $(N_t^i N_t^j - \int_0^t H_s^i H_s^j d\langle M^i, M^j \rangle_s)_{t \geq 0}$ ein stetiges lokales Martingal.

Der Fall allgemeiner progressiv messbarer H^1, H^2 , die die Integrierbarkeitsbedingung erfüllen, folgt durch die üblichen L^2 -Approximationssargumente.

Sind M^1 und M^2 unabhängig, so ist $\langle M^1, M^2 \rangle \equiv 0$. □

Wir betrachten im Folgenden Prozesse, die sich als Itô-Integral bezüglich einer Brown'schen Bewegung schreiben lassen, und geben für diese Prozesse einen detaillierteren Beweis von Satz 25.21 an.

Definition 25.23. Sei W eine Brown'sche Bewegung und σ und b progressiv messbare stochastische Prozesse mit $\int_0^t \sigma_s^2 + |b_s| ds < \infty$ fast sicher für alle $t \geq 0$. Dann nennen wir den Prozess X mit

$$X_t = \int_0^t \sigma_s dW_s + \int_0^t b_s ds \quad \text{für } t \geq 0$$

einen verallgemeinerten **Diffusionsprozess** (oder kurz: verallgemeinerte Diffusion) mit **Diffusionskoeffizienten** σ und **Drift** b . Oft wird X auch ein **Itô-Prozess** genannt.

Haben σ und b speziell die Gestalt $\sigma_s = \tilde{\sigma}(X_s)$ und $b_s = \tilde{b}(X_s)$ für gewisse Abbildungen $\tilde{\sigma} : \mathbb{R} \rightarrow [0, \infty)$ und $\tilde{b} : \mathbb{R} \rightarrow \mathbb{R}$, so nennen wir X eine Diffusion (im engeren Sinne).

Im Gegensatz zu verallgemeinerten Diffusionen sind Diffusionen im engeren Sinne unter gewissen Regularitätsannahmen an die Koeffizienten stets Markovprozesse, wie wir noch sehen werden (vergleiche Satz 26.8, 26.10 und 26.26).

Eine Diffusion X hat stets die Gestalt $X = M + A$, wobei $M_t = \int_0^t \sigma_s dW_s$ ein stetiges lokales Martingal mit quadratischer Variation $\langle M \rangle_t = \int_0^t \sigma_s^2 ds$ ist (nach Korollar 25.19) und $A_t = \int_0^t b_s ds$ ein stetiger Prozess von lokal endlicher Variation.

Offenbar ist für H aus (25.6)

$$\begin{aligned} \int_0^t H_s dM_s &= \sum_{i=1}^n h_{i-1} (M_{t_i \wedge t} - M_{t_{i-1} \wedge t}) \\ &= \sum_{i=1}^n h_{i-1} \int_{t_{i-1} \wedge t}^{t_i \wedge t} \sigma_s dW_s = \int_0^t (H_s \sigma_s) dW_s. \end{aligned}$$

Für progressiv messbares H mit $\int_0^T H_s^2 d\langle M \rangle_s = \int_0^T (H_s \sigma_s)^2 ds < \infty$ für alle $T \geq 0$ definieren wir daher das Itô-Integral

$$\int_0^t H_s dM_s := \int_0^t (H_s \sigma_s) dW_s.$$

Wir erhalten ohne Weiteres, speziell ohne auf Satz 25.21 zurückzugreifen, den folgenden Satz.

Satz 25.24. Sei $X = M + A$ eine verallgemeinerte Diffusion mit σ und b wie in Definition 25.23 und H progressiv messbar mit

$$\int_0^T H_s^2 \sigma_s^2 ds < \infty \quad f.s. \quad \text{für alle } T \geq 0 \quad (25.7)$$

und

$$\int_0^T |H_s b_s| ds < \infty \quad f.s. \quad \text{für alle } T \geq 0, \quad (25.8)$$

so ist der durch

$$Y_t := \int_0^t H_s dX_s := \int_0^t H_s dM_s + \int_0^t H_s dA_s := \int_0^t H_s \sigma_s dW_s + \int_0^t H_s b_s ds$$

definierte Prozess Y eine verallgemeinerte Diffusion mit Diffusionskoeffizienten $(H_s \sigma_s)_{s \geq 0}$ und Drift $(H_s b_s)_{s \geq 0}$. Speziell ist $N_t := \int_0^t H_s dM_s$ ein stetiges lokales Martingal mit Variationsprozess $\langle N \rangle_t = \int_0^t H_s^2 d\langle M \rangle_s = \int_0^t H_s^2 \sigma_s^2 ds$.

Übung 25.2.1. Sei M ein stetiges lokales Martingal mit absolutstetiger quadratischer Variation $\langle M \rangle$ (etwa eine verallgemeinerte Diffusion), und sei H progressiv messbar und stetig mit $\int_0^T H_s^2 d\langle M \rangle_s < \infty$ für jedes $T \geq 0$. Sei ferner $\mathcal{P} = (\mathcal{P}^{(n)})_{n \in \mathbb{N}}$ eine zulässige Zerlegungsfolge (siehe Definition 21.56). Zeige:

(i) Für alle $T \geq 0$ gilt im Sinne stochastischer Konvergenz

$$\int_0^T H_s dM_s = \lim_{n \rightarrow \infty} \sum_{t \in \mathcal{P}_T^n} H_t (M_{t'} - M_t). \quad (25.9)$$

(ii) Es gibt eine Teilfolge von \mathcal{P} , so dass (25.9) fast sicher für alle $T \geq 0$ gilt. ♣

25.3 Die Itô-Formel

Dieser und die beiden folgenden Abschnitte sind inhaltlich an ein Vorlesungsskript von Hans Föllmer angelehnt.

Ist $t \mapsto X_t$ eine differenzierbare Abbildung mit Ableitung X' und $F \in C^1(\mathbb{R})$ mit Ableitung F' , so gilt die klassische Substitutionsformel

$$F(X_t) - F(X_0) = \int_0^t F'(X_s) dX_s = \int_0^t F'(X_s) X'_s ds. \quad (25.10)$$

Diese Formel bleibt richtig, wenn X stetig und von lokal endlicher Variation ist (siehe Kapitel 21.10), also die Verteilungsfunktion eines absolutstetigen signierten Maßes auf $[0, \infty)$ ist. Dann existiert die Ableitung X' als Radon-Nikodym Ableitung fast überall, und man kann leicht zeigen, dass (25.10) auch in diesem Fall gilt.

Die Pfade der Brown'schen Bewegung W sind nirgends differenzierbar (Satz 21.17 von Paley, Wiener und Zygmund) und haben (folglich) überall lokal unendliche Variation. Wir können also eine einfache Substitutionsformel wie in (25.10) nicht erwarten, und in der Tat sieht man leicht ein, dass sie falsch sein *muss*: Wählen wir $F(x) = x^2$, so ist die rechte Seite in (25.10) (mit X durch W ersetzt) $\int_0^t 2W_s dW_s$, also ein Martingal. Die linke Seite ist W_t^2 , also ein Submartingal, das erst durch Subtraktion von t zu einem Martingal wird. In der Tat ist dieses fehlende t der zusätzliche Term, den wir in der Substitutionsformel für Itô-Integrale, der so genannten Itô-Formel, bekommen. Eine (etwas haarsträubende) Heuristik führt uns erstaunlicherweise auf die richtige Spur: Für kleine t ist W_t ungefähr von der Größe \sqrt{t} . Wenn wir nun formal $dW_t = \sqrt{dt}$ schreiben und für $F \in C^2(\mathbb{R})$ eine Taylor-Entwicklung bis zur zweiten Ordnung durchführen, so erhalten wir

$$dF(W_t) = F'(W_t) dW_t + \frac{1}{2} F''(W_t) (dW_t)^2 = F'(W_t) dW_t + \frac{1}{2} F''(W_t) dt,$$

oder als Integral geschrieben

$$F(W_t) - F(W_0) = \int_0^t F'(W_s) dW_s + \int_0^t \frac{1}{2} F''(W_s) ds. \quad (25.11)$$

(Für gewisse diskrete Martingale haben wir eine analoge Formel schon in Beispiel 10.9 hergeleitet.) Hauptanliegen dieses Abschnittes ist es zu zeigen, dass diese Formel, die **Itô-Formel** für die Brown'sche Bewegung genannt wird, in der Tat korrekt ist.

Die weitere Diskussion in diesem Abschnitt hängt nicht explizit davon ab, dass wir bezüglich der Brown'schen Bewegung integrieren, sondern benutzt lediglich, dass die Funktion, bezüglich der wir integrieren, stetige quadratische Variation hat (entlang einer geeigneten zulässigen Zerlegungsfolge $\mathcal{P} = (\mathcal{P}^n)_{n \in \mathbb{N}}$), für die Brown'sche Bewegung nämlich $\langle W \rangle_t = t$.

Sei im Folgenden also $\mathcal{P} = (\mathcal{P}^n)_{n \in \mathbb{N}}$ eine zulässige Zerlegungsfolge (siehe Definition 21.56 für die Definition und die Notation $\mathcal{C}_{qV} = \mathcal{C}_{qV}^{\mathcal{P}}$, \mathcal{P}_T^n , $\mathcal{P}_{S,T}^n$, t' und so weiter) und $X \in C([0, \infty))$ mit stetiger quadratischer Variation (entlang \mathcal{P})

$$T \mapsto \langle X \rangle_T = V_T^2(X) = \lim_{n \rightarrow \infty} \sum_{t \in \mathcal{P}_T^n} (X_{t'} - X_t)^2.$$

Für die Brown'sche Bewegung ist $W \in \mathcal{C}_{qV}^{\mathcal{P}}$ fast sicher für jede zulässige Zerlegungsfolge (Satz 21.64) und $\langle W \rangle_T = T$. Für stetige lokale Martingale M kann man immerhin durch Übergang zu einer geeigneten Teilfolge \mathcal{P}' von \mathcal{P} sicherstellen, dass $M \in \mathcal{C}_{qV}^{\mathcal{P}'}$ fast sicher gilt (Satz 21.70).

Sei also \mathcal{P} fest gewählt und $X \in \mathcal{C}_{qV}$ eine (deterministische) Funktion.

Satz 25.25 (Pfadweise Itô-Formel). Sei $X \in \mathcal{C}_{\text{qV}}$ und $F \in C^2(\mathbb{R})$. Dann existiert für alle $T \geq 0$ der Limes

$$\int_0^T F'(X_s) dX_s := \lim_{n \rightarrow \infty} \sum_{t \in \mathcal{P}_T^n} F'(X_t)(X_{t'} - X_t), \quad (25.12)$$

und es gilt die Itô-Formel

$$F(X_T) - F(X_0) = \int_0^T F'(X_s) dX_s + \frac{1}{2} \int_0^T F''(X_s) d\langle X \rangle_s. \quad (25.13)$$

Dabei ist das rechte Integral in (25.13) als klassisches (Lebesgue-Stieltjes-) Integral zu verstehen.

Bemerkung 25.26. Ist M ein stetiges lokales Martingal, so ist nach Übung 25.2.1 das Itô-Integral $\int_0^T F'(M_s) dM_s$ der stochastische Limes der Summen $\sum_{t \in \mathcal{P}_T^n} F'(M_t)(M_{t'} - M_t)$ für $n \rightarrow \infty$. Tatsächlich stimmt also für $X = M(\omega)$ das pfadweise Integral in (25.12) mit dem Itô-Integral (f.s.) überein. Speziell gilt für das Itô-Integral der Brown'schen Bewegung die Itô-Formel (25.11). \diamond

Beweis (von Satz 25.25). Wir müssen zeigen, dass der Limes in (25.12) existiert und dass (25.13) gilt.

Für $n \in \mathbb{N}$ und $t \in \mathcal{P}_T^n$ (mit Nachfolger $t' \in \mathcal{P}_T^n$) liefert die Taylor-Formel

$$F(X_{t'}) - F(X_t) = F'(X_t)(X_{t'} - X_t) + \frac{1}{2}F''(X_t) \cdot (X_{t'} - X_t)^2 + R_t^n, \quad (25.14)$$

wobei wir das Restglied

$$R_t^n = (F''(\xi) - F''(X_t)) \cdot \frac{1}{2}(X_{t'} - X_t)^2$$

(für eine geeignete Zwischenstelle ξ zwischen X_t und $X_{t'}$) wie folgt abschätzen. Da X stetig ist, ist $C := \{X_t : t \in [0, T]\}$ kompakt und $F''|_C$ gleichmäßig stetig. Zu jedem $\varepsilon > 0$ gibt es also ein $\delta > 0$ mit

$$|F''(X_r) - F''(X_s)| < \varepsilon \quad \text{für alle } r, s \in [0, T] \text{ mit } |X_r - X_s| < \delta.$$

Da auch X gleichmäßig stetig ist auf $[0, T]$ und die Feinheit $|\mathcal{P}^n|$ der Zerlegung nach 0 geht für $n \rightarrow \infty$, gibt es (zu jedem $\delta > 0$) ein N_δ , sodass

$$\sup_{n \geq N_\delta} \sup_{t \in \mathcal{P}_T^n} |X_{t'} - X_t| < \delta.$$

Also ist für $n \geq N_\delta$ und $t \in \mathcal{P}_T^n$

$$|R_t^n| \leq \frac{1}{2}\varepsilon (X_{t'} - X_t)^2.$$

Summieren wir in (25.14) über $t \in \mathcal{P}_T^n$, so erhalten wir

$$\sum_{t \in \mathcal{P}_T^n} (F(X_{t'}) - F(X_t)) = F(X_t) - F(X_0)$$

und

$$\sum_{t \in \mathcal{P}_T^n} |R_t^n| \leq \varepsilon \sum_{t \in \mathcal{P}_T^n} (X_{t'} - X_t)^2 \xrightarrow{n \rightarrow \infty} \varepsilon \langle X \rangle_t < \infty.$$

Da $\varepsilon > 0$ beliebig war, gilt also $\sum_{t \in \mathcal{P}_T^n} |R_t^n| \xrightarrow{n \rightarrow \infty} 0$. Es gilt (siehe Übung 21.10.2)

$$\sum_{t \in \mathcal{P}_T^n} \frac{1}{2} F''(X_t) (X_{t'} - X_t)^2 \xrightarrow{n \rightarrow \infty} \frac{1}{2} \int_0^T F''(X_s) d\langle X \rangle_s.$$

Daher muss auch die Summe des verbleibenden Terms in (25.14) konvergieren, das heißt, es existiert der Limes in (25.12). \square

Als direkte Folgerung erhalten wir die Itô-Formel für das Itô-Integral bezüglich Diffusionen.

Satz 25.27 (Itô-Formel für Diffusionen). Sei $Y = M + A$, wobei $M_t = \int_0^t \sigma_s dW_s$ und $A_t = \int_0^t b_s ds$, eine (verallgemeinerte) Diffusion ist (siehe Definition 25.23). Sei $F \in C^2(\mathbb{R})$. Dann gilt die Itô-Formel

$$\begin{aligned} F(Y_t) - F(Y_0) &= \int_0^t F'(Y_s) dM_s + \int_0^t F'(Y_s) dA_s + \frac{1}{2} \int_0^t F''(Y_s) d\langle M \rangle_s \\ &= \int_0^t F'(Y_s) \sigma_s dW_s + \int_0^t \left(F'(Y_s) b_s + \frac{1}{2} F''(Y_s) \sigma_s^2 \right) ds. \end{aligned} \quad (25.15)$$

Speziell gilt für die Brown'sche Bewegung

$$F(W_t) - F(W_0) = \int_0^t F'(W_s) dW_s + \frac{1}{2} \int_0^t F''(W_s) ds. \quad (25.16)$$

Als Anwendung der Itô-Formel bringen wir eine Charakterisierung der Brown'schen Bewegung als stetiges lokales Martingal mit einer bestimmten quadratischen Variation.

Satz 25.28 (Lévy'sche Charakterisierung der Brown'schen Bewegung).

Sei $X \in \mathcal{M}_{loc,c}$ mit $X_0 = 0$. Dann sind äquivalent

- (i) $(X_t^2 - t)_{t \geq 0}$ ist ein lokales Martingal,
- (ii) $\langle X \rangle_t = t$ für alle $t \geq 0$,
- (iii) X ist eine Brown'sche Bewegung.

Beweis. (iii) \Rightarrow (i) Das ist klar.

(i) \iff (ii) Das ist klar, weil der quadratische Variationsprozess eindeutig ist.

(ii) \Rightarrow (iii) Es reicht zu zeigen, dass $X_t - X_s \sim \mathcal{N}_{0,t-s}$ gegeben \mathcal{F}_s für $t > s \geq 0$. Wegen des Eindeutigkeitssatzes für charakteristische Funktionen reicht es zu zeigen, dass (mit $i = \sqrt{-1}$) für $A \in \mathcal{F}_s$ und $\lambda \in \mathbb{R}$ gilt:

$$\varphi_{A,\lambda}(t) := \mathbf{E}[e^{i\lambda(X_t - X_s)} \mathbf{1}_A] = \mathbf{P}[A] e^{-\lambda^2(t-s)/2}.$$

Wir wenden die Itô-Formel separat auf Real- und Imaginärteil an und erhalten

$$e^{i\lambda X_t} - e^{i\lambda X_s} = \int_s^t i \lambda e^{i\lambda X_r} dX_r - \frac{1}{2} \int_s^t \lambda^2 e^{i\lambda X_r} dr.$$

Es folgt

$$\begin{aligned} & \mathbf{E}[e^{i\lambda(X_t - X_s)} \mid \mathcal{F}_s] - 1 \\ &= \mathbf{E}\left[\int_s^t i \lambda e^{i\lambda(X_r - X_s)} dX_r \mid \mathcal{F}_s\right] - \frac{1}{2} \lambda^2 \mathbf{E}\left[\int_s^t e^{i\lambda(X_r - X_s)} dr \mid \mathcal{F}_s\right]. \end{aligned}$$

Nun sind $M_t := \operatorname{Re} \int_s^t i \lambda e^{i\lambda(X_r - X_s)} dX_r$ und $N_t := \operatorname{Im} \int_s^t i \lambda e^{i\lambda(X_r - X_s)} dX_r$, $t \geq s$, stetige lokale Martingale mit $\langle M \rangle_t = \int_s^t \lambda^2 \sin(\lambda(X_r - X_s))^2 dr \leq \lambda^2(t-s)$ und $\langle N \rangle_t = \int_s^t \lambda^2 \cos(\lambda(X_r - X_s))^2 dr \leq \lambda^2(t-s)$. Nach Korollar 21.76 sind M und N daher Martingale, also gilt

$$\mathbf{E}\left[\int_s^t i \lambda e^{i\lambda(X_r - X_s)} dX_r \mid \mathcal{F}_s\right] = 0.$$

Der Satz von Fubini liefert (wegen $A \in \mathcal{F}_s$)

$$\begin{aligned} \varphi_{A,\lambda}(t) - \varphi_{A,\lambda}(s) &= \mathbf{E}[e^{i\lambda(X_t - X_s)} \mathbf{1}_A] - \mathbf{P}[A] \\ &= -\frac{1}{2} \lambda^2 \int_s^t \mathbf{E}[e^{i\lambda(X_r - X_s)} \mathbf{1}_A] dr = -\frac{1}{2} \lambda^2 \int_s^t \varphi_{A,\lambda}(r) dr. \end{aligned}$$

Das heißt, $\varphi_{A,\lambda}$ ist die Lösung des linearen Anfangswertproblems

$$\varphi_{A,\lambda}(s) = \mathbf{P}[A] \quad \text{und} \quad \frac{d}{dt} \varphi_{A,\lambda}(t) = -\frac{1}{2} \lambda^2 \varphi_{A,\lambda}(t).$$

Die eindeutige Lösung hiervon ist $\varphi_{A,\lambda}(t) = \mathbf{P}[A] e^{-\lambda^2(t-s)/2}$. \square

Als Folgerung aus dem Satz erhalten wir, dass wir jedes stetige lokale Martingal, dessen quadratischer Variationsprozess absolutstetig (als Funktion der Zeit) ist, als Itô-Integral bezüglich einer Brown'schen Bewegung schreiben können.

Satz 25.29 (Itô'scher Martingal-Darstellungssatz). Sei M ein stetiges lokales Martingal mit $M_0 = 0$ und absolutstetiger quadratischer Variation $t \mapsto \langle M \rangle_t$. Dann gibt es, eventuell auf einer Erweiterung des Wahrscheinlichkeitsraums, eine Brown'sche Bewegung W mit

$$M_t = \int_0^t \sqrt{\frac{d\langle M \rangle_s}{ds}} dW_s \quad \text{für alle } t \geq 0.$$

Beweis. Wir nehmen an, dass auf dem Wahrscheinlichkeitsraum eine Brown'sche Bewegung \tilde{W} definiert ist, die unabhängig von M ist. (Gegebenenfalls muss der Wahrscheinlichkeitsraum hierzu erweitert werden.) Sei

$$f_t := \lim_{n \rightarrow \infty} n(\langle M \rangle_t - \langle M \rangle_{t-1/n}) \quad \text{für } t > 0.$$

Dann ist f eine progressiv messbare Version der Radon-Nikodym Ableitung $\frac{d\langle M \rangle_t}{dt}$. Klar ist $\int_0^T \mathbb{1}_{\{f_t > 0\}} f_t^{-1} d\langle M \rangle_t \leq T < \infty$ für jedes $T > 0$, also sind die folgenden Integrale wohldefiniert, und

$$W_t := \int_0^t \mathbb{1}_{\{f_s > 0\}} f_s^{-1/2} dM_s + \int_0^t \mathbb{1}_{\{f_s = 0\}} d\tilde{W}_s$$

ist als Summe stetiger lokaler Martingale selber eines. Nach Satz 25.22 ist

$$\begin{aligned} \langle W \rangle_t &= \int_0^t \mathbb{1}_{\{f_s > 0\}} f_s^{-1} d\langle M \rangle_s + \int_0^t \mathbb{1}_{\{f_s = 0\}} ds \\ &= \int_0^t \mathbb{1}_{\{f_s > 0\}} f_s^{-1} f_s ds + \int_0^t \mathbb{1}_{\{f_s = 0\}} ds \\ &= t. \end{aligned}$$

Nach Satz 25.28 ist W damit als Brown'sche Bewegung erkannt. Andererseits ist

$$\begin{aligned} \int_0^t f_s^{1/2} dW_s &= \int_0^t \mathbb{1}_{\{f_s > 0\}} f_s^{1/2} f_s^{-1/2} dM_s + \int_0^t \mathbb{1}_{\{f_s = 0\}} f_s^{1/2} d\tilde{W}_s \\ &= \int_0^t \mathbb{1}_{\{f_s > 0\}} dM_s. \end{aligned}$$

Nun ist aber $M_t - \int_0^t \mathbb{1}_{\{f_s > 0\}} dM_s = \int_0^t \mathbb{1}_{\{f_s = 0\}} dM_s$ ein stetiges lokales Martingal mit quadratischer Variation $\int_0^t \mathbb{1}_{\{f_s = 0\}} d\langle M \rangle_s = 0$, also fast sicher gleich Null. Also ist $M_t = \int_0^t f_s^{1/2} dW_s$, wie gewünscht. \square

Wir kommen nun zu einer mehrdimensionalen Verallgemeinerung der (pfadweisen) Itô-Formel. Sei hierzu C_{qV}^d der Raum der stetigen Abbildungen $X : [0, \infty) \rightarrow \mathbb{R}^d$, $t \mapsto X_t = (X_t^1, \dots, X_t^d)$, sodass für $k, l = 1, \dots, d$ die quadratische Kovariation (siehe Definition 21.58) $\langle X^k, X^l \rangle$ existiert und stetig ist. Ferner sei $C^2(\mathbb{R}^d)$ der Raum der zweimal stetig differenzierbaren Funktionen F auf \mathbb{R}^d mit partiellen Ableitungen $\partial_k F$ und $\partial_k \partial_l F$, $k, l = 1, \dots, d$. Mit ∇ bezeichnen wir den Gradienten und mit $\Delta = (\partial_1^2 + \dots + \partial_d^2)$ den Laplace-Operator.

Satz 25.30 (Mehrdimensionale pfadweise Itô-Formel). Sei $X \in \mathcal{C}_{\text{qV}}^d$ und $F \in C^2(\mathbb{R}^d)$. Dann gilt

$$F(X_T) - F(X_0) = \int_0^T \nabla F dX_s + \frac{1}{2} \sum_{k,l=1}^d \int_0^T \partial_k \partial_l F(X_s) d\langle X^k, X^l \rangle_s.$$

Dabei ist

$$\int_0^T \nabla F(X_s) dX_s := \lim_{n \rightarrow \infty} \sum_{t \in \mathcal{P}_T^n} \sum_{k=1}^d \partial_k F(X_t) (X_{t'}^k - X_t^k).$$

Beweis. Das geht wie im eindimensionalen Fall. Die Details verbleiben zur Übung. \square

Bemerkung 25.31. Falls jedes der Integrale $\int_0^T \partial_k F(X_s) dX_s^k$ existiert, so gilt

$$\int_0^T \nabla F(X_s) dX_s = \sum_{k=1}^d \int_0^T \partial_k F(X_s) dX_s^k.$$

Die Existenz der einzelnen Integrale folgt jedoch nicht aus der Existenz des Integrals $\int_0^T \nabla F(X_s) dX_s$. \diamond

Korollar 25.32 (Produktregel). Sind $X, Y, X - Y, X + Y \in \mathcal{C}_{\text{qV}}$, so gilt

$$X_T Y_T = X_0 Y_0 + \int_0^T Y_s dX_s + \int_0^T X_s dY_s + \langle X, Y \rangle_T \quad \text{für alle } T \geq 0,$$

falls die beiden Integrale existieren. Speziell gilt die Produktregel, wenn X und Y stetige lokale Martingale sind.

Beweis. Nach Voraussetzung (und der Polarisationsformel) existiert $\langle X, Y \rangle$. Nach Satz 25.30 mit $F(x, y) = xy$ folgt die Aussage.

Für stetige lokale Martingale existieren die Integrale (pfadweise), falls eine geeignete Zerlegungsfolge \mathcal{P} gewählt wird (Übung 25.2.1). \square

Sei nun $Y = M + A$ eine d -dimensionale verallgemeinerte Diffusion, also

$$M_t^k = \sum_{l=1}^d \int_0^t \sigma_s^{k,l} dW_s^l \quad \text{und} \quad A_t^k = \int_0^t b_s^k ds \quad \text{für } t \geq 0, k = 1, \dots, d.$$

Dabei ist $W = (W^1, \dots, W^d)$ eine d -dimensionale Brown'sche Bewegung und $\sigma^{k,l}$ (beziehungsweise b^k) sind progressiv messbare, lokal quadratisch integrierbare (beziehungsweise lokal integrierbare) stochastische Prozesse für $k, l = 1, \dots, d$. Wegen $\langle W^k, W^l \rangle_t = t \cdot \mathbb{1}_{\{k=l\}}$ ist $\langle Y^k, Y^l \rangle_t = \langle M^k, M^l \rangle_t = \int_0^t a_s^{k,l} ds$, wobei

$$a_s^{k,l} := \sum_{i=1}^d \sigma_s^{k,i} \sigma_s^{i,l}$$

die Kovarianzmatrix der Diffusion M ist. Speziell ist $M \in \mathcal{C}_{\text{qV}}^d$ fast sicher. Man beachte, dass die einzelnen Integrale $\int_0^t \sigma_s^{k,l} \partial_k F(Y_s) dW_s^l$ in (25.17) für eine geeignete Zerlegungsfolge \mathcal{P} sogar im pfadweisen Sinne existieren (Übung 25.2.1). Wir erhalten daher als Korollar zur mehrdimensionalen pfadweisen Itô-Formel (Satz 25.30 und Bemerkung 25.31):

Satz 25.33 (Mehrdimensionale Itô-Formel). Sei Y wie oben und $F \in C^2(\mathbb{R}^d)$. Dann gilt

$$\begin{aligned} F(Y_T) - F(Y_0) &= \int_0^T \nabla F(Y_s) dY_s + \frac{1}{2} \sum_{k,l=1}^d \int_0^T \partial_k \partial_l F(Y_s) d\langle M^k, M^l \rangle_s \\ &= \sum_{k,l=1}^d \int_0^t \sigma_s^{k,l} \partial_k F(Y_s) dW_s^l + \sum_{k=1}^d \int_0^t b_s^k \partial_k F(Y_s) ds \quad (25.17) \\ &\quad + \frac{1}{2} \sum_{k,l=1}^d \int_0^t a_s^{k,l} \partial_k \partial_l F(Y_s) ds. \end{aligned}$$

Speziell gilt für die Brown'sche Bewegung

$$F(W_t) - F(W_0) = \sum_{k=1}^d \int_0^t \partial_k F(W_s) dW_s^k + \frac{1}{2} \int_0^t \Delta F(W_s) ds. \quad (25.18)$$

Korollar 25.34. Der Prozess $(F(W_t))_{t \geq 0}$ ist genau dann ein stetiges lokales Martingal, wenn F harmonisch ist (also $\Delta F \equiv 0$ gilt).

Beweis. Ist F harmonisch, so ist $F(W_t) = F(W_0) + \sum_{k=1}^d \int_0^t \partial_k F(W_s) dW_s^k$ als Summe von Itô-Integralen ein stetiges lokales Martingal.

Ist andererseits F ein stetiges lokales Martingal, so ist auch $\int_0^t \Delta F(W_s) ds$ als Differenz von stetigen lokalen Martingalen ein stetiges lokales Martingal. Da $t \mapsto \int_0^t \Delta F(W_s) ds$ von endlicher Variation ist, ist $\int_0^t \Delta F(W_s) ds = 0$ für alle $t \geq 0$ fast sicher (nach Korollar 21.72). Also ist $\Delta F \equiv 0$. \square

Korollar 25.35 (Zeitabhängige Itô-Formel). Ist $F \in C^{2,1}(\mathbb{R}^d \times \mathbb{R})$, so gilt

$$\begin{aligned} F(W_T, T) - F(W_0, 0) &= \sum_{k=1}^d \int_0^T \partial_k F(W_s, s) dW_s^k + \int_0^T \left(\partial_{d+1} + \frac{1}{2} (\partial_1^2 + \dots + \partial_d^2) \right) F(W_s, s) ds. \end{aligned}$$

Beweis. Wende Satz 25.33 an auf $Y = (W_t^1, \dots, W_t^d, t)_{t \geq 0}$. \square

Übung 25.3.1 (Satz von Fubini für Itô-Integrale). Sei $X \in \mathcal{C}_{\text{qV}}$ und sei $g : [0, \infty)^2 \rightarrow \mathbb{R}$ stetig und im Inneren nach der zweiten Koordinate stetig differenzierbar mit Ableitung $\partial_2 g$. Man zeige mit Hilfe der Produktregel (Korollar 25.32)

$$\int_0^s \left(\int_0^t g(u, v) du \right) dX_v = \int_0^t \left(\int_0^s g(u, v) dX_v \right) du$$

und

$$\int_0^s \left(\int_0^v g(u, v) du \right) dX_v = \int_0^s \left(\int_u^s g(u, v) dX_v \right) du. \quad \clubsuit$$

Übung 25.3.2 (Stratonovich-Integral). Sei \mathcal{P} eine zulässige Zerlegungsfolge, $X \in \mathcal{C}_{\text{qV}}^{\mathcal{P}}$ und $f \in C^1(\mathbb{R})$ mit Stammfunktion F . Man zeige: Für jedes $t \geq 0$ ist das **Stratonovich-Integral**

$$\int_0^T f(X_t) \circ dX_t := \lim_{n \rightarrow \infty} \sum_{t \in \mathcal{P}_T^n} f\left(\frac{X_{t'} + X_t}{2}\right) (X_{t'} - X_t)$$

wohldefiniert, und es gilt die klassische Substitutionsregel

$$F(X_T) - F(X_0) = \int_0^T F'(X_t) \circ dX_t.$$

Man zeige, dass im Gegensatz zum Itô-Integral das Stratonovich-Integral bezüglich eines stetigen lokalen Martingals im Allgemeinen kein lokales Martingal ist. \clubsuit

25.4 Dirichlet-Problem und Brown'sche Bewegung

Ähnlich wie für diskrete Markovketten (vergleiche Kapitel 19.1) lässt sich die Lösung des Dirichlet-Problems in einem Gebiet $G \subset \mathbb{R}^d$ durch eine am Rande von G gestoppte d -dimensionale Brown'sche Bewegung beschreiben.

Sei im Folgenden $G \subset \mathbb{R}^d$ eine offene, beschränkte Menge.

Definition 25.36 (Dirichlet-Problem). Sei $f : \partial G \rightarrow \mathbb{R}$ stetig. Eine Funktion $u : \overline{G} \rightarrow \mathbb{R}$ heißt Lösung des Dirichlet-Problems auf G mit Randwert f , falls u stetig ist und in G zweimal stetig differenzierbar sowie

$$\begin{aligned} \Delta u(x) &= 0 && \text{für } x \in G, \\ u(x) &= f(x) && \text{für } x \in \partial G. \end{aligned} \tag{25.19}$$

Für hinreichend glatte Gebiete existiert stets eine Lösung des Dirichlet-Problems (siehe etwa [87, Kapitel 4.4]). Gibt es eine Lösung, so ist sie stets eindeutig (wie aus Satz 25.38 folgt).

Sei im Folgenden $W = (W^1, \dots, W^d)$ eine d -dimensionale Brown'sche Bewegung bezüglich der Filtration \mathbb{F} , die den üblichen Bedingungen genügt. Wir schreiben \mathbf{P}_x und \mathbf{E}_x für Wahrscheinlichkeiten und Erwartungswerte, wenn W in $W_0 = x = (x^1, \dots, x^d) \in \mathbb{R}^d$ gestartet wird. Ist $A \subset \mathbb{R}^d$ offen, so ist

$$\tau_{A^c} := \inf \{t > 0 : W_t \in A^c\}$$

eine \mathbb{F} -Stopzeit (siehe Übung 21.4.4). Da G beschränkt ist, ist $G \subset (-a, a) \times \mathbb{R}^{d-1}$ für gewisses $a > 0$. Also ist $\tau_{G^c} \leq \tau_{((-a, a) \times \mathbb{R}^{d-1})^c}$. Nach Übung 21.2.4 (angewandt auf W^1) ist für $x \in G$

$$\mathbf{E}_x[\tau_{G^c}] \leq \mathbf{E}_x[\tau_{((-a, a) \times \mathbb{R}^{d-1})^c}] = (a - x^1)(a + x^1) < \infty. \quad (25.20)$$

Speziell ist $\tau_{G^c} < \infty$ \mathbf{P}_x -fast sicher, also ist $W_{\tau_{G^c}}$ eine \mathbf{P}_x -fast sicher wohldefinierte Zufallsvariable mit Werten in ∂G .

Definition 25.37. Für $x \in G$ bezeichnen wir mit

$$\mu_{x,G} = \mathbf{P}_x \circ W_{\tau_{G^c}}^{-1}$$

das **harmonische Maß** auf ∂G .

Satz 25.38. Ist u eine Lösung des Dirichlet-Problems auf G mit Randwert f , so ist

$$u(x) = \mathbf{E}_x[f(W_{\tau_{G^c}})] = \int_{\partial G} f(y) \mu_{x,G}(dy) \quad \text{für } x \in G. \quad (25.21)$$

Insbesondere ist die Lösung des Dirichlet-Problems stets eindeutig.

Beweis. Sei $G_1 \subset G_2 \subset \dots$ eine Folge offener Mengen mit $x \in G_1$, $G_n \uparrow G$ und $\overline{G}_n \subset G$ für jedes $n \in \mathbb{N}$. Speziell ist also jedes \overline{G}_n kompakt und damit ∇u auf \overline{G}_n beschränkt. Wir schreiben kurz $\tau := \tau_{G^c}$ und $\tau_n := \tau_{G_n^c}$.

Da u harmonisch ist (das heißt, $\Delta u = 0$), ist nach der Itô-Formel für $t < \tau$

$$u(W_t) = u(W_0) + \int_0^t \nabla u(W_s) dW_s = u(W_0) + \sum_{k=1}^d \int_0^t \partial_k u(W_s) dW_s^k. \quad (25.22)$$

Speziell ist $M := (u(W_t))_{t \in [0, \tau)}$ ein lokales Martingal bis τ (aber im Allgemeinen kein Martingal). Für $t < \tau_n$ ist

$$(\partial_k u(W_s))^2 \leq C_n := \sup_{y \in \overline{G}_n} \|\nabla u(y)\|_2^2 < \infty \quad \text{für jedes } k = 1, \dots, d.$$

Also ist

$$\mathbf{E} \left[\int_0^{\tau_n} (\partial_k u(W_s))^2 ds \right] \leq C_n \mathbf{E}_x[\tau_n] \leq C_n \mathbf{E}[\tau] < \infty$$

nach (25.20). Nach Satz 25.17(ii) ist daher der gestoppte Prozess M^{τ_n} ein Martingal für jedes $n \in \mathbb{N}$ und damit

$$\mathbf{E}_x[u(W_{\tau_n})] = \mathbf{E}_x[M_{\tau_n}] = \mathbf{E}_x[M_0] = u(x). \quad (25.23)$$

Da W stetig ist und $\tau_n \uparrow \tau$ gilt, gilt $W_{\tau_n} \xrightarrow{n \rightarrow \infty} W_\tau \in \partial G$, aufgrund der Stetigkeit von u also auch

$$u(W_{\tau_n}) \xrightarrow{n \rightarrow \infty} u(W_\tau) = f(W_\tau). \quad (25.24)$$

Da u stetig ist und \overline{G} kompakt, ist u beschränkt. Nach dem Satz über majorisierte Konvergenz folgt aus (25.24) auch die Konvergenz der Erwartungswerte, also (zusammen mit (25.23))

$$u(x) = \lim_{n \rightarrow \infty} \mathbf{E}_x[u(W_{\tau_n})] = \mathbf{E}_x[f(W_\tau)]. \quad \square$$

Übung 25.4.1. Sei $G = \mathbb{R} \times (0, \infty)$ die offene, obere Halbebene des \mathbb{R}^2 und $x = (x_1, x_2) \in G$. Man zeige, dass $\tau_{G^c} < \infty$ fast sicher gilt und dass das harmonische Maß $\mu_{x,G}$ auf $\mathbb{R} \cong \partial G$ die um x_1 verschobene Cauchy-Verteilung mit Größenparameter x_2 ist: $\mu_{x,G} = \delta_{x_1} * \text{Cau}_{x_2}$. ♣

Übung 25.4.2. Sei $d \geq 3$ und $G = \mathbb{R}^{d-1} \times (0, \infty)$ ein offener Halbraum des \mathbb{R}^d sowie $x = (x_1, \dots, x_d) \in G$. Man zeige, dass $\tau_{G^c} < \infty$ fast sicher gilt und dass das harmonische Maß $\mu_{x,G}$ auf $\mathbb{R}^{d-1} \cong \partial G$ die folgende Dichte hat:

$$\frac{\mu_{x,G}(dy)}{dy} = \frac{\Gamma(d/2)}{\pi^{d/2}} \frac{x_d}{\sqrt{(x_1 - y_1)^2 + \dots + (x_{d-1} - y_{d-1})^2 + y_d^2}}. \quad \clubsuit$$

Übung 25.4.3. Sei $r > 0$ und $B_r(0) \subset \mathbb{R}^d$ die offene Kugel um den Ursprung mit Radius r . Man bestimme für $x \in B_r(0)$ das harmonische Maß $\mu_{x,B_r(0)}$. ♣

25.5 Rekurrenz und Transienz der Brown'schen Bewegung

Die symmetrische einfache Irrfahrt $(X_n)_{n \in \mathbb{N}}$ auf \mathbb{Z}^d ist nach dem Satz von Pólya (Satz 17.39) genau dann rekurrent (besucht also jeden Punkt unendlich oft), wenn $d \leq 2$ ist. Ist $d > 2$, so ist die Irrfahrt transient und verlässt jede endliche Menge $A \subset \mathbb{Z}^d$ schließlich. Wir können dieses Verhalten beschreiben durch

$$\liminf_{n \rightarrow \infty} \|X_n\| = 0 \text{ f.s.} \iff d \leq 2$$

und

$$\lim_{n \rightarrow \infty} \|X_n\| = \infty \text{ f.s.} \iff d > 2.$$

Hauptergebnis dieses Abschnitts ist es, dass eine ähnliche Dichotomie auch für die Brown'sche Bewegung gilt.

Satz 25.39. Sei $W = (W^1, \dots, W^d)$ eine d -dimensionale Brown'sche Bewegung.

(i) Ist $d \leq 2$, so ist W rekurrent in dem Sinne, dass

$$\liminf_{t \rightarrow \infty} \|W_t - y\| = 0 \text{ f.s. für jedes } y \in \mathbb{R}^d.$$

Insbesondere liegt der Pfad $\{W_t : t \geq 0\}$ dicht in \mathbb{R}^d fast sicher.

(ii) Ist $d > 2$, so ist W transient in dem Sinne, dass

$$\lim_{t \rightarrow \infty} \|W_t\| = \infty \text{ f.s.,}$$

und für jedes $y \in \mathbb{R}^d \setminus \{0\}$ ist $\inf\{\|W_t - y\| : t \geq 0\} > 0$ fast sicher.

Die Grundidee für den Beweis des Satzes besteht darin, mit Hilfe von geeigneten Dirichletproblemen und dem Ergebnis von Abschnitt 25.4 die Wahrscheinlichkeiten dafür auszurechnen, dass W gewisse Kugeln

$$B_R(x) := \{y \in \mathbb{R}^d : \|x - y\| < R\}$$

trifft. Sei $0 < r < R < \infty$ und $G_{r,R}$ der Anulus

$$G_{r,R} := B_R(0) \setminus \overline{B}_r(0) = \{x \in \mathbb{R}^d : r < \|x\| < R\}.$$

Zur Erinnerung: Für abgeschlossenes $A \subset \mathbb{R}^d$ schreiben wir $\tau_A = \inf\{t > 0 : W_t \in A\}$ für die Stoppzeit des ersten Eintretens in A . Wir schreiben

$$\tau_s := \inf\{t > 0 : \|W_t\| = s\} \quad \text{und} \quad \tau_{r,R} = \inf\{t > 0 : W_t \notin G_{r,R}\}.$$

Bei Start in $W_0 \in G_{r,R}$ ist offenbar $\tau_{r,R} = \tau_r \wedge \tau_R$. Auf dem Rand von $G_{r,R}$ definieren wir die Funktion f durch

$$f(x) = \begin{cases} 1, & \text{falls } \|x\| = r, \\ 0, & \text{falls } \|x\| = R. \end{cases} \quad (25.25)$$

Sei $u_{r,R} : \overline{G}_{r,R} \rightarrow \mathbb{R}$ definiert durch

$$u_{r,R}(x) = \frac{V(\|x\|) - V(R)}{V(r) - V(R)},$$

wobei $V : (0, \infty) \rightarrow \mathbb{R}$ die Newton'sche Potentialfunktion ist

$$V(s) = V_d(s) = \begin{cases} s, & \text{falls } d = 1, \\ \log(s), & \text{falls } d = 2, \\ -s^{2-d}, & \text{falls } d > 2. \end{cases} \quad (25.26)$$

Man prüft leicht nach, dass $\varphi : \mathbb{R}^d \setminus \{0\} \rightarrow \mathbb{R}$, $x \mapsto V_d(\|x\|)$ harmonisch ist (also $\Delta \varphi \equiv 0$ erfüllt). Also ist $u_{r,R}$ die Lösung des Dirichlet-Problems auf $G_{r,R}$ mit Randwert f . Nach Satz 25.38 ist für $x \in G_{r,R}$

$$\mathbf{P}_x[\tau_{r,R} = \tau_r] = \mathbf{P}_x[\|W_{\tau_{r,R}}\| = r] = \mathbf{E}_x[f(W_{\tau_{r,R}})] = u_{r,R}(x). \quad (25.27)$$

Satz 25.40. Für $r > 0$ und $x, y \in \mathbb{R}^d$ mit $\|x - y\| > r$ gilt

$$\mathbf{P}_x[W_t \in B_r(y) \text{ für ein } t > 0] = \begin{cases} 1, & \text{falls } d \leq 2, \\ \left(\frac{\|x-y\|}{r}\right)^{2-d}, & \text{falls } d > 2. \end{cases}$$

Beweis. Ohne Einschränkung sei $y = 0$. Dann ist

$$\begin{aligned} \mathbf{P}_x[\tau_r < \infty] &= \lim_{R \rightarrow \infty} \mathbf{P}_x[\tau_{r,R} = \tau_r] = \lim_{R \rightarrow \infty} \frac{V(\|x\|) - V(R)}{V(r) - V(R)} \\ &= \begin{cases} 1, & \text{falls } d = 2, \\ \frac{V_d(\|x\|)}{V_d(r)}, & \text{falls } d > 2, \end{cases} \end{aligned}$$

denn $\lim_{R \rightarrow \infty} V_d(R) = \infty$, falls $d \leq 2$ und $= 0$, falls $d > 2$. \square

Beweis (von Satz 25.39). Unter Verwendung der starken Markoveigenschaft der Brown'schen Bewegung erhalten wir für $r > 0$

$$\begin{aligned} \mathbf{P}_x\left[\liminf_{t \rightarrow \infty} \|W_t\| < r\right] &= \mathbf{P}_x\left[\bigcup_{s \in (0,r)} \bigcap_{R > \|x\|} \{\|W_t\| < r \text{ für ein } t > \tau_R\}\right] \\ &= \sup_{s \in (0,r)} \inf_{R > \|x\|} \mathbf{P}_x[\|W_t\| \leq s \text{ für ein } t > \tau_R] \\ &= \sup_{s \in (0,r)} \inf_{R > \|x\|} \mathbf{P}_x[\mathbf{P}_{W_{\tau_R}}[\tau_s < \infty]]. \end{aligned}$$

Nach Satz 25.40 ist aber (wegen $\|W_{\tau_R}\| = R$ für $R > \|x\|$)

$$\mathbf{P}_{W_{\tau_R}}[\tau_s < \infty] = \begin{cases} 1, & \text{falls } d \leq 2, \\ (s/R)^{d-2}, & \text{falls } d > 2. \end{cases}$$

Also ist

$$\mathbf{P}\left[\liminf_{t \rightarrow \infty} \|W_t\| < r\right] = \begin{cases} 1, & \text{falls } d \leq 2, \\ 0, & \text{falls } d > 2. \end{cases}$$

Hieraus folgt aber die Aussage des Satzes. \square

Definition 25.41 (Polare Menge). Eine Menge $A \subset \mathbb{R}^d$ heißt **polar**, falls

$$\mathbf{P}_x[W_t \notin A \text{ für alle } t > 0] = 1 \quad \text{für alle } x \in \mathbb{R}^d.$$

Satz 25.42. Ist $d = 1$, so ist nur die leere Menge polar. Ist $d \geq 2$, so ist $\{y\}$ polar für jedes $y \in \mathbb{R}^d$.

Beweis. Für $d = 1$ ist die Aussage klar, wegen

$$\limsup_{t \rightarrow \infty} W_t = \infty \quad \text{und} \quad \liminf_{t \rightarrow \infty} W_t = -\infty \quad \text{f.s.}$$

Aufgrund der Stetigkeit von W wird also jeder Punkt $y \in \mathbb{R}$ immer wieder getroffen.

Sei nun $d \geq 2$. Ohne Einschränkung sei $y = 0$. Ist $x \neq 0$, so ist

$$\begin{aligned} \mathbf{P}_x[\tau_{\{0\}} < \infty] &= \lim_{R \rightarrow \infty} \mathbf{P}_x[\tau_{\{0\}} < \tau_R] \\ &= \lim_{R \rightarrow \infty} \inf_{r > 0} \mathbf{P}_x[\tau_{r,R} = \tau_r] \\ &= \lim_{R \rightarrow \infty} \inf_{r > 0} u_{r,R}(x) = 0, \end{aligned} \tag{25.28}$$

weil $V_d(r) \xrightarrow{r \rightarrow 0} -\infty$, falls $d \geq 2$.

Ist hingegen $x = 0$, so gilt wegen der starken Markoveigenschaft der Brown'schen Bewegung (und weil $\mathbf{P}_0[W_t = 0] = 0$ ist für alle $t > 0$)

$$\begin{aligned} \mathbf{P}_0[\tau_{\{0\}} < \infty] &= \sup_{t > 0} \mathbf{P}_0[W_s = 0 \text{ für ein } s \geq t] \\ &= \sup_{t > 0} \mathbf{P}_0[\mathbf{P}_{W_t}[\tau_{\{0\}} < \infty]] = 0, \end{aligned}$$

wobei wir im letzten Schritt (25.28) ausgenutzt haben. \square

26 Stochastische Differentialgleichungen

Stochastische Differentialgleichungen beschreiben die zeitliche Entwicklung von gewissen stetigen Markovprozessen mit Werten in \mathbb{R}^n . Im Gegensatz zu klassischen Differentialgleichungen ist nicht nur die Ableitung einer Funktion angegeben, sondern zudem ein Term, der zufällige Fluktuationen beschreibt, die als Itô-Integral bezüglich einer Brown'schen Bewegung kodiert werden. Je nach dem, ob man die konkrete Brown'sche Bewegung als treibende Kraft des Rauschens ernst nimmt oder nicht, spricht man von starken oder schwachen Lösungen. Wir entwickeln im ersten Abschnitt die Theorie der starken Lösungen unter Lipschitz-Bedingungen an die Koeffizienten. Im zweiten Abschnitt lernen wir das (lokale) Martingalproblem als Methode zur Etablierung schwacher Lösungen kennen. Im dritten Abschnitt stellen wir die Methode der Dualität zur Sicherung der Eindeutigkeit von Lösungen an Beispielen vor.

Da die Theorie der stochastischen Differentialgleichungen ein sehr weites Feld ist und die Dinge sehr schnell sehr technisch werden, bringen wir nur kurзорisch ein paar der wichtigsten Ergebnisse, zum Teil ohne Beweis, um sie dann an Beispielen zu illustrieren.

26.1 Starke Lösungen

Wir betrachten eine stochastische Differentialgleichung (SDGL) von dem Typ

$$\begin{aligned} X_0 &= \xi, \\ dX_t &= \sigma(t, X_t) dW_t + b(t, X_t) dt. \end{aligned} \tag{26.1}$$

Dabei ist $W = (W^1, \dots, W^m)$ eine m -dimensionale Brown'sche Bewegung, ξ eine von W unabhängige \mathbb{R}^n -wertige Zufallsvariable mit Verteilung μ , $\sigma(t, x) = (\sigma_{ij}(t, x))_{\substack{i=1, \dots, n \\ j=1, \dots, m}}$ eine reelle $n \times m$ Matrix sowie $b(t, x) = (b_i(t, x))_{i=1, \dots, n}$ ein n -dimensionaler Vektor. Die Abbildungen $(t, x) \mapsto \sigma_{ij}(t, x)$ und $(t, x) \mapsto b_i(t, x)$ seien messbar.

Unter einer Lösung X von (26.1) wollen wir natürlich einen stetigen, adaptierten stochastischen Prozess X mit Werten in \mathbb{R}^n verstehen, der die folgende Integralgleichung erfüllt

$$X_t = \xi + \int_0^t \sigma(s, X_s) dW_s + \int_0^t b(s, X_s) ds \quad \mathbf{P}\text{-f.s. für alle } t \geq 0. \quad (26.2)$$

Koordinatenweise ausgeschrieben heißt dies

$$X_t^i = \xi^i + \sum_{j=1}^m \int_0^t \sigma_{ij}(s, X_s) dW_s^j + \int_0^t b_i(s, X_s) ds \quad \text{für alle } i = 1, \dots, n.$$

Nun ergibt sich folgendes Problem: An welche Filtration \mathbb{F} soll X adaptiert sein? Soll \mathbb{F} die Filtration sein, die von ξ und W erzeugt ist, oder darf \mathbb{F} eine größere Filtration sein? Aus der Theorie der gewöhnlichen Differentialgleichungen ist bekannt, dass es, je nach Differentialgleichung, Lösungen geben kann, die aber nicht eindeutig sind (beispielsweise für $f' = |f|^{1/3}$). Wenn \mathbb{F} größer als die von W erzeugte Filtration ist, können wir weitere Zufallsvariablen definieren, die unter mehreren Lösungen eine aussuchen. Wir haben also mehr Möglichkeiten, eine Lösung anzugeben als wenn $\mathbb{F} = \sigma(W)$ ist. In der Tat wird sich herausstellen, dass man in manchen Fällen überhaupt erst eine Lösung einer SDGL angeben kann, wenn man eine größere Filtration zulässt. Grob gesprochen nennen wir X eine starke Lösung von (26.1), wenn (26.2) gilt und X an $\mathbb{F} = \sigma(W)$ adaptiert ist, hingegen eine schwache Lösung, wenn X an eine größere Filtration \mathbb{F} adaptiert ist, bezüglich der W aber immer noch ein Martingal ist. Schwache Lösungen behandeln wir in Abschnitt 26.2.

Definition 26.1 (Starke Lösung). Wir sagen, dass die stochastische Differentialgleichung (SDGL) (26.1) eine **starke Lösung** X hat, falls es eine Abbildung $F : \mathbb{R}^n \times C([0, \infty); \mathbb{R}^m) \rightarrow C([0, \infty); \mathbb{R}^n)$ gibt mit den Eigenschaften

- (i) $(x, w) \mapsto F(x, w)$ ist für jedes $t \geq 0$ messbar bezüglich $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{G}_t^m - \mathcal{G}_t^n$, wobei (für $k = m$ oder $k = n$) $\mathcal{G}_t^k := \sigma(\pi_s : s \in [0, t])$ die von den Koordinatenabbildungen $\pi_s : C([0, \infty); \mathbb{R}^k) \rightarrow \mathbb{R}$, $w \mapsto w(s)$ erzeugte σ -Algebra ist.
- (ii) Der Prozess $X = F(\xi, W)$ erfüllt (26.2).

Bedingung (i) besagt, dass der Pfad $(X_s)_{s \in [0, t]}$ nur von ξ und $(W_s)_{s \in [0, t]}$ abhängt und sonst von keinen Informationen. Insbesondere ist X an $\mathcal{F}_t = \sigma(\xi, W_s : s \in [0, t])$ adaptiert und progressiv messbar, sodass das Itô-Integral in (26.2) wohldefiniert ist, falls σ und b nicht zu stark wachsen für große x .

Bemerkung 26.2. Offenbar ist eine starke Lösung einer SDGL stets eine verallgemeinerte n -dimensionale Diffusion. Sind die Koeffizienten σ und b unabhängig von t , so ist die Lösung eine n -dimensionale Diffusion. ◇

Bemerkung 26.3. Sei X eine starke Lösung und F wie in Definition 26.1. Ist W' eine m -dimensionale Brown'sche Bewegung auf einem Raum $(\Omega', \mathcal{F}', \mathcal{P}')$ mit Filtration \mathbb{F}' , und ist ξ' unabhängig von W' und \mathcal{F}'_0 -messbar, so erfüllt $X' = F(\xi', W')$ die Integralgleichung (26.2), ist also eine starke Lösung von (26.1) mit W' statt W .

Die Existenz einer starken Lösung hängt also nicht von der konkreten Realisierung der Brown'schen Bewegung oder der Filtration \mathbb{F} ab. \diamond

Definition 26.4. Wir sagen, dass die SDGL (26.1) eine eindeutige starke Lösung hat, falls es ein F wie in Definition 26.1 gibt, sodass gilt:

(i) Ist W eine m -dimensionale Brown'sche Bewegung auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ mit Filtration \mathbb{F} und ξ eine \mathcal{F}_0 -messbare von W unabhängige Zufallsvariable mit $\mathbf{P} \circ \xi^{-1} = \mu$, dann ist $X := F(\xi, W)$ eine Lösung von (26.2).

(ii) Für jede Lösung (X, W) von (26.2) gilt $X = F(\xi, W)$.

Beispiel 26.5. Seien $m = n = 1$ und $b \in \mathbb{R}$ sowie $\sigma > 0$. Der **Ornstein-Uhlenbeck Prozess**

$$X_t := e^{bt}\xi + \sigma \int_0^t e^{(t-s)b} dW_s, \quad t \geq 0, \quad (26.3)$$

ist eine starke Lösung der SDGL $X_0 = \xi$ und

$$dX_t = \sigma dW_t + b X_t dt.$$

In der Terminologie von Definition 26.1 ist (im Sinne des pfadweisen Itô-Integrals bezüglich w)

$$F(x, w) = \left(t \mapsto e^{bt}x + \int_0^t e^{(t-s)b} dw(s) \right)$$

für alle $w \in \mathcal{C}_{\text{qV}}$ (also mit stetiger quadratischer Variation). Wegen $\mathbf{P}[W \in \mathcal{C}_{\text{qV}}] = 1$, können wir $F(x, w) = 0$ setzen für $w \in C([0, \infty); \mathbb{R}) \setminus \mathcal{C}_{\text{qV}}$.

In der Tat gilt nach dem Satz von Fubini für Itô-Integrale (Übung 25.3.1)

$$\begin{aligned} \xi + \int_0^t \sigma dW_s + \int_0^t b X_s ds \\ = \xi + \sigma W_t + \int_0^t b e^{bs} \xi ds + \int_0^t \sigma b \left(\int_0^s e^{b(s-r)} dW_r \right) ds \\ = \xi + \sigma W_t + (e^{bt} - 1)\xi + \int_0^t \sigma \left(\int_r^t b e^{b(s-r)} ds \right) dW_r \\ = e^{bt}\xi + \int_0^t \left(\sigma + (e^{b(t-r)} - 1)\sigma \right) dW_r \\ = X_t. \end{aligned}$$

Man kann zeigen (siehe Satz 26.8), dass diese Lösung auch (stark) eindeutig ist. \diamond

Beispiel 26.6. Seien $\alpha, \beta \in \mathbb{R}$. Die eindimensionale SDGL $X_0 = \xi$ und

$$dX_t = \alpha X_t dW_t + \beta X_t dt \quad (26.4)$$

hat die starke Lösung

$$X_t = \xi \exp \left(\alpha W_t + \left(\beta - \frac{\alpha^2}{2} \right) t \right).$$

In der Terminologie von Definition 26.1 ist $\sigma(t, x) = \alpha x$, $b(t, x) = \beta x$ und

$$F(x, w) = \left(t \mapsto x \exp \left(\alpha w(t) + \left(\beta - \frac{\alpha^2}{2} \right) t \right) \right)$$

für alle $w \in C([0, \infty); \mathbb{R})$ und $x \in \mathbb{R}$. In der Tat ist nach der zeitabhängigen Itô-Formel (Korollar 25.35)

$$X_t = \xi + \int_0^t \alpha X_s dW_s + \int_0^t \left(\left(\beta - \frac{\alpha^2}{2} \right) + \frac{1}{2} \alpha^2 \right) X_s ds.$$

Auch in diesem Fall gilt starke Eindeutigkeit der Lösung (siehe Satz 26.8). Der Prozess X heißt **geometrische Brown'sche Bewegung** und dient beispielsweise zur Modellierung von Aktienkursen im so genannten **Black-Scholes Modell**. ◇

Wir geben nun ein einfaches Kriterium für die Existenz und Eindeutigkeit starker Lösungen an. Für eine $n \times m$ Matrix A definieren wir die **Hilbert-Schmidt Norm**

$$\|A\| = \sqrt{\text{Spur}(A A^T)} = \sqrt{\sum_{i=1}^n \sum_{j=1}^m A_{i,j}^2}. \quad (26.5)$$

Für $b \in \mathbb{R}^n$ verwenden wir die euklidische Norm $\|b\|$. Da alle Normen auf endlichdimensionalen Vektorräumen äquivalent sind, spielt es keine wesentliche Rolle, welche Norm wir genau benutzen. Allerdings vereinfacht die hier eingeführte Norm die Rechnungen, wie das folgende Lemma zeigt.

Lemma 26.7. Sei $t \mapsto H(t) = (H_{ij}(t))_{i=1, \dots, n, j=1, \dots, m}$ progressiv messbar und $\mathbf{E} \left[\int_0^T H_{ij}^2(t) dt \right] < \infty$ für alle i, j . Dann gilt

$$\mathbf{E} \left[\left\| \int_0^T H(t) dW_t \right\|^2 \right] = \mathbf{E} \left[\int_0^T \|H(t)\|^2 dt \right], \quad (26.6)$$

wobei $\|H\|$ die Hilbert-Schmidt Norm aus (26.5) bezeichnet.

Beweis. Für $i = 1, \dots, n$ ist $I_i(t) := \sum_{j=1}^m \int_0^t H_{ij}(s) dW_s^j$ ein stetiges Martingal mit Variationsprozess $\langle I_i \rangle_t = \int_0^t \sum_{j=1}^m H_{ij}^2(s) ds$. Daher ist

$$\mathbf{E}[(I_i(T))^2] = \mathbf{E} \left[\int_0^T \sum_{j=1}^m H_{ij}^2(s) ds \right].$$

Die linke Seite in (26.6) ist aber gleich

$$\sum_{i=1}^n \mathbf{E}[(I_i(T))^2] = \mathbf{E}\left[\int_0^T \sum_{i=1}^n \sum_{j=1}^m H_{ij}^2(s) ds\right].$$

Die Behauptung folgt nun aus der Definition von $\|H(s)\|^2$. \square

Satz 26.8. Seien b und σ Lipschitz-stetig in der ersten Koordinate. Das heißt, es existiere eine Konstante $K > 0$, sodass für alle $x, x' \in \mathbb{R}^n$ und $t \geq 0$ gilt, dass

$$\|\sigma(x, t) - \sigma(x', t)\| + \|b(x, t) - b(x', t)\| \leq K \|x - x'\|. \quad (26.7)$$

Ferner gelte die Wachstumsbedingung

$$\|\sigma(t, x)\|^2 + \|b(t, x)\|^2 \leq K^2 (1 + \|x\|^2) \quad \text{für alle } x \in \mathbb{R}^n, t \geq 0. \quad (26.8)$$

Dann existiert für jeden Anfangswert $X_0 = x \in \mathbb{R}^n$ eine eindeutige starke Lösung X der SDGL (26.1). Diese Lösung ist ein Markovprozess und im Falle, wo σ und b nicht von t abhängen, ein starker Markovprozess.

Als Hilfsmittel brauchen wir ein Lemma.

Lemma 26.9 (Gronwall). Seien $f, g : [0, T] \rightarrow \mathbb{R}$ integrierbar und $C > 0$ so, dass

$$f(t) \leq g(t) + C \int_0^t f(s) ds \quad \text{für alle } t \in [0, T]. \quad (26.9)$$

Dann ist

$$f(t) \leq g(t) + C \int_0^t e^{C(t-s)} g(s) ds \quad \text{für alle } t \in [0, T].$$

Ist speziell $g(t) \equiv G$ konstant, so ist $f(t) \leq Ge^{Ct}$ für alle $t \in [0, T]$.

Beweis. Seien $F(t) = \int_0^t f(s) ds$ und $h(t) = F(t) e^{-Ct}$. Dann ist nach (26.9)

$$\frac{d}{dt} h(t) = f(t) e^{-Ct} - CF(t) e^{-Ct} \leq g(t) e^{-Ct}.$$

Integration liefert

$$F(t) = e^{Ct} h(t) \leq \int_0^t e^{C(t-s)} g(s) ds.$$

Einsetzen in (26.9) liefert

$$f(t) \leq g(t) + CF(t) \leq g(t) + C \int_0^t g(s) e^{C(t-s)} ds. \quad \square$$

Beweis (von Satz 26.8). Es reicht zu zeigen, dass eine eindeutige starke Lösung bis T für jedes $T < \infty$ existiert.

Eindeutigkeit. Wir zeigen zunächst die Eindeutigkeit der Lösung. Seien X und X' Lösungen von (26.2). Dann ist

$$X_t - X'_t = \int_0^t (b(s, X_s) - b(s, X'_s)) ds + \int_0^t (\sigma(s, X_s) - \sigma(s, X'_s)) dW_s.$$

Daher ist

$$\begin{aligned} \|X_t - X'_t\|^2 &\leq 2 \left\| \int_0^t (b(s, X_s) - b(s, X'_s)) ds \right\|^2 \\ &\quad + 2 \left\| \int_0^t (\sigma(s, X_s) - \sigma(s, X'_s)) dW_s \right\|^2. \end{aligned} \quad (26.10)$$

Für den ersten Summanden in (26.10) verwenden wir die Cauchy-Schwarz'sche Ungleichung, für den zweiten hingegen Lemma 26.7 und erhalten so

$$\begin{aligned} \mathbf{E}[\|X_t - X'_t\|^2] &\leq 2t \int_0^t \mathbf{E}\left[\|b(s, X_s) - b(s, X'_s)\|^2\right] ds \\ &\quad + 2 \int_0^t \mathbf{E}\left[\|\sigma(s, X_s) - \sigma(s, X'_s)\|^2\right] ds. \end{aligned}$$

Schreiben wir $f(t) = \mathbf{E}[\|X_t - X'_t\|^2]$ und $C := 2(T+1)K^2$, so erhalten wir $f(t) \leq C \int_0^t f(s) ds$. Das Lemma von Gronwall (mit $g \equiv 0$) liefert daher $f \equiv 0$.

Existenz. Wir wenden eine Variante des Picard'schen Iterationsverfahrens an. Für jedes $N \in \mathbb{N}_0$ definieren wir iterativ einen Prozess X^N durch $X_t^0 \equiv x$ und

$$X_t^N := x + \int_0^t b(s, X_s^{N-1}) ds + \int_0^t \sigma(s, X_s^{N-1}) dW_s \quad \text{für } N \in \mathbb{N}. \quad (26.11)$$

Wegen der Wachstumsbedingung (26.8) kann man sukzessive zeigen:

$$\begin{aligned} \int_0^T \mathbf{E}[\|X_t^N\|^2] dt &\leq 2(T+1)K^2 \left(T + \int_0^T \mathbf{E}[\|X_t^{N-1}\|^2] dt \right) \\ &\leq (2T(T+1)K^2)^N (1 + \|x\|^2) < \infty, \quad N \in \mathbb{N}. \end{aligned}$$

Also ist in jedem Schritt das Itô-Integral wohldefiniert.

Wir betrachten nun die Differenzen

$$X_t^{N+1} - X_t^N = I_t + J_t,$$

wobei

$$I_t := \int_0^t (\sigma(s, X_s^N) - \sigma(s, X_s^{N-1})) dW_s$$

und

$$J_t := \int_0^t (b(s, X_s^N) - b(s, X_s^{N-1})) ds.$$

Indem wir die Doob'sche L^2 -Ungleichung auf das nichtnegative Submartingal $(\|I_t\|^2)_{t \geq 0}$, Lemma 26.7 sowie (26.7) anwenden, erhalten wir

$$\begin{aligned} \mathbf{E} \left[\sup_{s \leq t} \|I_s\|^2 \right] &\leq 4 \mathbf{E} [\|I_t\|^2] \\ &= 4 \mathbf{E} \left[\int_0^t \|\sigma(s, X_s^N) - \sigma(s, X_s^{N-1})\|^2 ds \right] \\ &\leq 4K^2 \int_0^t \mathbf{E} [\|X_s^N - X_s^{N-1}\|^2] ds. \end{aligned} \quad (26.12)$$

Für J_t bekommen wir mit der Cauchy-Schwarz'schen Ungleichung

$$\|J_t\|^2 \leq t \int_0^t \|b(s, X_s^N) - b(s, X_s^{N-1})\|^2 ds.$$

Also ist

$$\begin{aligned} \mathbf{E} \left[\sup_{s \leq t} \|J_s\|^2 \right] &\leq t \mathbf{E} \left[\int_0^t \|b(s, X_s^N) - b(s, X_s^{N-1})\|^2 ds \right] \\ &\leq tK^2 \int_0^t \mathbf{E} [\|X_s^N - X_s^{N-1}\|^2] ds. \end{aligned} \quad (26.13)$$

Setzen wir

$$\Delta^N(t) := \mathbf{E} \left[\sup_{s \leq t} \|X_s^N - X_s^{N-1}\|^2 \right],$$

so erhalten wir mit $C := 2K^2(4 + T) \vee 2(T + 1)K^2(1 + \|x\|^2)$

$$\Delta^{N+1}(t) \leq C \int_0^t \Delta^N(s) ds \quad \text{für } N \geq 1$$

und

$$\begin{aligned} \Delta^1(t) &\leq 2t \int_0^t \|b(s, x)\|^2 ds + 2 \int_0^t \|\sigma(s, x)\|^2 ds \\ &\leq 2(T + 1)K^2(1 + \|x\|^2) \cdot t \leq C t \end{aligned}$$

nach der Wachstumsvoraussetzung (26.8). Per Induktion folgt $\Delta^N(t) \leq \frac{(Ct)^N}{N!}$. Es folgt mit der Markov'schen Ungleichung

$$\begin{aligned} \sum_{N=1}^{\infty} \mathbf{P} \left[\sup_{s \leq t} \|X_s^N - X_s^{N-1}\|^2 > 2^{-N} \right] &\leq \sum_{N=1}^{\infty} 2^N \Delta^N(t) \\ &\leq \sum_{N=1}^{\infty} \frac{(2Ct)^N}{N!} \leq e^{2Ct} < \infty. \end{aligned}$$

Nach dem Lemma von Borel-Cantelli folgt $\sup_{s \leq t} \|X_s^N - X_s^{N-1}\|^2 \xrightarrow{N \rightarrow \infty} 0$ fast sicher. Mithin ist fast sicher $(X_s^N)_{N \in \mathbb{N}}$ eine Cauchy-Folge im Banachraum $(C([0, T]), \|\cdot\|_\infty)$. Also konvergiert X^N fast sicher gleichmäßig gegen ein X . Gleichmäßige Konvergenz impliziert Konvergenz der Integrale, also ist X eine starke Lösung von (26.2).

Markoveigenschaft. Die starke Markoveigenschaft folgt direkt aus der starken Markoveigenschaft der Brown'schen Bewegung, die die SDGL antreibt. \square

Wichtige Beispiele für diesen Satz haben wir oben schon kennen gelernt. Viele interessante Probleme führen jedoch auf stochastische Differentialgleichungen, bei denen die Koeffizienten nicht Lipschitz-stetig sind. Im eindimensionalen Fall kann man mit speziellen Vergleichsmethoden zeigen, dass es ausreicht, dass σ Hölder-stetig von der Ordnung $\frac{1}{2}$ in der Ortsvariablen ist.

Satz 26.10 (Yamada-Watanabe). *Wir betrachten die eindimensionale Situation $m = n = 1$. Es gebe $K < \infty$ und $\alpha \in [\frac{1}{2}, 1]$, sodass für alle $t \geq 0$ und $x, x' \in \mathbb{R}$ gilt*

$$|b(t, x) - b(t, x')| \leq K|x - x'| \quad \text{und} \quad |\sigma(t, x) - \sigma(t, x')| \leq |x - x'|^\alpha.$$

Dann besitzt die SDGL (26.1) für jedes $X_0 \in \mathbb{R}$ eine eindeutige starke Lösung X , und X ist ein starker Markovprozess.

Beweis. Siehe [184] oder [93, Proposition 5.2.13] und [54, Theorem 5.3.11] für Existenz und Eindeutigkeit der Lösung. Die starke Markov-Eigenschaft folgt aus Satz 26.26. \square

Beispiel 26.11. Betrachte die eindimensionale SDGL

$$dX_t = \sqrt{\gamma X_t^+} dW_t + a(b - X_t^+) dt \quad (26.14)$$

mit Startwert $X_0 = x \geq 0$, wobei $\gamma > 0$ und $a, b \geq 0$ Parameter sind. Die Bedingungen von Satz 26.10 sind mit $\alpha = \frac{1}{2}$ und $K = \sqrt{\gamma} + a$ erfüllt. Die eindeutige starke Lösung X hat offenbar die Eigenschaft, nichtnegativ zu bleiben, wenn $X_0 \geq 0$ ist. (Tatsächlich kann man sogar zeigen dass $X_t > 0$ für alle $t > 0$ gilt, falls $2ab/\gamma \geq 1$, und dass X_t die Null mit Wahrscheinlichkeit 1 beliebig häufig trifft, falls $2ab/\gamma < 1$. Siehe etwa [85, Beispiel IV.8.2, Seite 237]. Vergleiche Beispiel 26.16.)

Dieser Prozess wird je nach Kontext gelegentlich als **Feller'sche Verzweigungs-diffusion mit Immigration** oder als **Cox-Ingersoll-Ross Modell** für die zeitliche Entwicklung von Zinsraten bezeichnet.

Wir berechnen für den Fall $a = b = 0$ mit der Itô-Formel, dass

$$e^{-\lambda X_t} - e^{-\lambda x} - \gamma \frac{\lambda^2}{2} \int_0^t e^{-\lambda X_s} X_s ds = \lambda \int_0^t e^{-\lambda X_s} \sqrt{\gamma X_s} dW_s$$

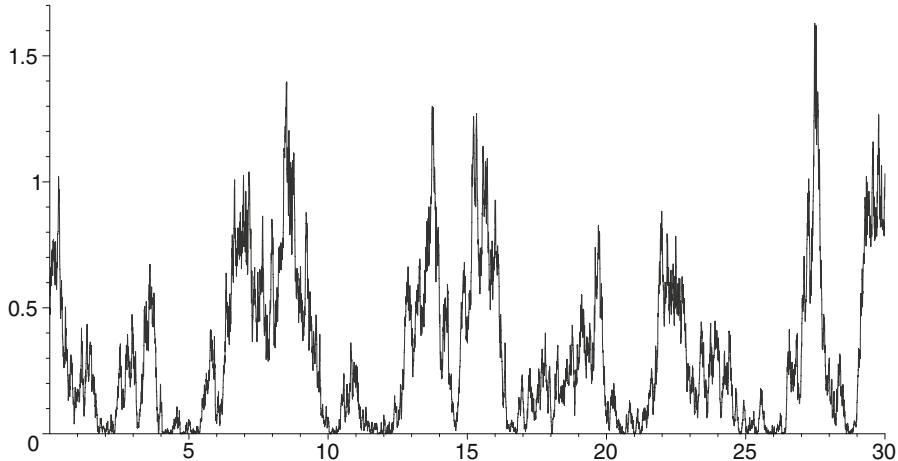


Abb. 26.1. Cox-Ingersoll-Ross Diffusion mit Parametern $\gamma = 1$, $b = 1$ und $a = 0.3$. Der Pfad trifft die Null immer wieder, da $2ab/\gamma = 0.6 < 1$ ist.

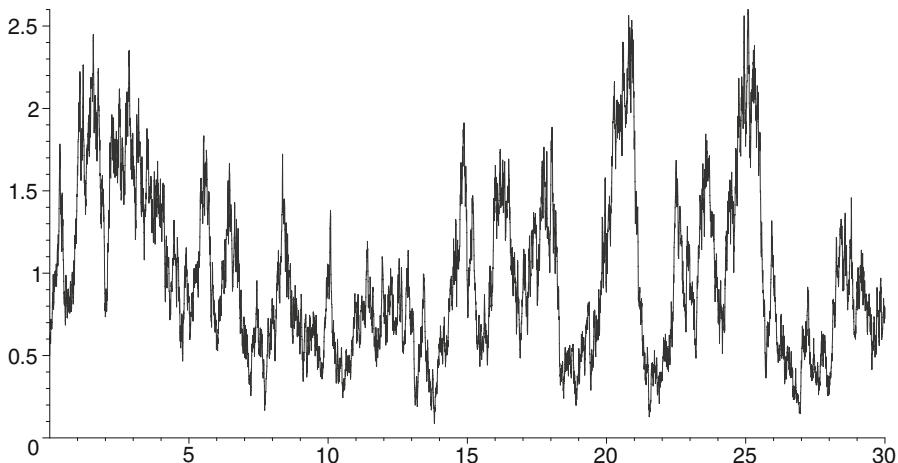


Abb. 26.2. Cox-Ingersoll-Ross Diffusion mit Parametern $\gamma = 1$, $b = 1$ und $a = 2$. Der Pfad trifft die Null nie, da $2ab/\gamma = 4 \geq 1$ ist.

ein Martingal ist. Indem wir Erwartungswerte bilden, erhalten wir für die Laplace-Transformierte $\varphi(t, \lambda, x) = \mathbf{E}_x[e^{-\lambda X_t}]$ die Differentialgleichung

$$\frac{d}{dt}\varphi(t, \lambda, x) = \gamma \frac{\lambda^2}{2} \mathbf{E}[X_t e^{-\lambda X_t}] = -\frac{\gamma \lambda^2}{2} \frac{d}{d\lambda}\varphi(t, \lambda, x).$$

Diese partielle Differentialgleichung hat mit dem Anfangswert $\varphi(0, \lambda, x) = e^{-\lambda x}$ die eindeutige Lösung

$$\varphi(t, \lambda, x) = \exp\left(-\frac{\lambda}{(\gamma/2)\lambda t + 1} x\right).$$

Dies ist aber (für $\gamma = 2$) genau die Laplace-Transformierte der Übergangswahrscheinlichkeiten des Markov-Prozesses, den wir in Satz 21.48 definiert hatten und den wir im Satz von Lindvall (Satz 21.51) als Grenzwert von reskalierten Galton-Watson Verzweigungsprozessen kennen gelernt haben. \diamond

Übung 26.1.1. Seien $a, b \in \mathbb{R}$. Man zeige, dass die stochastische Differentialgleichung

$$dX_t = \frac{b - X_t}{1 - t} dt + dW_t$$

mit Anfangswert $X_0 = a$ eine eindeutige starke Lösung für $t \in [0, 1)$ hat und dass $X_1 := \lim_{t \uparrow 1} X_t = b$ fast sicher gilt. Ferner zeige man, dass der Prozess $Y = (X_t - a - t(b - a))_{t \in [0, 1]}$ durch das Itô-Integral

$$Y_t = (1 - t) \int_0^t (1 - s)^{-1} dW_s, \quad t \in [0, 1),$$

beschrieben wird, also eine Brown'sche Brücke ist (vergleiche Übung 21.5.3). \clubsuit

26.2 Schwache Lösungen und Martingalproblem

Im letzten Abschnitt haben wir starke Lösungen der stochastischen Differentialgleichung

$$dX_t = \sigma(t, X_t) dW_t + b(t, X_t) dt \tag{26.15}$$

kennen gelernt. Diese zeichnen sich dadurch aus, dass jedem Pfad der Brown'schen Bewegung W genau ein Pfad der Lösung X zugeordnet wird. Wir wollen nun zum Begriff der schwachen Lösung kommen, bei der zusätzliche Information (das heißt zusätzlicher Zufall) in die Lösung mit eingehen kann.

Definition 26.12 (Schwache Lösung einer SDGL). Eine schwache Lösung von (26.15) mit Startverteilung $\mu \in \mathcal{M}_1(\mathbb{R}^n)$ ist ein Tripel

$$L = ((X, W), (\Omega, \mathcal{F}, \mathbf{P}), \mathbb{F}),$$

wobei gilt:

- $(\Omega, \mathcal{F}, \mathbf{P})$ ist ein Wahrscheinlichkeitsraum,
- $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ ist eine Filtration auf $(\Omega, \mathcal{F}, \mathbf{P})$, die die üblichen Bedingungen erfüllt,
- W ist eine Brown'sche Bewegung auf $(\Omega, \mathcal{F}, \mathbf{P})$ und bezüglich \mathbb{F} ein Martingal,

- X ist stetig und adaptiert (also progressiv messbar),
- $\mathbf{P} \circ (X_0)^{-1} = \mu$,
- (X, W) ist eine Lösung von

$$X_t = X_0 + \int_0^t \sigma(s, X_s) dW_s + \int_0^t b(s, X_s) ds \quad \mathbf{P}\text{-f.s.} \quad (26.16)$$

Eine schwache Lösung L heißt (schwach) eindeutig, falls für jede weitere Lösung L' mit Startverteilung μ gilt: $\mathbf{P}' \circ (X')^{-1} = \mathbf{P} \circ X^{-1}$.

Bemerkung 26.13. Offenbar ist eine schwache Lösung einer SDGL stets eine verallgemeinerte n -dimensionale Diffusion. Sind die Koeffizienten σ und b unabhängig von t , so ist die Lösung eine n -dimensionale Diffusion. \diamond

Bemerkung 26.14. Offenbar wird durch jede starke Lösung von (26.15) eine schwache Lösung definiert. Die Umkehrung ist falsch, wie wir im folgenden Beispiel sehen werden. \diamond

Beispiel 26.15. Betrachte die SDGL (mit Startwert $X_0 = 0$)

$$dX_t = \text{sign}(X_t) dW_t, \quad (26.17)$$

wobei $\text{sign} = \mathbb{1}_{(0,\infty)} - \mathbb{1}_{(-\infty,0)}$ die Vorzeichenfunktion ist. Es gilt genau dann

$$X_t = X_0 + \int_0^t \text{sign}(X_s) dW_s \quad \text{für alle } t \geq 0, \quad (26.18)$$

wenn

$$W_t = \int_0^t dW_s = \int_0^t \text{sign}(X_s) dX_s \quad \text{für alle } t \geq 0. \quad (26.19)$$

Folgendermaßen gelangen wir zu einer schwachen Lösung von (26.17). Sei X eine Brown'sche Bewegung auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ und $\mathbb{F} = \sigma(X)$. Definieren wir W durch (26.19), dann ist W ein stetiges \mathbb{F} -Martingal mit quadratischer Variation

$$\langle W \rangle_t = \int_0^1 (\text{sign}(X_s))^2 ds = t.$$

Nach der Lévy'schen Charakterisierung (Satz 25.28) ist W damit eine Brown'sche Bewegung. Also ist $((X, W), (\Omega, \mathcal{F}, \mathbf{P}), \mathbb{F})$ eine schwache Lösung von (26.17).

Um zu zeigen, dass es keine starke Lösung gibt, nehmen wir eine beliebige schwache Lösung her und zeigen, dass X nicht an $\sigma(W)$ adaptiert ist. Da X nach (26.18) ein stetiges Martingal mit quadratischer Variation $\langle X \rangle_t = t$ ist, ist X eine Brown'sche Bewegung.

Seien $F_n \in C^2(\mathbb{R})$ konvexe gerade Funktionen mit Ableitungen F'_n und F''_n , sodass

$$\sup_{x \in \mathbb{R}} |F_n(x) - |x|| \xrightarrow{n \rightarrow \infty} 0,$$

$|F'_n(x)| \leq 1$ für alle $x \in \mathbb{R}$ und $F'_n(x) = \text{sign}(x)$ für $|x| > \frac{1}{n}$. Insbesondere gilt

$$\int_0^t (F'_n(X_s) - \text{sign}(X_s))^2 ds \xrightarrow{n \rightarrow \infty} 0 \quad \text{f.s.}$$

und damit

$$\int_0^t F'_n(X_s) dX_s \xrightarrow{n \rightarrow \infty} \int_0^t \text{sign}(X_s) dX_s \quad \text{in } L^2. \quad (26.20)$$

Indem wir gegebenenfalls zu einer Teilfolge übergehen, können wir annehmen, dass in (26.20) fast sichere Konvergenz gilt.

Weil F''_n gerade ist, gilt

$$\begin{aligned} W_t &= \int_0^t \text{sign}(X_s) dX_s = \lim_{n \rightarrow \infty} \int_0^t F'_n(X_s) dX_s \\ &= \lim_{n \rightarrow \infty} \left(F_n(X_t) - F_n(0) - \frac{1}{2} \int_0^t F''_n(X_s) ds \right) \\ &= |X_t| - \lim_{n \rightarrow \infty} \frac{1}{2} \int_0^t F''_n(|X_s|) ds. \end{aligned}$$

Da die rechte Seite nur von $|X_s|$, $s \in [0, t]$ abhängt, ist W an $\mathbb{G} := (\sigma(|X_s| : s \in [0, t]))$ adaptiert. Also ist $\sigma(W) \subset \mathbb{G} \subsetneq \sigma(X)$, und damit ist X nicht an $\sigma(W)$ adaptiert. \diamond

Beispiel 26.16. Sei $n \in \mathbb{N}$ und $B = (B^1, \dots, B^n)$ eine n -dimensionale Brown'sche Bewegung mit Start in $y \in \mathbb{R}^n$. Setze $x := \|y\|^2$, $X_t := \|B_t\|^2 = (B_t^1)^2 + \dots + (B_t^n)^2$ und

$$W_t := \sum_{i=1}^n \int_0^t \frac{1}{\sqrt{X_s}} B_s^i dB_s^i.$$

Dann ist W ein stetiges lokales Martingal mit $\langle W \rangle_t = t$ für jedes $t \geq 0$ und

$$X_t = x + nt + \int_0^t \sqrt{X_s} dW_s.$$

Das heißt, (X, W) ist eine schwache Lösung der SDGL $dX_t = \sqrt{2X_t} dW_t + n dt$. X wird auch n -dimensionaler **Bessel-Prozess** genannt. Nach Satz 25.42 trifft B (und damit X) den Ursprung für ein $t > 0$ genau dann, wenn $n = 1$ ist. Offenbar kann man X auch für nicht-ganzzahlig $n \geq 0$ definieren. Man kann zeigen, dass X genau dann die Null trifft, wenn $n \leq 1$ ist. Vergleiche Beispiel 26.11. \diamond

Für den Zusammenhang von Existenz und Eindeutigkeit von schwachen Lösungen und starken Lösungen zitieren hier lediglich den Satz von Yamada und Watanabe.

Definition 26.17 (Pfadweise Eindeutigkeit). Wir sagen, dass die Lösung der SDGL (26.15) mit Startverteilung μ **pfadweise eindeutig** ist, falls für jedes $\mu \in \mathcal{M}_1(\mathbb{R}^n)$ und je zwei schwache Lösungen (X, W) und (X', W) auf dem selben Raum $(\Omega, \mathcal{F}, \mathbf{P})$ mit der selben Filtration \mathbb{F} gilt: $\mathbf{P}[X_t = X'_t \text{ für alle } t \geq 0] = 1$.

Satz 26.18 (Yamada und Watanabe). Es sind äquivalent:

- (i) Die SDGL (26.15) hat eine eindeutige starke Lösung.
- (ii) Für jedes $\mu \in \mathcal{M}_1(\mathbb{R}^n)$ hat (26.15) eine schwache Lösung, und es gilt pfadweise Eindeutigkeit.

Gelten (i) und (ii), so ist die Lösung schwach eindeutig.

Beweis. Siehe [184], [157, Seite 151ff] oder [85, Seite 163ff]. \square

Beispiel 26.19. Sei X eine schwache Lösung von (26.17). Dann ist auch $-X$ eine schwache Lösung, das heißt, es gilt keine pfadweise Eindeutigkeit (obwohl man zeigen kann, dass die Lösung schwach eindeutig ist, siehe Satz 26.25). \diamond

Wir betrachten den eindimensionalen Fall $m = n = 1$. Ist X eine Lösung (stark oder schwach) von (26.15), so ist

$$M_t := X_t - \int_0^t b(s, X_s) ds$$

ein stetiges lokales Martingal mit quadratischer Variation

$$\langle M \rangle_t = \int_0^t \sigma^2(s, X_s) ds.$$

Wir werden sehen, dass hierdurch eine schwache Lösung von (26.15) charakterisiert ist (jedenfalls unter milden Wachstumsbedingungen an σ und b).

Sei für alle $t \geq 0$ und $x \in \mathbb{R}^n$ die $n \times n$ Matrix $a(t, x)$ symmetrisch und nichtnegativ definit, und sei $(t, x) \mapsto a(t, x)$ messbar.

Definition 26.20. Wir sagen, dass ein n -dimensionaler stetiger Prozess X eine Lösung des **lokalen Martingalproblems** zu a und b mit Startverteilung $\mu \in \mathcal{M}_1(\mathbb{R}^n)$ (kurz: LMP(a, b, μ)) ist, falls $\mathbf{P} \circ X_0^{-1} = \mu$ ist und für jedes $i = 1, \dots, n$

$$M_t^i := X_t^i - \int_0^t b_i(s, X_s) ds, \quad t \geq 0,$$

ein stetiges lokales Martingal ist mit quadratischer Kovariation

$$\langle M^i, M^j \rangle_t = \int_0^t a_{ij}(s, X_s) ds \quad \text{für alle } t \geq 0, i, j = 1, \dots, n.$$

Wir sagen, dass die Lösung von $LMP(a, b, \mu)$ eindeutig ist, wenn für je zwei Lösungen X und X' gilt: $\mathbf{P} \circ X^{-1} = \mathbf{P} \circ (X')^{-1}$.

Mit σ^T bezeichnen wir die transponierte Matrix zu σ . Offenbar ist $a = \sigma\sigma^T$ dann eine nichtnegativ semidefinite symmetrische $n \times n$ Matrix.

Satz 26.21. X ist genau dann eine Lösung von $LMP(\sigma\sigma^T, b, \mu)$, wenn es (gegebenenfalls auf einer Erweiterung des Wahrscheinlichkeitsraums) eine Brown'sche Bewegung W gibt, sodass (X, W) eine schwache Lösung von (26.15) ist.

Insbesondere existiert genau dann eine eindeutige schwache Lösung der SDGL (26.15) mit Startverteilung μ , wenn $LMP(\sigma\sigma^T, b, \mu)$ eindeutig lösbar ist.

Beweis. Wir zeigen die Aussage nur für den Fall $m = n = 1$. Der allgemeine Fall erfordert ein paar Betrachtungen über Wurzeln von nichtnegativ semidefiniten symmetrischen Matrizen, die jedoch für die Stochastik keine tiefere Einsicht bringen. Wir verweisen hier lediglich auf [93, Proposition 5.4.6].

„ \Leftarrow “ Ist (X, W) eine schwache Lösung, dann löst X nach Korollar 25.19 das lokale Martingalproblem.

„ \Rightarrow “ Sei X eine Lösung von $LMP(\sigma^2, b, \mu)$. Nach Satz 25.29 existiert auf einer Erweiterung des Wahrscheinlichkeitsraums eine Brown'sche Bewegung \tilde{W} , sodass $M_t = \int_0^t |\sigma(s, X_s)| d\tilde{W}_s$ gilt. Setzen wir

$$W_t := \int_0^t \operatorname{sign}(\sigma(s, X_s)) d\tilde{W}_s,$$

so ist $M_t = \int_0^t \sigma(s, X_s) dW_s$, also (X, W) eine schwache Lösung von (26.15). \square

Ein lokales Martingalproblem ist in gewissem Sinne eine sehr natürliche Art und Weise, um eine stochastische Differentialgleichung zu schreiben, nämlich als:

X hat lokal die Ableitung (Drift) b und zusätzlich zufällige normalverteilte Fluktuationen von der Größenordnung σ .

Eine konkrete Brown'sche Bewegung taucht hier gar nicht mehr auf, und bei den meisten Problemen ist ihr Auftreten auch eher artifiziell. Genau wie man bei der Beschreibung von Markovketten meist nur die Übergangswahrscheinlichkeiten angibt, nicht aber die konkrete Realisierung, wie dies etwa in Satz 17.17 beschrieben wird, möchte man bei vielen stetigen (Zeit und Ort) Prozessen (neben der Drift) nur die Größe der Fluktuationen angeben, nicht aber eine konkrete Realisierung.

Technisch gesehen ist die Formulierung von stochastischen Differentialgleichungen als lokale Martingalprobleme sehr bequem, weil sie Zugang zu einer Reihe von

Techniken schafft wie Martingungleichungen und Approximationssätze für Martingale, mit denen sich Existenz und Eindeutigkeit von Lösungen etablieren lässt. Wir zitieren hier nur zwei wichtige Ergebnisse.

Satz 26.22 (Existenz von Lösungen). *Es seien $(t, x) \mapsto b(t, x)$ und $(t, x) \mapsto a(t, x)$ stetig und beschränkt. Dann existiert für jedes $\mu \in \mathcal{M}_1(\mathbb{R}^n)$ eine Lösung X des LMP(a, b, μ).*

Beweis. Siehe [157, Theorem V.23.5]. \square

Definition 26.23. *Wir sagen, dass das LMP(a, b) **gut gestellt** ist, wenn es für jedes $x \in \mathbb{R}^n$ eine eindeutige Lösung X von LMP(a, b, δ_x) gibt.*

Bemerkung 26.24. Erfüllen σ und b die Lipschitzbedingungen wie in Satz 26.8, so ist das LMP($\sigma\sigma^T, b$) gut gestellt. Dies folgt aus Satz 26.8 und Satz 26.18. \diamond

Im Folgenden gelte stets:

$$(t, x) \mapsto \sigma(t, x) \text{ bzw. } (t, x) \mapsto a(t, x) \text{ ist beschränkt auf kompakten Mengen.} \quad (26.21)$$

Diese Bedingung sichert die Äquivalenz des lokalen Martingalproblems zu dem etwas gebräuchlicheren Martingalproblem (siehe [93, Proposition 5.4.11]).

Satz 26.25 (Eindeutigkeit im Martingalproblem). *Es gelte (26.21). Für jedes $x \in \mathbb{R}^n$ existiere eine Lösung X^x von LMP(a, b, δ_x), deren Verteilung wir mit $\mathbf{P}_x := \mathbf{P} \circ (X^x)^{-1}$ bezeichnen.*

Für je zwei Lösungen X^x und Y^x von LMP(a, b, δ_x) gelte

$$\mathbf{P} \circ (X_T^x)^{-1} = \mathbf{P} \circ (Y_T^x)^{-1} \quad \text{für jedes } T \geq 0. \quad (26.22)$$

Dann ist LMP(a, b) gut gestellt, und der kanonische Prozess X ist ein starker Markovprozess bezüglich $(\mathbf{P}_x, x \in \mathbb{R}^n)$. Ist $a = \sigma\sigma^T$, so ist X unter \mathbf{P}_x die eindeutige schwache Lösung der SDGL (26.15).

Beweis. Siehe [54, Theorem 4.4.2 und Problem 49] und [93, Proposition 5.4.11]. \square

Eine wesentliche Stärke dieses Satzes liegt darin, dass wir die Eindeutigkeit nicht des gesamten Prozesses, sondern in (26.22) nur der eindimensionalen Randverteilungen prüfen müssen. Wir werden in Abschnitt 26.3 Beispiele dafür angeben, wie dies ausgenutzt werden kann.

Die Frage nach der Existenz von Lösungen einer stochastischen Differentialgleichung (oder äquivalent: eines lokalen Martingalproblems) ist leichter zu beantworten als die Frage nach der Eindeutigkeit von Lösungen. Wir wissen bereits, dass

Eindeutigkeit unter Lipschitzbedingungen an die Koeffizienten b und σ (nicht $\sigma\sigma^T$!) gilt (nach Satz 26.8 und Satz 26.18), da hier starke Eindeutigkeit der Lösungen gilt.

Eine vielleicht auf den ersten Blick verwirrende Erkenntnis ist, dass der Zufall stabilisierend wirken kann, dass also eine deterministische Differentialgleichung, deren Lösung nicht eindeutig ist, durch stochastische Störterme eindeutig lösbar werden kann. Dazu folgendes eindimensionale Beispiel:

$$dX_t = \text{sign}(X_t) |X_t|^{1/3} dt + \sigma dW_t, \quad X_0 = 0. \quad (26.23)$$

Ist $\sigma = 0$, so haben wir es mit einer deterministischen Differentialgleichung zu tun, die ein Kontinuum von Lösungen mit Parametern $v \in \{-1, +1\}$ und $T \geq 0$ hat, nämlich $X_t = v 2\sqrt{2} (t - T)^{3/2} \mathbb{1}_{\{t > T\}}$. Ist $\sigma > 0$, so wird die Instabilität der Gleichung (26.23) an $x = 0$ durch Verrauschen aufgelöst. Wir zitieren hier den folgenden Satz für den zeitunabhängigen Fall aus [157, Satz V.24.1] (siehe auch [171, Kapitel 10]).

Satz 26.26 (Stroock-Varadhan). Sei $a_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig und $b_i : \mathbb{R}^n \rightarrow \mathbb{R}$ messbar für $i, j = 1, \dots, n$. Es gelte

- (i) $a(x) = (a_{ij}(x))$ ist symmetrisch und strikt positiv definit für jedes $x \in \mathbb{R}^n$,
- (ii) es gibt ein $C < \infty$, sodass für alle $x \in \mathbb{R}^n$ und $i, j = 1, \dots, n$ gilt

$$|a_{ij}(x)| \leq C (1 + \|x\|^2) \quad \text{und} \quad |b_i(x)| \leq C (1 + \|x\|).$$

Dann ist LMP(a, b) gut gestellt, und die SDGL (26.15) hat eine eindeutige starke Lösung, die ein starker Markovprozess ist. Die Lösung X hat die starke Feller-Eigenschaft: $x \mapsto \mathbf{E}_x[f(X_t)]$ ist stetig für jedes $t > 0$ und jedes beschränkte, messbare $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Konkrete Beispiele geben wir im Abschnitt 26.3 an. Wir wollen hier nur festhalten, dass wir eine spezielle Methode entwickelt haben, um Markovprozesse zu konstruieren, nämlich als Lösung einer stochastischen Differentialgleichung oder eines lokalen Martingalproblems. Im Rahmen von Modellen in diskreter Zeit haben wir in Kapitel 17.2 und speziell in Übung 17.2.1 bereits Markovketten als Lösungen von Martingalproblemen charakterisiert. Dass dort die Angabe der Drift und der quadratischen Variation ausreichte, um den Prozess eindeutig zu bestimmen, lag daran, dass wir die Möglichkeiten für das Ziel eines Schrittes auf drei Punkte begrenzt hatten. Hier hingegen ist die entscheidende Begrenzung die Stetigkeit der Prozesse.

Übung 26.2.1. Sei der zeithomogene eindimensionale Fall ($m = n = 1$) betrachtet. Seien σ und b so, dass es für jedes $X_0 \in \mathbb{R}$ eine eindeutige schwache Lösung von

$$dX_t = \sigma(X_t) dW_t + b(X_t) dt$$

existiert und ein starker Markovprozess ist. Ferner gebe es ein $x_0 \in \mathbb{R}$ mit

$$C := \int_{-\infty}^{\infty} \frac{1}{\sigma^2(x)} \exp \left(\int_{x_0}^x \frac{2b(r)}{\sigma^2(r)} dr \right) dr < \infty.$$

(i) Man zeige: Das Maß $\pi \in \mathcal{M}_1(\mathbb{R})$ mit Dichte

$$\frac{\pi(dx)}{dx} = C^{-1} \frac{1}{\sigma^2(x)} \exp \left(\int_{x_0}^x \frac{2b(r)}{\sigma^2(r)} dr \right)$$

ist eine invariante Verteilung für X .

- (ii) Für welche Werte von b hat der Ornstein-Uhlenbeck Prozess $dX_t = \sigma dW_t + bX_t dt$ eine invariante Verteilung? Man bestimme diese Verteilung und vergleiche das Ergebnis mit dem, was nach expliziter Rechnung mit der Darstellung in (26.3) zu erwarten war.
- (iii) Man bestimme die invariante Verteilung der Cox-Ingersoll-Ross SDGL (26.14) (alias Feller'sche Verzweigungsdiffusion).
- (iv) Seien $\gamma, c > 0$ und $\theta \in (0, 1)$. Man zeige, dass die invariante Verteilung der Lösung X der folgenden SDGL auf $[0, 1]$

$$dX_t = \sqrt{\gamma X_t(1 - X_t)} dW_t + c(\theta - X_t) dt$$

gegeben ist durch die Betaverteilung $\beta_{2c\gamma/\theta, 2c\gamma/(1-\theta)}$.



Übung 26.2.2. Sei $\gamma > 0$. Seien X^1 und X^2 Lösungen von $dX_t^i = \sqrt{\gamma X_t^i} dW_t^i$, wo W^1 und W^2 zwei unabhängige Brown'sche Bewegungen sind, mit Startwerten $X_0^1 = x_0^1 > 0$ und $X_0^2 = x_0^2 > 0$. Man zeige, dass $Z := X^1 + X^2$ eine schwache Lösung ist von $Z_0 = 0$ und $dZ_t = \sqrt{\gamma Z_t} dW_t$.



26.3 Eindeutigkeit schwacher Lösungen via Dualität

Mit dem Satz von Stroock und Varadhan haben wir ein starkes Kriterium für die Existenz und Eindeutigkeit von Lösungen von stochastischen Differentialgleichungen. In vielen Fällen ist jedoch gerade die Bedingung der lokal gleichgradigen Elliptizität von a (Bedingung (i) in Satz 26.26) nicht erfüllt. Dies trifft insbesondere dann zu, wenn die Lösungen nur auf Teilmengen von \mathbb{R}^n definiert sind.

Wir werden hier ein mächtiges Hilfsmittel kennen lernen, das in vielen Spezialfällen schwache Eindeutigkeit von Lösungen sichert.

Definition 26.27 (Dualität). Seien $X = (X^x, x \in E)$ und $Y = (Y^y, y \in E')$ Familien von stochastischen Prozessen mit Werten in den Räumen E beziehungsweise E' und so, dass $X_0^x = x$ f.s. und $Y_0^y = y$ f.s. für alle $x \in E$ und $y \in E'$. Wir sagen, dass X und Y **dual** zueinander sind mit **Dualitätsfunktion** $H : E \times E' \rightarrow \mathbb{C}$,

falls für alle $x \in E$, $y \in E'$ und $t \geq 0$ die Erwartungswerte $\mathbf{E}[H(X_t^x, y)]$ und $\mathbf{E}[H(x, Y_t^y)]$ existieren und gleich sind:

$$\mathbf{E}[H(X_t^x, y)] = \mathbf{E}[H(x, Y_t^y)].$$

Wir nehmen im Folgenden an, dass $\sigma_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}$ und $b_i : \mathbb{R}^n \rightarrow \mathbb{R}$ beschränkt auf kompakten Mengen sind für alle $i = 1, \dots, n$, $j = 1, \dots, m$. Wir betrachten die zeithomogene stochastische Differentialgleichung

$$dX_t = \sigma(X_t) dW_t + b(X_t) dt. \quad (26.24)$$

Satz 26.28 (Eindeutigkeit via Dualität). Für jedes $x \in \mathbb{R}^n$ existiere eine Lösung des lokalen Martingalproblems zu $(\sigma\sigma^T, b, \delta_x)$. Es gebe eine Familie $(Y_t^y, y \in E')$ von Markovprozessen mit Werten in dem Messraum (E', \mathcal{E}') und eine messbare Abbildung $H : \mathbb{R}^n \times E' \rightarrow \mathbb{C}$, sodass für jedes $y \in E'$, $x \in \mathbb{R}^n$ und $t \geq 0$ der Erwartungswert $\mathbf{E}[H(x, Y_t^y)]$ existiert und endlich ist. Ferner sei $(H(\cdot, y), y \in E')$ eine trennende Funktionenklasse für $\mathcal{M}_1(\mathbb{R}^n)$ (siehe Definition 13.9).

Für jedes $x \in \mathbb{R}^n$ und jede Lösung X^x von $LMP(\sigma\sigma^T, b, \delta_x)$ gelte die Dualitätsgleichung

$$\mathbf{E}[H(X_t^x, y)] = \mathbf{E}[H(x, Y_t^y)] \quad \text{für alle } y \in E', t \geq 0. \quad (26.25)$$

Dann ist das lokale Martingalproblem zu $(\sigma\sigma^T, b)$ gut gestellt, also besitzt (26.24) eine eindeutige schwache Lösung und diese ist ein starker Markovprozess.

Beweis. Nach Satz 26.25 reicht es zu prüfen, dass für jedes $x \in \mathbb{R}^n$, jede Lösung X^x von $LMP(\sigma\sigma^T, b, \delta_x)$ und jedes $t \geq 0$ die Verteilung $\mathbf{P} \circ (X_t^x)^{-1}$ eindeutig ist. Da $(H(\cdot, y), y \in E')$ eine trennende Funktionenklasse ist, folgt dies aber aus (26.16). \square

Beispiel 26.29 (Wright-Fisher Diffusion). Betrachte die Wright-Fisher SDGL

$$dX_t = \mathbb{1}_{[0,1]}(X_t) \sqrt{\gamma X_t(1-X_t)} dW_t, \quad (26.26)$$

wobei $\gamma > 0$ ein Parameter ist. Nach Satz 26.22 existiert für jedes $x \in \mathbb{R}$ eine schwache Lösung (\tilde{X}, W) von (26.26). \tilde{X} ist ein stetiges lokales Martingal mit quadratischer Variation

$$\langle \tilde{X} \rangle_t = \int_0^t \gamma \tilde{X}_s (1 - \tilde{X}_s) \mathbb{1}_{[0,1]}(\tilde{X}_s) ds.$$

Sei $\tau := \inf\{t > 0 : \tilde{X}_t \notin [0, 1]\}$ und $X := \tilde{X}^\tau$ der in τ gestoppte Prozess. Dann ist X ein stetiges, beschränktes Martingal mit

$$\langle X \rangle_t = \int_0^t \gamma X_s (1 - X_s) \mathbb{1}_{[0,1]}(X_s) ds,$$

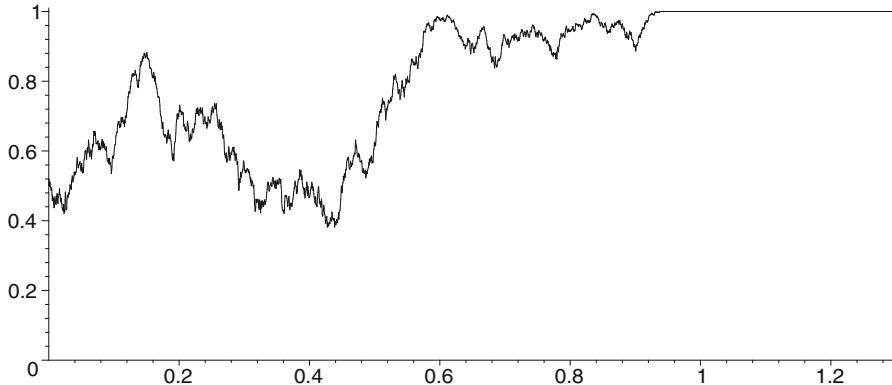


Abb. 26.3. Simulation einer Wright-Fisher Diffusion mit Parameter $\gamma = 1$.

also ist (X, W) eine Lösung von (26.26). Nach Konstruktion ist $X_t \in [0, 1]$ für alle $t \geq 0$, falls $X_0 = \tilde{X}_0 \in [0, 1]$ ist.

Sei

$$\tau' := \inf \{t > 0 : \tilde{X}_t \in [0, 1]\}.$$

Ist $\tilde{X}_0 \notin [0, 1]$, so ist $\tau' > 0$, weil \tilde{X} stetig ist. Weil $\tilde{X}^{\tau'}$ ein stetiges lokales Martingal ist mit $\langle \tilde{X}^{\tau'} \rangle \equiv 0$, ist $\tilde{X}_t^{\tau'} = \tilde{X}_0$ für alle $t \geq 0$. Daraus folgt aber, dass $\tilde{X}_t = \tilde{X}_0$ ist für alle $t < \tau'$. Wiederum aufgrund der Stetigkeit von \tilde{X} folgt, dass $\tau' = \infty$ ist und $\tilde{X}_t = \tilde{X}_0$ für alle $t \geq 0$.

Es reicht also, die Eindeutigkeit der Lösung für $\tilde{X}_0 = x \in [0, 1]$ zu zeigen. Sei hierzu $Y = (Y_t)_{t \geq 0}$ der Markovprozess auf \mathbb{N} mit Q -Matrix

$$q(m, n) = \begin{cases} \gamma \binom{m}{2}, & \text{falls } n = m - 1, \\ -\gamma \binom{m}{2}, & \text{falls } n = m, \\ 0, & \text{sonst.} \end{cases}$$

Wir zeigen die Dualität von X und Y bezüglich $H(x, n) = x^n$:

$$\mathbf{E}_x[X_t^n] = \mathbf{E}_n[x^{Y_t}] \quad \text{für alle } t \geq 0, x \in [0, 1], n \in \mathbb{N}. \quad (26.27)$$

Setze $m^{x,n}(t) = \mathbf{E}_x[X_t^n]$ und $g^{x,n}(t) = \mathbf{E}_n[x^{Y_t}]$. Nach der Itô-Formel ist

$$X_t^n - x^n - \int_0^t \gamma \binom{n}{2} X_s^{n-1} (1 - X_s) ds = \int_0^t n X_s^{n-1} \sqrt{\gamma X_s (1 - X_s)} dW_s$$

ein Martingal. Indem wir Erwartungswerte bilden, erhalten wir die folgende Rekursionsgleichung für die Momente von X

$$\begin{aligned} m^{x,1}(t) &= x, \\ m^{x,n}(t) &= x^n + \gamma \binom{n}{2} \int_0^t (m^{x,n-1}(s) - m^{x,n}(s)) ds \quad \text{für } n \geq 2. \end{aligned} \quad (26.28)$$

Dieses System linearer Differentialgleichungen ist offenbar eindeutig rekursiv (in n) lösbar.

Für $h > 0$ und $t \geq 0$ gilt wegen der Markov-Eigenschaft von Y

$$\begin{aligned} g^{x,n}(t+h) &= \mathbf{E}_n[x^{Y_{t+h}}] = \mathbf{E}_n[\mathbf{E}_{Y_h}[x^{Y_t}]] \\ &= \sum_{m=1}^n \mathbf{P}_n[Y_h = m] \mathbf{E}_m[x^{Y_t}] \\ &= \sum_{m=1}^n \mathbf{P}_n[Y_h = m] g^{x,m}(t). \end{aligned}$$

Es folgt

$$\begin{aligned} \frac{d}{dt}g^{x,n}(t) &= \lim_{h \downarrow 0} h^{-1} [g^{x,n}(t+h) - g^{x,n}(t)] \\ &= \lim_{h \downarrow 0} h^{-1} \sum_{m=1}^n \mathbf{P}_n[Y_h = m] (g^{x,m}(t) - g^{x,n}(t)) \\ &= \sum_{m=1}^n q(n, m) g^{x,m}(t) \\ &= \gamma \binom{n}{2} (g^{x,n-1}(t) - g^{x,n}(t)). \end{aligned} \tag{26.29}$$

Offenbar ist $g^{x,1}(t) = x$ für alle $x \in [0, 1]$ und $t \geq 0$ und $g^{x,n}(0) = x^n$. Das heißt, $g^{x,n}$ löst (26.28), und daher gilt (26.27).

Nach Satz 15.4 ist die Familie $(H(\cdot, n), n \in \mathbb{N}) \subset C([0, 1])$ trennend für $\mathcal{M}_1([0, 1])$, also sind die Bedingungen von Satz 26.28 erfüllt, und X ist die eindeutige schwache Lösung von (26.26) und ist ein starker Markovprozess. \diamond

Bemerkung 26.30. Das Martingalproblem für die Wright-Fisher Diffusion sieht fast genauso aus wie das diskrete Martingalproblem für das Moran-Modell (siehe Beispiel 17.22) $M^N = (M_n^N)_{n \in \mathbb{N}_0}$ mit Populationsgröße N : M^N ist ein Martingal mit Werten in der Menge $\{0, 1/N, \dots, (N-1)/N, 1\}$ und quadratischem Variationsprozess

$$\langle M^N \rangle_n = \frac{2}{N^2} \sum_{k=0}^{n-1} M_k^N (1 - M_k^N).$$

In jedem Schritt kann M^N nur entweder am Ort bleiben oder um $1/N$ nach oben oder unten springen. In Übung 17.2.1 hatten wir gesehen, dass dadurch der Prozess M^N schon eindeutig beschrieben ist. Man kann zeigen, ähnlich wie in Satz 21.51 für Verzweigungsprozesse, dass die zeitlich reskalierten Moran-Prozesse $\tilde{M}_t^N = M_{\lfloor N^2 t \rfloor}^N$ gegen die Wright-Fisher Diffusion mit $\gamma = 2$ konvergieren. Die Wright-Fisher Diffusion tritt also als Limes-Modell eines genealogischen Modells auf und

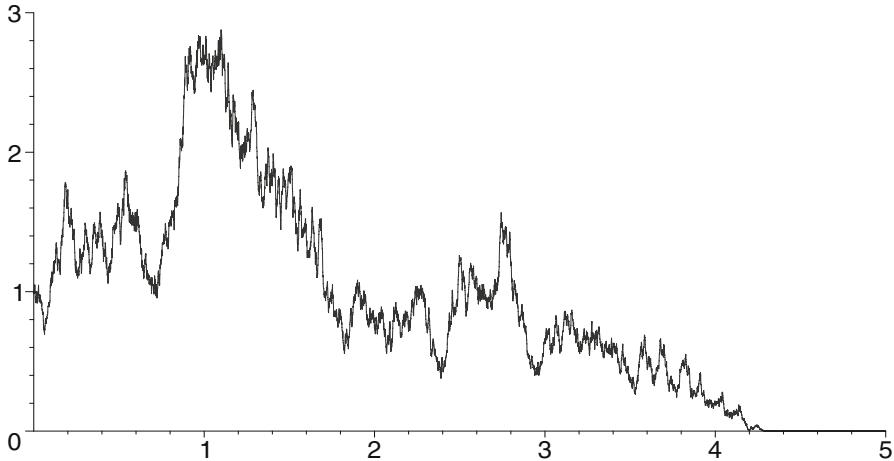


Abb. 26.4. Simulation einer Feller'schen Verzweigungsdiffusion mit Parameter $\gamma = 1$.

beschreibt die Genfrequenz (das heißt, den relativen Anteil) eines bestimmten Allels in einer Population, die durch die Generationenfolge in zufälliger Weise fluktuiert.

◇

Beispiel 26.31 (Feller'sche Verzweigungsdiffusion). Sei $(Z_n^N)_{n \in \mathbb{N}_0}$ ein Galton-Watson Verzweigungsprozess mit kritischer, geometrischer Nachkommenverteilung $p_k = 2^{-k-1}$, $k \in \mathbb{N}_0$ und $Z_0^N = N$ für jedes $N \in \mathbb{N}$. Dann ist Z^N ein diskretes Martingal, und es gilt

$$\mathbf{E}\left[\left(Z_n^N - Z_{n-1}^N\right)^2 \mid Z_{n-1}^N\right] = Z_{n-1}^N \left(\sum_{k=0}^{\infty} p_k k^2 - 1 \right) = 2 Z_{n-1}^N.$$

Daher hat Z^N die quadratische Variation

$$\langle Z^N \rangle_n = \sum_{k=0}^{n-1} 2Z_k^N.$$

Sei nun

$$Z_t^N := (t - N^{-1} \lfloor tN \rfloor) \left(Z_{\lfloor tN \rfloor + 1}^N - Z_{\lfloor tN \rfloor}^N \right) + \frac{1}{n} Z_{\lfloor tN \rfloor}^N$$

eine linear interpolierte Version von $N^{-1}Z_{\lfloor tN \rfloor}^N$. Nach dem Satz von Lindvall (Satz 21.51) gibt es einen stetigen Markovprozess Z , sodass $Z^N \xrightarrow{N \rightarrow \infty} Z$ in Verteilung gilt. Da man zeigen kann, dass auch die Momente konvergieren, gilt, dass Z ein stetiges Martingal ist und quadratische Variation

$$\langle Z \rangle_t = \int_0^t 2Z_s ds$$

hat. Tatsächlich hatten wir in Beispiel 26.11 bereits gezeigt, dass Z die Lösung der SDGL

$$dZ_t = \sqrt{2Z_t} dW_t \quad (26.30)$$

mit Start in $Z_0 = 1$ ist. Dort hatten wir auch gezeigt, dass Z zu $Y_t^y = \left(\frac{t\gamma}{2} + \frac{1}{y}\right)^{-1}$ dual ist mit $H(x, y) = e^{-xy}$. Hieraus folgt die Eindeutigkeit der Lösung von (26.30) und die starke Markoveigenschaft von Z . \diamond

Man kann einwenden, dass in den Beispielen 26.29 und 26.31 nur eindimensionale Situationen betrachtet wurden, für die wir nach dem Satz von Yamada-Watanabe (Satz 26.10) sowieso schon um die Eindeutigkeit sogar einer starken Lösung wissen. Die wahre Stärke der Methode der Dualität kann sich also erst in höherdimensionalen Problemen entfalten. Hierzu betrachten wir als Beispiel eine Erweiterung von Beispiel 26.29.

Beispiel 26.32 (Wechselwirkende Wright-Fisher Diffusionen).

Die Wright-Fisher Diffusion aus Beispiel 26.29 beschreibt die Fluktuationen der Genfrequenz eines Allels in *einer* großen Population. Wir wollen nun mehrere Populationen betrachten, die auf den Punkten $i \in S := \{1, \dots, N\}$ leben, und miteinander durch Migration, die durch Wechselwirkungsraten $r(i, j) \geq 0$ quantifiziert wird, in Wechselwirkung stehen. Als Modell für die Genfrequenzen $X_t(i)$ am Ort i zur Zeit t stellen wir daher die folgende N -dimensionale SDGL für $X = (X(1), \dots, X(N))$ auf:

$$dX_t(i) = \sqrt{\gamma X_t(i)(1 - X_t(i))} dW_t^i + \sum_{j=1}^N r(i, j)(X_t(j) - X_t(i)) dt. \quad (26.31)$$

Dabei ist $W = (W^1, \dots, W^N)$ eine N -dimensionale Brown'sche Bewegung. Diese SDGL hat nach Satz 26.22 schwache Lösungen, jedoch greift keines unserer allgemeinen Kriterien für schwache Eindeutigkeit. Wir werden daher die schwache Eindeutigkeit mittels Dualität zeigen.

Es ist, ähnlich wie in Beispiel 26.29, nicht schwer zu zeigen, dass Lösungen von (26.31), die in $X_0 = x \in E := [0, 1]^S$ starten, in $[0, 1]^S$ bleiben. Die Diagonalterme $r(i, i)$ tauchen in (26.31) nicht auf, daher können wir sie noch beliebig festsetzen und wählen $r(i, i) = -\sum_{j \neq i} r(i, j)$. Sei $Y = (Y_t)_{t \geq 0}$ der Markovprozess auf $E' := (\mathbb{N}_0)^S$ mit der folgenden Q -Matrix

$$q(\varphi, \eta) = \begin{cases} \varphi(i) r(i, j), & \text{falls } \eta = \varphi - \mathbf{1}_{\{i\}} + \mathbf{1}_{\{j\}} \text{ für gewisse } i, j \in S, i \neq j, \\ \gamma \binom{\varphi(i)}{2}, & \text{falls } \eta = \varphi - \mathbf{1}_{\{i\}} \text{ für ein } i \in S, \\ \sum_{i \in S} \left(\varphi(i) r(i, i) - \gamma \binom{\varphi(i)}{2} \right), & \text{falls } \eta = \varphi, \\ 0, & \text{sonst.} \end{cases}$$

Dabei bezeichnet $\varphi \in E'$ einen generischen Zustand mit $\varphi(i)$ Teilchen am Ort $i \in S$, und $\mathbb{1}_{\{i\}} \in E'$ bezeichnet den Zustand mit genau einem Teilchen am Ort i . Der Prozess Y beschreibt ein System von Teilchen, die unabhängig voneinander mit Rate $r(i, j)$ vom Ort i zum Ort j springen. Sind mehrere Teilchen an einem Ort i , so verschmilzt jedes der $\binom{\varphi(i)}{2}$ Paare von Teilchen mit der selben Rate γ zu einem Teilchen. Die gängige genealogische Interpretation dieses Prozesses ist, dass er (in umgekehrter Zeit) die Ahnenlinien einer Stichprobe von je $Y_0(i)$ Individuen an den Orten $i \in S$, beschreibt. Durch Migration wechseln die Linien den Ort. Haben zwei Individuen den selben Vorfahren, so verschmelzen zwei Linien. Offenbar ist für einen gemeinsamen Vorfahren notwendig aber nicht hinreichend, dass beide Linien am selben Ort sind.

Für $x \in \mathbb{R}^n$ und $\varphi \in E'$ schreiben wir $x^\varphi := \prod_{i \in S} x(i)^{\varphi(i)}$. Wir zeigen, dass X und Y dual zueinander sind mit der Dualitätsfunktion $H(x, \varphi) = x^\varphi$:

$$\mathbf{E}_x[X_t^\varphi] = \mathbf{E}_\varphi[x^{Y_t}] \quad \text{für alle } \varphi \in (\mathbb{N}_0)^S, x \in [0, 1]^S, t \geq 0. \quad (26.32)$$

Sei $m^{x, \varphi}(t) := \mathbf{E}_x[X_t^\varphi]$ und $g^{x, \varphi}(t) := \mathbf{E}_\varphi[x^{Y_t}]$. Offenbar hat H die Ableitungen

$$\partial_i H(\cdot, \varphi)(x) = \varphi(i) x^{\varphi - \mathbb{1}_{\{i\}}}$$

und

$$\partial_i \partial_i H(\cdot, \varphi)(x) = 2 \binom{\varphi(i)}{2} x^{\varphi - 2 \mathbb{1}_{\{i\}}}.$$

Nach der Itô-Formel ist

$$\begin{aligned} X_t^\varphi - X_0^\varphi &= \int_0^t \sum_{i,j \in S} \varphi(i) r(i, j) (X_s(j) - X_s(i)) X_t^{\varphi - \mathbb{1}_{\{i\}}} ds \\ &\quad - \sum_{i \in S} \int_0^t \gamma \binom{\varphi(i)}{2} (X_s(i)(1 - X_s(i))) X_s^{\varphi - 2 \mathbb{1}_{\{i\}}} ds \end{aligned}$$

ein Martingal. Indem wir Erwartungswerte bilden, erhalten wir ein System von linearen Integralgleichungen

$$\begin{aligned} m^{x, 0}(t) &= 1, \\ m^{x, \varphi}(t) &= x^\varphi + \int_0^t \sum_{i,j \in S} \varphi(i) r(i, j) (m^{x, \varphi + \mathbb{1}_{\{j\}} - \mathbb{1}_{\{i\}}}(s) - m^{x, \varphi}(s)) ds \\ &\quad + \int_0^t \gamma \sum_{i \in S} \binom{\varphi(i)}{2} (m^{x, \varphi - \mathbb{1}_{\{i\}}}(s) - m^{x, \varphi}(s)) ds. \end{aligned} \quad (26.33)$$

Dieses System von Gleichungen lässt sich per Induktion über $n = \sum_{i \in I} \varphi(i)$ eindeutig lösen. Wir wollen die Lösung jedoch nicht explizit ausrechnen, sondern nur zeigen, dass sie mit $g^{x, \varphi}(t)$ übereinstimmt, indem wir zeigen, dass g ein äquivalentes System von Differentialgleichungen löst.

Für g erhalten wir wie in (26.29)

$$\begin{aligned} \frac{d}{dt} g^{x,\varphi}(t) &= \sum_{\eta \in E'} q(\varphi, \eta) g^{x,\varphi}(t) \\ &= \sum_{i,j \in S} r(i,j) \left(g^{x,\varphi + \mathbb{1}_{\{j\}} - \mathbb{1}_{\{i\}}}(t) - g^{x,\varphi}(t) \right) \\ &\quad + \sum_{i \in S} \gamma \binom{\varphi(i)}{2} \left(g^{x,\varphi - \mathbb{1}_{\{i\}}}(t) - g^{x,\varphi}(t) \right). \end{aligned} \quad (26.34)$$

Zusammen mit dem Startwert $g^{x,0}(t) = 1$ und $g^{x,\varphi}(0) = x^\varphi$ ist das System (26.34) von Differentialgleichungen äquivalent zu (26.33). Also gilt die Dualität (26.32), und damit ist die SDGL (26.31) eindeutig schwach lösbar. (Tatsächlich kann man zeigen, dass es eine eindeutige starke Lösung gibt, sogar wenn S abzählbar unendlich ist und r gewisse Regularitätsannahmen erfüllt, beispielsweise die Q -Matrix einer Irrfahrt auf $S = \mathbb{Z}^d$ ist, siehe [163].) \diamond

Übung 26.3.1 (Aussterbewahrscheinlichkeit der Feller'schen Verzweigungsdifusion). Sei $\gamma > 0$ und Z die Lösung von $dZ_t := \sqrt{\gamma Z_t} dW_t$ mit Anfangswert $Z_0 = z > 0$. Man zeige mit Hilfe der Dualität

$$\mathbf{P}_z[Z_t = 0] = \exp \left(-\frac{2z}{\gamma t} \right). \quad (26.35)$$

Man bestimme mit Hilfe von Lemma 21.44 die Wahrscheinlichkeit, dass ein Galton-Watson Verzweigungsprozess X mit kritischer, geometrischer Nachkommenverteilung und $X_0 = N \in \mathbb{N}$ bis zur Zeit $n \in \mathbb{N}$ ausgestorben ist und vergleiche das Ergebnis mit (26.35). \clubsuit

Literatur

1. Robert J. Adler. *An introduction to continuity, extrema, and related topics for general Gaussian processes*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 12. Institute of Mathematical Statistics, Hayward, CA, 1990.
2. M. Aizenman, H. Kesten und C. M. Newman. Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation. *Comm. Math. Phys.*, 111(4):505–531, 1987.
3. M. Aizenman, H. Kesten und C. M. Newman. Uniqueness of the infinite cluster and related results in percolation. In *Percolation theory and ergodic theory of infinite particle systems (Minneapolis, Minn., 1984–1985)*, Band 8 aus *IMA Volumes in Mathematics and Its Applications*, Seiten 13–20. Springer, New York, 1987.
4. David J. Aldous. Exchangeability and related topics. In *École d’été de probabilités de Saint-Flour, XIII—1983*, Band 1117 aus *Lecture Notes in Mathematics*, Seiten 1–198. Springer, Berlin, 1985.
5. Krishna B. Athreya und Peter E. Ney. *Branching Processes*. Springer-Verlag, Berlin, 1972.
6. Jacques Azéma und Marc Yor. Le problème de Skorokhod: compléments à “Une solution simple au problème de Skorokhod”. In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, Band 721 aus *Lecture Notes in Mathematics*, Seiten 625–633. Springer, Berlin, 1979.
7. Jacques Azéma und Marc Yor. Une solution simple au problème de Skorokhod. In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, Band 721 aus *Lecture Notes in Mathematics*, Seiten 90–115. Springer, Berlin, 1979.
8. Martin Barner und Friedrich Flohr. *Analysis. II*. de Gruyter Lehrbuch. Walter de Gruyter & Co., Berlin, 2. Auflage, 1989.
9. Heinz Bauer. *Maß - und Integrationstheorie*. de Gruyter Lehrbuch. Walter de Gruyter & Co., Berlin, 2. Auflage, 1992.
10. Heinz Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter Lehrbuch. Walter de Gruyter & Co., Berlin, 5. Auflage, 2002.
11. Leonard E. Baum und Melvin Katz. Convergence rates in the law of large numbers. *Trans. Amer. Math. Soc.*, 120:108–123, 1965.
12. M. Baxter und R. Rennie. *Financial Calculus*. Cambridge University Press, Cambridge, 1997.
13. Andrew C. Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, 49:122–136, 1941.
14. Albrecht Beutelspacher. *Kryptologie*. Vieweg + Teubner, Wiesbaden, 9. Auflage, 2009.
15. Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.

16. Patrick Billingsley. *Weak convergence of measures: Applications in probability*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1971. Conference Board of the Mathematical Sciences Regional Conference Series in Appl. Mathematics, No. 5.
17. Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 2. Auflage, 1999. A Wiley-Interscience Publication.
18. K. Binder und D. W. Heermann. *Monte Carlo simulation in statistical physics: An introduction*, Band 80 aus *Springer Series in Solid-State Sciences*. Springer-Verlag, Berlin, 3. Auflage, 1997.
19. G.D. Birkhoff. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci.*, 17:656–660, 1931.
20. David Blackwell und David Kendall. The martin boundary of Pólya's urn scheme, and an application to stochastic population growth. *J. Appl. Probability*, 1:284–296, 1964.
21. R. M. Blumenthal. An extended Markov property. *Trans. Amer. Math. Soc.*, 85:52–72, 1957.
22. Salomon Bochner. *Vorlesungen über Fouriersche Integrale*. Chelsea Publishing Company, New York, 1932. Nachdruck von 1948.
23. Leo Breiman. *Probability*. Addison-Wesley Publishing Company, Reading, Mass., 1968.
24. Pierre Brémaud. *Markov chains*, Band 31 aus *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues.
25. Brockhaus. *Die Enzyklopädie in 24 Bänden*. F.A. Brockhaus, Mannheim, 20. Auflage, 1998.
26. Jörg Brüdern. *Einführung in die analytische Zahlentheorie*. Springer-Verlag, Berlin Heidelberg, 1995.
27. Dirk Brüggemann. *Starke Gesetze der großen Zahlen bei blockweisen Unabhängigkeitsbedingungen*. Dissertation, Universität zu Köln, 2002.
28. R. M. Burton und M. Keane. Density and uniqueness in percolation. *Comm. Math. Phys.*, 121(3):501–505, 1989.
29. Gustave Choquet und Jacques Deny. Sur l'équation de convolution $\mu = \mu * \sigma$. *C. R. Acad. Sci. Paris*, 250:799–801, 1960.
30. Yuan Shih Chow und Henry Teicher. *Probability theory: Independence, interchangeability, martingales*. Springer Texts in Statistics. Springer-Verlag, New York, 3. Auflage, 1997.
31. K. L. Chung und W. H. J. Fuchs. On the distribution of values of sums of random variables. *Mem. Amer. Math. Soc.*, 6, 1951.
32. Kai Lai Chung. *Markov chains with stationary transition probabilities*. Die Grundlehren der mathematischen Wissenschaften, Bd. 104. Springer-Verlag, Berlin, 1960.
33. Peter Clifford und Aidan Sudbury. A model for spatial conflict. *Biometrika*, 60:581–588, 1973.
34. Harald Cramér. Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles*, 763:5–23, 1938. Colloque consacré à la théorie des probabilités.
35. Freddy Delbaen und Walter Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(3):463–520, 1994.
36. Amir Dembo und Ofer Zeitouni. *Large deviations techniques and applications*, Band 38 aus *Applications of Mathematics*. Springer-Verlag, New York, 2. Auflage, 1998.
37. Jean-Dominique Deuschel und Daniel W. Stroock. *Large deviations*, Band 137 aus *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA, 1989.
38. P. Diaconis und D. Freedman. Finite exchangeable sequences. *Ann. Probab.*, 8(4):745–764, 1980.

39. J. Dieudonné. *Foundations of Modern Analysis*, Band X aus *Pure and Applied Mathematics*. Academic Press, New York und London, 1960.
40. Monroe D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 6, 1951.
41. Peter G. Doyle und J. Laurie Snell. *Random walks and electric networks*, Band 22 aus *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 1984.
42. R. M. Dudley. *Real analysis and probability*, Band 74 aus *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revidierter Nachdruck der Originalausgabe von 1989.
43. Nelson Dunford und Jacob T. Schwartz. *Linear Operators. I. General Theory*. With the assistance of W. G. Bade and R. G. Bartle. Pure and Applied Mathematics, Vol. 7. Interscience Publishers, Inc., New York, 1958.
44. Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 4. Auflage, 2010.
45. Aryeh Dvoretzky, Paul Erdős und Shizuo Kakutani. Nonincrease everywhere of the Brownian motion process. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, Vol. II*, Seiten 103–116. University of California Press, Berkeley, CA, 1961.
46. Dmitri Egoroff. Sur les suites des fonctions mesurables. *C. R. Acad. Sci., Paris*, 152:135–157, 1911.
47. Robert J. Elliott und P. Ekkehard Kopp. *Mathematics of financial markets*. Springer Finance. Springer-Verlag, New York, 1999.
48. Richard S. Ellis. *Entropy, large deviations, and statistical mechanics*, Band 271 aus *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, New York, 1985.
49. Jürgen Elstrodt. *Maß- und Integrationstheorie*. Springer-Verlag, New York, 7. Auflage, 2011.
50. P. Erdős und R. L. Graham. On a linear diophantine problem of Frobenius. *Acta Arith.*, 21:399–408, 1972.
51. Carl-Gustav Esseen. On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. och Fys.*, 28A(9):1–19, 1942.
52. Nasrollah Etemadi. An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete*, 55(1):119–122, 1981.
53. Alison Etheridge. *A course in financial calculus*. Cambridge University Press, Cambridge, 2002.
54. Stewart N. Ethier und Thomas G. Kurtz. *Markov processes: Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.
55. Steven N. Evans und Xiaowen Zhou. Identifiability of exchangeable sequences with identically distributed partial sums. *Electron. Comm. Probab.*, 4:9–13 (electronic), 1999.
56. William Feller. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitstheorie I. *Math. Zeit.*, 40:521–559, 1935.
57. William Feller. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitstheorie II. *Math. Zeit.*, 42:301–312, 1937.
58. William Feller. *An introduction to probability theory and its applications. Vol. I*. John Wiley & Sons Inc., New York, 3. Auflage, 1968.
59. William Feller. *An introduction to probability theory and its applications. Vol. II*. John Wiley & Sons Inc., New York, 2. Auflage, 1971.

60. James Allen Fill. An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.*, 8(1):131–162, 1998.
61. James Allen Fill, Motoya Machida, Duncan J. Murdoch und Jeffrey S. Rosenthal. Extension of Fill’s perfect rejection sampling algorithm to general chains. *Random Structures Algorithms*, 17(3-4):290–316, 2000. Proceedings of the Ninth International Conference “Random Structures and Algorithms” (Poznan, 1999).
62. Hans Föllmer und Alexander Schied. *Stochastic finance*, Band 27 aus *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2. Auflage, 2004.
63. David A. Freedman. Bernard Friedman’s urn. *Ann. Math. Statist.*, 36:956–970, 1965.
64. Peter Gänssler und Winfried Stute. *Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin, 1977.
65. Hans-Otto Georgii. *Stochastik*. de Gruyter Lehrbuch. Walter de Gruyter & Co., Berlin, 4. Auflage, 2009.
66. Alison L. Gibbs und Francis Edward Su. On choosing and bounding probability metrics. *Int. Stat. Rev.*, 70(3):419–435, 2002.
67. M. L. Glasser und I. J. Zucker. Extended Watson integrals for the cubic lattices. *Proc. Nat. Acad. Sci. U.S.A.*, 74(5):1800–1801, 1977.
68. B. V. Gnedenko und A. N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills., Ont., 1968.
69. Geoffrey Grimmett. *Percolation*, Band 321 aus *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2. Auflage, 1999.
70. Geoffrey R. Grimmett und David R. Stirzaker. *Probability and random processes*. Oxford University Press, New York, 3. Auflage, 2001.
71. E. Grosswald. The Student t -distribution of any degree of freedom is infinitely divisible. *Z. Wahrsch. Verw. Gebiete*, 36(2):103–109, 1976.
72. Olle Häggström. *Finite Markov chains and algorithmic applications*, Band 52 aus *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2002.
73. Takashi Hara und Gordon Slade. Mean-field critical behaviour for percolation in high dimensions. *Comm. Math. Phys.*, 128(2):333–391, 1990.
74. J. Michael Harrison und Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Process. Appl.*, 11(3):215–260, 1981.
75. Philip Hartman und Aurel Wintner. On the law of the iterated logarithm. *Amer. J. Math.*, 63:169–176, 1941.
76. W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.
77. Edwin Hewitt und Kenneth A. Ross. *Abstract harmonic analysis. Vol. II: Structure and analysis for compact groups. Analysis on locally compact Abelian groups*. Die Grundlehren der mathematischen Wissenschaften, Band 152. Springer-Verlag, New York, 1970.
78. Edwin Hewitt und Leonard J. Savage. Symmetric measures on Cartesian products. *Trans. Math. Soc.*, 80:470–501, 1955.
79. C. C. Heyde. On a property of the lognormal distribution. *J. Royal Stat. Soc. B*, 29:392–393, 1963.
80. Friedrich Hirzebruch und Winfried Scharlau. *Einführung in die Funktionalanalysis*. Bibliographisches Institut, Mannheim, 1971. B. I.-Hochschultaschenbücher, No. 296.
81. Frank den Hollander. *Large deviations*, Band 14 aus *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000.

82. Richard A. Holley und Thomas M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.*, 3(4):643–663, 1975.
83. Barry D. Hughes. *Random walks and random environments. Vol. 1.* Oxford Science Publications. The Clarendon Press/Oxford University Press, New York, 1995. Random walks.
84. Barry D. Hughes. *Random walks and random environments. Vol. 2.* Oxford Science Publications. The Clarendon Press/Oxford University Press, New York, 1996. Random environments.
85. Nobuyuki Ikeda und Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, Band 24 aus *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, 2. Auflage, 1989.
86. Albrecht Irle. *Finanzmathematik*. Teubner Studienbücher Mathematik. B. G. Teubner, Stuttgart, 1998.
87. Jürgen Jost. *Partial differential equations*, Band 214 aus *Graduate Texts in Mathematics*. Springer-Verlag, New York, 3. Auflage, 2013.
88. G. S. Joyce. Singular behaviour of the lattice Green function for the d -dimensional hypercubic lattice. *J. Phys. A*, 36(4):911–921, 2003.
89. Shizuo Kakutani. Examples of ergodic measure preserving transformations which are weakly mising but not strongly mixing. In *Recent advances in topological dynamics (Proceedings of the Conference at Yale University, New Haven, CT, 1972, in honor of Gustav Arnold Hedlund)*, Band 318 aus *Lecture Notes in Mathematics*, Seiten 143–149. Springer, Berlin, 1973.
90. Olav Kallenberg. *Random measures*. Akademie-Verlag, Berlin, 4. Auflage, 1986.
91. Olav Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer-Verlag, New York, Berlin, 2. Auflage, 2002.
92. L. V. Kantorovič und G. Š. Rubinštejn. On a space of completely additive functions. *Vestnik Leningrad Univ.*, 13(7):52–59, 1958.
93. Ioannis Karatzas und Steven E. Shreve. *Brownian motion and stochastic calculus*, Band 113 aus *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2. Auflage, 1991.
94. Ioannis Karatzas und Steven E. Shreve. *Methods of mathematical finance*, Band 39 aus *Applications of Mathematics*. Springer-Verlag, New York, 1998.
95. Gerhard Keller. *Equilibrium states in ergodic theory*, Band 42 aus *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1998.
96. Gerhard Keller. *Wahrscheinlichkeitstheorie*. Vorlesungsskript. Universität Erlangen, 2003.
97. John L. Kelley. *General topology*, Band 27 aus *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1975. Nachdruck der Ausgabe von 1955 [Van Nostrand, Toronto, Ontario].
98. John G. Kemeny und J. Laurie Snell. *Finite Markov chains*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1976. Reprinting of the 1960 original.
99. Richard W. Kenyon, James G. Propp und David B. Wilson. Trees and matchings. *Electron. J. Combin.*, 7:Research Paper 25, 34 pp. (electronic), 2000.
100. H. Kesten, M. V. Kozlov und F. Spitzer. A limit law for random walk in a random environment. *Compositio Math.*, 30:145–168, 1975.
101. Harry Kesten. Sums of stationary sequences cannot grow slower than linearly. *Proc. Amer. Math. Soc.*, 49:205–211, 1975.
102. Harry Kesten. The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. *Comm. Math. Phys.*, 74(1):41–59, 1980.
103. Harry Kesten und Bernt P. Stigum. A limit theorem for multidimensional Galton-Watson processes. *Ann. Math. Statist.*, 37:1211–1223, 1966.

104. Aleksandr Khintchine. Über dyadische Brüche. *Mathematische Zeitschrift*, 18:109–116, 1923.
105. J. F. C. Kingman. Uses of exchangeability. *Ann. Probability*, 6(2):183–197, 1978.
106. J. F. C. Kingman. *Poisson processes*, Band 3 aus *Oxford Studies in Probability*. The Clarendon Press/Oxford University Press, New York, 1993. Oxford Science Publications.
107. Achim Klenke und Lutz Mattner. Stochastic ordering of classical discrete distributions. *Adv. in Appl. Probab.*, 42(2):392–410, 2010.
108. A. N. Kolmogorov. Sulla determinazione empirica di una legge di distibuzione. *Gior- nale Istituto Italiano degli Attuari*, 4:83–91, 1933.
109. Ralf Korn und Elke Korn. *Optionsbewertung und Portfolio-Optimierung*. Friedr. Vieweg & Sohn, Braunschweig, 1999.
110. Ulrich Krengel. *Ergodic theorems*, Band 6 aus *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1985.
111. Ulrich Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Friedr. Vieweg & Sohn, Braunschweig, 2005.
112. S. Kullback und R. A. Leibler. On information and sufficiency. *Ann. Math. Statistics*, 22:79–86, 1951.
113. Thomas Kurtz, Russell Lyons, Robin Pemantle und Yuval Peres. A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes. In *Classical and modern branching processes (Minneapolis, MN, 1994)*, Band 84 aus *IMA Volumes in Mathematics and Its Applications*, Seiten 181–185. Springer, New York, 1997.
114. Steffen L. Lauritzen. *Extremal families and systems of sufficient statistics*, Band 49 aus *Lecture Notes in Statistics*. Springer-Verlag, New York, 1988.
115. Paul Lévy. *Théorie de l'Addition des Variables Aléatoires*. Gauthier-Villars, Paris, 1937.
116. Paul Lévy. *Processus Stochastiques et Mouvement Brownien. Suivi d'une note de M. Loève*. Gauthier-Villars, Paris, 1948.
117. Jarl Waldemar Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15:211–225, 1922.
118. Jarl Waldemar Lindeberg. Sur la loi de Gauss. *C.R. Acad. Sci. Paris*, 174:1400–1402, 1922.
119. Torgny Lindvall. Convergence of critical Galton-Watson branching processes. *J. Appl. Probability*, 9:445–450, 1972.
120. Russell Lyons, Robin Pemantle und Yuval Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23(3):1125–1138, 1995.
121. Russell Lyons und Yuval Peres. *Probability on Trees*. 2012. Vorabversion im Internet unter <http://mypage.iu.edu/~rdlyons/>.
122. Neal Madras. *Lectures on Monte Carlo methods*, Band 16 aus *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2002.
123. Dimitri E. Menchoff. Sur les séries des fonctions orthogonales (première partie). *Fund. Math.*, 4:92–105, 1923.
124. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller und E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.
125. Paul-A. Meyer. *Probability and potentials*. Blaisdell Publishing Co. Ginn and Co., Waltham, Mass.-Toronto, Ont.-London, 1966.
126. S. P. Meyn und R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

127. Ferenc Móricz und Károly Tandori. An improved Menshov-Rademacher theorem. *Proc. Amer. Math. Soc.*, 124(3):877–885, 1996.
128. Peter Mörters und Yuval Peres. *Brownian motion*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010. Mit einem Anhang von Oded Schramm und Wendelin Werner.
129. Rajeev Motwani und Prabhakar Raghavan. *Randomized algorithms*. Cambridge University Press, Cambridge, 1995.
130. Alfred Müller und Dietrich Stoyan. *Comparison methods for stochastic models and risks*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2002.
131. John von Neumann. Proof of the quasi-ergodic hypothesis. *Proc. Nat. Acad. Sci.*, 18:70–82, 1932.
132. J. R. Norris. *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998. Nachdruck der Originalausgabe von 1997.
133. Esa Nummelin. *General irreducible Markov chains and nonnegative operators*, Band 83 aus *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1984.
134. Raymond E. A. C. Paley und Norbert Wiener. *Fourier transforms in the complex domain*, Band 19 aus *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1987. Reprint of the 1934 original.
135. Raymond E. A. C. Paley, Norbert Wiener und Antoni Zygmund. Note on random functions. *Math. Zeit.*, 37:647–668, 1933.
136. Ronald F. Peierls. On Ising’s model of ferromagnetism. *Proc. Cambridge Phil. Soc.*, 32:477–481, 1936.
137. Valentin V. Petrov. *Sums of independent random variables*, Band 82 aus *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, New York, 1975.
138. Jim Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields*, 102(2):145–158, 1995.
139. Jim Pitman. *Combinatorial stochastic processes*, Band 1875 aus *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Vorlesungen von der 32ten Sommerschule über Wahrscheinlichkeitstheorie in Saint-Flour, 7.–24. Juli 2002, Mit einem Vorwort von Jean Picard.
140. Jim Pitman und Marc Yor. Bessel processes and infinitely divisible laws. In *Stochastic integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, Band 851 aus *Lecture Notes in Mathematics*, Seiten 285–370. Springer, Berlin, 1981.
141. Jim Pitman und Marc Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900, 1997.
142. Jim Pitman und Marc Yor. On the distribution of ranked heights of excursions of a Brownian bridge. *Ann. Probab.*, 29(1):361–384, 2001.
143. George Pólya. Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. *Math. Ann.*, 84:149–160, 1921.
144. George Pólya. Sur quelques points de la théorie de probabilités. *Ann. Inst. H. Poincaré*, 1:117–161, 1931.
145. Yu. V. Prohorov. Convergence of random processes and limit theorems in probability theory. *Teor. Veroyatnost. i Primenen.*, 1:177–238, 1956. Russisch mit englischer Zusammenfassung.
146. James Propp und David Wilson. Coupling from the past: A user’s guide. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, Band 41 aus *DIMACS Series in*

- Discrete Mathematics and Theoretical Computer Science*, Seiten 181–192. American Mathematical Society, Providence, RI, 1998.
147. James Gary Propp und David Bruce Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms*, 9(1–2):223–252, 1996.
 148. James Gary Propp und David Bruce Wilson. How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *J. Algorithms*, 27(2):170–217, 1998. 7th Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996).
 149. Philip E. Protter. *Stochastic integration and differential equations*, Band 21 aus *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 2. Auflage, 2004. Stochastic Modelling and Applied Probability.
 150. Boto von Querenburg. *Mengentheoretische Topologie*. Springer-Verlag, Berlin, 1976. Korrigierter Nachdruck der ersten Auflage, Hochschultext.
 151. Hans Rademacher. Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen. *Math. Ann.*, 87:112–138, 1922.
 152. Pál Révész. *Random walk in random and non-random environments*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second Auflage, 2005.
 153. D. Revuz. *Markov chains*, Band 11 aus *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, 2. Auflage, 1984.
 154. Daniel Revuz und Marc Yor. *Continuous martingales and Brownian motion*, Band 293 aus *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 3. Auflage, 1999.
 155. R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, NJ, 1970.
 156. L. C. G. Rogers und David Williams. *Diffusions, Markov processes, and martingales. Vol. 1: Foundations*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Nachdruck der zweiten Auflage von 1994.
 157. L. C. G. Rogers und David Williams. *Diffusions, Markov processes, and martingales. Vol. 2: Itô calculus*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Nachdruck der zweiten Auflage von 1994.
 158. Walter Rudin. *Analysis*. Oldenbourg Verlag, München, Wien, 1988.
 159. Ivan Nicolaevich Sanov. On the probability of large deviations of random magnitudes. *Mat. Sb. N. S.*, 42 (84):11–44, 1957. Russisch.
 160. Ivan Nicolaevich Sanov. On the probability of large deviations of random variables. In *Select. Transl. Math. Statist. and Probability, Vol. 1*, Seiten 213–244. Inst. Math. Statist. and Amer. Math. Soc., Providence, RI, 1961.
 161. René L. Schilling und Lothar Partzsch. *Brownian motion*. De Gruyter, Berlin, 2012. An introduction to stochastic processes, With a chapter on simulation by Björn Böttcher.
 162. E. Seneta. *Non-negative matrices and Markov chains*. Springer Series in Statistics. Springer, New York, 2006. Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544].
 163. Tokuzo Shiga und Akinobu Shimizu. Infinite-dimensional stochastic differential equations and their applications. *J. Math. Kyoto Univ.*, 20(3):395–416, 1980.
 164. Albert N. Shiryaev. *Probability*, Band 95 aus *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2. Auflage, 1996. Übersetzung der ersten russischen Ausgabe von 1980.
 165. Ja. Sinaĭ. On the concept of entropy for a dynamic system. *Dokl. Akad. Nauk SSSR*, 124:768–771, 1959.

166. N. V. Smirnov. Sur les écarts de la courbe de distribution empirique. *Matematicheskiy Sbornik, Rossijskaya Akademiya Nauk, Moscow*, 2:3–16, 1939. Russisch mit französischer Zusammenfassung.
167. Fred Solomon. Random walks in a random environment. *Ann. Probab.*, 3:1–31, 1975.
168. Frank Spitzer. *Principles of random walks*, Band 34 aus *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2. Auflage, 1976.
169. V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
170. Daniel W. Stroock und S. R. Srinivasa Varadhan. Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.*, 24:147–225, 1971.
171. Daniel W. Stroock und S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*, Band 233 aus *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1979.
172. J. J. Sylvester. Mathematical questions with their solutions. *Educ. Times*, 41:171–178, 1884.
173. Károly Tandori. Über die orthogonalen Funktionen. I. *Acta Sci. Math. Szeged*, 18:57–130, 1957.
174. Károly Tandori. Über die Divergenz der Orthogonalreihen. *Publ. Math. Debrecen*, 8:291–307, 1961.
175. Károly Tandori. Bemerkung über die paarweise unabhängigen zufälligen Größen. *Acta Math. Hungar.*, 48(3–4):357–359, 1986.
176. S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286, 1966.
177. Peter Walters. *An introduction to ergodic theory*, Band 79 aus *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.
178. G. N. Watson. Three triple integrals. *Quart. J. Math., Oxford Ser.*, 10:266–276, 1939.
179. Dirk Werner. *Funktionalanalysis*. Springer-Verlag, Heidelberg, 7. Auflage, 2011.
180. David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.
181. David Bruce Wilson und James Gary Propp. How to get an exact sample from a generic Markov chain and sample a random spanning tree from a directed graph, both within the cover time. In *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996)*, Seiten 448–457, ACM, New York, 1996.
182. Sewall Wright. Evolution in Mendelian populations. *Genetics*, 16:97–159, 1931.
183. A. M. Yaglom. Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)*, 56:795–798, 1947.
184. Toshio Yamada und Shinzo Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11:155–167, 1971.
185. Kōsaku Yosida. *Functional analysis*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Nachdruck der sechsten Auflage von 1980.
186. Ofer Zeitouni. Random walks in random environment. In *Lectures on probability theory and statistics*, Band 1837 aus *Lecture Notes in Mathematics*, Seiten 189–312. Springer, Berlin, 2004.

Notation

$\mathbb{1}_A$	Indikatorfunktion der Menge A
2^Ω	Potenzmenge, 1
$\#A$	Kardinalität der Menge A
A^c	Komplement $\Omega \setminus A$ der Menge $A \subset \Omega$, 1
$A \cap B$	Schnittmenge
$A \cup B$	Vereinigungsmenge
$A \uplus B$	disjunkte Vereinigungsmenge (eigentlich ist hierin eine Aussage enthalten)
$A \subset B$	A ist (nicht notwendigerweise echte) Teilmenge von B
$A \setminus B$	Differenzmenge
$A \triangle B$	symmetrische Differenz zweier Mengen, 30
$A \times B$	kartesisches Produkt von A und B
\mathcal{A}	Teilmenge von 2^Ω , typischerweise eine σ -Algebra, 1
$\mathcal{A} _B$	Spur-Mengensystem auf B , 11
$\mathcal{A} \otimes \mathcal{A}'$	Produkt der σ -Algebren \mathcal{A} und \mathcal{A}' , 280
$\mathcal{B}(E)$	Borel'sche σ -Algebra von E , 9
Ber_p	Bernoulliverteilung, 44
$\beta_{r,s}$	Beta-Verteilung mit Parametern r und s , 47
$b_{n,p}$	Binomialverteilung, 44, 309
$b_{r,p}^-$	negative Binomialverteilung, 45, 309
$C(E), C_b(E), C_c(E)$	Raum der stetigen (beschränkten) Funktionen, bzw. mit kompakten Träger, 252
\mathcal{C}_{qV}	Funktionen mit stetiger quadratischer Variation, 509
\mathbb{C}	Menge der komplexen Zahlen, 82
Cau_a	Cauchy Verteilung, 309

$\text{Cov}[X, Y]$	Kovarianz der Zufallsvariablen X und Y , 104
CPoi_ν	zusammengesetzte Poisson-Verteilung, 339
δ_x	Dirac-Verteilung, 12
$\mathbf{E}[X]$	Erwartungswert der Zufallsvariablen X , 103
$\mathbf{E}[X; A]$	$= \mathbf{E}[X \mathbb{1}_A]$, 175
$\mathbf{E}[X \mathcal{F}]$	bedingter Erwartungswert, 177
\exp_θ	Exponentialverteilung, 46, 309
$\mathbb{F} = (\mathcal{F}_t)_{t \in I}$	Filtration, 195
f.s, f.ü.	fast sicher und fast überall, 32
$G(x, y)$	Greenfunktion einer Markovkette, 375
$\Gamma_{\theta, r}$	Gammaverteilung mit Größenparameter $\theta > 0$ und Formparameter $r > 0$, 47, 309
$\gamma_p = b_{1,p}^-$	geometrische Verteilung mit Parameter p , 44
$\text{ggT}(M)$	größter gemeinsamer Teiler aller $m \in M \subset \mathbb{N}$, 398
$H \cdot X$	diskretes stochastisches Integral von H bezüglich X , 202
\mathcal{I}	Menge der invarianten Verteilungen einer Markovkette, 384
<i>i.i.d.</i>	<i>independent and identically distributed</i> , 57
$\text{Im}(z)$	Imaginärteil von $z \in \mathbb{C}$, 301
λ, λ^n	Lebesgue-Maß, n -dimensionales, 26
$\text{Lip}(E)$	Raum der Lipschitz-stetigen Funktionen auf E , 253
\mathcal{L}^p, L^p	Lebesgue'sche Räume p -fach integrierbarer Funktionen, 93, 147, 148
$\mathcal{L}(X)$	Verteilung der Zufallsvariablen X
$\mathcal{M}(E), \mathcal{M}_f(E), \mathcal{M}_{\leq 1}, \mathcal{M}_1(E)$	Menge der (endlichen bzw. (Sub-)W-) Maße auf E , 17, 252
$\mathcal{M}_{loc,c}$	Raum der stetigen lokalen Martingale, 511
$\mu \otimes \nu$	Produkt der Maße μ und ν , 28, 284
$\mu * \nu$	Faltung der Maße μ und ν , 62, 286
$\mu^{\otimes n}$	n -faches Produktmaß, 284
μ^{*n}	n -fache Faltungspotenz, 62
$\mu \ll \nu$	μ ist absolutstetig bezüglich ν , 160

$\mu \perp \nu$	μ ist singulär bezüglich ν , 160
$\mu \approx \nu$	μ und ν sind äquivalent, 160
$\mu \leq_{\text{st}} \nu$	μ ist stochastisch kleiner oder gleich ν , 393
\mathbb{N}, \mathbb{N}_0	$\mathbb{N} = \{1, 2, 3, \dots\}, \mathbb{N}_0 = \mathbb{N} \cup \{0\}$
$\mathcal{N}_{\mu, \sigma^2}$	Normalverteilung, 46, 309
$d\mu/d\nu$	Radon-Nikodym-Ableitung, 161
Ω	Raum der Elementarereignisse, auf dem \mathbf{P} definiert ist
\mathbf{P}	generisches Wahrscheinlichkeitsmaß
$\mathbf{P}[A B], \mathbf{P}[A \mathcal{F}]$	bedingte Wahrscheinlichkeiten, 174, 177
$\mathbf{P}_X = \mathbf{P} \circ X^{-1}$	Verteilung der Zufallsvariablen X , 43
Poi_λ	Poissonverteilung mit Parameter $\lambda \geq 0$, 45, 309
$p^n(x, y) = p^{(n)}(x, y)$	n -Schritt-Übergangswahrscheinlichkeiten einer Markovkette, 364
$\mathcal{P}_{S,T}^n, \mathcal{P}_T^n$	siehe Seite 509
φ_X	charakteristische Funktion der Zufallsvariablen X , 308
ψ_X	Erzeugendenfunktion der Zufallsvariablen X , 79
\mathbb{Q}	Menge der rationalen Zahlen
\mathbb{R}	Menge der reellen Zahlen
$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$	Zweipunktkompaktifizierung der reellen Zahlen
$\text{Re}(z)$	Realteil von $z \in \mathbb{C}$, 301
$\text{sign}(x)$	$= \mathbb{1}_{(0,\infty)}(x) - \mathbb{1}_{(-\infty,0)}(x)$, Vorzeichen von $x \in \mathbb{R}$, 38
$\sigma(\cdot)$	von \cdot erzeugte σ -Algebra oder Filtration, 6, 35, 195
τ_x^k	Zeit des k -ten Besuches einer Markovkette in x , 373
$\mathcal{T}(\cdot)$	terminale σ -Algebra, 64
\mathcal{U}_A	uniforme Verteilung auf A , 13, 33, 309
u.i.v.	unabhängig und identisch verteilt, 57
$V^1(G), V^2(G)$	Variation und quadratische Variation von G , 507, 509
$\text{Var}[X]$	Varianz der Zufallsvariablen X , 103
v-lim	vager Limes, 257
w-lim	schwacher Limes, 257
X^τ	in τ gestoppter Prozess, 214

$\langle X \rangle$	quadratischer Variationsprozess von X , 210, 509, 512, 516
$f(t) \sim g(t), t \rightarrow a$: $\iff \lim_{t \rightarrow a} f(t)/g(t) = 1$
$X \sim \mu$	Die Zufallsvariable X hat Verteilung μ , 43
$x \vee y, x \wedge y, x^+, x^-$	Maximum, Minimum, Positivteil, Negativteil reeller Zahlen, 38
$\lfloor x \rfloor, \lceil x \rceil$	Abgerundetes und Aufgerundetes von x , 37
\bar{z}	komplex konjugierte Zahl zu $z \in \mathbb{C}$, 301
\mathbb{Z}	Menge der ganzen Zahlen
$\stackrel{\mathcal{D}}{=}$	Gleichheit in Verteilung, 43
$\xrightarrow[n \rightarrow \infty]{\mathcal{D}}, \xrightarrow[n \rightarrow \infty]{}$	Konvergenz der Verteilungen, 260
$\xrightarrow[\text{fdd}]{n \rightarrow \infty}, \xrightarrow[\text{fdd}]{}$	Konvergenz der endlichdimensionalen Verteilungen, 494
$\xrightarrow{\text{stoch}}, \xrightarrow{\text{f.s.}}, \xrightarrow{\text{f.ü.}}$	stochastische Konvergenz, fast sichere, und fast überall, 134

Glossar englischer Ausdrücke

a.a. = almost all	fast alle
a.e. = almost everywhere	fast überall
a.s. = almost surely	fast sicher
array (of random variables)	Schema von Zufallsvariablen
backward martingale	Rückwärtsmartingal
bond, edge	Kante (eines Graphen)
Brownian motion	Brown'sche Bewegung
central limit theorem	Zentraler Grenzwertsatz
completion	Vervollständigung
compound Poisson	zusammengesetzt Poisson
conductivity	Leitfähigkeit
continuous	stetig
convolution	Faltung
decomposition	Zerlegung
density	Dichte
derivative	Ableitung
distribution	Verteilung
dominated convergence	majorisierte Konvergenz
dynamical system	Dynamisches System
expectation (conditional)	Erwartungswert (bedingter)
ergodic theorem	Ergodensatz
event	Ereignis
exchangeable	austauschbar
extension theorem	Fortsetzungssatz, Erweiterungssatz
flow (electric)	Fluss (elektrischer)
iff = if and only if	dann und nur dann, wenn
i.i.d. = independent and identically distributed	unabhängig und identisch verteilt
increment	Zuwachs
indistinguishable	ununterscheidbar
integer (number)	ganze Zahl
joint distribution	gemeinsame Verteilung
large deviation	große Abweichung

law	Verteilung
level set	Niveaumenge
Markov chain	Markovkette
(strong) Markov property	(starke) Markoveigenschaft
map	Abbildung
marginal (distribution)	Randverteilung
mean	Mittelwert
measurable space	Messraum
measure	Maß
measure preserving	maßerhaltend
mixing	mischend
modulus (of a number)	Absolutbetrag (einer Zahl)
modulus of continuity	Stetigkeitsmodul
null array	asymptotisch vernachlässigbares Schema
partition function	Zustandssumme
p.d.f. = probability distribution function	Verteilungsfunktion
p.g.f. = probability generating function	Erzeugendenfunktion
phase transition	Phasenübergang
predictable, previsible	previsibel, vorhersagbar
probability	Wahrscheinlichkeit
random walk	Irrfahrt
random variable	Zufallsvariable
representation	Darstellung
semigroup	Halbgruppe
σ -field	σ -Algebra
size-biased (sampling)	größenverzerrtes Ziehen einer Stichprobe
tight	straff
trace	Spur
transition kernel	Übergangskern
uniform distribution	Gleichverteilung
uniformly integrable	gleichgradig integrierbar
urn model	Urnensmodell
(probability) weight	(Wahrscheinlichkeits-)gewicht
vertex	Punkt/Knoten eines Graphen
w.p. = with probability	

Namensregister

- Banach, Stefan, 1892 (Krakau) – 1945
(Lemberg, Ukraine), 155
- Bayes, Thomas, 1702 (London) – 1761
(Tunbridge Wells, England), 174
- Bernoulli, Jakob, 1654 (Basel) – 1705
(Basel), 18
- Bienaymé, Irénée-Jules, 1796 (Paris) – 1878
(Paris), 106
- Blackwell, David, 1919, 109
- Bochner, Salomon, 1899 (Krakau) – 1982
(Houston, Texas), 319
- Boltzmann, Ludwig, 1844 (Wien) – 1906
(Duino bei Triest), 408
- Borel, Emile, 1871 (Saint-Affrique,
Frankreich) – 1956 (Paris), 9
- Brown, Robert, 1773 (Montrose, Scotland)
– 1858 (London), 474
- Cantelli, Francesco Paolo, 1875 (Palermo) –
1966 (Rom), 53
- Carathéodory, Constantin, 1873 (Berlin) –
1950 (München), 19
- Cauchy, Augustin Louis, 1789 (Paris) –
1857 (bei Paris), 107
- Cesàro, Ernesto, 1859 (Neapel) – 1906
(Torre Annunziata, Italien), 65
- Chebyshev, Pafnutij Lvovich (Чебышёв,
Пафнутий Льович), 1821 (Okatovo,
Russland) – 1894 (Sankt Petersburg),
110
- Cramér, Harald, 1893 (Stockholm) – 1985
(Stockholm), 335
- Curie, Pierre, 1859 (Paris) – 1906 (Paris),
549
- Dieudonné, Jean Alexandre 1906 (Lille,
Frankreich) – 1992 (Paris), 302
- Dirac, Paul Adrien Maurice, 1902 (Bristol)
– 1984 (Tallahassee, Florida), 12
- Dirichlet, Lejeune, 1805 (Düren) – 1859
(Göttingen), 421
- Doob, Joseph Leo, 1910 (Cincinnati, Ohio)
– 2004 (Urbana, Illinois), 209
- Dynkin, Eugene, 1924 (Petrograd, heute
Sankt Petersburg), 4
- Egorov, Dmitrij Fedorovich
(Егоров, Дмитрий Фёдорович),
1869 (Moskau) – 1931 (Kasan), 138
- Esseen, Carl-Gustav, 1918 (Linköping,
Schweden) – 2001 (Uppsala ?), 333
- Euler, Leonard, 1707 (Basel) – 1783 (Sankt
Petersburg), 52
- Fatou, Pierre, 1878 (Lorient, Frankreich) –
1929 (Pornichet, Frankreich), 95
- Feller, William, 1906 (Zagreb) – 1970 (New
York), 328
- Fischer, Ernst, 1875 (Wien) – 1954 (Köln),
155
- Fourier, Jean Baptiste Joseph, 1768
(Auxerre, Frankreich) – 1830 (Paris),
306
- Fréchet, Maurice René, 1878 (Maligny,
Frankreich) – 1973 (Paris), 155
- Fubini, Guido, 1879 (Venedig) – 1943 (New
York), 284
- Galton, Francis, 1822 (bei Birmingham) –
1911 (Grayshott House, England), 85
- Gauß, Carl-Friedrich, 1777 (Braunschweig)
– 1855 (Göttingen), 46
- Gibbs, Josiah Willard, 1839 (New Haven,
Connecticut) – 1903 (New Haven,
Connecticut), 411

- Green, George, 1793 (Nottingham) – 1841 (Nottingham), 375
- Hölder, Otto Ludwig, 1859 (Stuttgart) – 1937 (Leipzig), 154
- Hahn, Hans, 1879 (Wien) – 1934 (Wien), 165
- Helly, Eduard, 1884 (Wien) – 1943 (Chicago), 268
- Hesse, Ludwig Otto, 1814 (Königsberg, heute Kaliningrad, Russland) – 1874 (München), 152
- Hewitt, Edwin, 1920 (Everett, Washington), 242
- Hilbert, David, 1862 (Königsberg, heute Kaliningrad, Russland) – 1943 (Göttingen), 155
- Hopf, Eberhard, 1902 (Salzburg) – 1983, 453
- Ionescu-Tulcea, Cassius, 1923, 293
- Ising, Ernst, 1900 (Köln) – 1988 (Peoria, Illinois), 407
- Itô, Kiyoshi, 1915 (Hokusei-cho, Japan), 490
- Jensen, Johan Ludwig, 1859 (Nakskov, Dänemark) – 1925 (Kopenhagen), 152
- Jordan, Camille, 1838 (bei Lyon) – 1922 (Paris), 166
- Kesten, Harry, 1931, 73
- Khinchin, Aleksandr Jakovlevich (Хинчин, Александр Яковлевич) 1894 (Kondrovo, Russland) – 1959 (Moskau), 342
- Kirchhoff, Gustav Robert, 1824 (Königsberg) – 1887 (Berlin), 425
- Kolmogorov, Andrej Nikolaevich (Колмогоров, Андрей Николаевич), 1903 (Tambow, Russland) – 1987 (Moskau), 65
- Laplace, Pierre-Simon, 1749 (Beaumont-en-Auge, Normandie) – 1827 (Paris), 146
- Lebesgue, Henri Léon, 1875 (Beauvais, Oise, Frankreich) – 1941 (Paris), 18
- Legendre, Adrien-Marie, 1752 (Paris) – 1833 (Paris), 532
- Levi, Beppo, 1875 (Turin, Italien) – 1961 (Rosario, Santa Fe, Argentinien), 95
- Lévy, Paul Pierre, 1886 (Paris) – 1971 (Paris), 316, 522
- Lindeberg, Jarl Waldemar, 1876 – 1932, 327
- Lipschitz, Rudolph, 1832 (Königsberg, heute Kaliningrad, Russland) – 1903 (Bonn), 253
- Lusin, Nikolai Nikolaevich (Лусин, Николай Николаевич), 1883 (Irkutsk, Russland) – 1950 (Moskau), 255
- Lyapunov, Aleksandr Mikhajlovich (Ляпунов Александр Михайлович), 1857 (Jaroslavl, Russland) – 1918 (Odessa), 327
- Markov, Andrej Andreevich (Марков, Андрей Андреевич), 1856 (Ryazan, Russland) – 1922 (Petrograd, heute Sankt Petersburg), 110
- Menshov, Dmitrij Evgen'evich (Меншов, Дмитрий Евгеньевич), 1892 (Moskau) – 1988 (Moskau), 124
- Minkowski, Hermann, 1864 (Alexotas, heute: Kaunas, Litauen) – 1909 (Göttingen), 154
- Neumann, John von, 1903 (Budapest) – 1957 (Washington, D.C.), 161
- Nikodym, Otton Marcin, 1889 (Zablotow, Galizien, Ukraine) – 1974 (Utica, New York), 161
- Ohm, Georg Simon, 1789 (Erlangen) – 1854 (München), 425
- Ornstein, Leonard Salomon, 1880 (Nijmegen) – 1941 (Utrecht), 599
- Paley, Raymond E. A. C., 1907 (Bournemouth, England) – 1933 (Banff, Alberta), 477
- Parseval, Marc-Antoine, 1755 (Rosières-aux-Salines, Frankreich) – 1836 (Paris), 485
- Pascal, Blaise, 1623 (Clermont-Ferrand, Frankreich) – 1662 (Paris), 45

- Plancherel, Michel, 1885 (Bussy (Fribourg), Schweiz) – 1967 (Zürich), 307
- Poisson, Siméon Denis, 1781 (Pithiviers, Frankreich) – 1840 (bei Paris), 45
- Pólya, George, 1887 (Budapest) – 1985 (Palo Alto), 318
- Prohorov, Jurij Vasil'evich (Прохоров, Юрий Васильевич), 1929, 266
- Rademacher, Hans, 1892 (Hamburg-Wandsbek) – 1969 (Haverford, Pennsylvania), 124
- Radon, Johann, 1887 (Tetschen, Böhmen) – 1956 (Wien), 161
- Riemann, Georg Friedrich Bernhard, 1826 (Breselenz, Kreis Lüchow-Dannenberg) – 1866 (Selasca, Italien), 52
- Riesz, Frigyes, 1880 (Györ, Ungarn) – 1956 (Budapest), 155
- Saks, Stanislav (Сакс, Станислав), 1897 (Kalish, Russland (heute Polen)) – 1942 (Warschau, von der Gestapo ermordet), 234
- Savage, Jimmie Leonard, 1917 (Detroit, Michigan) – 1971 (New Haven, Connecticut), 242
- Schwarz, Hermann Amandus, 1843 (Hermsdorf, Schlesien) – 1921 (Berlin), 107
- Skorohod, Anatolii Volodymyrovych (Скороход, Анатолій Володимирович), 1930 (Nikopo, Ukraine) – 2011 (Lansing, Michigan), 392
- Slutzky, Evgenij Evgen'evich (Слуцкий, Евгений Евгеньевич), 1880 (Novoe, Gouvernement Jaroslavl, Russland) – 1948 (Moskau), 260
- Stieltjes, Thomas Jan, 1856 (Zwolle, Overijssel) – 1894 (Toulouse), 26
- Stone, Marshall Harvey, 1903 (New York) – 1989 (Madras, Indien), 302
- Thomson, William (Lord Kelvin), 1824 (Belfast) – 1907 (Largs, Ayrshire, Schottland), 429
- Uhlenbeck, George Eugene, 1900 (Batavia, heutiges Jakarta) – 1988 (Boulder, Colorado), 599
- Varadhan, S.R. Srinivasa, 1945 (Madras, Indien), 543
- Watson, George Neville, 1886 (Westward Ho, England) – 1965 (Leamington Spa, England), 382
- Watson, Henry William, 1827 (bei London) – 1903 (bei Coventry), 85
- Weierstraß, Karl, 1815 (Ostenfelde, Westfalen) – 1897 (Berlin), 302
- Weiss, Pierre-Ernest, 1865 (Mulhouse, Frankreich) – 1940 (Lyon), 546
- Wiener, Norbert, 1894 (Columbia, Missouri) – 1964 (Stockholm), 494
- Wintner, Aurel Friedrich, 1903 (Budapest) – 1958 (Baltimore), 527
- Wright, Sewall, 1889 (Melrose, Massachusetts) – 1988 (Madison, Wisconsin), 367
- Yaglom, Akiva Moiseevich (Яглом, Аківа Моїсеєвич), 1921 (Kharkov), 234
- Zygmund, Antoni, 1900 (Warschau) – 1992 (Chicago), 477

Sachregister

- 0-1 Gesetze
 - Blumenthal 476
 - für invariante Ereignisse 461
 - Hewitt-Savage 242
 - Kolmogorov 65
 - Ø-stetig 16
- abgeschlossen 8
- Abschluss 250
- absolutstetig 160
- absorbierend 374
- adaptiert 195
- additiv 12
- Algebra 3, 302
- Anziehungsbereich einer Verteilung 353
- aperiodisch 398
- Approximationssatz für Maße 30
- äquivalente Maße 160
- äquivalentes Martingalmaß 207
- Arbitrage 206
- Arkussinus-Gesetz 480
- asymptotisch vernachlässigbar 327
- Aufkreuzung 223
- äußeres Maß 21
- austauschbar 235
- austauschbare σ -Algebra 237
- Auswertungsabbildung 492
- Azuma'sche Ungleichung 202
- Banachraum 155
- Bayes'sche Formel 174, 183
- bedingte
 - Erwartung 177
 - Unabhängigkeit 243
 - Verteilung 186
 - Wahrscheinlichkeit 174, 177
- Benford'sches Gesetz 457
- Bernoulli-Maß 29
- Bernoulli-Verteilung 44
- Bernstein-Chernov Abschätzung 112
- Bernstein-Polynom 112
- Berry-Esseen, Satz von 333
- beschränkt in L^p 141
- Bessel-Prozess 608
- Beta-Verteilung 47, 247, 324, 562
 - Momente 110
- Bienaymé-Gleichung 106
- Bildmaß 41
- binär aufspaltender stochastischer Prozess 205
- binäres Modell 205
- Binomialverteilung 44
- Black-Scholes Formel 207
- Black-Scholes Modell 600
- Blackwell-Girshick 109
- Blumenthal'sches 0-1 Gesetz 476
- Bochner 319
- Boltzmann-Verteilung 408, 546
- Borel-Cantelli Lemma 53
 - bedingte Version 232
- Borel-Maß 251
- Borel'scher Raum 189
- Borel'sche σ -Algebra 9
- Borel'sches Paradoxon 191
- Box-Muller Methode 63
- Brown'sche Bewegung 299, 474
 - Existenzsatz 474
 - kanonische 494
 - Karhunen-Loëve Darstellung 489
 - Lévy Charakterisierung 586
 - Paley-Wiener Darstellung 489
 - Skalierungseigenschaft 475
- Brown'sche Brücke 475, 491, 500, 606
- Brown'sches Blatt 491
- càdlàg 482

- Call 206
- Carathéodory 19
- Cauchy-Schwarz'sche Ungleichung 107
 - bedingte 183
- Cauchy-Verteilung 47, 309, 593
- Cesàro-Limes 65
- CFW 337
- Chapman-Kolmogorov'sche Gleichung 297, 364
 - charakteristische Funktion 305, 553
 - Inversionsformel 306
- Chebyshev Polynom 417
- Chebyshev'sche Ungleichung 110
- Chernov Abschätzung *siehe* Bernstein-Chernov Abschätzung
- China-Restaurant Prozess 567
- Cholesky-Faktorisierung 336
- Chung-Fuchs, Satz von 382, 458
- Claim 206
- Continuous Mapping Theorem 262
- Cox-Ingersoll-Ross Modell 604
- Cox-Ross-Rubinstein'sches Modell 207
- Cramér-Lundberg'sche Ungleichung 217
- Cramér-Transformierte 533
- Cramér-Wold Device 335
- Curie-Temperatur 408, 549
- Curie-Weiss'sches Gesetz 549
 - detaillierte Balance 423
 - Diagonalfolgenargument 269
 - dicht 250
 - Dichte 13, 27, 46, 59, 93, 159
 - Dichtetransformationsformel
 - mehrdimensional 41
 - Differentiationslemma 145
 - Diffusionsprozess 582
 - Dirac-Maß 12
 - Dirichlet-Problem 591
 - diskretes 421
 - Dirichlet'sches Prinzip 429
 - Dirichlet-Verteilung 562
 - domain of attraction 353
 - Donsker, Satz von 498
 - Doob'sche Regularisierung 482
 - Doob'sche Ungleichung 222
 - Doob-Zerlegung 209
 - Dreireihensatz 332
 - Drift 582
 - Dualität 613
 - Dualraum 169
 - dynamisches System 450
 - Dynkin-System 4
 - einfache Irrfahrt 424
 - Einheitsmasse 12
 - Einschluss-Ausschlussformel 15
 - Einschränkung 11
 - Eintrittszeit 373
 - elektrischer Fluss 425
 - Elementarfunktion 39
 - empirische Verteilung 245
 - empirische Verteilungsfunktion 117
 - Entropie 118, 120, 539
 - eines dynamischen Systems 463, 465
 - Kolmogorov-Sinai 465
 - relative 539
 - Ereignis 17, 43
 - invariantes 74
 - Ergodensatz
 - Individueller (Birkhoff) 453
 - Statistischer (von Neumann) 455
 - ergodisch 450
 - Erwartungswert 103
 - Erzeugendenfunktion 79
 - Erzeuger 6
 - erzeugte σ -Algebra 6, 34
 - Etemadi
 - Ungleichung von 125
 - Euler'sche Primzahlformel 52
 - Explosion 371
 - Exponentialverteilung 46
 - f.a. *siehe* fast alle
 - Faktorisierungslemma 40
 - Falle 420
 - Faltung
 - Dichten 286
 - diskrete Verteilungen 61
 - Maße auf \mathbb{R}^n 62, 286
 - Faltungshalbgruppe 300
 - Färbungssatz 560
 - fast alle 32
 - fast sicher 32
 - fast überall 32
 - Fatou, Lemma von 95
 - Feinheit 509
 - Feller-Eigenschaft 483
 - starke 612

- Feller-Prozess 484
- Feller'sche Halbgruppe 483
- Feller'sche Verzweigungsdiffusion 504, 604, 617
- Filtration 195
 - rechtsstetige 482
 - übliche Bedingungen 482
- de Finetti, Satz von 244, 275
- Fischer-Riesz, Satz von 155
- Fluchtwahrscheinlichkeit 431
- Fluss 425
- Fortsetzungssatz für Maße 19, **24**
- Fourier-Inversionsformel 306
- Fourier-Reihen 158
- Fréchet-Shohat, Satz von 325
- freie Energie 546
- Frobenius Problem 399
- f.s. *siehe* fast sicher
- f.ü. *siehe* fast überall
- Fubini, Satz von 285
 - für Itô-Integrale 591
 - für Übergangskerne 290
- Funktionaler Zentraler Grenzwertsatz 497
- Galton-Watson-Prozess 85
 - Reskalierung 501
- Gambler's Ruin 216, 415
- Gamma-Verteilung 47
 - Lévy-Maß 344
 - Subordinator 563
- Gedächtnislosigkeit der Exponentialverteilung 176
- GEM-Verteilung 566, 567
- gemeinsame Dichte 59
- gemeinsame Verteilung 58
- gemeinsame Verteilungsfunktion 58
- Generator 369
- geometrische Brown'sche Bewegung 600
- geometrische Verteilung 45
- Gesetz der großen Zahl
 - Konvergenzraten 122
 - schwaches 110
 - starkes 110, **114**, 241
- gestoppter Prozess 214
- Gewichtsfunktion 13
- Gibbs-Sampler 411
- gitterverteilt 315
- gleichgradig gleichmäßig stetig 315
- gleichgradig integrierbar 138
- Gleichverteilung 13, 33
- gleitendes Mittel 195, 450
- größenverzerrte Verteilung 274
- Graph 67
- Greenfunktion 375, 421
 - Tabelle 383
- Gronwall Lemma 601
- große Abweichungen 531
- Haar-Funktionen 486
- Hahn'scher Zerlegungssatz 165
- Halbring 3
- halbstetig von unten 535
- haploid 367
- harmonische Funktion 384, **420**
- harmonisches Maß 592
- Hartman-Wintner, Satz von 527
- Hauptsatz der Differential- und Integralrechnung 256
- heat bath algorithm 411
- Hedge 206
- Helly, Satz von 268
- Helmholtz-Potential 546
- Hilbertraum 155
- Hilbert-Schmidt Norm 600
- Hilbert-Schmidt Operator 291
- Hölder'sche Ungleichung 154
- Hölder-stetig 468
- Hopf 453
- hypergeometrische Verteilung **45**
- identisch verteilt 43
- i.i.d. *siehe* u.i.v.
- Indikatorfunktion 5
- Inhalt 12
- Innernes 250
- integrierbar 90
- Integral 87, **88**, **90**, 91
 - Lebesgue 93, 98
 - Riemann 98
 - stochastisches 490
- integrierbar 103
 - quadrat 103
 - stochastischer Prozess 194
- Intensitätsmaß 552
- invariantes Ereignis 450
- Invarianzprinzip von Donsker 498
- inverse Temperatur 546
- Inversionsformel 306

- Irrfahrt 358
 - auf einem Graphen 424
 - Greenfunktion (Tabelle) 383
 - in zufälliger Umgebung 447
 - Range 457
 - Rekurrenz 377
 - Satz von Chung-Fuchs 458
 - Satz von Pólya 378
 - symmetrische 194
- Ising-Modell 407, 412
- Iterierter Logarithmus
 - Brown'sche Bewegung 519
 - Hartman-Wintner 527
- Itô-Formel 584
 - diskrete 212
 - mehrdimensionale 590
 - pfadweise 585
- Itô-Integral 575
 - Produktregel 589
 - Satz von Fubini 591
- Itô-Prozess 582
- Jensen'sche Ungleichung 152, 181
- Jordan, Satz von 166
- kanonische Brown'sche Bewegung 494
- kanonischer Prozess 281
- kanonisches Maß 342, 345, 560
- Kantenperkolation 68, 419
- Karhunen-Loëve Darstellung der Brown'schen Bewegung 489
- Kaufoption 206
- Kelvin *siehe* Thomson
- Kesten-Stigum, Satz von 234
- Khinchin'sches Gesetz vom iterierten Logarithmus 527
- Kirchhoff'sches Gesetz 425
- Kolmogorov-Chentsov, Satz von 470
- Kolmogorov-Sinai Entropie 465
- Kolmogorov-Sinai, Satz von 465
- Kolmogorov'sche Ungleichung 123
- Kolmogorov'scher Dreireihensatz 332
- Kolmogorov'scher Erweiterungssatz 295
- Kolmogorov'sches 0-1 Gesetz 65
- Kolmogorov'sches Kriterium für schwache Relativkompaktheit 496
- Kolmogorov-Smirnov Test 500
- komplementstabil 1
- konkave Funktion 150
- Kontraktionsprinzip 543
- Konvergenz
 - dem Maße nach 134
 - fast sichere 134
 - fast überall 134
 - im Mittel 135
 - im p -ten Mittel 148
 - in Verteilung 260
 - majorisierte 144
 - schnelle 136
 - schwache 83, **256**
 - stochastische 134
 - vage 256
 - von Verteilungsfunktionen 261
- konvexe Funktion 150
- konvexe Menge 149
- Koordinatenabbildung 280
- Kopplung 70, 391
- Kopplung aus der Vergangenheit 413
- korreliert 104
- Kovarianz 104
- Kovarianzfunktion 475
- Kullback-Leibler Information 539
- Ladungsverteilung 164
- λ -System *siehe* Dynkin-System
- langsam variierend 354
- Laplace-Operator 588
- Laplace-Raum 13
- Laplace-Transformation 146, **304**, 502, 553
- Large Deviations *siehe* Prinzip großer Abweichungen
- LDP *siehe* Prinzip großer Abweichungen
- Lebesgue-Borel-Maß *siehe* Lebesgue-Maß
- Lebesgue-Integral 93
- Lebesgue-Maß **26**, 33
- Lebesgue'scher Konvergenzsatz 144
- Lebesgue'scher Zerlegungssatz 160
- Lebesgue-Stieltjes Integral 507
- Lebesgue-Stieltjes-Maß 26
- Legendre-Transformierte 532
- Leistung (elektrisches Netzwerk) 429
- Leitfähigkeit 424
- Lévy-Abstand 263
- Lévy-Khinchin Formel 342, 345
 - für zufällige Maße 560

- Lévy-Konstruktion der Brown'schen Bewegung 486
- Lévy-Maß 342, 345
 - allgemeine stabile Verteilung 350
 - Cauchy-Verteilung 348
 - Gamma-Verteilung 344
 - symmetrische stabile Verteilung 349
- Lévy'scher Stetigkeitsmodul 522
- Lévy'scher Stetigkeitssatz 316
- Limes inferior 5
- Lindeberg-Bedingung 327
- Lindvall, Satz von 506
- Lipschitz-stetig 253
- logarithmische momentenerzeugende Funktion 532
- Log-Normalverteilung 304
- lokal beschränkt 203
- lokal endlich 251
- lokales Martingal 511
- lokalisierende Folge 511
- lokalkompakt 250
- Lokalzeit 211
- L^p -beschränkt 141
- Lusin 255
- Lusin, Satz von 42
- $\mathcal{L}V$ 164
- Lyapunov-Bedingung 327

- Markoveigenschaft
 - elementare 357
 - schwache 358
 - starke 362
- Markovkern 185
- Markovkette 358
 - aperiodische 398
 - diskrete 364
 - invariante Verteilung 384
 - invariantes Maß 384
 - irreduzibel 377
 - Konvergenzgeschwindigkeit 413
 - Konvergenzsatz 405
 - Kopplung 401
 - Monte Carlo Methode 406
 - nullrekurrent 374
 - Periode eines Punktes 398
 - positiv rekurrent 374
 - rekurrent 374
 - reversible 423
 - schwach irreduzibel 377

- transient 374
- unabängiges Verschmelzen 401
- Markovprozess 358
- Markov'sche Halbgruppe 297
- Markov'sche Ungleichung 110
 - bedingte 183
- Martingal 198
 - Konvergenzsatz (L^1) 225
 - Konvergenzsatz (L^p) 226
 - Konvergenzsatz (f.s.) 224
 - Konvergenzsatz (rückwärts) 241
 - Konvergenzsätze (RCLL) 484
 - lokales 511
 - quadratische Variation 210
 - Rückwärts- 240
- Martingaldarstellungssatz 588
- Martingalproblem 609
 - diskretes 368
 - gut gestelltes 611
- Martingaltransformierte 202
- Maß 12
 - äußeres 21
 - Bernoulli 29
 - Borel 251
 - Einschränkung 33
 - harmonisches 592
 - invariantes 384
 - Lebesgue 26
 - lokal endliches 251
 - Produkt- 29, 296
 - Radon 251
 - reguläres 251
 - σ -endliches 12
 - signiertes 164
 - stationäres 384
 - Wahrscheinlichkeits- 12
- Maßraum **17**
- maßtreue Abbildung 450
- Maximal-Ergodenlemma 453
- MCMC 406
- mean field 546
- mehrstufiges Binomialmodell 207
- Mellin-Transformierte 307
- messbar
 - Abbildung 34
 - Borel 9
 - Lebesgue 33
 - μ - 22

- Menge 17
- Messraum **17**
 - Isomorphie 189
- Metrik
 - auf $C([0, \infty))$ 492
 - Lévy 263
 - Prohorov 257
 - stochastische Konvergenz 135
 - vollständige 250
 - Wasserstein 391
- metrisierbar 250
- Metropolis-Algorithmus 407
- Minkowski'sche Ungleichung 154
- mischend 461
- Modifikation 467
- Momente 103
 - absolute 103
- Momentenproblem 322
- monoton 12
- Monotonieprinzip von Rayleigh 428
- Monte Carlo Simulation 117
- Moran-Gamma-Subordinator 563
- Moran-Modell 367
- de Morgan'sche Regeln 2
- moving average 450
- Multinomialkoeffizient 63
- Multinomialverteilung 63

- negative Binomialverteilung 45, 81
 - stochastische Ordnung 562
- Niveaumenge 535
- Normalverteilung 46
 - mehrdimensionale 46, 334
- Nullmenge 32
- nullrekurrent 374

- offen 8
- Ohm'sches Gesetz 425
- Optional Sampling Theorem **213**, 218
 - stetige Zeit 473
- Optional Stopping Theorem **214**
 - stetige Zeit 473
- Ornstein-Uhlenbeck Prozess 599
- orthogonale Polynome 418
- orthogonales Komplement 156

- Paley Wiener Darstellung der Brown'schen Bewegung 489
- Parseval'sche Gleichung 485

- partiell stetig 316
- Partitionsfunktion 408, 546
- Pascal-Verteilung 45
- perfekte Simulation 412
- Periode 398
- Perkolation **67**, 419
- Petersburger Spiel 96, 195, 203
- Pfad 469
- pfadweise eindeutig 609
- Phasenübergang 408, 548
- π -System *siehe* schnittstabil
- Plancherel'sche Gleichung 307
- Poisson-Approximation 83
- Poisson-Dirichlet-Verteilung 564, 567
- Poissonprozess 127, 359
- Poisson'sche Summationsformel 481
- Poisson'scher Punktprozess 553
- Poisson-Verteilung 45
 - zusammengesetzte 339
- polare Menge 596
- Polarisationsformel 510
- polnischer Raum 189, 251
- Pólya, Satz von 318, 378
- Pólya'sches Urnenmodell 246, 296, 562
 - verallgemeinertes 371, 373
- Portemanteau-Theorem 259
- positiv rekurrent 374
- positiv semidefinit 318
- Präfixcode 119
- Prämaß 12
- previsibel 195, 574
- Prinzip großer Abweichungen 535
- Produktmaß 28, 29, 284, 294, 296
- produktmessbar 574
- Produktraum 280
- Produkt- σ -Algebra 280
- Produkttopologie 280
- progressiv messbar 574
- Prohorov 266
- Prohorov-Metrik 257, 405
- projektive Familie 294
- projektiver Limes 295
- Propp-Wilson Algorithmus 412
- Punkte trennend 302
- Punktperkolation 68

- Q-Q-Plot 334
- Q -Matrix 369
- Quader 9

- quadratintegrierbar 103
- quadratische Variation 509
- quadratischer Kovariationsprozess 516
- quadratischer Variationsprozess 210, 513
- Quellenkodierungssatz 120

- Radon-Maß 251
- Radon-Nikodym-Ableitung 161
- Rand 250
- random walk in random environment 447
- Ratenfunktion 530, **535**
- Rayleigh'sches Monotonieprinzip 428
- RCLL 482
- Rechteckzylinder 282
- Reflexionsprinzip 363
 - Brown'sche Bewegung 480
- reguläre Version der bedingten Verteilung 186
- Regularität von Maßen 31, 251
- Rejection Sampling 192
- rekurrent 374
- relativ kompakt 250
- replizierbar 206
- reversibel 407, 423
- Riemann-Integral 98
- Riemann'sche Zetafunktion 52
- Ring 3
- risikoneutral 207
- Rückwärtsmartingal 240

- Satz
 - Approximation von Maßen 30
 - Arzelà-Ascoli 495
 - Bepo Levi 95
 - Berry-Esseen 333
 - Bochner 319
 - Borel-Cantelli Lemma 53
 - bedingte Version 232
 - Carathéodory 19, 24
 - Choquet-Deny 404
 - Chung-Fuchs 382, 458
 - Continuous Mapping Theorem 262
 - Cramér 531, 538
 - Donsker 498
 - Dreireihen 332
 - Egorov 138
 - Etemadi 114
 - Fatou'sches Lemma 95
 - de Finetti 244, 275

- Fischer-Riesz 155
- Fortsetzung zu Maßen 19, **24**
- Fréchet-Shohat 325
- Fubini 285
- Fubini für Itô-Integrale 591
- Fubini für Übergangskerne 290
- Glivenko-Cantelli 117
- große Abweichungen 531
- Hahn'scher Zerlegungssatz 165
- Hartman-Wintner 527
- Hauptsatz der Differential- und Integralrechnung 256
- Helly 268
- Hewitt-Savage 242
- Ionescu-Tulcea 293
- iterierter Logarithmus 520, 527
- Jordan'scher Zerlegungssatz 166
- Kantorovich-Rubinstein 391
- Kesten-Stigum 234
- Kolmogorov-Chentsov 470
- Kolmogorov-Sinai 465
- Kolmogorov'sche Ungleichung 122
- Kolmogorov'scher Dreireihensatz 332
- Kolmogorov'scher Erweiterungssatz 295
- Kolmogorov'sches Kriterium für schwache Relativkompaktheit 496
- Lebesgue'scher Zerlegungssatz 160
- Lévy-Khinchin 342, 345
- Lévy'scher Stetigkeitssatz 316
- Lindeberg-Feller 328
- Lindvall 506
- Lusin 42, 255
- majorisierte Konvergenz 144
- Martingalsdarstellung 588
- monotone Konvergenz 95
- Optional Sampling 213, 218
- Optional Sampling, stetige Zeit 473
- Optional Stopping 214
- Optional Stopping, stetige Zeit 473
- Paley-Wiener-Zygmund 477
- $\pi-\lambda$ 7
- Poisson-Approximation 83
- Pólya 318, 378
- Portemanteau 259
- Prohorov 266
- Quellenkodierungssatz 120
- Rademacher-Menshov 124

- Radon-Nikodym 161, 230
- Rayleigh'sches Monotonieprinzip 428
- reguläre bedingte Verteilungen 186, **189**
- Sanov 540
- Shannon 118
- Skorohod Kopplung 392
- Skorohod'sche Einbettung 522
- Slutsky 260
- Solomon 447
- Stetigkeitssatz von Lévy 316
- Stone-Weierstraß 302
- Strassen 393
- Stroock-Varadhan 612
- Thomson'sches Prinzip 429
- Varadhan'sches Lemma 543
- Yamada-Watanabe 604
- Zentraler Grenzwertsatz 326
- Schauderfunktionen 486
- Schema von Zufallsvariablen 327
- schnittstabil 1
- schwache Konvergenz 257
- schwache Lösung 606
- schwache Topologie 257
- SDGL *siehe* stochastische Differentialgleichung
- Semiring 3
- separabel 250
- Shannon 118
- Shift 452
- σ -additiv 12
- σ -Algebra 2
 - austauschbare 237
 - der τ -Vergangenheit 197
 - invariante 450
 - Produkt- 280
 - terminale 63, 238
- σ -kompakt 250
- σ -Ring 3
- σ -subadditiv 12
- signiertes Maß 164
- singulär 160
- Skalarprodukt 155
- Skorohod Kopplung 392
- Skorohod'scher Einbettungssatz 522
- Slutzky, Satz von 260
- Spannung 425
- Spektrallücke 414
- Spiegelungsprinzip 363
- Spielstrategie 203
- Spin 407
- Spur 11
- stabile Verteilung 318, 349, **350**
 - im weiteren Sinne 350
- Standardabweichung 103
- starke Lösung 598
- starke Markoveigenschaft 362
- stationär 449
- stetig von oben/ unten 16
- Stetigkeitslemma 145
- Stetigkeitsmodul, Lévy'scher 522
- Stetigkeitssatz, Lévy'scher 316
- Stirling'sche Formel 322, 531
- stochastisch größer 392
- Stochastische Differentialgleichung
 - pfadweise Eindeutigkeit 609
 - schwache Lösung 606
 - starke Lösung unter Lipschitz-Bedingungen 601
 - starke Lösung 598
- stochastische Differentialgleichung 597
- stochastische Kerne
 - Produkt 288
- stochastische Matrix 365
- stochastische Ordnung 392
- negative Binomialverteilung 562
- unbegrenzt teilbare Verteilungen 558
- stochastischer Kern 185
 - Halbgruppe 297
 - konsistente Familie 297
 - Verkettung 289
- stochastischer Prozess 193
 - adaptiert 195
 - binär aufspaltender 205
 - Dualität 613
 - Explosion 371
 - Galton-Watson 85, 233
 - Gauß'scher 194, 475
 - gestoppter 214
 - integrierbarer 194
 - Markoveigenschaft 357
 - Modifikation 467
 - Pfad 469
 - Poisson 359
 - previsibler 195, 574
 - produktmessbarer 574

- progressiv messbarer 574
- starke Markoveigenschaft 362
- stationärer 194
- stationäre Zuwächse 194
- unabhängige Zuwächse 194
- ununterscheidbar 467
- Version 467
- vorhersagbarer 195, 574
- stochastisches Integral 490
 - diskretes 202
- Stone-Weierstraß, Satz von 302
- Stoppzeit 196
- straff 266
- Strassen, Satz von 393
- Stratonovich-Integral 591
- Streuung 103
- Stromstärke 425
- Student'sche t -Verteilung 338
- Sub-Wahrscheinlichkeitsmaße 252
- subadditiv 12
 - Folge 466
- subharmonisch 384
- Submartingal 198
- Subordinator 559
- Supermartingal 198
- symmetrische Differenz 30
- symmetrische einfache Irrfahrt 194
- tail σ -field *siehe* terminale σ -Algebra
- terminale σ -Algebra 63, 238
- Thomson'sches Prinzip 429
- Topologie 8
 - schwache 257
 - vage 258
- topologischer Raum 8
- total beschränkt 251
- totale Wahrscheinlichkeit 174
- totalstetig 162
- Totalvariationsnorm 167
- Transformationsformel 41
- transient 374
- translationsinvariant 366
- trennende Familie 254
- Treppenfunktion 97
- Tschebyscheff *siehe* Chebyshev
- Turmeigenschaft 178
- t -Verteilung 338
- Übergangskern 185
- Übergangsmatrix 364
- Übergangswahrscheinlichkeiten 358
- übliche Bedingungen 482
- u.i.v. 57
- unabhängige Inkremente *siehe* unabhängige Zuwächse
- unabhängige Kopie 391
- unabhängige Zuwächse 553
- Unabhängigkeit
 - bedingte 243
 - von Ereignissen 51
 - von Mengensystemen 55
 - von Zufallsvariablen 57
- unbegrenzt teilbar 337
- zufälliges Maß 559
- unbegrenzt teilbare Verteilung
 - stochastische Ordnung 558
- Ungleichung
 - Azuma 202
 - Bernstein-Chernov 112
 - Cauchy-Schwarz 107
 - Chebyshev 110
 - Chernov *siehe* Bernstein-Chernov
 - Doob 222
 - Etemadi 125
 - Hölder 154
 - Jensen 152
 - Kolmogorov 122
 - Markov *siehe* Chebyshev
 - Minkowski 154
 - Young 154
- uniforme Verteilung 33
- unkorriktiert 104
- Unstetigkeitsstellen 11
- ununterscheidbar 467
- vage Konvergenz 257
- vage Topologie 258
- Varadhan'sches Lemma 543
- Varianz 103
- Variation 507
 - p - 509
 - quadratische 509
- Verkettung von Kernen 289
- Version 467
- Verteilung 43
 - Anziehungsbereich 353
 - Bernoulli 44
 - Beta 47, 247, 324, 562

- binomial 44
- Boltzmann 408
- Cauchy 47, 309, 593
- compound Poisson 339
- Exponential- 46
- Gamma 47, 324
- GEM 566, 567
- geometrische 45
- hypergeometrische 45
- negativ binomial 45, 81
- Normal 46
- Pascal 45, 81
- Poisson 45
- Poisson-Dirichlet 562, 564, 567
- stabile 349, **350**
- *t*- 338
- uniforme 12, 33
- zusammengesetzt Poisson 339
- zweiseitig exponential 309
- Verteilungsfunktion 21, 27
 - einer Zufallsvariablen 43
 - empirische 117
- Vervollständigung 32
- Verwerfungsmethode 192
- Verzweigungsprozess 85, 233
- Vitali-Menge 9
- vollständig 32, 250
- vorhersagbar 195, 574
- voter model *siehe* Wählermodell
- Wählermodell 228
- Wahrscheinlichkeitsmaß 12
- Wahrscheinlichkeitsraum **17**
- Wahrscheinlichkeitsvektor 13
- Wald'sche Identität 105
- Wasserstein Metrik 391
- Watson Integral 382
- Weierstraß'scher Approximationssatz 112
- weißes Rauschen 490
- Weiss'scher Ferromagnet 546
- Widerstand 424
- Wiener-Prozess 494
- W-Maß *siehe* Wahrscheinlichkeitsmaß
- Wright'sches Evolutionsmodell 367
- Wright-Fisher Diffusion 614
 - wechselwirkende 618
- Young'sche Ungleichung 154
- Zählmaß 13
- Zentraler Grenzwertsatz 326
 - Berry-Esseen 333
 - Lindeberg-Feller 328
 - mehrdimensional 336
- zentriert 103
- Zerlegungsfolge, zulässige 508
- zufälliges Maß 552
- Zufallsvariable 43
- zulässige Zerlegungsfolge 508
- zusammengesetzte Poissonverteilung 339
- Zustandssumme 408, 546
- Zweistufenexperiment 279
- Zylindermenge 18, 282