Wahrscheinlichkeitstheorie
Sommersemester 07

Erwin Bolthausen

1 Masstheoretische Grundlagen und Ergänzungen

Nachfolgend wird eine Zusammenstellung der für die Wahrscheinlichkeitstheorie wichtigsten Begriffe und Sätze aus der Mass- und Integrationstheorie gegeben. Für ausführliche Darstellungen sind die folgenden Bücher empfehlenswert:

- H. Bauer: Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie, de Gruyter 1978 (3. neubearbeitete Auflage)
- H. Bauer: Mass- und Integrationstheorie, de Gruyter 1990
- D. L. Cohn: Measure Theory, Birkhäuser 1980

Verweise in diesem Kapitel beziehen sich auf das zweitgenannte Buch von Bauer.

1.1 Masse und Erweiterungen

Definition 1.1
*Sei Ω eine Menge. Eine nichtleere Familie $\mathcal{F}$ von Teilmengen von Ω heisst **Algebra** auf Ω, falls für alle $A, B \in \mathcal{F}$ auch $A^c := \Omega \backslash A$, $A \cap B$ und $A \cup B$ in $\mathcal{F}$ sind. Eine Algebra heisst **σ-Algebra**, wenn zusätzlich für jede Folge $\{A_n\}_{n \in \mathbb{N}}$, $A_n \in \mathcal{F}$ $\forall n$, auch $\bigcup_{n=1}^{\infty} A_n$ in $\mathcal{F}$ ist.*

Jede Algebra enthält $\emptyset$ und Ω, weil $\emptyset = A \cap A^c$ für $A \in \mathcal{F} \neq \emptyset$ und $\Omega = \emptyset^c$ gelten. Ferner enthält jede σ-Algebra mit abzählbar vielen Mengen $A_n, n \in \mathbb{N}$, auch deren Durchschnitt $\bigcap_n A_n = \left(\bigcup_n A_n^c\right)^c$.

Lemma 1.2
a) Sind $\mathcal{A}_i$, $i \in I$, Algebren in Ω, wobei I eine beliebige nicht leere Menge ist, so ist $\bigcap_{i \in I} \mathcal{A}_i$ eine Algebra.

b) Sind $\mathcal{F}_i$, $i \in I$, σ-Algebren in Ω, wobei I eine beliebige nicht leere Menge ist, so ist $\bigcap_{i \in I} \mathcal{F}_i$ eine σ-Algebra.

Beweis. Offensichtlich nach den Definitionen. ■

Beispiele 1.3

a) $\{\emptyset, \Omega\}$ ist eine σ-Algebra

b) Die Potenzmenge von Ω, die wir mit $\mathcal{P}(\Omega)$ bezeichnen, ist eine σ-Algebra.

c) Sei $\Omega = \mathbb{R}$. Wir betrachten die Menge $\mathcal{J}$ der rechts offenen, links abgeschlossenen Intervalle (inklusive $\mathbb{R}$ und der leeren Menge):

$$\mathcal{J} := \{\emptyset, \mathbb{R}\} \cup \{(a,b] : -\infty < a < b < \infty\} \cup \{(-\infty, b] : b \in \mathbb{R}\} \cup \{(a,\infty) : a \in \mathbb{R}\}. \tag{1.1}$$

$\mathcal{A}$ sei die Menge aller endlichen Vereinigungen paarweise disjunkter dieser Intervalle. Man überlegt sich leicht, dass $\mathcal{A}$ eine Algebra aber keine σ-Algebra ist.

Bemerkung 1.4
Ein Mengensystem $\mathcal{F}$ ist genau dann eine σ-Algebra, wenn die folgenden drei Eigenschaften erfüllt sind:

1. $\Omega \in \mathcal{F}$,
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,
3. Ist $\{A_n\}_{n\in\mathbb{N}}$ eine Folge mit $A_n \in \mathcal{F}\ \forall n$, so gilt $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Beweis. Einfache Übungsaufgabe. ■

Proposition 1.5
Zu jedem Mengensystem $\mathcal{C}$ in Ω gibt es eine kleinste Algebra $a(\mathcal{C})$ und eine kleinste σ-Algebra $\sigma(\mathcal{C})$, die $\mathcal{C}$ enthalten, d.h. $a(\mathcal{C})$ hat die folgenden Eigenschaften:

a) $a(\mathcal{C})$ ist eine Algebra.

b) $a(\mathcal{C}) \supset \mathcal{C}$

c) Ist $\mathcal{A}$ eine Algebra mit $\mathcal{A} \supset \mathcal{C}$ so gilt $\mathcal{A} \supset a(\mathcal{C})$.

Entsprechende Aussagen gelten für $\sigma(\mathcal{C})$.

Beweis. Wir diskutieren den Fall von $a(\mathcal{C})$: Es gibt mindestens eine Algebra, die $\mathcal{C}$ enthält, nämlich die Potenzmenge $\mathcal{P}(\Omega)$.

$$a(\mathcal{C}) := \bigcap \{\mathcal{A} : \mathcal{A} \text{ Algebra}, \mathcal{A} \supset \mathcal{C}\}$$

erfüllt offensichtlich b) und nach Lemma 1.2 auch a). c) folgt aus der Definitions von $a(\mathcal{C})$. Analog konstruiert man $\sigma(\mathcal{C})$. ■

Zur Erinnerung: Ist Ξ eine Menge von Mengen so ist $\bigcap \Xi$ der Durchschnitt der Mengen in Ξ:

$$\bigcap \Xi := \{x : x \in \mathcal{A},\ \forall \mathcal{A} \in \div\}.$$

Ist $\mathcal{F}$ eine σ-Algebra, so nennt man ein Mengensystem $\mathcal{C}$ mit $\mathcal{F} = \sigma(\mathcal{C})$ ein **Erzeugendensystem** von $\mathcal{F}$. In der Regel haben die uns interessierenden σ-Algebren viele Erzeugendensysteme.

Bemerkung 1.6

a) Ist $\{\mathcal{A}_i\}_{i\in I}$ eine beliebige Familie von σ-Algebren auf derselben Menge Ω, so ist $\bigcap_{i\in I}\mathcal{A}_i$ nach Lemma 1.2 wieder eine σ-Algebra. Die Vereinigung $\bigcup_{i\in I}\mathcal{A}_i$ ist jedoch im allgemeinen keine σ-Algebra.

b) Ist $\mathcal{A}$ eine Algebra, so gilt $a(\mathcal{A}) = \mathcal{A}$. Ist $\mathcal{F}$ eine σ-Algebra, so gilt $\sigma(\mathcal{F}) = \mathcal{F}$. Ist $\mathcal{C}$ ein beliebiges Mengensystem auf Ω so gelten demzufolge $a(a(\mathcal{C})) = a(\mathcal{C})$ sowie $\sigma(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$.

c) Sind $\mathcal{C}$ und $\mathcal{C}'$ Mengensysteme auf Ω mit $\mathcal{C} \subset \mathcal{C}'$, so folgen $a(\mathcal{C}) \subset a(\mathcal{C}')$, $\sigma(\mathcal{C}) \subset \sigma(\mathcal{C}')$. Somit folgt aus $\mathcal{C} \subset \sigma(\mathcal{C}')$ die Inklusion $\sigma(\mathcal{C}) \subset \sigma(\mathcal{C}')$. Um die Gleichung $\sigma(\mathcal{C}) = \sigma(\mathcal{C}')$ nachzuweisen, genügt es daher, die beiden Inklusionen $\mathcal{C} \subset \sigma(\mathcal{C}')$ und $\mathcal{C}' \subset \sigma(\mathcal{C})$ zu zeigen.

Mit $\bigvee_{i\in I}\mathcal{A}_i$ wird die σ-Algebra $\sigma(\bigcup_{i\in I}\mathcal{A}_i)$ bezeichnet. Im Fall $I = \{1,\ldots,n\}$ ist auch die Schreibweise $\mathcal{A}_1 \vee \cdots \vee \mathcal{A}_n$ gebräuchlich.

Es ist häufig wichtig, Erzeugendensysteme mit speziellen Eigenschaften zu haben. Von besonderer Bedeutung in der Wahrscheinlichkeitstheorie sind **durchschnittstabile** Erzeugendensysteme. Eine Familie $\mathcal{A}$ von Teilmengen von Ω heisst **durchschnittstabil**, falls $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$. In diesem Zusammenhang sind Mengensysteme wichtig, die etwas allgemeiner als die σ-Algebren sind, d.h. von denen wir weniger verlangen.

Definition 1.7

Eine Familie $\mathcal{D}$ von Teilmengen von Ω heisst ***Dynkin-System****, falls die folgenden Bedingungen erfüllt sind:*

1. $\Omega \in \mathcal{D}$
2. $D \in \mathcal{D} \Rightarrow D^c \in \mathcal{D}$
3. *Für jede Folge $\{D_n\}_{n\in\mathbb{N}}$ von paarweise disjunkten Mengen aus $\mathcal{D}$, ist $\bigcup_{n=1}^{\infty} D_n$ ebenfalls in $\mathcal{D}$.*

Aus 1. und 2. folgt, dass auch $\emptyset \in \mathcal{D}$ gilt. Aus 3. folgt daher auch, dass $\mathcal{D}$ abgeschlossen gegenüber Vereinigungsbildung von endlich vielen paarweise disjunkten Mengen ist.

Ist $\mathcal{C} \subset \mathcal{P}(\Omega)$, so gibt es analog wie bei den Algebren und σ-Algebren ein kleinstes Dynkin-System $d(\mathcal{C})$, das $\mathcal{C}$ enthält.

Der springende Punkt ist, dass für Dynkin-Systeme nur die Abgeschlossenheit des Systems gegenüber Vereinigungen *paarweise disjunkter* Folgen von Mengen verlangt wird. Dies gestattet es oft, von gewissen Mengensystemen nachzuweisen, dass sie Dynkin sind, wohingegen ein (direkter) Nachweis, dass es sich um eine σ-Algebra handelt, schwierig ist. Wir werden bald Beispiele dazu kennenlernen.

Jede σ-Algebra ist ein Dynkin-System, die Umkehrung gilt jedoch nicht: Man kann leicht Dynkin-Systeme angeben, die keine σ-Algebren sind (siehe Übungen). Es gilt jedoch:

Lemma 1.8
Ist ein Dynkin-System durchschnittsstabil, so ist es eine σ-Algebra.

Beweis. Sei $\mathcal{D}$ ein durchschnittstabiles Dynkin-System. Wir müssen nur zeigen, dass $\mathcal{D}$ abgeschlossen gegenüber abzählbaren Vereinigungen ist.

Sei $\{A_i\}_{i\in\mathbb{N}}$ eine Folge von Elementen in $\mathcal{D}$. Wir definieren die Folge $\{B_n\}_{n\in\mathbb{N}}$ durch

$$\begin{aligned} &B_1 := A_1, \\ &B_n := A_n \setminus (A_1 \cup \dots \cup A_{n-1}), \quad n \geq 2. \end{aligned}$$

Wir zeigen mit Induktion nach n, dass B_n und $A_1 \cup \dots \cup A_n$ zu $\mathcal{D}$ gehören. Für $n = 1$ ist nichts zu zeigen.

Sei $n \geq 2$. B_n lässt sich wie folgt darstellen:

$$B_n = A_n \cap ((A_1 \cup \dots \cup A_{n-1})^c).$$

Per Induktionsvorraussetzung ist $A_1 \cup \dots \cup A_{n-1}$ in $\mathcal{D}$ und somit auch das Komplement. Da $\mathcal{D}$ als durchschnittstabil vorausgesetzt ist, folgt $B_n \in \mathcal{D}$. $A_1 \cup \dots \cup A_{n-1}$ und B_n sind disjunkt, und es gilt $A_1 \cup \dots \cup A_n = (A_1 \cup \dots \cup A_{n-1}) \cup B_n$. Nach der Dynkin-Eigenschaft gilt dann $A_1 \cup \dots \cup A_n \in \mathcal{D}$. Wir haben somit gezeigt, dass alle B_n in $\mathcal{D}$ liegen. Die B_n sind jedoch paarweise disjunkt, und es gilt

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n.$$

Somit folgt $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{D}$. ■

Satz 1.9
Ist $\mathcal{C}$ ein durchschnittstabiles Mengensystem, so gilt $d(\mathcal{C}) = \sigma(\mathcal{C})$.

Beweis. Da jede σ-Algebra auch ein Dynkin-System ist, folgt sofort $d(\mathcal{C}) \subset \sigma(\mathcal{C})$. Um Gleichheit nachzuweisen, müssen wir daher nur noch zeigen, dass $d(\mathcal{C})$ eine σ-Algebra ist. Gemäss Lemma 1.8 genügt es zu zeigen, dass $d(\mathcal{C})$ durchschnittstabil ist. Wir definieren

$$\mathcal{A} := \{\, A \subset \Omega : A \cap C \in d(\mathcal{C}) \ \forall C \in \mathcal{C} \,\}.$$

Da $\mathcal{C}$ als durchschnittstabil vorausgesetzt war, folgt $\mathcal{A} \supset \mathcal{C}$.

Wir zeigen nun, dass $\mathcal{A}$ die Dynkin-Eigenschaften hat, indem wir die drei Eigenschaften in der Definition 1.7 nachweisen:

1. klar.
2. $A \in \mathcal{A} \Rightarrow A \cap C \in d(\mathcal{C}) \ \forall C \in \mathcal{C} \Rightarrow A^c \cap C = (C^c \cup (A \cap C))^c \in d(\mathcal{C}) \ \forall C \in \mathcal{C} \Rightarrow A^c \in \mathcal{A}$. (Beachte: C^c und $A \cap C$ sind disjunkt!)
3. $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, seien paarweise disjunkt. Wegen $A_n \cap C \in d(\mathcal{C}) \ \forall C \in \mathcal{C}$ folgt $\left(\bigcup_{n\in\mathbb{N}} A_n\right) \cap C \in d(\mathcal{C}) \ \forall C \in \mathcal{C}$, d.h. $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Somit ist gezeigt, dass $\mathcal{A}$ ein Dynkin-System ist, also gilt $\mathcal{A} \supset d(\mathcal{C})$. Wir definieren nun

$$\bar{\mathcal{A}} := \{\, A \subset \Omega : A \cap A' \in d(\mathcal{C}) \text{ für alle } A' \in d(\mathcal{C}) \,\}.$$

Nach dem vorangegangenen Schritt gilt $\bar{\mathcal{A}} \supset \mathcal{C}$. Man zeigt nun genau gleich wie oben für $\mathcal{A}$, dass $\bar{\mathcal{A}}$ Dynkin ist. Somit folgt $\bar{\mathcal{A}} \supset d(\mathcal{C})$. Dies besagt jedoch nichts anderes, als dass $d(\mathcal{C})$ durchschnittstabil ist. ∎

Definition 1.10

*Ein **Inhalt** μ auf einer Algebra $\mathcal{A}$ ist eine Abbildung von $\mathcal{A} \to [0, \infty]$ mit den Eigenschaften $\mu(\emptyset) = 0$ und $\mu(A \cup B) = \mu(A) + \mu(B)$ für alle $A, B \in \mathcal{A}$ mit $A \cap B = \emptyset$.*

*Ein Inhalt μ heisst **σ-endlich**, falls eine Folge $\{A_n\}_{n\in\mathbb{N}}$ von Mengen aus $\mathcal{A}$ existiert, für die $\Omega = \bigcup_{n=1}^{\infty} A_n$ und $\mu(A_n) < \infty$ für alle $n \in N$ gelten. μ heisst **endlich** falls $\mu(\Omega) < \infty$ gilt.*

*Ein Inhalt μ heisst **σ-additiv**, falls für jede Folge $\{A_n\}_{n\in\mathbb{N}}$ von paarweise disjunkten Mengen aus $\mathcal{A}$, für die $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ gilt,*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \tag{1.2}$$

*erfüllt ist. (Ein σ-additiver Inhalt heisst auch **Prämass**.)*

*Ein σ-additiver Inhalt, der auf einer σ-Algebra definiert ist, heisst **Mass**. Ein Mass μ mit $\mu(\Omega) = 1$ heisst ein **Wahrscheinlichkeitsmass**.*

Bemerkung: Der Unterschied zwischen Prämassen und Massen besteht nur im Mengensystem auf dem diese definiert sind: Bei Massen ist dies eine σ-Algebra und bei Prämassen nur eine Algebra. Jedes Mass ist daher auch ein Prämass. Da wir bei einer Algebra nicht voraussetzen, dass sie abgeschlossen gegenüber abzählbaren Vereinigungen ist, müssen wir in der Gleichung (1.2) für Prämasse voraussetzen, dass $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ gilt. Für Masse, die ja von vornherein auf einer σ-Algebra definiert sind, ist diese Bedingung automatisch erfüllt.

Konvention: Es werden im folgenden nur σ-endliche Inhalte und Masse betrachtet; dies wird stets stillschweigend vorausgesetzt.

Es sind hier einige Bemerkungen angebracht: Der entscheidende Aufgabe ist die Konstruktion von Massen auf geeigneten σ-Algebren. Eine der Schwierigkeiten dabei ist es, dass man die Mengen in σ-Algebren in der Regel nicht direkt beschreiben kann. Eine direkte konkrete Beschreibung dieser Masse ist daher in der Regel nicht möglich. Die von uns anvisierten σ-Algebren besitzen jedoch konkrete Erzeugendensysteme, die Algebren sind. Man versucht daher, die gewünschten Masse auf diesen Erzeugendensysteme zu konstruieren und zwar in der Form von Prämassen. Nur mit den Prämassen zu arbeiten ist jedoch nicht ausreichend, denn Algebren sind nicht abgeschlossen gegenüber abzählbaren Mengenoperationen. Wir benötigen deshalb einen Satz, der es uns erlaubt, Prämasse auf Algebren zu Massen auf den erzeugten σ-Algebren hochzuziehen. Dies ist der Erweiterungsatz von Caratheodory.

Lemma 1.11
Sei μ ein Inhalt auf einer Algebra $\mathcal{A}$. Dann gelten die folgenden Eigenschaften:

a) *μ ist monoton, das heisst für $A \supset B$ gilt $\mu(A) \geq \mu(B)$.*

b) *μ ist endlich additiv, d.h. sind $A_1, \ldots, A_n \in \mathcal{A}$ paarweise disjunkt, so gilt*

$$\mu\left(\bigcup\nolimits_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu\left(A_i\right).$$

c) *μ ist endlich subadditiv, d.h. sind $A_1, \ldots, A_n \in \mathcal{A}$, so gilt*

$$\mu\left(\bigcup\nolimits_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mu\left(A_i\right).$$

Beweis. Sofort aus der Definition. ■

Satz 1.12
Es seien μ ein Mass auf der σ-Algebra $\mathcal{F}$. Dann gelten die folgenden Eigenschaften:

a) *Sei $A_n, n \in \mathbb{N}$ eine fallende Folge von Mengen in $\mathcal{F}$, und $A := \bigcap_n A_n$. (Wir schreiben dafür auch $A_n \downarrow A$). Falls ein n existiert mit $\mu\left(A_n\right) < \infty$ so gilt $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.*

b) *Sei $A_n, n \in \mathbb{N}$ eine ansteigende Folge von Mengen in $\mathcal{F}$, und $A := \bigcup_n A_n$. (Wir schreiben dafür auch $A_n \uparrow A$). Dann gilt $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.*

c) *Für eine beliebige Folge $A_n \in \mathcal{F}$, $n \in \mathbb{N}$ gilt*

$$\mu\left(\bigcup\nolimits_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

Beweis. Wir beweisen b). Die anderen Teile seien dem Leser als Übungsaufgaben überlassen (siehe auch Bauer §3).

Wir definieren $B_1 := A_1$ und für $n \geq 2 : B_n := A_n \backslash A_{n-1}$. Die B_n sind paarweise disjunkt und es gilt $\bigcup_n A_n = \bigcup_n B_n$. Demzufolge gilt

$$\begin{aligned}\mu\left(\bigcup\nolimits_n A_n\right) &= \mu\left(\bigcup\nolimits_n B_n\right) = \sum_n \mu\left(B_n\right)\\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \mu\left(B_n\right) = \lim_{N\to\infty} \mu\left(\bigcup\nolimits_{n=1}^{N} B_n\right) = \lim_{N\to\infty} \mu\left(A_N\right).\end{aligned}$$

■

Lemma 1.13
μ sei ein endlicher Inhalt auf einer Algebra $\mathcal{A}$. Dann sind die folgenden zwei Bedingungen äquivalent:

a) μ *ist* σ*-additiv (d.h.* μ *ist ein Prämass).*

b)

$$A_n \in \mathcal{A},\ n \in \mathbb{N}, A_n \downarrow \emptyset \Longrightarrow \mu\left(A_n\right) \downarrow 0.$$

Beweis. I) Sei μ ein σ-additiver endlicher Inhalt, und sei A_n eine Folge wie in b). Wir definieren $B_n := A_n \setminus A_{n+1}$ für $n \in \mathbb{N}$. Die B_n sind paarweise disjunkt, und wegen $\bigcap_n A_n = \emptyset$ gilt

$$A_n = \bigcup_{m=n}^{\infty} B_m$$

für alle $n \in \mathbb{N}$. Somit gilt

$$\mu(A_n) = \sum_{m=n}^{\infty} \mu(B_m).$$

Die Summe auf der rechten Seite ist konvergent und somit folgt $\lim_{n\to\infty} \mu\left(A_n\right) = 0$.

II) μ erfülle b) und $\{B_n\}_{n\in\mathbb{N}}$ sei eine Folge paarweise disjunkter Mengen $\in \mathcal{A}$, mit $B := \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$. Dann sind für jedes n die Mengen $B_1, \ldots, B_n, A_{n+1} := \bigcup_{m=n+1}^{\infty} B_m = B \setminus (\bigcup_{j=1}^{m} B_j)$ paarweise disjunkt und in $\mathcal{A}$. Wegen der endlichen Additivität von μ gilt

$$\mu(B) = \sum_{j=1}^{n} \mu(B_j) + \mu(A_{n+1}).$$

Es gilt aber $A_{n+1} \downarrow \emptyset$ für $n \to \infty$ und demzufolge $\mu(A_{n+1}) \downarrow 0$. Somit folgt

$$\mu(B) = \sum_{j=1}^{\infty} \mu(B_j).$$

■

Satz 1.14 (Caratheodory)
Es sei μ_0 *ein Prämass (stets* σ*-endlich vorausgesetzt!) auf einer Algebra* $\mathcal{A}$*. Dann gibt es genau ein Mass* μ *auf* $\sigma(\mathcal{A})$*, das* μ_0 *erweitert, das heisst, das auf* $\mathcal{A}$ *mit* μ_0 *übereinstimmt.*

Wir beweisen den Satz hier nicht. Für einen Beweis sei auf das Buch von Bauer verwiesen. Die Eindeutigkeit ist einfach und folgt aus dem folgenden Resultat, das wir noch sehr häufig verwenden werden:

Satz 1.15
Stimmen zwei Masse μ *und* ν*, die auf einer* σ*-Algebra* $\mathcal{F}$ *definiert sind, auf einem durchschnittstabilen Erzeugendensystem* $\mathcal{D}$ *von* $\mathcal{F}$ *überein, und existiert eine Folge* $\Omega_n \in \mathcal{D},\ n \in \mathbb{N}$ *mit* $\Omega_n \uparrow \Omega$ *und* $\mu(\Omega_n) = \nu(\Omega_n) < \infty$*, so gilt* $\mu = \nu$ *auf* $\mathcal{F}$*. Insbesonder stimmen zwei endliche Masse, die auf einem durchschnittstabilen Erzeugendensystem übereinstimmen und für die* $\mu\left(\Omega\right) = \nu\left(\Omega\right)$ *gilt, auf der erzeugten* σ*-Algebra überein.*

Beweis. Wir beweisen zunächst den Spezialfall, wo $\mu(\Omega) = \nu(\Omega) < \infty$ gilt. Sei $\mathcal{F}' = \{ A \in \mathcal{F} : \mu(A) = \nu(A) \}$. Dann ist $\mathcal{D} \subset \mathcal{F}'$, und $\mathcal{F}'$ ist ein Dynkin-System. Wir zeigen die drei Bedingungen in der Definition 1.7):

1. $\Omega \in \mathcal{F}'$.
2. $D \in \mathcal{F}' \Rightarrow \mu(D^c) = \mu(\Omega \setminus D) = \mu(\Omega) - \mu(D) = \nu(\Omega) - \nu(D) = \nu(D^c) \Rightarrow D^c \in \mathcal{F}'$.
3. Für jede Folge $\{D_n\}_{n\in\mathbb{N}}$ von paarweise disjunkten Mengen aus $\mathcal{F}'$ gilt
$$\mu\left(\bigcup\nolimits_{n=1}^{\infty} D_n\right) = \sum_{n=1}^{\infty} \mu(D_n) = \sum_{n=1}^{\infty} \nu(D_n) = \nu\left(\bigcup\nolimits_{n=1}^{\infty} D_n\right),$$
woraus $\bigcup_{n=1}^{\infty} D_n \in \mathcal{F}'$ folgt.

Somit folgt $\mathcal{F}' \supset d(\mathcal{D}) = \sigma(\mathcal{D}) = \mathcal{F}$ aus Satz 1.9.

Für den allgemeinen Fall betrachten wir die endlichen Masse μ_n, ν_n definiert durch

$$\mu_n(A) = \mu(A \cap \Omega_n), \quad \nu_n(A) = \nu(A \cap \Omega_n),$$

$A \in \mathcal{F}$. Nach der vorangegangenen Überlegung gilt $\mu_n = \nu_n$ für alle $n \in \mathbb{N}$. Somit folgt

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap \Omega_n) = \lim_{n\to\infty} \nu(A \cap \Omega_n) = \nu(A)$$

für alle $A \in \mathcal{F}$, das heisst $\mu = \nu$. ■

Definition 1.16

a) *Ist $\mathcal{F}$ eine σ-Algebra auf einer Menge Ω, so heisst $(\Omega, \mathcal{F})$* ***messbarer Raum*** *und die Mengen in $\mathcal{F}$ heissen* ***messbar****.*

b) *Ist $(\Omega, \mathcal{F})$ ein messbarer Raum und μ ein Mass auf $\mathcal{F}$, so heisst $(\Omega, \mathcal{F}, \mu)$ ein* ***Massraum****. Ist $\mu(\Omega) < \infty$, so heisst der Massraum endlich. Gilt $\mu(\Omega) = 1$, so spricht man von einem* ***Wahrscheinlichkeitsraum****, und μ heisst* ***Wahrscheinlichkeitsmass****.*

Definition 1.17

Sei $(\Omega, \mathcal{F}, \mu)$ ein Massraum. Eine Menge $A \subset \Omega$ heisst eine ***μ-Nullmenge****, wenn eine Menge $F \in \mathcal{F}$ existiert mit $A \subset F$ und $\mu(F) = 0$. (Man beachte, dass wir nicht verlangen, dass $A \in \mathcal{F}$ gilt). Ein Massraum heisst* ***vollständig****, wenn $\mathcal{F}$ alle μ-Nullmengen enthält.*

Lemma 1.18

Sind A_i, $i \in \mathbb{N}$, μ-Nullmgen, so ist auch $\bigcup_i A_i$ eine μ-Nullmenge.

Beweis. Seien $B_i \in \mathcal{F}$ mit $A_i \subset B_i$, $\mu(B_i) = 0$. Dann gilt natürlich $\bigcup_i A_i \subset \bigcup_i B_i \in \mathcal{F}$ und

$$\mu\left(\bigcup\nolimits_i B_i\right) \leq \sum_i \mu(B_i) = 0.$$

■

Bemerkung 1.19
Jeder Massraum lässt sich sehr einfach vervollständigen. Ist nämlich $(\Omega, \mathcal{F}, \mu)$ ein beliebiger Massraum, so betrachten wir

$$\overline{\mathcal{F}}_\mu := \{A \subset \Omega : \exists B \in \mathcal{F},\ A\Delta B \text{ ist } \mu\text{-Nullmenge}\} \supset \mathcal{F}.$$

Man weist sehr einfach nach, dass $\overline{\mathcal{F}}_\mu$ eine σ-Algebra ist. Ferner lässt sich μ sehr einfach auf $\overline{\mathcal{F}}_\mu$ erweitern: Ist $B \in \overline{\mathcal{F}}_\mu$, so existiert nach Definition eine Menge $A \in \mathcal{F}$ mit der Eigenschaft, dass $A\Delta B$ eine Nullmenge ist. Wir setzen $\overline{\mu}(B) := \mu(A)$. Natürlich muss man nachweisen, dass die Festlegung nicht von der gewählten Menge A abhängt. Es stellt sich heraus, dass $\left(\Omega, \overline{\mathcal{F}}_\mu, \overline{\mu}\right)$ ein vollständiger Massraum ist. Man nennt ihn die **Vervollständigung** von $(\Omega, \mathcal{F}, \mu)$. Die obigen Eigenschaften sind alle sehr einfach nachzuprüfen.

Mit vollständigen Massräumen zu arbeiten hat manchmal Vorteile. Von daher ist man versucht, Massräume immer automatisch zu vervollständigen. Anderseits muss man bedenken, dass die vervollständigte σ-Algebra von dem vorliegenden Mass abhängt, was nachteilig ist, wenn man mit mehreren Massen gleichzeitig arbeitet.

In diesem Zusammenhang ist auch die σ-Algebra der **universell messbaren** Mengen interessant. Sie ist definiert durch

$$\hat{\mathcal{F}} := \bigcap_{\mu:\text{endliches Mass}} \overline{\mathcal{F}}_\mu.$$

$\hat{\mathcal{F}}$ ist offensichtlich eine σ-Algebra, die $\mathcal{F}$ enthält. Die Mengen in $\hat{\mathcal{F}}$ heissen auch **universell messbar.**

1.2 Beispiele von messbaren Räumen und Massräumen

Definition 1.20

a) *Es seien $\Omega = \mathbb{R}$ und sei $\mathcal{J}$ die Menge der rechts abgeschlossenen und links offenen Intervalle (1.1). $\mathcal{B} = \sigma(\mathcal{J})$ heisst* ***Borel-σ-Algebra*** *in $\mathbb{R}$. Die Mengen in $\mathcal{B}$ heissen* ***Borelmengen.***

b) *Es sei $\Omega = \mathbb{R}^n$ und $\mathcal{J}_n$ sei die Menge aller "Hyperkuben" der Form $I_1 \times \cdots \times I_n$, $I_j \in \mathcal{J}$. Dann heisst $\mathcal{B}_n := \sigma(\mathcal{J}_n)$ die* ***Borel-σ-Algebra*** *in $\mathbb{R}^n$. Die Mengen in $\mathcal{B}_n$ heissen (n-dimensionale)* ***Borelmengen.***

Es sollte hier bemerkt werden, dass Borel-Mengen nicht durch irgendwelche "Eigenschaften" charakterisiert werden. Eine offene Menge etwa ist durch die Eigenschaft, dass jeder Punkt der Menge eine Umgebung in der Menge besitzt, charakterisiert. Etwas Ähnliches ist bei Borel-Mengen nicht möglich.

Die Borel-σ-Algebra hat eine Reihe von anderen nützlichen Erzeugendensysteme.

Lemma 1.21
Die folgenden Mengensysteme sind Erzeugendensysteme der Borel-σ-Algebra $\mathcal{B}$ in $\mathbb{R}$:

a) $\{(-\infty, t] : t \in \mathbb{R}\}$.

b) Die Menge aller Intervalle in $\mathbb{R}$.

Die folgenden Mengensysteme sind Erzeugendensysteme der Borel-σ-Algebra $\mathcal{B}_n$ in $\mathbb{R}^n$.

c) Die Menge der offenen Teilmengen von $\mathbb{R}^n$.

d) Die Menge der abgeschlossenen Teilmengen von $\mathbb{R}^n$.

e) Die Menge der kompakten Teilmengen von $\mathbb{R}^n$.

Beweis. Wir zeigen, dass die Menge $\mathcal{O}$ der offenen Teilmengen von $\mathbb{R}^n$ die Borel-σ-Algebra $\mathcal{B}_n$ erzeugt. Die Beweise der anderen Teile des Lemmas folgen demselben Muster und seien dem Leser als Übungsaufgabe überlassen.

Sei $U \in \mathcal{O}$. Zu jedem Punkt $x \in U$ existiert ein Kubus $Q := (s_1, t_1] \times \cdots \times (s_n, t_n] \subset U$, mit $x \in Q$ und $s_i, t_i \in \mathbb{Q}$. (Falls nicht bekannt: einfache Übungsaufgabe). Daraus folgt, dass sich jede offene Menge als abzählbare Vereinigung von derartigen Kuben darstellen lässt. (Abzählbar, da es insgesamt nur abzählbar viele derartiger Kuben gibt). Da diese Kuben alle in $\mathcal{J}_n$ sind folgt $U \in \sigma(\mathcal{J}_n) = \mathcal{B}_n$. Somit gilt $\mathcal{O} \subset \mathcal{B}_n$, woraus auch $\sigma(\mathcal{O}) \subset \mathcal{B}_n$ folgt. (Siehe Bemerkung 1.6).

Wir zeigen umgekehrt, dass $\mathcal{B}_n \subset \sigma(\mathcal{O})$ gilt. Dazu beachte man, dass

$$(s_i, t_i] = \bigcap_{m=1}^{\infty} (s_i, t_i + 1/m)$$

und somit

$$(s_1, t_1] \times \cdots \times (s_n, t_n] = \bigcap_{m=1}^{\infty} (s_1, t_1 + 1/m) \times \cdots \times (s_1, t_1 + 1/m)$$

gilt. Es folgt $\mathcal{J}_n \subset \sigma(\mathcal{O})$ und somit $\mathcal{B}_n \subset \sigma(\mathcal{O})$. ■

Sind $(S_i, \mathcal{S}_i)$, $i \in I$, beliebige messbare Räume, wobei I eine beliebige Indexmenge ist, so kann man stets die Produktmenge $\prod_{i \in I} S_i$ mit einer Produkt-σ-Algebra auf die folgende Weise versehen: Für $k \in I$, $A \in$S$_k$, sei

$$A^{(k)} := \left\{ x = (x_i)_{i \in I} : x_k \in A \right\}. \tag{1.3}$$

Definition 1.22
Die σ-Algebra auf $\prod_{i \in I} S_i$, die von den Mengen der Form (1.3) erzeugt wird, nennt man die ***Produkt-σ-Algebra****, und bezeichnet sie mit $\bigotimes_{i \in I} \mathcal{S}_i$. Den messbaren Raum $\left(\prod_{i \in I} S_i, \bigotimes_{i \in I} \mathcal{S}_i\right)$ bezeichnet man als den* ***Produktraum*** *der $(S_i, \mathcal{S}_i)$. Sind alle $(S_i, \mathcal{S}_i)$ gleich: $(S_i, \mathcal{S}_i) = (S, \mathcal{S})$, so schreiben wir auch einfach $\left(S^I, \mathcal{S}^{\otimes I}\right)$ für den Produktraum.*

Die Mengen der Form (1.3) bilden *kein* durchschnittstabiles Erzeugendensystem. Deshalb arbeitet man oft mit dem Erzeugendensystem von $\bigotimes_{i \in I} \mathcal{S}_i$, das aus den endlichen Durchschnitten von Mengen der Form (1.3) besteht. Für eine endliche Teilmenge

$J \subset I$ und $A_j \in \mathcal{S}_j$, $j \in J$, setzen wir $\mathbf{A}_J := \left\{(x_i)_{i\in I} \in \prod_{i\in I} S : x_j \in A_j \text{ für } j \in J\right\}$ und definieren

$$\mathcal{D} := \{\mathbf{A}_J : J \subset I \text{ endlich}, A_j \in \mathcal{S}_j,\ j \in J\}.$$

Dann gilt offensichtlich $\bigotimes_{i\in I} \mathcal{S}_i = \sigma(\mathcal{D})$ und $\mathcal{D}$ ist durchschnittstabil. Die Mengen der Form $\mathbf{A}_J$ nennt man auch endlichdimensionale Zylinder. Nach Satz 1.15 stimmen zwei Wahrscheinlichkeitsmasse auf $\bigotimes_{i\in I} \mathcal{S}_i$ überein, wenn sie auf $\mathcal{D}$ übereinstimmen. Diese Aussage ist für das Erzeugendensystem in der Definition 1.22 nicht richtig.

Ein einfaches Beispiel ist der unendliche Produktraum für den Münzwurf. Wir setzen $S := \{K, Z\}$, $i \in \mathbb{N}$. Für S nehmen wir natürlich die Potenzmenge auf $\{K, Z\}$: $\mathcal{S} = \mathcal{P}(\{K, Z\}) = \{\emptyset, \{K, Z\}, \{K\}, \{Z\}\}$. Der unendliche Produktraum ist dann $(\{K, Z\}^{\mathbb{N}}, (\mathcal{P}(\{K, Z\}))^{\otimes\mathbb{N}})$. Man beachte, dass $(\mathcal{P}(\{K, Z\}))^{\otimes\mathbb{N}} \neq \mathcal{P}\left(\{K, Z\}^{\mathbb{N}}\right)$ ist, was jedoch nicht ganz einfach einzusehen ist.[1].

Nun zu Massen: Zunächst auf $\mathbb{R}$. Wir betrachten Funktionen $F : \mathbb{R} \to \mathbb{R}$ mit den folgenden Eigenschaften:

$$F \text{ ist rechtsseitig stetig.} \tag{1.4}$$

$$F \text{ ist nicht fallend, d.h. } s < t \Longrightarrow F(s) \leq F(t). \tag{1.5}$$

Wir definieren $F(-\infty) := \lim_{s\to-\infty} F(s)$, selbst im Fall, wo der Grenzwert in $\mathbb{R}$ nicht existiert. In diesem Fall setzen wir $F(-\infty) := -\infty$. Analog definieren wir $F(\infty) := \lim_{s\to\infty} F(s)$, was gleich $+\infty$ sein kann. $\mathcal{J}$ die Menge der rechts abgeschlossenen, links offenen Intervalle gemäss (1.1). Sei $a(\mathcal{J})$ die davon erzeugte Algebra. Nach Beispiel 1.3 c) besteht diese Algebra einfach aus den endlichen disjunkten Vereinigungen von Intervallen in $\mathcal{J}$. Für ein endliches Intervall $(s, t]$, $-\infty \leq s < t < \infty$ setzen wir

$$\mu((s, t]) := F(t) - F(s), \tag{1.6}$$

(was ∞ ist, wenn $s = -\infty$ und $F(-\infty) = -\infty$ ist). Ferner setzen wir

$$\mu((s, \infty)) := F(\infty) - F(s). \tag{1.7}$$

Wir können μ natürlich sofort auf die erzeugte Algebra $a(\mathcal{J})$ ausdehnen: Eine disjunkte Vereinigung von Intervallen erhält als μ-Wert einfach die Summe der μ-Werte der Intervalle. Es ist evident, dass μ ein Inhalt auf $a(\mathcal{J})$ ist. Offensichtlich ist μ ein σ-endlicher Inhalt: $\mathbb{R} = \bigcup_n (-n, n]$, und $\mu((-n, n]) = F(n) - F(-n) < \infty$.

Ein Spezialfall ist $F(t) = t$. In diesem Fall ist μ einfach die übliche Länge.

---

[1] Eine „billige“ Möglichkeit, das zu sehen verwendet etwas Kardinalzahlarithmetik. $\{K, Z\}^{\mathbb{N}}$ hat bekanntlich die Mächtigkeit $\aleph_c$ des Kontinuums. Demzufolge hat $\mathcal{P}\left(\{K, Z\}^{\mathbb{N}}\right)$ die Mächtigkeit $2^{\aleph_c} \neq \aleph_c$. Man kann jedoch zeigen, dass $(\mathcal{P}(\{K, Z\}))^{\otimes\mathbb{N}}$ nur die Mächtigkeit des Kontinuums hat. Demzufolge gibt es „sehr viele“ Teilmengen von $\{K, Z\}^{\mathbb{N}}$, die nicht in $(\mathcal{P}(\{K, Z\}))^{\otimes\mathbb{N}}$ sind. Allerdings haben wir mit diesem Argument keine dieser Mengen konkret konstruiert.

Lemma 1.23
μ ist ein Prämass.

Beweis. Das sollte aus der Vorlesung Analysis III bekannt sein (wenigstens im wichtigen Spezialfall $F(t) = t$). ■

Aus dem Satz von Caratheodory folgt, dass sich μ eindeutig zu einem Mass auf der von $\mathcal{J}$ erzeugten σ-Algebra, d.h. der Borel-σ-Algebra erweitern lässt. In Falle $F(t) = t$ ist dies das Lebesgue-Mass auf $\mathcal{B}$. Wenn man zu Wahrscheinlichkeitsmassen gelangen will. muss offensichtlich $F(\infty) - F(-\infty) = 1$ sein. Da nur die Zuwächse von F für das Mass wichtig sind, können wir annehmen, dass

$$\lim_{t\to\infty} F(t) = 1, \ \lim_{t\to-\infty} F(t) = 0 \tag{1.8}$$

Definition 1.24
*Eine Funktion $F : \mathbb{R} \to \mathbb{R}$, die (1.4), (1.5) und (1.8) erfüllt, heisst **Verteilungsfunktion.***

Satz 1.25
Zu jeder Verteilungsfunktion F existiert genau ein Wahrscheinlichkeitsmass μ auf $(\mathbb{R}, \mathcal{B})$ mit

$$\mu((-\infty, t]) = F(t), \ t \in \mathbb{R}.$$

Ist umgekehrt ein beliebiges Wahrscheinlichkeitsmass μ auf $(\mathbb{R}, \mathcal{B})$ gegeben, so ist die Funktion $t \to \mu((-\infty, t])$ eine Verteilungsfunktion.

Beweis. Wegen $\lim_{t\to-\infty} F(t) = 0$ impliziert $\mu((-\infty, t]) = F(t), \ t \in \mathbb{R}$, auch (1.6) und (1.7), wenn μ ein Mass sein soll. Nach Lemma 1.23 und dem Satz von Caratheodory folgt die Existenz des Masses μ auf $(\mathbb{R}, \mathcal{B})$. Die Eindeutigkeit folgt aus der Tatsache, dass $\{(-\infty, t] : t \in \mathbb{R}\}$ ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}$ ist.

Ist umgekehrt μ ein Wahrscheinlichkeitsmass auf $(\mathbb{R}, \mathcal{B})$, so hat die Funktion $t \to F(t) := \mu((-\infty, t])$ die verlangten Eigenschaften: Die Monotonie ist klar. Wir zeigen die Rechtsstetigkeit. Wegen der Monotonie genügt es zu zeigen, dass $\lim_{n\to\infty} F(t + 1/n) = F(t)$ ist. Nun ist die Folge $(-\infty, t + 1/n]$ eine monoton fallende Folge von Teilmengen von $\mathbb{R}$ mit $(-\infty, t] = \bigcap_{n\in\mathbb{N}} (-\infty, t + 1/n]$. Nach Satz 1.12 folgt $\mu((-\infty, t]) = \lim_{n\to\infty} \mu((-\infty, t + 1/n])$. Die Eigenschaften (1.8) können in gleicher Weise verifiziert werden. ■

Analog zum eindimensionalen Lebesgue-Mass definiert man das n-dimensionale Lebesgue-Mass λ_n auf $(\mathbb{R}^n, \mathcal{B}_n)$. Wir wollen jedoch darauf im Moment nicht eingehen, da wir weiter unten des n-dimensionale Lebesgue-Mass als Produktmass des eindimensionalen einführen werden.

Besonders einfach sind sogenannte diskrete Masse:

Beispiel 1.26 (Diskrete Masse)
Es seien Ω eine beliebige Menge, $\{x_i\}_{i\in I}$ eine höchstens abzählbare Menge von verschiedenen Punkten in Ω und $a_i \in [0, \infty)$ für alle $i \in I$. Für jede σ-Algebra $\mathcal{F}$ auf Ω sei

$\mu = \sum_{i \in I} a_i \delta_{x_i}$ definiert durch

$$\mu(A) = \sum_{i \in I} a_i 1_A(x_i), \quad A \in \mathcal{F}.$$

Dies definiert ein Mass μ, jedoch nicht in jedem Fall ein σ-endliches, wie das Beispiel $a_i = 1$ für alle $i \in I := \mathbb{N}$ und $\mathcal{F} = \{\emptyset, \Omega\}$ mit $\Omega := \mathbb{R}$ zeigt. Falls jedoch $\{x_i\} \in \mathcal{F}$ für jedes $i \in I$ gilt, so ist μ offenbar σ-endlich. Ein Mass dieser Gestalt heisst **diskret**. Die a_i heissen *Gewichte*auf den Punkten x_i. Im Fall eines Wahrscheinlichkeitsmasses gilt $\sum_{i \in I} a_i = 1$. Wir schreiben dieses Mass auch als

$$\mu = \sum_{i \in I} a_i \delta_{x_i}$$

Ein einfacher Spezialfall ist $\Omega = \mathbb{N}$ mit dem Zählmass $\sum_n \delta_n$. $\mu(A)$ zählt einfach die Anzahl der Punkte in A.

Ein Spezialfall für diskrete Wahrscheinlichkeitsmasse ist der Fall, wo Ω abzählbar ist, $\mathcal{F} := \mathcal{P}(\Omega)$ und $\mu = \sum_{i \in \Omega} p_i \delta_i$, mit $\sum_i p_i = 1$. Das sind die altbekannten Wahrscheinlichkeitsmasse von früher.

Der unendliche Münzwurf ist nicht mehr ganz so einfach. In diesem Fall nehmen wir $\Omega := \{K, Z\}^{\mathbb{N}}$ und $\mathcal{F}$ die Produkt-σ-Algebra.

$$\mathcal{A} := \left\{ \left\{ \omega = (\omega_n)_{n \in \mathbb{N}} \in \Omega : (\omega_1, \ldots, \omega_m) \in A \right\} : m \in \mathbb{N},\ A \subset \{K, Z\}^m \right\}$$

ist offensichtlich eine Algebra, die $\mathcal{F}$ erzeugt, und wir definieren μ auf $\mathcal{A}$ durch

$$\mu\left(\{\omega \in \Omega : (\omega_1, \ldots, \omega_m) \in A\}\right) := 2^{-m} |A|.$$

Proposition 1.27
μ ist ein Prämass.

Ein Beweis dafür ist nicht sehr schwer. Da wir aber später einen sehr viel allgemeineren Satz beweisen werden (Satz von Ionescu-Tulcea), der die Proposition als Spezialfall enthält, wollen wir im Moment nicht darauf eingehen. Aus der obigen Proposition und dem Satz von Caratheodory folgt also, dass sich μ auf $\mathcal{A}$ zu einem eindeutigen Wahrscheinlichkeitsmass auf der Produkt-σ-Algebra erweitern lässt.

1.3 Messbare Abbildungen

Definition 1.28

a) *Es seien $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$ zwei messbare Räume. Eine Abbildung f von Ω nach Ω' heisst **$\mathcal{F}$-$\mathcal{F}'$-messbar**, falls $f^{-1}(\mathcal{F}') := \{ f^{-1}(A) : A \in \mathcal{F}' \} \subset \mathcal{F}$ gilt. Ist aus dem Zusammenhang klar, welche σ-Algebren gemeint sind, so spricht man auch einfach von einer messbaren Abbildung.*

b) Ist $(\Omega, \mathcal{F})$ ein messbarer Raum, so ist eine $\mathcal{F}$-messbare Funktion eine Abbildung $f : \Omega \to \mathbb{R}$, die $\mathcal{F} - \mathcal{B}$-messbar ist.

Manchmal ist es bequem, Funktionen zuzulassen, die Werte in $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ annehmen. Eine solche Funktion nennt man manchmal eine *numerische* Funktion. Obwohl dies eine ziemliche unsinnige Bezeichnung ist, wollen wir sie hier (mangels einer besseren) ebenfalls verwenden. Auf $\bar{\mathbb{R}}$ betrachten wir die σ-Algebra $\overline{\mathcal{B}}$, die von allen Borelmengen in $\mathbb{R}$, und $\{\infty\}$ und $\{-\infty\}$ erzeugt wird. Eine $\mathcal{F}$-$\overline{\mathcal{B}}$ messbare numerische Funktion nennen wir dann einfach messbare numerische Funktion.

Lemma 1.29
Ist $\mathcal{C}$ ein Erzeugendensystem der σ-Algebra $\mathcal{F}'$, so ist f genau dann $\mathcal{F}$-$\mathcal{F}'$-messbar, wenn $f^{-1}(\mathcal{C}) \subset \mathcal{F}$ gilt.

Beweis. Man zeigt, dass

$$\mathcal{A}_f := \{ A \subset \Omega' : f^{-1}(A) \in \mathcal{F} \}$$

eine σ-Algebra ist. Wir weisen dazu die Eigenschaften 1.–3. aus Bemerkung 1.4 nach:

1. $\Omega' \in \mathcal{A}_f$, wegen $f^{-1}(\Omega') = \Omega$.
2. $A \in \mathcal{A}_f \Rightarrow f^{-1}(A^c) = (f^{-1}(A))^c \in \mathcal{F} \Rightarrow A^c \in \mathcal{A}_f$.
3. Sind $A_n \in \mathcal{A}_f$ für $n \in \mathbb{N}$, so gilt

$$f^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(A_n) \in \mathcal{F}, \quad \text{also} \quad \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_f.$$

Damit ist gezeigt, dass $\mathcal{A}_f$ eine σ-Algebra ist. Nach Voraussetzung ist $\mathcal{A}_f \supset \mathcal{C}$. Somit folgt $\mathcal{A}_f \supset \sigma(\mathcal{C}) = \mathcal{A}'$, das heisst $f^{-1}(\mathcal{A}') \subset \mathcal{A}$. ■

Eine Funktion $f : \Omega \to \mathbb{R}$ ist genau dann $\mathcal{F}$-$\mathcal{B}$-messbar, wenn $\{\omega : f(\omega) \leq t\} \in \mathcal{F}$ für alle $t \in \mathbb{R}$ gilt. Dies folgt sofort aus der Tatsache, dass $\{(-\infty, t] : t \in \mathbb{R}\}$ ein Erzeugendensystem von $\mathcal{B}$ ist. Wie man leicht nachweisen kann, ist $\{[-\infty, t] : t \in \mathbb{R}\}$ ein Erzeugendensystem von $\overline{\mathcal{B}}$. Eine numerische Funktion ist daher genau dann $\mathcal{F}$-$\overline{\mathcal{B}}$-messbar ist, wenn $\{\omega : f(\omega) \leq t\} \in \mathcal{F}$ für alle $t \in \mathbb{R}$ gilt.

Lemma 1.30
Sind $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$, $(\Omega_3, \mathcal{F}_3)$ drei messbare Räume und $f_1 : \Omega_1 \to \Omega_2$, $f_2 : \Omega_2 \to \Omega_3$ zwei messbare Abbildungen, so ist $f_2 \circ f_1 : \Omega_1 \to \Omega_3$ messbar.

Beweis. Für $A \in \mathcal{F}_3$ gilt

$$(f_2 \circ f_1)^{-1}(A) = f_1^{-1}(f_2^{-1}(A)) \in \mathcal{F}_1.$$

■

Satz 1.31
Es sei $(\Omega, \mathcal{F})$ ein messbarer Raum. Ist $\{f_n\}_{n\in\mathbb{N}}$ eine Folge messbarer numerischer Funktionen, so sind $\liminf_{n\to\infty} f_n$ und $\limsup_{n\to\infty} f_n$. Sind die f_n reellwertig und existiert $\liminf_{n\to\infty} f_n(\omega)$ in$\mathbb{R}$ für alle ω, so ist diese Funktion messbar. Gleiches gilt für $\limsup_{n\to\infty} f_n$. Insbesonder gilt für eine Folge von Funktionen, dass wenn $f(\omega) := \lim_{n\to\infty} f_n(\omega)$ für alle $\omega \in \Omega$ existiert, diese Grenzfunktion messbar ist.

Beweis. Wir zeigen die Aussage für $\liminf_{n\to\infty} f_n$. Nach der Bemerkung im Anschluss an Lemma 1.29 genügt es zu zeigen, dass $\{\omega : \liminf_{n\to\infty} f_n \le t\} \in \mathcal{F}$ für alle $t \in \mathbb{R}$. Nun ist $\liminf_{n\to\infty} f_n = \lim_{n\to\infty} \inf_{m\ge n} f_m$, wobei der Limes ein ansteigender ist. Somit gilt

$$\left\{\omega : \lim_{n\to\infty} \inf_{m\ge n} f_m \le t\right\} = \bigcap_{n\in\mathbb{N}} \left\{\omega : \inf_{m\ge n} f_m(\omega) \le t\right\} = \bigcap_{n\in\mathbb{N}} \bigcap_{k\in\mathbb{N}} \left\{\omega : \inf_{m\ge n} f_m(\omega) < t + \frac{1}{k}\right\}$$
$$= \bigcap_{n\in\mathbb{N}} \bigcap_{k\in\mathbb{N}} \bigcup_{m\in\mathbb{N}:m\ge n} \left\{\omega : f_m(\omega) < t + \frac{1}{k}\right\} \in \mathcal{F},$$

da die f_m als messbar vorausgesetzt sind.

Die anderen Behauptungen beweist man auf dieselbe Weise. ■

Die Menge der messbaren Funktionen ist auch abgeschlossen gegenüber den üblichen arithmetischen Operationen ist. Der Beweis sei dem Leser als Übungsaufgabe überlassen (siehe auch Bauer §9)

Satz 1.32

a) *Sind f, g messbare Funktionen, so sind auch $f + g$ (punktweise definiert durch $(f+g)(\omega) = f(\omega) + g(\omega)$ für alle $\omega \in \Omega$) und $f \cdot g$ sowie af für $a \in \mathbb{R}$ messbar.*

b) *Sind f, g messbare Funktionen und gilt $g(\omega) \neq 0$ für alle $\omega \in \Omega$, so ist f/g messbar.*

c) *Jede konstante Funktion ist messbar.*

d) *Ist $A \in \mathcal{F}$, so ist die Indikatorfunktion $1_A : \Omega \to \mathbb{R}$ messbar.*

Satz 1.33
Jede stetige Abbildung $f : \mathbb{R}^n \to \mathbb{R}^m$ ist $\mathcal{B}_n$-$\mathcal{B}_m$-messbar.

Beweis. Wir benutzen die Tatsache, dass $\mathcal{B}_n$ von den offenen Mengen erzeugt wird (Lemma 1.21 c)). Da das inverse Bild einer offenen Mengen unter einer stetigen Abbildung wieder offen ist, folgt die Behauptung aus Lemma 1.29. ■

Aus Satz 1.32 , Lemma 1.30 und Satz 1.33 folgt:

Satz 1.34
Ist f eine messbare Funktion, so sind $f^+ := \max(f, 0)$, $f^- := \max(-f, 0)$ und $|f|$ messbar.

Definition 1.35
Sei $(\Omega, \mathcal{F})$ ein messbarer Raum. Funktionen der Form $\sum_{i=1}^n a_i 1_{A_i}$ mit $n \in \mathbb{N}$, $a_i \in \mathbb{R}$ und $A_i \in \mathcal{F}$ für $i \in \{1, \dots, n\}$ bezeichnet man als ***einfache Funktionen****.*

Die Menge der einfachen Funktionen ist offensichtlich abgeschlossen gegenüber den üblichen Operationen: Sind f, g einfache Funktionen, so sind αf für $\alpha \in \mathbb{R}$, $f + g$, $f \cdot g$, $\max(f, g)$, $\min(f, g)$ einfache Funktionen.

Satz 1.36
Jede nichtnegative, messbare numerische Funktion f ist punktweiser Limes einer monoton ansteigenden Folge $\{f_n\}_{n\in\mathbb{N}}$ nichtnegativer, einfacher Funktionen. (Im Fall $f(\omega) = \infty$ bedeutet $\lim_{n\to\infty} f_n(\omega) = \infty$ dass für alle $K > 0$ ein $n_0 \in \mathbb{N}$ existiert mit $f_n(\omega) \geq K$ für alle $n \geq n_0$.)

Beweis. Wähle

$$f_n := \sum_{k=1}^{n2^n} (k-1)2^{-n} 1_{\{(k-1)2^{-n} \leq f < k2^{-n}\}} + n1_{\{f \geq n\}}.$$

Dann gilt $f_n(\omega) \uparrow f(\omega)$. (Im Falle $f(\omega) = \infty$ ist $f_n(\omega) = n$ für alle n). ■

Aus diesem Satz folgt sofort die folgende Charakterisierung der nichtnegativen, messbaren Funktionen:

Satz 1.37
Es sei $(\Omega, \mathcal{F})$ ein messbarer Raum, und Γ sei eine Menge nichtnegativer, messbarer numerischer Funktionen, für die folgende Bedingungen erfüllt sind:

(i) $f, g \in \Gamma$ und $a, b \in \mathbb{R}^+ \Rightarrow af + bg \in \Gamma$

(ii) $f_n \in \Gamma$ für alle $n \in N$, $f_n(\omega) \uparrow f(\omega)$ für alle $\omega \in \Omega \Rightarrow f \in \Gamma$

(iii) $1_A \in \Gamma$ für alle $A \in F$.

Dann ist Γ die Menge aller nichtnegativen, messbaren numerischen Funktionen.

Beweis. Aus (i) und (iii) folgt, dass Γ alle nichtnegativen einfachen Funktionen enthält. Aus (ii) und Satz 1.36 folgt, dass jede nichtnegative messbare numerische Funktion in Γ ist. ■

Notation: Ist $f : \Omega \to \Omega'$ eine Abbildung und $\mathcal{F}'$ eine σ-Algebra auf Ω', so bezeichnet $\sigma(f) := f^{-1}(\mathcal{F}')$ die von f auf Ω erzeugte σ-Algebra. Sind $f_i : \Omega \to \Omega_i$ Abbildungen und $\mathcal{F}_i$ σ-Algebren auf Ω_i für alle i aus einer beliebigen Indexmenge I, so bezeichnet $\sigma(f_i : i \in I) := \bigvee_{i\in I} f_i^{-1}(\mathcal{F}_i)$ die von den $\{f_i\}_{i\in I}$ auf Ω erzeugte σ-Algebra.

1.4 Integration

Für den ganzen Abschnitt sei ein fester Massraum $(\Omega, \mathcal{F}, \mu)$ vorgegeben (stets σ-endlich)

Definition 1.38

a) *Sei $f = \sum_{i=1}^n a_i 1_{A_i}$ eine nichtnegative (das heisst $a_i \geq 0$), messbare, einfache Funktion. Dann wird das Integral $\int f \, d\mu = \int f(\omega)\mu(d\omega)$ von f definiert durch $\int f \, d\mu = \sum_{i=1}^n a_i \mu(A_i)$.*

b) *Sei $f : \Omega \to [0, \infty]$ nichtnegativ und messbar. Nach Satz 1.36 existiert eine Folge von nichtnegativen, einfachen Funktionen $\{f_n\}_{n\in\mathbb{N}}$ mit $f_n \uparrow f$. Dann ist $\int f\, d\mu = \int f(\omega)\, \mu(d\omega) \in [0, \infty]$ definiert durch $\lim_{n\to\infty} \int f_n\, d\mu$.*

Bemerkung 1.39
Für detaillierte Beweise der nachfolgenden Bemerkungen, siehe Bauer §10.

a) Damit die Festlegung a) wohldefiniert ist, muss man nachweisen, dass $\sum_{i=1}^n a_i \mu(A_i)$ nicht von der speziellen Darstellung von f als $\sum_{i=1}^n a_i 1_{A_i}$, d.h. gilt $\sum_{i=1}^n a_i 1_{A_i}(\omega) = \sum_{i=1}^{n'} a_i' 1_{A_i'}(\omega)$ für alle $\omega \in \Omega$, so gilt $\sum_{i=1}^n a_i \mu(A_i) = \sum_{i=1}^{n'} a_i' \mu(A_i')$. Das ist einfach und sei dem Leser überlassen.

b) In b) der obigen Definition hat man ein ähnliches Problem, das allerdings etwas schwieriger ist: Man muss nachweisen, dass das Integral nicht von der speziell gewählten Folge von einfachen Funktionen $\{f_n\}$ abhängt.

c) Ist $\mu(\{f = \infty\}) > 0$, so ist $\int f\, d\mu = \infty$.

d) Sind f und g zwei nicht negative, einfache Funktionen und sind $a, b \in \mathbb{R},\ a, b \geq 0$, so ist $af + bg$ wieder eine nicht negative, einfache Funktion und es gilt

$$\int (af + bg)\, d\mu = a \int f\, d\mu + b \int g\, d\mu.$$

Gilt für zwei nichtnegative einfache Funktionen $f(\omega) \leq g(\omega)\ \forall\omega$, so gilt $\int f\, d\mu \leq \int g\, d\mu$. Dies sieht man einfach daraus, dass $g - f$ unter dieser Voraussetzung eine nichtnegative einfache Funktion ist.

e) Durch Limesbildung übertragen sich diese Eigenschaften sofort auf nichtnegative messbare numerische Funktionen.

f) Für nichtnegative, messbare numerische Funktionen ist das Integral stets definiert; es kann aber unendlich sein.

Definition 1.40
*Eine messbare reellwertige Funktion f heisst μ-**integrierbar**, falls $\int |f| d\mu < \infty$ ist.*

Wegen $|f| = f^+ + f^-$ mit $f^+ := \max(f, 0)$ und $f^- := \max(-f, 0)$ ist das gleichbedeutend damit, dass $\int f^+\, d\mu < \infty$ und $\int f^-\, d\mu < \infty$ gelten.

Definition 1.41
Sei f eine μ-integrierbare Funktion.

a) *Das Integral $\int f\, d\mu = \int f(\omega)\mu(d\omega)$ ist definiert durch $\int f^+\, d\mu - \int f^-\, d\mu$.*

b) *Ist $A \in F$, so ist $\int_A f\, d\mu := \int (1_A f)\, d\mu$ das Integral von f über A.*

Notation: Wir schreiben $f \in \mathcal{L}_1(\Omega, \mathcal{F}, \mu)$ beziehungsweise kurz $f \in \mathcal{L}_1(\mu)$ oder $f \in \mathcal{L}_1$, wenn f μ-integrierbar ist.

Satz 1.42
Seien $f, g \in \mathcal{L}_1$. Dann gelten:

a)
$$f \geq g \Rightarrow \int f \, d\mu \geq \int g \, d\mu.$$

b)
$$a, b \in \mathbb{R} \Longrightarrow af + bg \in \mathcal{L}_1 \text{ und } \int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu.$$

c)
$$A, B \in \mathcal{F} \text{ mit } A \cap B = \emptyset \Longrightarrow \int_{A \cup B} f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu$$

d) *Ist f eine messbare Funktion, $g \in \mathcal{L}_1,\ g \geq 0$ und gilt $|f(\omega)| \leq g(\omega)$ für alle ω, so ist $f \in \mathcal{L}_1$.*

Beweis. Die Aussagen sind ganz einfache Folgerungen aus der Definition. (siehe auch Bauer §10). ■

Für die Formulierung der nachfolgenden Konvergenzsätze benötigt man folgende Begriffsbildung: Sei $(\Omega, \mathcal{F}, \mu)$ ein Massraum. Eine Eigenschaft in Bezug auf die Elemente von Ω gilt μ-**fast überall** (Abkürzung μ-f.ü.) falls die Menge A der $\omega \in \Omega$, für die die Eigenschaft nicht gilt, in einer messbaren Menge vom μ-Mass null enthalten ist. Ist A selbst messbar, so bedeutet das einfach $\mu(A) = 0$. Wir setzen jedoch im allgemeinen nicht voraus, dass $A \in \mathcal{F}$ ist, obwohl das in den meisten betrachteten Fällen zutrifft. Im Spezialfall, wo μ ein Wahrscheinlichkeitsmass ist, sagt man meist μ-**fast sicher** (Abkürzung μ-f.s.). Beispiel: Sind f und g zwei messbare Funktionen, so bedeutet $f = g$ μ-f.ü.: $\mu(\{\omega \in \Omega : f(\omega) \neq g(\omega)\}) = 0$.

Lemma 1.43
Seien $f, g \in \mathcal{L}_1$. Dann gelten:

a) $f = g$ *μ-f.ü* $\Rightarrow \int f \, d\mu = \int g \, d\mu$.

b) $f \geq g$ *und* $\int f \, d\mu = \int g \, d\mu \Rightarrow f = g$ *μ-f.ü.*

Beweis. a). Durch Zerlegung in Positiv- und Negativteil können wir annehmen, dass $f, g \geq 0$ gelten. Wir wählen eine Folge von einfachen Funktionen $\{f_n\}$ mit $f_n \uparrow f$. Wir setzen $D := \{\omega : f(\omega) = g(\omega)\}$. Dann gelten $\mu(D^c) = 0$ und $1_D f_n \uparrow 1_D f = 1_D g$. Für eine einfache Funktion f_n gilt jedoch offensichtlich $\int 1_D f_n \, d\mu = \int f_n \, d\mu$. Somit folgt $\int 1_D f \, d\mu = \int f \, d\mu$ und mit demselben Argument $\int 1_D g \, d\mu = \int g \, d\mu$, woraus sich wegen $1_D f = 1_D g$ die Behauptung ergibt.

b) Nach Voraussetzung gelten $f - g \geq 0$ und $\int (f-g)\, d\mu = 0$. Es genügt also, den Fall $g \equiv 0$ zu betrachten. Für $n \in \mathbb{N}$ sei $A_n = \{f \geq 1/n\}$. Dann ist $f \geq (\frac{1}{n}) 1_{A_n}$. Aus $\int f\, d\mu = 0$ folgt somit $\int (\frac{1}{n}) 1_{A_n}\, d\mu = (\frac{1}{n})\mu(A_n) = 0$, das heisst $\mu(A_n) = 0$. Nun ist $\{f > 0\} = \bigcup_{n\in\mathbb{N}} A_n$, also $\mu(f > 0) \leq \sum_{n=1}^{\infty} \mu(A_n) = 0$. ■

Wir kommen nun zu den wichtigen Konvergenzsätzen:

Satz 1.44 (Monotoner Konvergenzsatz)
Sei f_n, $n \in \mathbb{N}$, eine Folge nicht negativer, messbarer, numerischer Funktionen mit $f_n \uparrow f$ μ−f.ü. Dann gilt

$$\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu.$$

Beweis. Durch eine Modifikation der Funktionen auf einer Nullmenge können wir annehmen, dass $f_n(\omega) \uparrow f(\omega)$ für alle $\omega \in \Omega$ gilt. (Diese Modifikation ändert an den Integralen nach Lemma 1.43 nichts). Wegen der Monotonie des Integrals folgt natürlich $\int f_n\, d\mu \leq \int f\, d\mu$ für alle n. Wir müssen also nur noch $\lim_{n\to\infty} \int f_n\, d\mu \geq \int f\, d\mu$ nachweisen. Der Limes existiert als monoton wachsende Folge, zumindest in $[0,\infty])$. Dazu wählen wir für jedes n eine Folge $\{f_{n,m}\}_{m\in\mathbb{N}}$ nicht negativer einfacher Funktionen mit $f_{n,m} \uparrow f_n$ für $m \to \infty$. Wir definieren

$$g_{n,m} = \max\{f_{1,m}, \ldots, f_{n,m}\}.$$

Dies sind ebenfalls einfach Funktionen, und es gilt ebenfalls $g_{n,m} \uparrow f_n$ für $m \to \infty$. (Wegen $f_n \geq g_{n,m} \geq f_{n,m}$, und $g_{n,m} \uparrow$ für $m \uparrow$). Anderseits gilt nach der Konstruktion der g's: $g_{n+1,m} \geq g_{n,m}$ für alle m, n. Daraus folgt sehr einfach $g_{n,n} \uparrow f$ für $n \to \infty$. Damit ergibt sich

$$\int f\, d\mu = \lim_{n\to\infty} \int g_{n,n}\, d\mu \leq \lim_{n\to\infty} \int f_n\, d\mu,$$

wegen $g_{n,n} \leq f_n$. ■

Bemerkung 1.45
Es muss hier allerdings bemerkt werden, dass wir die eigentliche Schwierigkeit des Beweises unter den Tisch gekehrt haben. Wir haben ganz wesentlich benutzt, dass das Integral nicht von der Folge der einfachen Funktionen abhängt, die eine Funktion von unten approximiert (Bemerkung 1.39 b)). Das haben wir hier nicht bewiesen.

Korollar 1.46
Sei f eine nicht negative, messbare, numerische Funktion und seien $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, paarweise disjunkte Mengen. Dann gilt

$$\int_{\bigcup_{i\in\mathbb{N}} A_i} f\, d\mu = \sum_{i\in\mathbb{N}} \int_{A_i} f\, d\mu.$$

Satz 1.47 (Lemma von Fatou)
Sei f_n, $n \in \mathbb{N}$, eine Folge nicht negativer, messbarer, numerischer Funktionen. Dann gilt

$$\int \liminf_{n\to\infty} f_n\, d\mu \leq \liminf_{n\to\infty} \int f_n\, d\mu.$$

Beweis. $\inf_{m:m\geq n} f_m \uparrow \liminf_{n\to\infty} f_n$. Nach dem Monotonen Konvergenzsatz folgt:

$$\begin{aligned}\int \liminf_{n\to\infty} f_n \, d\mu &= \lim_{n\to\infty} \int \inf_{m:m\geq n} f_m \, d\mu \\ &\leq \lim_{n\to\infty} \inf_{m:m\geq n} \int f_m \, d\mu = \liminf_{n\to\infty} \int f_n \, d\mu.\end{aligned}$$

■

Satz 1.48 (Satz von Lebesgue)
Sei f_n, $n \in \mathbb{N}$, eine Folge von μ-integrierbaren Funktionen, und es existiere $g \geq 0,\ g \in \mathcal{L}_1$ mit $|f_n(\omega)| \leq g(\omega)$ für alle $\omega \in \Omega$. Falls $f(\omega) = \lim_{n\to\infty} f_n(\omega)$ für fast alle ω existiert, so gilt

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Beweis. Durch eine Änderung der Funktionen auf einer Nullmenge können wir annehmen, dass der Limes überall existiert. Es gilt $f_n + g \geq 0$ und aus dem Lemma von Fatou folgt daraus:

$$\begin{aligned}\int f \, d\mu &= \int (f+g) \, d\mu - \int g \, d\mu \leq \liminf_{n\to\infty} \int (f_n - g) \, d\mu - \int g \, d\mu \\ &\leq \liminf_{n\to\infty} \int f_n \, d\mu.\end{aligned}$$

Derselbe Trick auf $g - f_n \geq 0$ angewandt liefert

$$\int f \, d\mu \geq \limsup_{n\to\infty} \int f_n \, d\mu.$$

■

Im Spezialfall $\Omega = \mathbb{N}$, $\mathcal{F} = \mathcal{P}(\mathbb{N})$ und μ das Zählmass $\sum_n \delta_n$ ergeben sich die bekannten Sätze über reelle Zahlenfolgen. Eine Funktion ist einfach eine reelle Zahlenfolge. Der Satz von Lebesgue besagt, dass wenn eine Doppelfolge $(a_{i,n})_{i,n\in\mathbb{N}}$ die Bedingungen

$$a_i = \lim_{n\to\infty} a_{i,n} \text{ existiert für alle } i,$$

$$|a_{i,n}| \leq b_i,\ \forall i, n, \text{ mit } \sum_i b_i < \infty,$$

erfüllt, die Summation mit dem Limes vertauschbar ist:

$$\lim_{n\to\infty} \sum_i a_{i,n} = \sum_i a_i.$$

1.5 Bildmasse

Sei $(\Omega, \mathcal{F}, \mu)$ ein Massraum, $(\Omega', \mathcal{F}')$ ein zweiter messbarer Raum und $\phi : \Omega \to \Omega'$ eine $\mathcal{F} - \mathcal{F}'$-m.b. Abbildung. Wir definieren eine Abbildung $\mathcal{F}' \to [0, \infty]$, die wir mit $\mu\phi^{-1}$ bezeichnen:

$$\mu\phi^{-1}(B) \overset{\text{def}}{=} \mu\left(\phi^{-1}(B)\right), \tag{1.9}$$

wobei $\phi^{-1}(B)$ wie üblich als $\{\omega \in \Omega : \phi(\omega) \in B\}$ definiert ist.

Proposition 1.49
$\mu\phi^{-1}$ ist ein Mass auf $(\Omega', \mathcal{F}')$. Ist μ ein Wahrscheinlichkeitsmass, so ist $\mu\phi^{-1}$ ebenfalls ein Wahrscheinlichkeitsmass.

Beweis. Wegen $\phi^{-1}(\emptyset) = \emptyset$ folgt $\mu\phi^{-1}(\emptyset) = \mu(\emptyset) = 0$.

Sind die $A_n \in \mathcal{F}'$, $n \in \mathbb{N}$, paarweise disjunkt, so sind auch die Mengen $\phi^{-1}(A_n) \in \mathcal{F}$ paarweise disjunkt und es gilt daher

$$\begin{aligned} \mu\phi^{-1}\left(\bigcup\nolimits_n A_n\right) &= \mu\left(\phi^{-1}\left(\bigcup\nolimits_n A_n\right)\right) = \mu\left(\bigcup\nolimits_n \phi^{-1}(A_n)\right) \\ &= \sum_n \mu\left(\phi^{-1}(A_n)\right) = \sum_n \mu\phi^{-1}(A_n). \end{aligned}$$

Somit ist gezeigt, dass $\mu\phi^{-1}$ ein Mass ist.

Der Zusatz über Wahrscheinlichkeitsmasse ist wegen $\phi^{-1}(\Omega') = \Omega$ offensichtlich. ∎

Es sollte hier allerdings bemerkt werden, dass die σ-Endlichkeit von μ nicht automatisch impliziert, dass $\mu\phi^{-1}$ σ-endlich ist. Ist z.B. $\mathcal{F}' = \{\emptyset, \Omega'\}$ und μ nicht endlich so ist $\mu\phi^{-1}$ offensichtlich nicht σ-endlich.

Satz 1.50

a) *In der obigen Situation sei $f : \Omega' \to [0, \infty]$ eine messbare numerische Funktion. Dann gilt*

$$\int f \, d\left(\mu\phi^{-1}\right) = \int f \circ \phi \, d\mu. \tag{1.10}$$

b) *Sei $f : \Omega \to \mathbb{R}$ messbar. f ist genau dann $\mu\phi^{-1}$-integrierbar, wenn $f \circ \phi : \Omega \to \mathbb{R}$ integrierbar bezüglich μ ist. Ist dies erfüllt, so gilt ebenfalls die Gleichung (1.10).*

Beweis. a): Wir bezeichnen mit Γ die Menge der messbaren numerischen Funktionen $f : \Omega' \to [0, \infty]$, für die (1.10) richtig ist und weisen die Bedingungen (i)-(iii) von Satz 1.37 nach. (iii): Für $A \in \mathcal{F}'$ gilt

$$\begin{aligned} \int 1_A \, d\left(\mu\phi^{-1}\right) &= \mu\phi^{-1}(A) = \mu\left(\phi^{-1}(A)\right) \\ &= \int 1_{\phi^{-1}(A)} \, d\mu = \int (1_A \circ \phi) \, d\mu. \end{aligned}$$

(i) folgt mit Bemerkung 1.39 e) und (ii) aus Satz 1.44. Demzufolge ist Γ die Menge aller nicht negativen, messbaren, numberischen Funktionen $f : \Omega' \to [0, \infty]$. Damit ist a) gezeigt.

b) folgt durch Zerlegung von f in Positiv- und Negativteil und a). ∎

1.6 Produktmasse, Satz von Fubini

Es seien $(\Omega_1, \mathcal{F}_1)$ und $(\Omega_2, \mathcal{F}_2)$ zwei messbare Räume.

Definition 1.51

1. *Sei $\Omega = \Omega_1 \times \Omega_2$ der Produktraum. Die* ***Produkt-σ-Algebra*** *$\mathcal{F}_1 \otimes \mathcal{F}_2$ ist die σ-Algebra, die vom Mengensystem $\mathcal{C} = \{ A_1 \times A_2 : A_i \in \mathcal{F}_i \}$ erzeugt wird. (Das Mengensystem $\mathcal{C}$ ist selbst keine σ-Algebra, es ist aber offensichtlich durchschnittstabil.)*

2. *Sei $f : \Omega_1 \times \Omega_2 \to R$ eine Abbildung. Für $\omega_1 \in \Omega_1$ sei $f^{(1)}_{\omega_1} : \Omega_2 \to R$ definiert durch $f^{(1)}_{\omega_1}(\omega_2) = f(\omega_1, \omega_2)$ für alle $\omega_2 \in \Omega_2$, und analog für $\omega_2 \in \Omega_2$ sei $f^{(2)}_{\omega_2}(\omega_1) = f(\omega_1, \omega_2)$ für alle $\omega_1 \in \Omega_1$ definiert. Die Funktionen $f^{(1)}_{\omega_1}$ und $f^{(2)}_{\omega_2}$ heissen* ***Schnitte*** *von f. Für $A \subset \Omega_1 \times \Omega_2$ betrachten wir die Schnitte $A^{(2)}_{\omega_2} := \{ \omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A \}$, $\omega_2 \in \Omega_2$, und analog $A^{(1)}_{\omega_1} \in \Omega_2$ für $\omega_1 \in \Omega_1$. Offensichtlich gilt $1_{A^{(2)}_{\omega_2}} = (1_A)^{(2)}_{\omega_2}$ und analog für $A^{(1)}_{\omega_1}$.*

Satz 1.52
Ist $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ eine $\mathcal{F}_1 \otimes \mathcal{F}_2$-messbare Funktion, so sind alle Schnitte messbar bezüglich der jeweiligen σ-Algebra, das heisst, für alle $\omega_2 \in \Omega_2$ ist $f^{(2)}_{\omega_2}$ $\mathcal{F}_1$-messbar und für alle $\omega_1 \in \Omega_1$ ist $f^{(1)}_{\omega_1}$ $\mathcal{F}_2$-messbar.

Beweis. Es genügt, die erste Aussage zu zeigen. Wir betrachten zunächst das Mengensystem

$$\mathcal{G} := \{ A \in \mathcal{F}_1 \otimes \mathcal{F}_2 : A^{(2)}_{\omega_2} \in \mathcal{F}_1 \text{ für alle } \omega_1 \in \Omega_1 \}.$$

$\mathcal{G}$ enthält alle Mengen der Form $A_1 \times A_2$, $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$, denn

$$(A_1 \times A_2)^{(2)}_{\omega_2} = \begin{cases} A_1 & \text{falls } \omega_2 \in A_2, \\ \emptyset & \text{falls } \omega_2 \notin A_2. \end{cases}$$

Wegen $(A^c)^{(2)}_{\omega_2} = (A^{(2)}_{\omega_2})^c$, $(\bigcup_{n=1}^{\infty} A_n)^{(2)}_{\omega_2} = \bigcup_{n=1}^{\infty} (A_n)^{(2)}_{\omega_2}$ ist $\mathcal{G}$ eine σ-Algebra. Somit gilt $\mathcal{G} = \mathcal{A}_1 \otimes \mathcal{A}_2$. Damit ist die Aussage des Satzes für Funktionen f der Form $f = 1_A$, $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$ gezeigt. Der Rest folgt dann nach Satz 1.37. ∎

Satz 1.53
Seien μ_1, μ_2 σ-endliche Masse auf $\mathcal{F}_1$ beziehungsweise $\mathcal{F}_2$. Dann gibt es genau ein Mass μ auf $\mathcal{F}_1 \otimes \mathcal{F}_2$ mit der Eigenschaft, dass $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ für alle $A_1 \in \mathcal{F}_1$ und $A_2 \in \mathcal{F}_2$ gilt.

Beweis. $\mathcal{C} = \{ A_1 \times A_2 : A_i \in \mathcal{F}_i \}$ ist ein durchschnittstabiles Erzeugendensystem von $\mathcal{F}_1 \otimes \mathcal{F}_2$. Die Eindeutigkeit ergibt sich dann sofort aus Bemerkung .1.15 Zur Existenz: Sei $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$. Nach dem vorangegangenen Satz ist $A^{(2)}_{\omega_2} \in \mathcal{F}_1$ für alle $\omega_2 \in \mathcal{F}_2$, und damit ist $\mu_1(A^{(2)}_{\omega_2}) \in [0,\infty]$. Wir müssen nun nachweisen, dass $\omega_2 \mapsto \mu_1(A^{(2)}_{\omega_2})$ eine $\mathcal{F}_2$-messbare, numerische Funktion ist. Da μ_1 als σ-endlich vorausgesetzt wird, existiert eine Folge $\{B_n\}_{n\in\mathbb{N}}$ von messbaren Teilmengen von Ω_1 mit $B_n \uparrow \Omega_1$, $\mu_1(B_n) < \infty$. Wegen

$$\mu_1(A^{(2)}_{\omega_2}) = \lim_{n\to\infty} \mu_1(A^{(2)}_{\omega_2} \cap B_n)$$

genügt es daher zu zeigen, dass

$$\omega_2 \mapsto \mu_1(A^{(2)}_{\omega_2} \cap B_n)$$

für jedes n $\mathcal{F}_2$-messbar ist. Eine einfach Überlegung zeigt, dass

$$\{\, A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : \omega_2 \mapsto \mu_1(A^{(2)}_{\omega_2} \cap B_n) \text{ ist } \mathcal{F}_2\text{-messbar} \,\}$$

ein Dynkin-System ist, das $\mathcal{C}$ umfasst. Somit ist es gleich $\mathcal{A}_1 \otimes \mathcal{A}_2$.

Wir haben somit gezeigt, dass $\mu_1(A^{(2)}_{\omega_2})$ als (numerische) Funktion von ω_2 $\mathcal{F}_2$-messbar ist.

Wir definieren nun für $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$:

$$\mu(A) := \int \mu_1(A^{(2)}_{\omega_2})\mu_2(d\omega_2).$$

Eine Anwendung von Korollar 1.46 zeigt, dass $\mu_1 \otimes \mu_2$ ein Mass ist, und natürlich gilt

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

für $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$. ■

Definition 1.54
*Dieses Mass μ heisst das **Produktmass** von μ_1 und μ_2 und wird mit $\mu = \mu_1 \otimes \mu_2$ bezeichnet.*

Satz 1.55 (Satz von Tonelli)
Es sei $f : \Omega_1 \times \Omega_2 \to [0,\infty]$ eine messbare, nichtnegative numerische Funktion. Dann gilt

$$\begin{aligned}\int f\, d(\mu_1 \otimes \mu_2) &= \int \left(\int f^{(1)}_{\omega_1}(\omega_2)\mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\ &= \int \left(\int f^{(2)}_{\omega_2}(\omega_1)\mu_1(d\omega_1) \right) \mu_2(d\omega_2).\end{aligned}$$

Beweis. Die Aussage ist richtig – nach Konstruktion des Masses $\mu_1 \otimes \mu_2$ – für Funktionen f der Form $f = 1_A$, $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$. Für allgemeine nicht negative numerische Funktionen folgt die Aussage mit Hilfe von Satz 1.37. ■

Satz 1.56 (Satz von Fubini)
Sei $f:\Omega_1\times\Omega_2\to\mathbb{R}$ in $\mathcal{L}_1(\Omega_1\times\Omega_2,\mathcal{F}_1\otimes\mathcal{F}_2,\mu_1\otimes\mu_2)$. Dann ist für μ_1-fast alle $\omega_1\in\Omega_1$ die Funktion $f^{(1)}_{\omega_1}$ in $\mathcal{L}_1(\Omega_2,\mathcal{F}_2,\mu_2)$ und für μ_2-fast alle $\omega_2\in\Omega_2$ die Funktion $f^{(2)}_{\omega_2}$ in $\mathcal{L}_1(\Omega_1,\mathcal{F}_1,\mu_1)$. Setzt man

$$f_1(\omega_1) := \begin{cases} \int f^{(1)}_{\omega_1}(\omega_2)\,\mu_2(d\omega_2) & \text{falls } f^{(1)}_{\omega_1}\in\mathcal{L}_1 \\ 0 & \text{sonst,}\end{cases}$$

und $f_2(\omega_2)$ analog, so gilt

$$\int f\,d(\mu_1\otimes\mu_2) = \int f_1(\omega_1)\,\mu_1(d\omega_1) = \int f_2(\omega_2)\,\mu_2(d\omega_2).$$

In einfacherer Sprechweise (die nicht ganz präzise ist) besagt der obige Satz: Ist $f\in\mathcal{L}_1(\mu_1\otimes\mu_2)$, so darf man die Integrationsreihenfolge vertauschen

$$\begin{aligned}\int f\,d(\mu_1\otimes\mu_2) &= \int\left(\int f(\omega_1,\omega_2)\;\mu_1(d\omega_1)\right)\mu_2(d\omega_2)\\ &= \int\left(\int f(\omega_1,\omega_2)\;\mu_2(d\omega_2)\right)\mu_1(d\omega_1).\end{aligned}$$

Der Satz folgt aus dem Satz von Tonelli durch Zerlegung von f in Positiv- und Negativteil. Die Details seien dem Leser überlassen.

Bemerkung 1.57
Die obigen Betrachtungen können leicht auf endlich viele Faktoren ausgedehnt werden: Sind $(\Omega_i,\mathcal{F}_i,\mu_i)$ für $1\le i\le n$ Massräume, so definiert man die Produkt-σ-Algebra $\mathcal{F}_1\otimes\cdots\otimes\mathcal{F}_n$ durch $(\ldots((\mathcal{F}_1\otimes\mathcal{F}_2)\otimes\mathcal{F}_3)\otimes\cdots\otimes\mathcal{F}_{n-1})\otimes\mathcal{F}_n$, wobei der Produktraum $\Omega_1\times\cdots\times\Omega_n$ mit $(\ldots((\Omega_1\times\Omega_2)\times\Omega_3)\times\cdots\times\Omega_{n-1})\times\Omega_n$ identifiziert wird. Analog wird das Produktmass $\mu_1\otimes\cdots\otimes\mu_n$ durch $(\ldots((\mu_1\otimes\mu_2)\otimes\mu_3)\otimes\cdots\otimes\mu_{n-1})\otimes\mu_n$ definiert. Man weist leicht nach, dass man beliebig umklammern kann: Ist $1\le k<n$, so gelten die Beziehungen $(\mathcal{F}_1\otimes\cdots\otimes\mathcal{F}_k)\otimes(\mathcal{F}_{k+1}\otimes\cdots\otimes\mathcal{F}_n)=\mathcal{F}_1\otimes\cdots\otimes\mathcal{F}_n$ und $(\mu_1\otimes\cdots\otimes\mu_k)\otimes(\mu_{k+1}\otimes\cdots\otimes\mu_n)=\mu_1\otimes\cdots\otimes\mu_n$ beim Identifizieren von $(\Omega_1\times\cdots\times\Omega_k)\times(\Omega_{k+1}\times\cdots\times\Omega_n)$ mit $\Omega_1\times\cdots\times\Omega_n$. Wir schreiben auch $\mathcal{F}_1^n$ für das n-fache Produkt $\mathcal{F}_1\otimes\cdots\otimes\mathcal{F}_1$ und μ_1^n für das n-fache Produktmass $\mu_1\otimes\cdots\otimes\mu_1$.

Beispiel 1.58
Es seien $(\Omega_i,\mathcal{F}_i,\mu_i)=(\mathbb{R},\mathcal{B},\lambda)$ mit dem Lebesgue-Mass λ für alle $i\in\{1,\ldots,n\}$. Man zeigt, dass $\mathcal{B}_n=\mathcal{B}\otimes\cdots\otimes\mathcal{B}$ gilt, und $\lambda_n=\lambda\otimes\cdots\otimes\lambda$ heisst n-dimensionales Lebesgue-Mass.

1.7 Der Satz von Radon-Nikodym

Nachfolgend sei $(\Omega,\mathcal{F},\mu)$ ein σ-endlicher Massraum.

Proposition 1.59
Sei $f : \Omega \to [0, \infty)$ messbar. Dann wird durch $\mathcal{F} \ni A \mapsto \nu(A) := \int_A f \, d\mu$ ein (stets σ-endliches!) Mass auf $\mathcal{F}$ definiert. Für jede Menge $A \in \mathcal{F}$ mit $\mu(A) = 0$ gilt $\nu(A) = 0$.

Beweis. Offensichtlich gilt $\nu(\emptyset) = 0$. Ist $\{A_n\}_{n \in \mathbb{N}}$ eine Folge paarweise disjunkter Mengen aus $\mathcal{F}$, und ist $A = \bigcup_{n=1}^{\infty} A_n$, so gilt

$$\begin{aligned} \nu(A) &= \int_A f \, d\mu = \lim_{n \to \infty} \int_{\bigcup_{k=1}^n A_k} f \, d\mu = \lim_{n \to \infty} \sum_{k=1}^n \int_{A_k} f \, d\mu \\ &= \sum_{k=1}^{\infty} \nu(A_k). \end{aligned}$$

Es bleibt zu zeigen, dass ν σ-endlich ist. Sei $\{\Omega_n\}_{n \in \mathbb{N}}$ eine Folge von messbaren Teilmengen von Ω mit $\Omega_n \uparrow \Omega$ und $\mu(\Omega_n) < \infty$. Sei ferner $A_n := \{f \leq n\}$. Dann gilt $\Omega_n \cap A_n \uparrow \Omega$ und $\nu(A_n \cap \Omega_n) \leq n\mu(\Omega_n) < \infty$ für alle n.

Ist $A \in \mathcal{F}$ eine μ-Nullmenge, so ist $1_A f$ fast überall 0. Daraus folgt $\nu(A) = 0$. ■

Eine wichtige Frage ist, ob sich ein Mass ν aus einem Mass μ auf diese Weise mit Hilfe einer Funktion f darstellen lässt. In diesem Fall nennt man f eine **Dichte** von ν bezüglich μ. Es ist nicht schwer zu sehen, dass f, wenn es überhaupt existiert, eindeutig bis auf μ-f.s.-Gleichheit ist:

Sind $f, g \geq 0$ zwei Dichten von ν bezüglich μ, so gilt

$$\int_A f \; d\mu = \int_A g \; d\mu$$

für alle $A \in \mathcal{F}$. Sei $\Omega_n \in \mathcal{F}$ eine Folge mit $\Omega_n \uparrow \Omega$ und $\mu(\Omega_n) < \infty$ für alle n. Sei $A_n := \Omega_n \cap \{f < g \leq n\}$. Dann ist

$$\int_{A_n} (g - f) \; d\mu = 0.$$

Wegen $g > f$ auf A_n folgt $\mu(A_n) = 0$. Wegen $\Omega_n \cap \{f < g \leq n\} \uparrow \{f < g\}$ folgt $\mu(f < g) = 0$, d.h. $f \geq g$ μ-f.s.. Analog folgt $g \geq f$ μ-f.s.. Damit ist gezeigt, dass eine Dichte, falls sie überhaupt existiert, eindeutig ist bis auf μ-f.s.-Gleichheit.

Wann existiert eine derartige Dichte? Nach der obigen Proposition ist eine notwendige Bedingung, dass alle μ-Nullmengen auch ν-Nullmengen sind. Eine erstaunliche Tatsache ist, dass dies auch eine hinreichende Bedingung ist.

Definition 1.60
*Es seien $(\Omega, \mathcal{F})$ ein messbarer Raum und μ, ν zwei Masse auf $\mathcal{F}$. ν heisst **absolutstetig** bezüglich μ (Notation: $\nu \ll \mu$), falls folgende Bedingung erfüllt ist:*

$$\forall A \in \mathcal{F} : \; \mu(A) = 0 \Rightarrow \nu(A) = 0.$$

Satz 1.61 (Satz von Radon-Nikodym)
ν, μ seien zwei σ-endliche Masse auf $(\Omega, \mathcal{F})$. Es gilt $\nu \ll \mu$ genau dann, wenn eine nichtnegative, messbare Funktion $f : \Omega \to \mathbb{R}^+$ existiert mit $\nu(A) = \int_A f \, d\mu$ für alle $A \in \mathcal{F}$. Diese Funktion ist eindeutig bis auf μ-f.ü.-Gleichheit.

Für einen Beweis, siehe: Bauer: Mass- und Integrationstheorie, §17.

Wir schreiben

$$f = \frac{d\nu}{d\mu},$$

wenn eine derartige Dichte existiert.

Es ist naheliegend, dass in einem derartigen Fall Integrale bezüglich ν in Integrale bezüglich μ umgeschrieben werden können:

Proposition 1.62
Sei ν absolut stetig bezüglich μ mit Dichte f.

a) *Ist $\phi : \Omega \to [0, \infty]$ eine messbare numerische Funktion, so gilt*

$$\int \phi \; d\nu = \int \phi f \; d\mu. \tag{1.11}$$

b) *Ist $\phi : \Omega \to \mathbb{R}$ messbar, so gilt $\phi \in \mathcal{L}_1(\nu)$ genau dann, wenn $\phi f \in \mathcal{L}_1(\mu)$ ist. In diesem Fall gilt ebenfalls die obige Gleichung.*

Beweis. Für $\phi = 1_A$, $A \in \mathcal{F}$, ist (1.11) einfach die Definition von ν. Der allgemeine Fall in a) folgt einfach mit 1.37. In b) folgt $\phi \in \mathcal{L}_1(\nu) \Longleftrightarrow \phi f \in \mathcal{L}_1(\mu)$ aus a) angewandt auf $|\phi|$. Die Gleichung (1.11) folgt einfach aus einer Zerlegung in Positiv- und Negativteil. ■

Bemerkung 1.63
Der Beweis des Satzes von Radon-Nikodym ist nicht allzu schwierig. Es ist jedoch wichtig zu bemerken, dass er nicht konstruktiv ist, d.h. er liefert kein Verfahren, wie die Dichte tatsächlich konstruiert werden kann. Erinnert man sich an den Fundamentalsatz der Analysis, so ist eine naheliegende Vermutung, dass man Dichten als eine Art Differentialquotient durch eine Limesbildung etwa als

$$f(x) = \lim_{n\to\infty} \frac{\nu(U_n)}{\mu(U_n)}$$

gewinnen kann, wobei $\{U_n\}$ eine Folge von messbaren Mengen mit $x \in U_n$, $U_n \downarrow \{x\}$ ist. Das ist jedoch in dem ganz allgemeinen Rahmen des obigen Satzes nicht möglich, stimmt jedoch in Spezialfällen, wobei man aber die Folge $\{U_n\}$ sehr sorgfältig wählen muss. Eine Diskussion solcher Fragen ist viel schwieriger als der Beweis des Satzes von Radon-Nikodym.

2 Grundlegende Begriffsbildungen der Wahrscheinlichkeitstheorie

Es sei $(\Omega, \mathcal{F})$ ein messbarer Raum. Wahrscheinlichkeitsmasse auf $\mathcal{F}$ werden oft (aber nicht immer) mit P, Q usw. statt mit μ, ν usw. bezeichnet (P für "probability"). Die Elemente von $\mathcal{F}$ bezeichnet man in der Wahrscheinlichkeitstheorie meist als **Ereignisse**. Die einzelnen Elemente ω von Ω nennt man die **Elementarereignisse**. Da für ein Wahrscheinlichkeitsmass $P(\Omega) = 1$ gilt, folgt $0 \leq P(A) \leq 1$ für jedes $A \in \mathcal{F}$. Ferner gilt $P(A^c) = 1 - P(A)$. Statt P-fast überall sagt man meist P-*fast sicher* (Abkürzung: P-f. s.). $(\Omega, \mathcal{F}, P)$ nennt man einen **Wahrscheinlichkeitsraum.**

Ist $\{A_n\}_{n \in \mathbb{N}}$ eine Folge von Ereignissen, so verwenden wir die folgenden Bezeichnungen:

$$\liminf_{n\to\infty} A_n := \bigcup_n \bigcap_{m:m\geq n} A_m \in \mathcal{F},$$
$$\limsup_{n\to\infty} A_n := \bigcap_n \bigcup_{m:m\geq n} A_m \in \mathcal{F}.$$

Die Vereinigung im ersten Ausdruck ist offenbar ansteigend, d.h. es gilt

$$\bigcap_{m:m\geq n} A_m \uparrow \liminf_{n\to\infty} A_n.$$

Analog gilt

$$\bigcup_{m:m\geq n} A_m \downarrow \limsup_{n\to\infty} A_n.$$

Ein Element $\omega \in \Omega$ ist genau dann in $\liminf_{n\to\infty} A_n$, wenn es in allen A_n bis auf endlich vielen liegt. Man sagt dann auch, dass ω in "fast allen" A_n ist. $\omega \in \limsup_{n\to\infty} A_n$ gilt genau dann, wenn ω in unendlich vielen der A_n ist. Statt $P(\limsup_{n\to\infty} A_n)$ schreibt man dann auch $P(A_n \text{ unendlich oft})$, bzw für $P(\liminf_{n\to\infty} A_n)$: $P(A_n, \text{ fast alle } n)$.

2.1 Zufallsgrössen und ihre Verteilungen

$(\Omega, \mathcal{F}, P)$ sei ein Wahrscheinlichkeitsraum.

Definition 2.1
*Es sei $(S,\mathcal{S})$ ein messbarer Raum. Eine $(S,\mathcal{S})$-**wertige Zufallsgrösse** (kurz S-wertige Zufallsgrösse) ist eine $\mathcal{F}$-$\mathcal{S}$-messbare Abbildung $X : \Omega \to S$. Im Spezialfall $(S,\mathcal{S}) = (\mathbb{R}, \mathcal{B})$ spricht man auch einfach von einer **Zufallsgrösse**. Im Falle $(S,\mathcal{S}) = (\mathbb{R}^n, \mathcal{B}_n)$ nennt man X einen n-dimensionalen **Zufallsvektor** und schreibt diesen oft als $X = (X_1, \ldots, X_n)$, wobei die Abbildungen $X_i : \Omega \to \mathbb{R}$ die Koordinatenabbildungen sind.*

Ein Zufallsvektor $X = (X_1, \ldots, X_n)$ ist nichts anderes als eine Kollektion von n (eindimensionalen) Zufallsgrössen. $X : \Omega \to \mathbb{R}^n$ ist nämlich genau dann $\mathcal{F}$-$\mathcal{B}_n$-m.b., wenn die Komponenten $X_i : \Omega \to \mathbb{R}$ alle $\mathcal{F}$-$\mathcal{B}$-m.b. sind. Das folgt sofort aus Lemma 1.29.

Wir können ohne Schwierigkeiten auch unendliche Familien von Zufallsgrössen betrachten. Ist I eine beliebige Menge, und sind die X_i für jedes $i \in I$ eine (S_i,S_i)-wertige Zufallsgrösse, die auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert sind, so bezeichnet man mit $\sigma(X_i : i \in I)$ die von den Zufallsgrössen $\{X_i\}_{i\in I}$ erzeugte σ-Algebra $\bigvee_{i\in I} X_i^{-1}(\mathrm{S}_i) = \sigma\left(\bigcup_i X_i^{-1}(\mathcal{S}_i)\right) \subset \mathcal{F}$.

Jede (S,S)-wertige Zufallsgrösse X definiert gemäss (1.9) ein Wahrscheinlichkeitsmass PX^{-1} auf S. Man nennt das auch die **Verteilung** von X. Im Spezialfall, wo X reellwertig ist, ist die Verteilung ein Wahrscheinlichkeitsmass auf der Borel-σ-Algebra $\mathcal{B}$.

Haben zwei (S,S)-wertige Zufallsgrössen X und X' dieselbe Verteilung, so schreiben wir $\mathcal{L}(X) = \mathcal{L}(X')$ (dabei steht $\mathcal{L}$ für „law“, den englischen Begriff für Verteilung); X und X' müssen dabei nicht auf demselben Wahrscheinlichkeitsraum definiert sein.

Für eine reellwertige Zufallsgrösse ist die Verteilung ein Wahrscheinlichkeitsmass auf $(\mathbb{R}, \mathcal{B})$. Wie wir schon wissen (Satz 1.25), werden diese durch ihre Verteilungsfunktion eindeutig festgelegt. Die Verteilungsfunktion der Verteilung von X bezeichnen wir einfach als die Verteilungsfunktion von X :

Definition 2.2
*Sei X eine $\mathbb{R}$-wertige Zufallsgrösse. Die Abbildung $\mathbb{R} \ni t \mapsto F_X(t) = P(\{\omega : X(\omega) \le t\})$ heisst **Verteilungsfunktion** von X.*

Notation: Im Umgang mit Zufallsgrössen bedient man sich meist einer gewissen Kurzsschreibweise. Z.B. lässt man die ω in der Notation meist weg. So schreibt man $\{X \le t\}$ kurz für $\{\omega : X(\omega) \le t\} = X^{-1}((-\infty, t])$. Innerhalb von Wahrscheinlichkeiten lässt man dann auch die geschweiften Klammern weg und schreibt einfach $P(X \le t)$.

Eine wichtige Klasse von Wahrscheinlichkeitsmassen auf $(\mathbb{R}, \mathcal{B})$ sind diejenigen, die über ein Dichtefunktion bezüglich des Lebesgue-Masses definiert sind. Eine messbare Funktion $f : \mathbb{R} \to \mathbb{R}^+$ heisst **Wahrscheinlichkeitsdichte**, wenn $\int f\, d\lambda = 1$ gilt, wobei λ das Lebesgue-Mass ist. Eine solche Dichtefunktion definiert nach Proposition 1.59 ein Mass

$$\mathcal{B} \ni A \to \mu(A) := \int_A f\, d\lambda. \tag{2.1}$$

auf $(\mathbb{R}, \mathcal{B})$, das natürlich wegen $\int f\, d\lambda = 1$ ein Wahrscheinlichkeitsmass ist. Hat die Verteilung einer reellen Zufallsgrösse X, d.h. das Wahrscheinlichkeitsmass PX^{-1} eine Dichte bezüglich des Lebesgue-Masses, so sagt man, X habe eine Dichte.

Neben den Verteilungen auf $\mathcal{B}$ mit Dichten sind die einfachsten Verteilungen die **diskreten**, d.h. Wahrscheinlichkeitsmasse der Form $\sum_{n\in\mathbb{N}} p_n\, \delta_{x_n}$ mit $x_n \in \mathbb{R}$ und $p_n \ge 0$, $\sum_n p_n = 1$. Diskrete Verteilungen haben offenbar keine Dichten, denn $\lambda(\{x\}) = 0$ gilt für alle $x \in \mathbb{R}$. Die Verteilungsfunktion einer diskreten Verteilung ist nicht stetig. Offenbar hat sie Sprünge der Höhe p_n in den Punkten x_n. Neben diskreten Verteilungen und solchen mit Dichten gibt es jedoch noch andere, die für uns im Moment nicht besonders wichtig sind.

Hier einige der wichtigsten Verteilungen auf $(\mathbb{R}, \mathcal{B})$:

Beispiel 2.3 (Verteilungen auf $\mathbb{R}$)

a) *Poisson-Verteilung zum Parameter $\lambda > 0$:*

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda} \delta_n.$$

b) *Standardnormalverteilung: Sie hat die Dichte*

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}, \quad x \in \mathbb{R}.$$

c) *Normalverteilung mit Mittelwert $a \in \mathbb{R}$ und Varianz $\sigma^2 > 0$: Sie hat die Dichte*

$$\varphi\left(x; a, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}.$$

d) *Cauchy-Verteilung zum Parameter $c > 0$ mit Dichte*

$$f(x) = \frac{c}{\pi(c^2 + x^2)}, \quad x \in \mathbb{R}.$$

Man spricht von Poisson-verteilten Zufallsgrössen, normalverteilten Zufallsgrössen etc., wenn die Verteilung der Zufallsgrösse wie oben gegeben ist.

Nun zu Verteilungen von $(\mathbb{R}^n, \mathcal{B}_n)$. Ein Wahrscheinlichkeitsmass μ auf $(\mathbb{R}^n, \mathcal{B}_n)$ hat die Dichte $f : \mathbb{R}^n \to \mathbb{R}^+$ bezüglich des n-dimensionalen Lebesgue-Masses λ_n, wenn für jede Menge $A \in \mathcal{B}_n$ die Gleichung

$$\mu(A) = \int_A f \, d\lambda_n$$

gilt. f wird dabei als messbar vorausgesetzt (bezüglich der Borel-σ-Algebren). Eine besonders wichtige Klasse von Wahrscheinlichkeitsmassen auf $(\mathbb{R}^n, \mathcal{B}_n)$ sind die n-dimensionalen Normalverteilungen.

Definition 2.4

a) *Das Wahrscheinlichkeitsmass auf $(\mathbb{R}^n, \mathcal{B}_n)$, das durch die Dichte*

$$\varphi(x_1, \ldots, x_n) := (2\pi)^{-n/2} \exp\left[-\frac{1}{2} \sum\nolimits_{i=1}^{n} x_i^2\right], \quad (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

definiert wird, heisst ***Standardnormalverteilung*** *auf $\mathbb{R}^n$.*

b) *Ein Wahrscheinlichkeitsmass μ auf $(\mathbb{R}^n, \mathcal{B}_n)$ heisst* ***Normalverteilung****, wenn eine reelle $n \times n$-Matrix A und $b \in \mathbb{R}^n$ existieren, sodass $\mu = \mu_{\mathrm{st}} \phi^{-1}$ ist, wobei ϕ die affine Abbildung $\mathbb{R}^n \ni x \mapsto \phi(x) = Ax + b \in \mathbb{R}^n$ und μ_{st} die Standardnormalverteilung sind.*

Wir wollen noch kurz diskutieren, wann die Normalverteilung eine Dichte besitzt.

Proposition 2.5
Das Wahrscheinlichkeitsmass μ der obigen Definition besitzt genau dann eine Dichte, wenn A eine invertierbare Matrix ist. In diesem Fall ist die Dichte gegeben durch

$$\varphi(x;b,\Sigma) := \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left[-\frac{1}{2}(x-b)^T\,\Sigma^{-1}(x-b)\right], \quad x \in \mathbb{R}^n,$$

mit $\Sigma = AA^T$ (A^T die Transponierte von A).

Beweis. Sei die Matrix A regulär. Dann ist ϕ eine invertierbare Abbildung, und es gilt für $B \in \mathcal{B}_n$

$$\begin{aligned}\mu(B) = \mu_{\text{st}}(\phi^{-1}(B)) &= \int_{\phi^{-1}(B)} \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2} \lambda_n(dx) \\ &= \int 1_B(\phi(x)) \frac{1}{(2\pi)^{-n/2}} \exp\left[-\frac{1}{2}|\phi^{-1}\phi(x)|^2\right] \lambda_n(dx) \\ &= \int 1_B(y) \frac{1}{(2\pi)^{-n/2}} \exp\left[-\frac{1}{2}|\phi^{-1}(y)|^2\right] \left(\lambda_n\phi^{-1}\right)(dy).\end{aligned}$$

Nun verwenden wir, dass $\lambda_n\phi^{-1}$ die Dichte $\left|\det\left(A^{-1}\right)\right|$ bezüglich λ_n hat, was aus der Analysis bekannt sein sollte. Mit Proposition 1.62 ergibt sich

$$\begin{aligned}\mu(B) &= \int 1_B(y) \frac{1}{(2\pi)^{-n/2}} \exp\left[-\frac{1}{2}|\phi^{-1}(y)|^2\right] \left|\det\left(A^{-1}\right)\right| \lambda_n(dy) \\ &= \int_B \frac{1}{(2\pi)^{-n/2}\sqrt{\det \Sigma}} \exp\left[-\frac{1}{2}(y-b)^T\Sigma^{-1}(y-b)\right] \lambda_n(dx).\end{aligned}$$

Damit ist die behauptete Form der Dichte nachgewiesen.

Wenn ϕ nicht invertierbar ist, so hat $\operatorname{im}(\phi) := \{\phi(x) : x \in \mathbb{R}^n\}$ Lebesgue Mass Null. Somit gilt $\mu(\operatorname{im}(\phi)) = 1$ und $\lambda_n(\operatorname{im}(\phi)) = 0$. In diesem Fall kann μ natürlich keine Dichte besitzen. ■

2.2 Erwartungswerte

Definition 2.6
Sei X eine reelle Zufallsgrösse, die auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert ist. Ist $X \geq 0$ oder $X \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$, so heisst $EX = \int X\,dP$ der ***Erwartungswert von** X*.*

Der Erwartungswert ist also für positive Zufallsgrössen stets definiert, kann in diesem Fall jedoch gleich unendlich sein.

Lemma 2.7
Es seien X eine $(S,\mathcal{S})$-wertige Zufallsgrösse und $\varphi : S \to \mathbb{R}$ eine $\mathcal{S}$-$\mathcal{B}$-messbare Abbildung. Dann ist $\varphi \circ X$ (meist $\varphi(X)$ geschrieben) eine reelle Zufallsgrösse. Sei $\mu := PX^{-1}$ die Verteilung von X auf $(S,\mathcal{S})$.

a) *Ist* $\varphi \geq 0$, *so gilt* $E(\varphi(X)) = \int \varphi \, d\mu$.

b) *Es gilt* $\varphi(X) \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ *genau dann, wenn* $\varphi \in \mathcal{L}_1(S, \mathrm{S}, \mu)$ *ist, und in diesem Fall gilt ebenfalls die obige Gleichung.*

Beweis. Das Lemma ist eine Umformulierung von Satz 1.50. ■

Bemerkung 2.8
Sei X eine (S,S)-wertige Zufallsgrösse auf $(\Omega, \mathcal{F}, P)$ und $\mu := PX^{-1}$ ihre Verteilung.

1. Aus Lemma ergibt sich, dass $E(\varphi(X))$ nur von der Verteilung von X (und natürlich von φ) abhängt. Wir betrachten den Spezialfall $(S,\mathrm{S}) = (\mathbb{R}, \mathcal{B})$ und wählen φ als die Identität auf $\mathbb{R}$. Dann folgt aus Lemma 2.7, dass X genau dann in $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ liegt, wenn $\int |x| \, \mu(dx) < \infty$ gilt. In diesem Fall ist $E(X) = \int x \, \mu(dx)$.

2. Ist μ diskret, das heisst, $\mu = \sum_{i \in I} p_i \delta_{x_i}$ mit $x_i \in \mathbb{R}$, $p_i \geq 0$ für alle i aus einer höchstens abzählbaren Menge I und $\sum_{i \in I} p_i = 1$, so gilt $\int |x| \, \mu(dx) = \sum_{i \in I} p_i |x_i|$. Somit gilt $X \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ genau dann, wenn $\sum_{i \in I} p_i |x_i| < \infty$, und in diesem Fall ist $E(X) = \sum_{i \in I} x_i p_i = \sum_{i \in I} x_i P(X = x_i)$.

3. Besitzt μ eine Dichte f bezüglich des Lebesgue-Masses (2.1), so gilt nach Proposition 1.62
$$X \in \mathcal{L}_1(\Omega, \mathcal{F}, P) \iff \int |x| \, \mu(dx) = \int |x| f(x) \, \lambda(dx) < \infty,$$
und in diesem Fall gilt $E(X) = \int x f(x) \, \lambda(dx)$.

Beispiel 2.9
Sei X eine Zufallsgrösse auf $(\Omega, \mathcal{F}, P)$.

1. Ist X standardnormalverteilt, so gelten
$$\frac{1}{\sqrt{2\pi}} \int |x| \mathrm{e}^{-x^2/2} \, \lambda(dx) < \infty \quad \text{und} \quad E(X) = \frac{1}{\sqrt{2\pi}} \int x \mathrm{e}^{-x^2/2} \, \lambda(dx) = 0.$$

2. Ist X normalverteilt mit Mittelwert a und Varianz $\sigma^2 > 0$, so ergibt sich $X \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ und $E(X) = a$.

3. Ist X Cauchy-verteilt zum Parameter $c > 0$, so gilt
$$\int |x| f(x) \, \lambda(dx) = \int \frac{c}{\pi} \frac{|x|}{c^2 + x^2} \, \lambda(dx) = \infty,$$
das heisst $X \notin \mathcal{L}_1(\Omega, \mathcal{F}, P)$.

Das folgende Lemma ergibt sich aus den Eigenschaften des Integrals aus Kapitel 1.

Lemma 2.10

a) *Sind $X, Y \in \mathcal{L}_1(\Omega, F, P)$ und $a, b \in \mathbb{R}$, so gelten $aX + bY \in \mathcal{L}_1(\Omega, F, P)$ und $E(aX + bY) = aE(X) + bE(Y)$ (Linearität des Erwartungswertes).*

b) *Ist $X \geq 0$ mit $E(X) = 0$, so folgt $X = 0$ P-fast sicher.*

c) *Ist $X \in \mathcal{L}_1(\Omega, F, P)$ so gilt $|EX| \leq E(|X|)$.*

Beweis. a) ist einfach die Linearität des Integrals, b) folgt aus Lemma 1.43 b). c) ist offensichtlich durch eine Zerlegung in Positiv- und Negativteil. ■

Definition 2.11
Ist X eine Zufallsgrösse und $p > 0$, so ist $|X|^p$ eine nicht negative Zufallsgrösse. Es bezeichne $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ die Menge der auf $(\Omega, \mathcal{F}, P)$ definierten Zufallsgrössen mit $E(|X|^p) < \infty$.

Lemma 2.12
Für $p \geq p' > 0$ gilt $\mathcal{L}_p(\Omega, \mathcal{F}, P) \subset \mathcal{L}_{p'}(\Omega, \mathcal{F}, P)$.

Beweis. Sei $X \in \mathcal{L}_p$ und $A = \{|X| \leq 1\}$. Dann gilt

$$E(|X|^{p'}) = \int_A |X|^{p'} dP + \int_{A^c} |X|^{p'} dP \leq P(A) + \int_{A^c} |X|^p dP \leq 1 + E(|X|^p) < \infty.$$

■

Die nachfolgenden Ungleichungen sollen hier nicht bewiesen werden, weil die Beweise praktisch identisch mit aus der Stochastik oder der Analysis schon bekannten sind.

Satz 2.13

a) ***Markoff-Ungleichung:*** *Sei $X \in \mathcal{L}_p$ mit $p > 0$. Dann gilt für alle $a > 0$ die Abschätzung $P(|X| \geq a) \leq a^{-p} E(|X|^p)$.*

b) ***Schwarzsche Ungleichung:*** *Sind $X, Y \in \mathcal{L}_2$, so gelten $XY \in \mathcal{L}_1$ und $E|XY| \leq (E(X^2)E(Y^2))^{1/2}$.*

c) ***Höldersche Ungleichung:*** *Seien $p, q \geq 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Für $X \in \mathcal{L}_p$, $Y \in \mathcal{L}_q$ gelten*

$$XY \in \mathcal{L}_1 \quad \text{und} \quad E|XY| \leq (E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}.$$

Definition 2.14
Sei $X \in \mathcal{L}_1$. Die ***Varianz*** *von X ist definiert durch*

$$\operatorname{var}(X) = E((X - EX)^2) \in [0, \infty].$$

Bemerkung 2.15
Die folgenden Eigenschaften sind einfache Folgerungen aus der Definition: Sei $X \in \mathcal{L}_1$.

1. $\operatorname{var}(X) = E(X^2 - 2XEX + (EX)^2) = E(X^2) - (EX)^2$.

2. Es gilt $\operatorname{var}(X) < \infty \iff X \in \mathcal{L}_2$.

3. $\operatorname{var}(X) = 0 \iff X = EX$ fast sicher.

4. Die Markoff-Ungleichung angewandt auf $X - EX$ mit $p = 2$ ergibt $P(|X - EX| \geq a) \leq \frac{1}{a^2} \operatorname{var}(X)$. (Tschebyscheff-Ungleichung)

Ist X standard normalverteilt, so gilt $\operatorname{var}(X) = 1$. Ist X normalverteilt mit normalverteilt mit Mittel a und Varianz σ^2 (siehe Beispiel 2.3 c), so ist die Varianz natürlich tatsächlich σ^2, was man ebenfalls sofort nachrechnet.

Wir kommen nun noch zu den analogen Begriffsbildungen für mehrdimensionale Zufallsgrössen

Definition 2.16
Ist $X = (X_1, \ldots, X_n)$ ein Zufallsvektor, so definiert man seinen Erwartungswert $EX \in \mathbb{R}^n$ komponentenweise durch $EX = (EX_1, \ldots, EX_n)$ (falls dies existiert).

An die Stelle der Varianz treten die Kovarianzen:

Definition 2.17

a) *Sind X und Y zwei Zufallsgrössen aus $\mathcal{L}_1$ mit $XY \in \mathcal{L}_1$, so ist ihre* ***Kovarianz*** $\operatorname{cov}(X, Y)$ *definiert durch*

$$\operatorname{cov}(X, Y) = E(XY) - E(X)E(Y) = E((X - EX)(Y - EY)).$$

b) *Ist $X = (X_1, \ldots, X_n)$ ein Zufallsvektor mit $X_i \in \mathcal{L}_1$ und $X_i X_j \in \mathcal{L}_1$ für alle $i, j \in \{1, \ldots, n\}$, so ist die* ***Kovarianzmatrix*** $\Sigma(X) = (\sigma_{ij}(X))$ *definiert durch* $\sigma_{ij}(X) = \operatorname{cov}(X_i, X_j)$.

Offenbar ist für eine eindimensionale Zufallsgrösse X : $\operatorname{var}(X) = \operatorname{cov}(X, X)$. Ist X ein Zufallsvektor, als Spaltenvektor geschrieben, so ist $\Sigma(X) = E((X - EX)(X - EX)^T)$.

Lemma 2.18

a) *Sind $X, Y \in \mathcal{L}_2$, so ist die Kovarianz $\operatorname{cov}(X, Y)$ definiert.*

b) *Die Kovarianzmatrix ist symmetrisch und positiv semidefinit.*

Beweis. a) folgt aus der Schwarzschen Ungleichung. b): Die Symmetrie ist offensichtlich. Ferner gilt alle $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$

$$0 \leq E\left(\sum_{i=1}^n \lambda_i (X_i - E(X_i))\right)^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \operatorname{cov}(X_i, X_j).$$

Daraus folgt die Definitheit. ■

Beispiel 2.19

a) Sei X standardnormalverteilt, das heisst, die Verteilung μ von X auf $\mathbb{R}^n$ hat die Dichte $x = (x_1, \ldots, x_n) \mapsto (2\pi)^{-n/2} \exp(-\frac{1}{2} \sum_{k=1}^n x_k^2)$. Dann gilt für $i \in \{1, \ldots, n\}$

$$E(X_i) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_i \exp\left[-\frac{1}{2} \sum\nolimits_{k=1}^n x_k^2\right] \lambda_n(dx) = 0.$$

Für alle $i, j \in \{1, \ldots, n\}$ mit $i \neq j$ gelten

$$E(X_i X_j) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_i x_j \exp\left[-\frac{1}{2} \sum\nolimits_{k=1}^n x_k^2\right] \lambda_n(dx) = 0$$

und

$$\begin{aligned} E(X_i^2) &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_i^2 \exp\left[-\frac{1}{2} \sum\nolimits_{k=1}^n x_k^2\right] \lambda_n(dx) \\ &= (2\pi)^{-1/2} \int_{\mathbb{R}} x_i^2 \mathrm{e}^{-x_i^2/2} \, \lambda(dx_i) = 1, \end{aligned}$$

das heisst, $\Sigma(X)$ ist die Einheitsmatrix.

b) Sei X ein n-dimensionaler Zufallsvektor mit Kovarianzmatrix $\Sigma(X)$ und Erwartungswert $a \in \mathbb{R}^n$. Sei ferner A eine $m \times n$-Matrix und $b \in \mathbb{R}^m$. Wir definieren den m-dimensionalen Zufallsvektor Y durch $Y = AX + b$. Dann gelten

$$EY = Aa + b$$

$$\Sigma(Y) = E((Y - EY)(Y - EY)^T) = E(A\,(X - a)\,(X - a)^T\, A^T) = A\Sigma\,(X)\,A^T.$$

Speziell sehen wir für die in Definition 2.4 b) eingeführte allgemeine Normalverteilung, dass die Kovarianzmatrix gleich AA^T und der Vektor der Erwartungswerte gleich b ist.

2.3 Charakteristische Funktionen

Definition 2.20

Sei μ ein Wahrscheinlichkeitsmass auf $(\mathbb{R}^n, \mathcal{B}_n)$. Die ***charakteristische Funktion*** *$\hat{\mu}$ von μ ist die Abbildung von $\mathbb{R}^n$ nach $\mathbb{C}$, die durch*

$$\hat{\mu}(t) = \int \mathrm{e}^{i\langle t,x\rangle} \, \mu(dx) = \int \cos(\langle t, x\rangle) \, \mu(dx) + i \int \sin(\langle t, x\rangle) \, \mu(dx), \quad t \in \mathbb{R}^n,$$

definiert wird. Dabei bezeichnet i die imaginäre Einheit und $\langle t, x\rangle = \sum_{j=1}^n t_j x_j$ ist das Skalarprodukt von t und x. Die charakteristische Funktion eines Zufallsvektors X ist die charakteristische Funktion der Verteilung von X; sie kann nach Lemma 2.7 als $E(\exp(i\langle t, X\rangle))$ geschrieben werden. Die charakteristische Funktion eines Zufallsvektors X (oder einer reellen Zufallsgrösse X) bezeichnen wir oft mit χ_X.

Die charakteristische Funktion ist offenbar für alle $t \in \mathbb{R}^n$ definiert, da Sinus und Cosinus beschränkt sind, und sie ist stetig in t nach dem Satz von Lebesgue. Die Theorie charakteristischer Funktionen soll hier nur rudimentär behandelt werden. Als einziges Ergebnis werden wir das folgende benötigen:

Satz 2.21
Es seien μ, ν zwei Wahrscheinlichkeitsmasse auf $(\mathbb{R}^n, \mathcal{B}_n)$. Gilt $\hat{\mu}(t) = \hat{\nu}(t)$ für alle $t \in \mathbb{R}^n$, so gilt $\mu = \nu$.

Beweis. Da die Familie der kompakten Mengen in $\mathbb{R}^n$ ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_n$ ist (Lemma 1.21 e)), genügt es nach Satz 1.15 nachzuweisen, dass $\mu(K) = \nu(K)$ für alle kompakten Mengen K gilt. Für eine derartige Menge K und $m \in \mathbb{N}$ sei

$$f_m(x) = \begin{cases} 1 & \text{falls } x \in K, \\ 0 & \text{falls } d(x, K) := \inf\{\, |x - y| : y \in K \,\} \geq 1/m, \\ 1 - m\, d(x, K) & \text{sonst.} \end{cases}$$

Dann hat f_m die folgenden Eigenschaften:

1. $0 \leq f_m(x) \leq 1$ für alle $x \in \mathbb{R}^n$,
2. f_m ist stetig,
3. f_m hat kompakten Träger,
4. $f_m(x) \downarrow 1_K(x)$ für $m \to \infty$.

Falls $\int f_m \, d\mu = \int f_m \, d\nu$ für alle $m \in \mathbb{N}$ gilt, so folgt $\mu(K) = \nu(K)$ mit dem Satz von Lebesgue. Es genügt also nachzuweisen, dass $\int f \, d\mu = \int f \, d\nu$ für alle f gilt, die die obigen Bedingungen 1.-3. erfüllen.

Sei also f eine derartige Funktion. Für $\varepsilon > 0$ sei $N > 0$ so gross gewählt, dass

$$B_N := [-N, N]^n \supset \{\, x : f(x) \neq 0 \,\}$$

und $\max\{\mu(B_N^c), \nu(B_N^c)\} \leq \varepsilon$ gelten. Nach dem Weierstrassschen Approximationssatz gibt es eine Funktion $g : \mathbb{R}^n \to \mathbb{C}$ der Form $g(x) = \sum_{j=1}^m c_j \exp(i\langle \frac{\pi}{N} t_j, x\rangle)$ mit $c_j \in \mathbb{C}$ und $t_j \in \mathbb{Z}^n$, die periodisch in jeder Komponente ist und f in B_N bis auf ε approximiert, das heisst, $\sup\{\, |f(x) - g(x)| : x \in B_N \,\} \leq \varepsilon$. Es folgen $\sup_{x \in \mathbb{R}^n} |g(x)| \leq 1 + \varepsilon$ und

$$\left| \int f \, d\mu - \int f \, d\nu \right| \leq \left| \int f \, d\mu - \int g \, d\mu \right| + \left| \int g \, d\mu - \int g \, d\nu \right| + \left| \int g \, d\nu - \int f \, d\nu \right|.$$

Der zweite Summand ist nach der Voraussetzung $\hat{\mu} = \hat{\nu}$ gleich null. Der erste Summand kann wegen $|g(x)| \leq 1 + \varepsilon$ und $|f(x)| \leq 1$ für alle $x \in \mathbb{R}$ folgendermassen abgeschätzt werden:

$$\begin{aligned}\left|\int f\,d\mu - \int g\,d\mu\right| &\le \left|\int_{B_N} f\,d\mu - \int_{B_N} g\,d\mu\right| + \int_{B_N^c} |f|\;d\mu + \int_{B_N^c} |g|\;d\mu \\ &\le \int_{B_N} |f-g|\,d\mu + (1+\varepsilon)\mu(B_N^c) \\ &\le \varepsilon\mu(B_N) + (1+\varepsilon)\mu(B_N^c) \\ &\le \varepsilon(2+\varepsilon).\end{aligned}$$

Der dritte Summand wird analog abgeschätzt. Da $\varepsilon > 0$ beliebig war, folgt $\int f\,d\mu = \int f\,d\nu$. ■

Beispiel 2.22

a) Sei μ die Standardnormalverteilung. Dann gilt

$$\hat{\mu}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{itx}\mathrm{e}^{-x^2/2}\,dx = \mathrm{e}^{-t^2/2},\ t \in \mathbb{R}.$$

Zum Beweis beachten wir zuerst, dass mit einer Substitution $x \to -x$

$$\hat{\mu}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-itx}\mathrm{e}^{-x^2/2}\,dx$$

also

$$\hat{\mu}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx)\,\mathrm{e}^{-x^2/2}\,dx$$

folgt. $\hat{\mu}(t)$ ist also reellwertig. Wir differentieren nach t : (Der einfache Beweis, dass die Ableitung existiert und man Integration mit Differentiation vertauschen kann, sei dem Leser überlassen)

$$\begin{aligned}\frac{d\hat{\mu}(t)}{dt} &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\sin(tx)\,\mathrm{e}^{-x^2/2}\,dx \\ &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t\cos(tx)\,\mathrm{e}^{-x^2/2}\,dx,\end{aligned}$$

die zweite Gleichung mit partieller Integration. $\hat{\mu}(t)$ erfüllt daher die Differentialgleichung

$$\frac{d\hat{\mu}(t)}{dt} = -t\hat{\mu}(t).$$

Die einzige Lösung dieser Differentialgleichung, die $\hat{\mu}(0) = 1$ erfüllt, ist

$$\hat{\mu}(t) = \mathrm{e}^{-t^2/2}.$$

b) Sei μ die Cauchy-Verteilung zum Parameter $c > 0$. Dann gilt

$$\hat{\mu}(t) = \frac{c}{\pi} \int_{-\infty}^{\infty} \mathrm{e}^{itx} \frac{dx}{c^2 + x^2}, \quad t \in \mathbb{R}.$$

Die Funktion $\mathbb{C} \ni z \mapsto \frac{1}{c^2+z^2}$ hat Pole in $\pm ic$. Eine Anwendung des Residuensatzes ergibt $\hat{\mu}(t) = \mathrm{e}^{-c|t|}$.

c) Sei μ die Standardnormalverteilung in $(\mathbb{R}^n, \mathcal{B}_n)$. Dann folgt

$$\hat{\mu}(t) = \exp\left[-\frac{1}{2} \sum_{j=1}^{n} t_j^2\right] = \mathrm{e}^{-\langle t,t\rangle/2} \quad \text{für alle } t = (t_1, \ldots, t_n) \in \mathbb{R}^n.$$

d) Die allgemeine Normalverteilung ist das Bildmass $\nu = \mu\phi^{-1}$ der Standardnormalverteilung μ unter einer affinen Transformation $\mathbb{R}^n \ni x \mapsto \phi(x) = Ax + b \in \mathbb{R}^n$. Bezeichnet A^T die Transponierte von A, so gilt

$$\begin{aligned} \hat{\nu}(t) &= \int \mathrm{e}^{i\langle t,x\rangle}\, \nu(dx) = \int \mathrm{e}^{i\langle t,\phi(x)\rangle}\, \mu(dx) = \mathrm{e}^{i\langle t,b\rangle} \int \mathrm{e}^{i\langle A^T t,x\rangle}\, \mu(dx) \\ &= \mathrm{e}^{i\langle t,b\rangle} \hat{\mu}(A^T t) = \mathrm{e}^{i\langle t,b\rangle} \mathrm{e}^{-\langle A^T t, A^T t\rangle/2} = \exp\left[i\langle t,b\rangle - \frac{1}{2}\langle t, \Sigma t\rangle\right], \end{aligned}$$

mit $\Sigma = AA^T$ als der Kovarianzmatrix von ν (siehe Beispiel 2.19 b)).

Satz 2.23
Für jedes $b \in \mathbb{R}^n$ und jede positiv semidefinite, symmetrische $n \times n$-Matrix Σ gibt es genau eine Normalverteilung μ auf $\mathbb{R}^n$ mit b als Erwartungswert und Σ als Kovarianzmatrix.

Beweis. Die Eindeutigkeit folgt aus Satz 2.21 und der Rechnung im obigen Beispiel. Die Existenz folgt daraus, dass mindestens eine $n \times n$-Matrix A existiert mit $AA^T = \Sigma$, wenn Σ eine nicht negative symmetrische Matrix ist. ■

Korollar 2.24
Sei μ die Normalverteilung auf $\mathbb{R}^n$ mit Kovarianzmatrix Σ und $a \in \mathbb{R}^n$ als Vektor der Erwartungswerte, und sei $\phi : \mathbb{R}^n \to \mathbb{R}^m$ eine affine Abbildung, d.h. eine Abbildung der Form $x \to \phi(x) := Ax + b$, A eine $m \times n$-Matrix und $b \in \mathbb{R}^m$. Dann ist $\mu\phi^{-1}$ die Normalverteilung auf $\mathbb{R}^m$ mit Erwartungswert $Aa + b$ und der Kovarianzmatrix $A\Sigma A^T$.

Beweis.

$$\begin{aligned} \widehat{\mu\varphi^{-1}}(t) &= \int \mathrm{e}^{i\langle t,x\rangle}\, \mu\varphi^{-1}(dx) = \int \mathrm{e}^{i\langle t,Ax+b\rangle}\, \mu(dx) \\ &= \mathrm{e}^{i\langle t,b\rangle} \int \mathrm{e}^{i\langle A^T t,x\rangle}\, \mu(dx) = \mathrm{e}^{i\langle t,b\rangle} \exp\left[i\langle A^T t, a\rangle - \frac{1}{2}\langle A^T t, \Sigma A^T t\rangle\right] \\ &= \exp\left[i\langle t, Aa + b\rangle - \frac{1}{2}\langle t, A\Sigma A^T t\rangle\right]. \end{aligned}$$

Nun folgt die Aussage aus dem vorangegangen Satz und Beispiel 2.22 d). ■

2.4 Konvergenz von Folgen von Zufallsgrössen

Im folgenden sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallsgrössen, die auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert sind. In der Wahrscheinlichkeitstheorie sind drei Konvergenzbegriffe besonders wichtig.

Definition 2.25

1. *Die Folge $\{X_n\}_{n\in\mathbb{N}}$ konvergiert **fast sicher** gegen eine Zufallsgrösse X, falls*

$$P(\{\, \omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\,\}) = 1$$

gilt (Notation: $X_n \to X$ P-fast sicher).

2. *Die Folge $\{X_n\}_{n\in\mathbb{N}} \subset \mathcal{L}_p(\Omega, F, P)$ konvergiert **im p-ten Mittel** $(p > 0)$ gegen eine Zufallsgrösse X, falls $X \in \mathcal{L}_p(\Omega, F, P)$ und*

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0$$

gilt.

3. *Die Folge $\{X_n\}_{n\in\mathbb{N}}$ konvergiert **in Wahrscheinlichkeit** gegen eine Zufallsgrösse X, falls*

$$P(|X_n - X| \geq \varepsilon) \to 0 \quad \text{für } n \to \infty$$

für alle $\varepsilon > 0$ gilt.

Satz 2.26

a) *Fast sichere Konvergenz impliziert Konvergenz in Wahrscheinlichkeit.*

b) *Konvergenz im p-ten Mittel impliziert Konvergenz in Wahrscheinlichkeit.*

Beweis. Der Beweis von b) folgt sofort aus der Markoff-Ungleichung.

a): Sei $Y_n = 1_{\{|X_n - X| \geq \varepsilon\}}$ für $\varepsilon > 0$. Gilt $X_n \to X$ fast sicher, so gilt $Y_n \to 0$ fast sicher. Wegen $|Y_n| \leq 1$ folgt aus dem Satz von Lebesgue

$$P(|X_n - X| \geq \varepsilon) = E(Y_n) \to 0.$$

■

Die anderen denkbaren Implikationen sind nicht richtig, wie die folgenden zwei Beispiele belegen:

Beispiel 2.27

Sei $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}_{[0,1]}, \lambda)$.

a) Wähle $X_n = n^{1/p} 1_{[0,1/n]}$ für $p > 0$. Dann gilt $X_n \to 0$ fast sicher und in Wahrscheinlichkeit, aber $E(|X_n|^p) = 1$ für alle $n \in \mathbb{N}$, das heisst, $\{X_n\}_{n\in\mathbb{N}}$ konvergiert nicht im p-ten Mittel gegen null.

b) Ist $n = 2^m + k$ für $m \in \mathbb{N}_0$ und $0 \le k < 2^m$, so setzt man $X_n = 1_{[k2^{-m},(k+1)2^{-m}]}$. Offenbar konvergiert die Folge $\{X_n(\omega)\}_{n\in\mathbb{N}}$ für kein $\omega \in [0,1]$. Andererseits gelten $P(|X_n| \ge \varepsilon) \le 2^{-m}$ für alle $\varepsilon > 0$ und $E(|X_n|^p) = 2^{-m}$ für $p > 0$, das heisst $\{X_n\}_{n\in\mathbb{N}}$ konvergiert gegen null in Wahrscheinlichkeit und im p-ten Mittel.

Unter Zusatzbedingungen impliziert die fast sichere Konvergenz die Konvergenz im p-ten Mittel:

Satz 2.28
Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallsgrössen, die fast sicher gegen X konvergiert. Gilt $|X_n| \le Y$ fast sicher für eine Zufallsgrösse $Y \in \mathcal{L}_p$ (für $p > 0$), so gilt $X_n \to X$ im p-ten Mittel.

Beweis. Es gelten $|X_n - X|^p \le (|X_n| + |X|)^p \le (2Y)^p \le 2^p Y^p \in \mathcal{L}_1$ und $|X_n - X|^p \to 0$ fast sicher. Daher folgt aus dem Satz von Lebesgue $E(|X_n - X|^p) \to 0$. ■

Wie aus Beispiel 2.27 b) hervorgeht, folgt aus der Konvergenz in Wahrscheinlichkeit nicht die fast sichere Konvergenz. Es gilt aber der folgende Satz:

Satz 2.29
Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallsgrössen, die in Wahrscheinlichkeit gegen X konvergiert. Dann existiert eine Teilfolge $\{X_{n_k}\}_{k\in\mathbb{N}}$ mit $\lim_{k\to\infty} X_{n_k} = X$ fast sicher.

Zum Beweis benötigt man das folgende sehr einfache, aber wichtige Lemma.

Lemma 2.30 (1. Borel-Cantelli-Lemma)
Sei $\{A_n\}_{n\in\mathbb{N}}$ eine Folge von Ereignissen mit $\sum_{n=1}^\infty P(A_n) < \infty$. Dann gilt $P(\limsup_{n\to\infty} A_n) = 0$.

Beweis. Aus $B_k := \bigcup_{n=k}^\infty A_n \downarrow \limsup_{n\to\infty} A_n$ und Satz 1.12 a) und c) folgt

$$P(\limsup_{n\to\infty} A_n) = \lim_{k\to\infty} P(B_k) \le \lim_{k\to\infty} \sum_{n=k}^\infty P(A_n) = 0.$$

■

Beweis von Satz 2.29. Zu jedem $k \in \mathbb{N}$ existiert nach Voraussetzung ein $n_k \in \mathbb{N}$ mit $P(|X_{n_k} - X| \ge 1/k) \le 1/k^2$. Wir können $n_{k+1} > n_k$ für alle $k \in \mathbb{N}$ annehmen. Da $\sum_{k=1}^\infty 1/k^2 < \infty$ gilt, folgt $P(\limsup_{k\to\infty}\{\,|X_{n_k} - X| \ge 1/k\,\}) = 0$ aus Lemma 2.30. Für $\omega \notin \limsup_{k\to\infty}\{\,|X_{n_k} - X| \ge 1/k\,\}$ gilt $|X_{n_k}(\omega) - X(\omega)| < 1/k$ für genügend grosse k, das heisst $\lim_{k\to\infty} X_{n_k}(\omega) = X(\omega)$. ■

Bemerkung 2.31
Alle drei Konvergenztypen sind *vollständig,* das heisst, dass jede Cauchy-Folge konvergiert. Für die fast sichere Konvergenz ist das klar, denn wenn $X_n - X_m \to 0$ fast sicher für $n, m \to \infty$ gilt, dann folgt aufgrund der Vollständigkeit von $\mathbb{R}$, dass $\{X_n(\omega)\}_{n\in\mathbb{N}}$ für fast alle $\omega \in \Omega$ konvergiert. Mit Hilfe von Lemma 2.30 folgt das Entsprechende für die Konvergenz in Wahrscheinlichkeit:

Satz 2.32
Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallsgrössen mit

$$\lim_{n,m\to\infty} P(|X_n - X_m| \geq \varepsilon) = 0$$

für alle $\varepsilon > 0$. Dann existiert eine Zufallsgrösse X mit $X_n \to X$ in Wahrscheinlichkeit.

Beweis. Wähle wie im Beweis des vorangegangenen Satzes eine Teilfolge $\{n_k\}_{k\in\mathbb{N}}$ nun aber mit

$$P(\{\, |X_{n_k} - X_{n_{k+1}}| \geq 1/k^2 \,\}) \leq 1/k^2.$$

Aus dem Borel-Cantelli-Lemma folgt

$$P\left(\limsup_{k\to\infty}\{\, |X_{n_k} - X_{n_{k+1}}| \geq 1/k^2 \,\}\right) = 0.$$

Für $\omega \notin \limsup_{k\to\infty}\{\, |X_{n_k} - X_{n_{k+1}}| \geq 1/k^2 \,\}$ ist $\{X_{n_k}(\omega)\}_{k\in\mathbb{N}}$ offenbar eine Cauchy-Folge in $\mathbb{R}$, das heisst, X_{n_k} konvergiert für $k \to \infty$ fast sicher gegen eine Zufallsgrösse X, also nach Satz 2.26 auch in Wahrscheinlichkeit. Für $\varepsilon > 0$ gilt

$$P(|X_m - X| \geq \varepsilon) \leq P(|X_m - X_{n_k}| \geq \varepsilon/2) + P(|X_{n_k} - X| \geq \varepsilon/2)$$

für alle m und k. Wählt man k als die kleinste Zahl mit $n_k \geq m$, dann folgt

$$\lim_{m\to\infty} P(|X_m - X| \geq \varepsilon) = 0.$$

■

Für die Konvergenz im p-ten Mittel gilt die Vollständigkeit auch, soll aber hier nicht bewiesen werden. (Dies sollte aus der Analysis bekannt sein, zumindest für das Lebesgue Mass.)

2.5 Unabhängigkeit

Es sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum. Im folgenden wird von Familien von Teilmengen von Ω *stets* stillschweigend vorausgesetzt, dass sie Ω enthalten.

Definition 2.33

a) *Teilmengen $\mathcal{E}_1, \dots, \mathcal{E}_n$ von $\mathcal{F}$ (mit $\Omega \in \mathcal{E}_i$!) heissen* ***unabhängig****, wenn für $A_i \in \mathcal{E}_i$, $1 \leq i \leq n$, die folgende Gleichung gilt:*

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n). \tag{2.2}$$

b) *Seien eine Indexmenge I und $\mathcal{E}_i$ für $i \in I$ Teilmengen von $\mathcal{F}$. Sie heissen unabhängig, wenn je endlich viele unabhängig sind.*

c) *Ereignisse A_i für $i \in I$ heissen unabhängig, wenn die Mengensysteme $\{A_i, \Omega\}$, $i \in I$, unabhängig sind.*

Notation: Für zwei unabhängige Teilmengen $\mathcal{E}_1$, $\mathcal{E}_2$ von $\mathcal{F}$ schreiben wir $\mathcal{E}_1 \perp \mathcal{E}_2$.

Die Voraussetzung, dass die Mengensysteme stets Ω enthalten, dient nur der bequemen Notation. Dies hat nämlich zur Folge, dass für unabhängige Mengensysteme $\mathcal{E}_1, \ldots, \mathcal{E}_n$ auch stets

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j}) \tag{2.3}$$

für $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ und $A_{i_j} \in \mathcal{E}_{i_j}$ ist. Setzt man $\Omega \in \mathcal{E}_i$ *nicht* voraus, so muss man (2.3) als Definition verwenden, was offensichtlich stets einen grösseren Schreibaufwand mit Doppelindices erfordert.

Lemma 2.34

a) *Sind die $\mathcal{E}_i$ für $i \in I$ unabhängig und gilt $\mathcal{D}_i \subset \mathcal{E}_i$ für $i \in I$, so sind die $\mathcal{D}_i$ für $i \in I$ unabhängig.*

b) *Gilt $\mathcal{D} \perp \mathcal{E}_i$ für $i \in I$, so gilt $\mathcal{D} \perp \bigcup_{i \in I} \mathcal{E}_i$.*

Beweis. a) ist klar.

b) Für $A \in \mathcal{D}$ und $B \in \bigcup_{i \in I} \mathcal{E}_i$ existiert ein $i \in I$ mit $B \in \mathcal{E}_i$, das heisst, dass $P(A \cap B) = P(A)P(B)$ gilt. ■

Wir diskutieren nun einige Möglichkeiten, Unabhängigkeitsaussagen von Mengensystemen auf grössere Mengensysteme hochzuziehen.

Satz 2.35

Es seien $\mathcal{D}_i$ für $i \in I$ unabhängige Teilmengen von $\mathcal{F}$ (stets $\Omega \in \mathcal{D}_i$). Sind die $\mathcal{D}_i$ durchschnittstabil, so sind die $\sigma(\mathcal{D}_i)$ für $i \in I$ unabhängig.

Beweis. Es genügt den Satz zu zeigen, wenn I endlich ist. Sei etwa $I = \{1, \ldots, n\}$. Wir müssen (2.2) für $A_i \in \sigma(\mathcal{D}_i)$ nachweisen. Für $0 \leq k \leq n$ sei L_k die folgende Aussage:

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n), \ \forall A_i \in \sigma(\mathcal{D}_i) \text{ für } i \leq k, \ \forall A_i \in \mathcal{D}_i \text{ für } i > k.$$

Die Aussage L_0 gilt wegen der Unabhängigkeit der $\mathcal{D}_i$.

Wir zeigen $L_k \Rightarrow L_{k+1}$ für $0 \leq k \leq n-1$. Wir betrachten dazu das Mengensystem $\mathcal{A}_{k+1}$ bestehend aus den Mengen $A_{k+1} \in \sigma(\mathcal{D}_{k+1})$, die die Eigenschaft haben, dass die Gleichung (2.2) für $\forall A_1 \in \sigma(\mathcal{D}_1), \ldots, \forall A_k \in \sigma(\mathcal{D}_k), \forall A_{k+2} \in \mathcal{D}_{k+2}, \ldots, \forall A_n \in \mathcal{D}_n$ gilt

Aus L_k folgt $\mathcal{A}_{k+1} \supset \mathcal{D}_{k+1}$. Wir zeigen, dass $\mathcal{A}_{k+1}$ ein Dynkin-System ist.

1. $\Omega \in \mathcal{A}_{k+1}$ gilt wegen $\Omega \in \mathcal{D}_{k+1}$.

2. Für $D \in \mathcal{A}_{k+1}$ gilt

$$
\begin{aligned}
P\left(\bigcap_{j=1}^{k} A_j \cap D^c \cap \bigcap_{j=k+2}^{n} A_j\right) &= P\left(\bigcap_{j=1}^{k} A_j \cap \bigcap_{j=k+2}^{n} A_j\right) - P\left(\bigcap_{j=1}^{k} A_j \cap D \cap \bigcap_{j=k+2}^{n} A_j\right) \\
&= \prod_{j:j\neq k} P\left(A_j\right) - P(D) \prod_{j:j\neq k} P\left(A_j\right) \\
&= \prod_{j:j\neq k} P\left(A_j\right) P(D^c).
\end{aligned}
$$

   für alle A_i gemäss den obigen Bedingungen, das heisst $D^c \in \mathcal{A}_{k+1}$.

3. Für paarweise disjunkte $D_i \in \mathcal{A}_{k+1}$, $i \in \mathbb{N}$, folgt analog $\bigcup_{i=1}^{\infty} D_i \in \mathcal{A}_{k+1}$.

Nach Satz 1.9 folgt $\mathcal{A}_{k+1} = \sigma(\mathcal{D}_{k+1})$, das heisst, dass L_{k+1} gilt. ■

Bemerkung 2.36
Da das Mengensystem $\{A, \Omega\}$ durchschnittstabil ist, folgt, wenn die Ereignisse A_i für $i \in I$ unabhängig sind, dass auch die σ-Algebren $\{\emptyset, A_i, A_i^c, \Omega\}$ unabhängig sind; insbesondere dann auch die Komplemente A_i^c.

Korollar 2.37
Es seien $\mathcal{D}_i \subset \mathcal{F}$ für $i \in I$ unabhängig und durchschnittstabil. Es sei $(I_k)_{k\in K}$ eine Familie von paarweise disjunkten Teilmengen von I. Dann sind die $\sigma(\bigcup_{j\in I_k} \mathcal{D}_j)$ für $k \in K$ unabhängig.

Beweis. Für $k \in K$ sei $\hat{\mathcal{D}}_k$ die Familie der endlichen Durchschnitte von Elementen aus $\mathcal{D}_j$ für $j \in I_k$. Das Mengensystem $\hat{\mathcal{D}}_k$ ist offenbar durchschnittstabil, und da die $\mathcal{D}_j$ durchschnittstabil sind, hat jedes Element aus $\hat{\mathcal{D}}_k$ die Gestalt $A_{j_1} \cap \cdots \cap A_{j_n}$ mit $n \in \mathbb{N}$, $A_j \in \mathcal{D}_j$ und verschiedenen $j_1, \ldots, j_n \in I_k$. Daraus folgt sofort, dass die $\hat{\mathcal{D}}_k$ für $k \in K$ unabhängig sind. Da $\hat{\mathcal{D}}_k \supset \mathcal{D}_j$ für alle $j \in I_k$ ist, gilt $\sigma(\bigcup_{j\in I_k} \mathcal{D}_j) = \sigma(\hat{\mathcal{D}}_k)$. Das Lemma folgt nun aus Satz 2.35. ■

Als Folgerung ergibt sich das folgende verblüffende Resultat mit einem ebenso verblüffenden Beweis:

Satz 2.38 (Kolmogoroffs 0-1-Gesetz)
Sei $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen Teil-σ-Algebren $\mathcal{F}_n$ von $\mathcal{F}$. Seien $\check{\mathcal{F}}_n := \bigvee_{k=n}^{\infty} \mathcal{F}_k$ und $\mathcal{T}_\infty = \bigcap_{n=1}^{\infty} \check{\mathcal{F}}_n$. Für $A \in \mathcal{T}_\infty$ gilt $P(A) \in \{0, 1\}$.

$\mathcal{T}_\infty$ heisst die σ-Algebra der **terminalen Ereignisse** der Folge $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ oder auch terminale σ-Algebra der $\mathcal{F}_n$.
Beweis. Nach Korollar 2.37 gilt $\check{\mathcal{F}}_{n+1} \perp \bigvee_{k=1}^{n} \mathcal{F}_k$, also nach Lemma 2.34 a): $\mathcal{T}_\infty \perp \bigvee_{k=1}^{n} \mathcal{F}_k$ für alle $n \in \mathbb{N}$. Somit gilt nach Teil b) desselben Lemmas $\mathcal{T}_\infty \perp \bigcup_{n=1}^{\infty} \bigvee_{k=1}^{n} \mathcal{F}_k$.

Da die rechte Seite als Vereinigung einer aufsteigenden Folge von σ-Algebren durchschnittstabil ist, folgt nach Satz 2.35

$$\mathcal{T}_\infty \perp \sigma\left(\bigcup_{n=1}^{\infty} \bigvee_{k=1}^{n} \mathcal{F}_k\right) = \bigvee_{n=1}^{\infty} \mathcal{F}_n.$$

Nun ist aber $\check{\mathcal{F}}_n \subset \bigvee_{k=1}^{\infty} \mathcal{F}_k$ für alle $n \in \mathbb{N}$, also auch $\mathcal{T}_\infty \subset \bigvee_{n=1}^{\infty} \mathcal{F}_n$. Nach Lemma 2.34 folgt also $\mathcal{T}_\infty \perp \mathcal{T}_\infty$, das heisst, für $A \in \mathcal{T}_\infty$ gilt $P(A) = P(A \cap A) = P(A)^2$, das heisst $P(A) \in \{0, 1\}$. ■

Teil-σ-Algebren $\mathcal{T} \subset \mathcal{F}$ in einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$, die die Eigenschaft haben, dass $P(A) \in \{0, 1\}$ für alle $A \in \mathcal{T}$ ist, spielen in der Wahrscheinlichkeitstheorie eine grosse Rolle. Die einfachste derartige Teil-σ-Algebra ist natürlich $\{\emptyset, \Omega\}$. $\mathcal{T}_\infty$ aus dem obigen Satz ist jedoch im allgemeinen *sehr* viel grösser als $\{\emptyset, \Omega\}$. Dennoch ist P eingeschränkt auf $\mathcal{T}_\infty$ gewissermassen trivial. Ereignisse, die in $\mathcal{T}_\infty$ liegen, sind gewissermassen „nicht mehr zufällig“. Dieser Aspekt spiegelt sich auch in dem nachfolgenden Lemma über Zufallsgrössen wieder:

Lemma 2.39
Es sei $\mathcal{T} \subset \mathcal{F}$ eine σ-Algebra mit $P(A) \in \{0, 1\}$ für alle $A \in \mathcal{T}$. Ist Z eine $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$-wertige, $\mathcal{T}$-messbare Zufallsgrösse, so existiert ein $c \in \overline{\mathbb{R}}$ mit $P(Z = c) = 1$.

Beweis. Sei $F(t) = P(Z \leq t)$, so ist $F(t) \in \{0, 1\}$ für $t \in \mathbb{R}$. Die Funktion F ist nichtfallend. Demzufolge sind drei Fälle möglich:

1. $F(t) = 0$ für alle $t \in \mathbb{R} \Rightarrow P(Z > n) = 1$ für alle $n \Rightarrow P(Z = \infty) = 1$.
2. $F(t) = 1$ für alle $t \in \mathbb{R} \Rightarrow P(Z \leq n) = 1$ für alle $n \Rightarrow P(Z = -\infty) = 1$.
3. F springt an einer Stelle $t_0 \in \mathbb{R}$ von 0 nach 1. Dann gilt

$$F\left(t_0 + \frac{1}{n}\right) - F\left(t_0 - \frac{1}{n}\right) = P\left(Z \in \left(t_0 - \frac{1}{n},\ t_0 + \frac{1}{n}\right]\right) = 1 \quad \text{für alle } n \in \mathbb{N}.$$

   Somit ist dann $P(Z = t_0) = \lim_{n\to\infty} P(Z \in (t_0 - \frac{1}{n},\ t_0 + \frac{1}{n}]) = 1$.

■

Wir werden Anwendungen von Satz 2.38 weiter unten diskutieren.

Definition 2.40
Seien X_i, $i \in I$, auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definierte $(E_i, \mathcal{E}_i)$-wertige Zufallsgrössen. (Die messbaren Räume $(E_i, \mathcal{E}_i)$ können verschieden sein.) Die X_i heissen ***unabhängig****, wenn die Teil-σ-Algebren $X_i^{-1}(\mathcal{E}_i)$ unabhängig sind. (Die X_i müssen auf demselben Wahrscheinlichkeitsraum definiert sein, damit die Aussage einen Sinn hat.)*

Notation: Sind zwei Zufallsgrössen unabhängig, so schreiben wir $X \perp Y$.

Lemma 2.41
X_i, $i \in I$ seien auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definierte Zufallsgrössen mit Werten in $(E_i, \mathcal{E}_i)$.

a) Sind $\mathcal{D}_i \subset \mathcal{F}$ für $i \in I$ unabhängig und sind die X_i $\mathcal{D}_i$-$\mathcal{E}_i$-messbar, so sind diese unabhängig.

b) Sind die X_i unabhängig und sind $\varphi_i : E_i \to E_i'$ $\mathcal{E}_i$-$\mathcal{E}_i'$-messbare Abbildungen ($\mathcal{E}_i'$ σ-Algebren auf E_i), so sind die $\varphi_i \circ X_i$ ebenfalls unabhängig.

c) Die X_i sind genau dann unabhängig, wenn für $n \in \mathbb{N}$ und $i_1, \ldots, i_n \in I$ sowie $A_1 \in \mathcal{E}_{i_1}, \ldots, A_n \in \mathcal{E}_{i_n}$ die Gleichung

$$P(X_{i_1} \in A_1, \ldots, X_{i_n} \in A_n) = \prod_{j=1}^{n} P(X_{i_j} \in A_j)$$

gilt.

Beweis. Die Aussagen folgen alle unmittelbar aus den Definitionen. ■

Im Spezialfall reeller Zufallsgrössen X_i ergibt sich das folgende Kriterium für die Unabhängigkeit:

Lemma 2.42
Zufallsgrössen X_i, $i \in I$, sind genau dann unabhängig, wenn für alle $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I$ und $t_1, \ldots, t_n \in \mathbb{R}$

$$P(X_{i_1} \le t_1, \ldots, X_{i_n} \le t_n) = \prod_{j=1}^{n} P(X_{i_j} \le t_j)$$

gilt.

Beweis. $\{\, X_i^{-1}((-\infty, t]) : t \in \mathbb{R} \,\} \cup \{\Omega\}$ ist ein durchschnittstabiles Erzeugendensystem von $X_i^{-1}(\mathcal{B})$. Die Behauptung folgt aus Satz 2.35. ■

Eine andere Folgerung aus der Unabhängigkeit ist

Satz 2.43
Es seien X, Y zwei unabhängige reelle Zufallsgrössen.

a) Falls X und Y nichtnegativ sind, so gilt $E(XY) = E(X)E(Y)$.

b) Sind $X, Y \in \mathcal{L}_1$, so gilt $XY \in \mathcal{L}_1$ und $E(XY) = E(X)E(Y)$.

Beweis. a): Es seien $\mathcal{F}_1 := X^{-1}(\mathcal{B})$ und $\mathcal{F}_2 := Y^{-1}(\mathcal{B})$. Dann gilt $\mathcal{F}_1 \perp \mathcal{F}_2$. Für $A \in \mathcal{F}_1$ erfüllt die Menge der nichtnegativen $\mathcal{F}_2$-messbaren Zufallsgrössen Y' mit $E(1_A Y') = P(A)E(Y')$ die Eigenschaften 1.–3. von Satz 1.37. Demzufolge gilt diese Gleichung für alle nichtnegativen $\mathcal{F}_2$-messbaren Y', also insbesondere für Y selbst. Die Menge der nichtnegativen $\mathcal{F}_1$-messbaren Zufallsgrössen X' mit $E(X'Y) = E(X')E(Y)$ erfüllt ebenfalls die Bedingungen von Satz 1.37. Das gleiche Argument wie oben belegt, dass diese Gleichung für $X' = X$ gilt.

b): Aus $X \perp Y$ folgt $|X| \perp |Y|$. Somit folgt aus Teil a), dass gilt: $E(|XY|) = E(|X|)E(|Y|) < \infty$, wenn $X, Y \in \mathcal{L}_1$ sind, das heisst $XY \in \mathcal{L}_1$. Die Gleichung $E(XY) = E(X)E(Y)$ folgt, indem X und Y in Positiv- und Negativteil zerlegt werden. ■

Besonders nützlich für die Untersuchung von unabhängigen Zufallsgrössen sind charakteristischen Funktionen.

Satz 2.44
Es seien X, Y zwei unabhängige Zufallsgrössen mit charakteristischen Funktionen $\mathcal{X}_X$ beziehungsweise $\mathcal{X}_Y$. Dann ist $\mathcal{X}_X \cdot \mathcal{X}_Y$ die charakteristische Funktion von $X + Y$.

Beweis. Für alle $t \in \mathbb{R}$ gilt

$$E(\mathrm{e}^{it(X+Y)}) = E(\mathrm{e}^{itX}\mathrm{e}^{itY}) = E(\mathrm{e}^{itX})E(\mathrm{e}^{itY}),$$

da $\mathrm{e}^{itX} \perp \mathrm{e}^{itY}$ gilt. Der Beweis ist insofern unvollständig, als Satz 2.43 nur für reellwertige Zufallsgrössen bewiesen wurde. Eine Zerlegung in Real- und Imaginärteil liefert jedoch sofort die entsprechende Aussage für komplexwertige Zufallsgrössen. ■

Unabhängigkeit von Zufallsgrössen ist äquivalent damit, dass ihre Verteilung die Produktwahrscheinlichkeit mit den Einzelverteilungen ist. Der Einfachheit halber formulieren wir das im Fall, dass alle Zufallsgrössen denselben Wertebereich haben.

Proposition 2.45
Seien X_i, $i \in I$, $(S, \mathcal{S})$-wertige Zufallsgrössen. Seien μ_i die Verteilungen der X_i : $\mu_i := PX_i^{-1}$. Die Zufallsgrössen sind genau dann unabhängig, wenn für jedes n und $i_1, \ldots, i_n \in I$ (alle verschieden) die Gleichung

$$P\left(X_{i_1}, \ldots, X_{i_n}\right)^{-1} = \bigotimes\nolimits_{k=1}^{n} \mu_{i_k}$$

gilt.

Beweis. Dies ist eine unmittelbare Folgerung aus Lemma 2.41 und der Definition des Produktmasses (siehe Definition 1.54 und Bemerkung 1.57). ■

Aus Satz 2.44 und Satz 2.21 folgt , dass die Verteilung von $X + Y$ für unabhängige Zufallsgrössen X, Y nur von den Verteilungen von X und Y abhängt. Man nennt diese Verteilung auch die **Faltung** der einzelnen Verteilungen.

Die Verteilungsfunktion von $X + Y$ kann auch wie folgt berechnet werden: Seien $\mu = PX^{-1}$, $\nu = PY^{-1}$. Dann ist nach der obigen Proposition die Verteilung von (X, Y) gleich $\mu \otimes \nu$ und es gilt:

$$P(X + Y \leq t) = \int_{\mathbb{R}^2} f_t(x, y)(\mu \otimes \nu)(d(x, y))$$

mit $f_t(x, y) = 1_{\{x+y \leq t\}} = 1_{(-\infty, t-y]}(x)$. Nach dem Satz von Fubini ist die rechte Seite gleich

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} 1_{(-\infty, t-y]}(x)\mu(dx) \right) \nu(dy) = \int_{\mathbb{R}} F_X(t - y)\nu(dy),$$

wobei F_X die Verteilungsfunktion von X ist. Hat die Verteilung von X die Dichte f und diejenige von Y die Dichte g bezüglich des Lebesgue Masses, so ergibt sich

$$\begin{aligned} P(X+Y \leq t) &= \int_{\mathbb{R}} \left(\int_{(-\infty,t-y]} f(x) \lambda\,(dx)\, g(y) \right) \lambda\,(dy) = \int_{\mathbb{R}} \left(\int_{(-\infty,t]} f(x-y) \lambda\,(dx)\, g(y) \right) \lambda\,(dy) \\ &= \int_{(-\infty,t]} \left(\int_{\mathbb{R}} f(x-y) g(y) \lambda\,(dy) \right) \lambda\,(dx)\,. \end{aligned}$$

Demzufolge hat dann auch die Verteilung von $X+Y$ eine Dichte bezüglich λ, nämlich die Abbildung

$$x \mapsto \int_{\mathbb{R}} f(x-y) g(y) \lambda\,(dy) = \int_{\mathbb{R}} f(y) g(x-y) \lambda\,(dy)\,.$$

Charakteristische Funktionen sind für die Berechnung jedoch oft einfacher.

Beispiel 2.46

a) *Cauchy-Verteilung:*
Behauptung: Sind X, Y unabhängig und Cauchy-verteilt zum Parameter $c > 0$, so ist für $\lambda \in (0,1)$ die Zufallsgrösse $\lambda X + (1-\lambda)Y$ auch Cauchy-verteilt zum Parameter $c > 0$.
Beweis: Für $\lambda \in (0,1)$ und $t \in \mathbb{R}$ gilt:

$$\begin{aligned} \chi_{\lambda X+(1-\lambda)Y}(t) &= E(\exp(it(\lambda X + (1-\lambda)Y))) = \chi_X(\lambda t) \cdot \chi_Y((1-\lambda)t) \\ &= \exp(-c|\lambda t|) \exp(-c|(1-\lambda)t|) = \mathrm{e}^{-c|t|}. \end{aligned}$$

b) *Normalverteilung:*

Behauptung: Ist X normalverteilt mit Mittelwert a und Varianz σ^2, Y normalverteilt mit Parametern a', σ'^2, und gilt $X \perp Y$, so ist $X+Y$ normalverteilt mit Parametern $a+a'$ und $\sigma^2 + \sigma'^2$.
Beweis: Für $t \in \mathbb{R}$ gilt:

$$\begin{aligned} \chi_{X+Y}(t) = \chi_X(t)\chi_Y(t) &= \exp\left[iat - \frac{\sigma^2}{2} t^2\right] \cdot \exp\left[ia't - \frac{\sigma'^2}{2} t^2\right] \\ &= \exp\left[i(a+a')t - \frac{1}{2}(\sigma^2 + \sigma'^2)t^2\right] \end{aligned}$$

2.6 Gesetz der grossen Zahlen

Das Kolmogoroffsche 0-1-Gesetz (Satz 2.38) soll nun auf unabhängige Zufallsgrössen angewandt werden. Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsgrössen, und $\mathcal{F}_n := X_n^{-1}(\mathcal{B})$. Es sei $\mathcal{T}_\infty$ die terminale σ-Algebra der $\mathcal{F}_n$ wie in Satz 2.38. Mit S_n sei $\sum_{j=1}^n X_j$ bezeichnet.

Lemma 2.47
Sei $\{a_n\}_{n\in\mathbb{N}}$ eine Folge positiver Zahlen mit $a_n \to \infty$. Dann sind $\overline{Y} := \limsup_{n\to\infty} \frac{S_n}{a_n}$ und $\underline{Y} := \liminf_{n\to\infty} \frac{S_n}{a_n}$ $\mathcal{T}_\infty$-messbare $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$-wertige Zufallsgrössen.

Beweis. Für jedes $m \in \mathbb{N}$ gilt

$$\overline{Y} = \limsup_{n\to\infty} \frac{S_n}{a_n} = \limsup_{n\to\infty} \frac{1}{a_n} \left(\sum\nolimits_{j=1}^{m} X_j + \sum\nolimits_{j=m+1}^{n} X_j\right) = \limsup_{n\to\infty} \frac{1}{a_n} \sum_{j=m+1}^{n} X_j,$$

was offenbar $\check{\mathcal{F}}_m$-messbar ist. Daher ist $\overline{Y}$ $\mathcal{T}_\infty$-messbar. Für $\underline{Y}$ geht der Beweis gleich. ■

Als Folgerung aus Lemma , dem Kolmogoroff 0-1-Gesetz und Lemma 2.47 ergibt sich:

Satz 2.48
Es sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsgrössen und $\{a_n\}_{n\in\mathbb{N}}$ eine positive Zahlenfolge mit $\lim_{n\to\infty} a_n = \infty$. Dann sind $\overline{Y} = \limsup_{n\to\infty} \frac{S_n}{a_n}$ und $\underline{Y} = \liminf_{n\to\infty} \frac{S_n}{a_n}$ fast sicher konstant in $[-\infty, \infty]$.

Beispiel 2.49

a) Die X_n seien unabhängig und Cauchy-verteilt zu $c = 1$ und sei $a_n = n$. Aus Beispiel 2.46a) folgt sofort, dass S_n/n auch Cauchy-verteilt ist zu $c = 1$. Somit ist für $a \in \mathbb{R}$:

$$\begin{aligned} 0 < \int_a^\infty \frac{1}{\pi} \frac{1}{1+x^2} dx &= P\left(\frac{S_n}{n} \geq a\right) \leq P\left(\sup_{k\geq n} \frac{S_k}{k} \geq a\right) \\ &\to P\left(\limsup_{n\to\infty} \frac{S_n}{n} \geq a\right) = P(\overline{Y} \geq a). \end{aligned}$$

Daraus folgt

$$P(\overline{Y} \geq a) = \lim_{n\to\infty} P\left(\sup_{k\geq n} \frac{S_k}{k} \geq a\right) > 0$$

Nach Satz 2.48 folgt $P(\overline{Y} \geq a) = 1$ für alle $a \in \mathbb{R}$, das heisst $P(\overline{Y} = \infty) = 1$. Analog zeigt man $P(\underline{Y} = -\infty) = 1$.

b) Die Zufallsgrössen X_n seien unabhängig und standard-normalverteilt. Nach Beispiel 2.46 b) ist auch $S_n/\sqrt{n}$ standard-normalverteilt. Wie oben folgt dann, dass $\limsup_{n\to\infty} S_n/\sqrt{n} = \infty$ fast sicher und $\liminf_{n\to\infty} S_n/\sqrt{n} = -\infty$ fast sicher gilt.

c) Das Beispiel b) kann mit Hilfe des Zentralen Grenzwertsatzes (siehe Stochastik und Abschnitt 2.7 unten) verallgemeinert werden. Seien X_n, $n \in \mathbb{N}$, unabhängige und identisch verteilte Zufallsgrössen in $\mathcal{L}_2$ mit $\operatorname{var}(X_n) > 0$. Dann gelten $\limsup_{n\to\infty} S_n/\sqrt{n} = \infty$ fast sicher und $\liminf_{n\to\infty} S_n/\sqrt{n} = -\infty$ fast sicher. Der Beweis sei dem Leser überlassen.

Aus diesem Beispiel ist ersichtlich, dass S_n/n im Cauchy-verteilten Fall nicht fast sicher konvergiert. Dies widerspricht nicht dem Gesetz der grossen Zahlen, weil Cauchy-verteilte Zufallsgrössen keinen Erwartungswert besitzen.

Satz 2.50 (Starkes Gesetz der grossen Zahlen)
Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsgrössen, die alle dieselbe Verteilung haben. Es gelte $X_i \in \mathcal{L}_1$ für alle $i \in \mathbb{N}$, und es sei $S_n = \sum_{i=1}^n X_i$. Dann gilt $\lim_{n\to\infty} S_n/n = EX_1$ fast sicher.

Beweis. Ein vollständiger Beweis ergibt sich als Spezialfall des Ergodensatzes. Unter der Zusatzbedingung $X_i \in \mathcal{L}_4$ kann ein einfacher Beweis wie folgt geführt werden: Sei $a = EX_n$ (unabhängig von n), $X_n' := X_n - a$. Es gilt $EX_n' = 0$, und $S_n/n \to a$ fast sicher gilt genau dann, wenn $\sum_{i=1}^n X_i'/n \to 0$ fast sicher gilt. Man kann also annehmen, dass $a = 0$ ist.

Sei $A_n = \left\{\left|\frac{S_n}{n}\right| \geq \frac{1}{n^{1/8}}\right\}$; $\overline{A} = \limsup_{n\to\infty} A_n$. Wir schätzen $P(A_n)$ mit der Markoff-Ungleichung:

$$P(A_n) \leq \frac{n^{1/2}}{n^4} E(S_n^4).$$

Nun ist

$$\begin{aligned} E(S_n^4) &= E\left(\left(\sum\nolimits_{i=1}^n X_i\right)^4\right) = E\left(\sum_{i_1,i_2,i_3,i_4=1}^n X_{i_1}X_{i_2}X_{i_3}X_{i_4}\right) \\ &= \sum_{i_1,i_2,i_3,i_4=1}^n E(X_{i_1}X_{i_2}X_{i_3}X_{i_4}) = \sum_{i=1}^n E(X_i^4) + 3\sum_{i\neq j} E(X_i^2 X_j^2). \end{aligned}$$

Um die letzte Gleichung einzusehen beachte man, dass die Terme der Summe mit einem Index, der verschieden von den anderen ist, wegen $EX = 0$ und der Unabhängigkeit alle verschwinden. Somit folgt $E(S_n^4) = nE(X_1^4) + n(n-1)(E(X_1^2))^2$. Daraus folgt folgt $\sum_{n=1}^\infty P(A_n) < \infty$, und aus dem 1. Borel-Cantelli-Lemma (Lemma 2.30) folgt $P(\overline{A}) = 0$. Für $\omega \notin \overline{A}$ konvergiert $S_n(\omega)/n$ offenbar gegen null. ■

Der Satz hat viele Verallgemeinerungen. Eine wichtige ist der Ergodensatz, den wir später diskutieren werden.

Ein Gesetz der grossen Zahlen für nicht notwendig identisch verteilte Zufallsgrössen ist der folgende Satz.

Satz 2.51
Seien X_n für $n \in \mathbb{N}$ unabhängige Zufallsgrössen in $\mathcal{L}_2$ mit $EX_n = 0$. Ist $\{a_n\}_n$ eine Folge positiver Zahlen mit $a_n \uparrow \infty$ und

$$\sum_{n=1}^\infty \frac{EX_n^2}{a_n^2} < \infty \tag{2.4}$$

so gilt $\lim_{n\to\infty} S_n/a_n = 0$ fast sicher.

Der Beweis basiert auf der folgenden Ungleichung, die auf Kolmogoroff zurückgeht.

Proposition 2.52 (Kolmogoroff)
Seien $X_1, \ldots, X_n$ unabhängige Zufallsgrössen in $\mathcal{L}_2$ mit $EX_i = 0$, $i = 1, \ldots, n$. und sei $S_k := \sum_{i=1}^k X_i$, $k = 1, \ldots, n$. Dann gilt

$$P\left(\max_{1\le k\le n} |S_k| \ge \lambda\right) \le \lambda^{-2} \sum_{k=1}^n \operatorname{var}(X_k)$$

Beweis. Wir betrachten die Ereignisse

$$\begin{aligned} A_k &: = \{|S_1| < \lambda, \ldots, |S_{k-1}| < \lambda, |S_k| \ge \lambda\}, \ 1 \le k \le n \\ A_{n+1} &: = \{|S_1| < \lambda, \ldots, |S_n| < \lambda\} \end{aligned}$$

Offensichtlich gilt

$$\left\{\max_{1\le k\le n} |S_k| \ge \lambda\right\} = \bigcup\nolimits_{k=1}^n A_k = A_{n+1}^c.$$

wobei die A_k paarweise disjunkt sind. Somit gilt

$$\begin{aligned} P\left(\max_{1\le k\le n} |S_k| \ge \lambda\right) &= \sum_{k=1}^n P(A_k) \\ &\le \lambda^{-2} \sum_{k=1}^n E\left(S_k^2; A_k\right) \\ &\le \lambda^{-2} \left[\sum_{k=1}^n E\left(S_k^2; A_k\right) + E\left(S_n^2; A_{n+1}\right)\right] \end{aligned}$$

$$\begin{aligned} E\left(S_k^2; A_k\right) &= \sum_{j=1}^k E\left(X_j^2; A_k\right) + 2\sum_{j=2}^k \sum_{i=1}^{j-1} E\left(X_i X_j; A_k\right) \\ \sum_{k=1}^n E\left(S_k^2; A_k\right) &= \sum_{j=1}^n \sum_{k=j}^n E\left(X_j^2; A_k\right) + 2\sum_{j=2}^n \sum_{i=1}^{j-1} \sum_{k=j}^n E\left(X_i X_j; A_k\right) \\ E\left(S_n^2; A_{n+1}\right) &= \sum_{j=1}^n E\left(X_j^2; A_{n+1}\right) + 2\sum_{j=2}^n \sum_{i=1}^{j-1} E\left(X_i X_j; A_{n+1}\right). \end{aligned}$$

Somit folgt

$$\begin{aligned} \sum_{k=1}^n E\left(S_k^2; A_k\right) + E\left(S_n^2; A_{n+1}\right) &= \sum_{j=1}^n E\left(X_j^2; \bigcup\nolimits_{k=j}^{n+1} A_k\right) + 2\sum_{j=2}^n \sum_{i=1}^{j-1} E\left(X_i X_j; \bigcup\nolimits_{k=j}^{n+1} A_k\right) \\ &\le \sum_{j=1}^n E\left(X_j^2\right) + 2\sum_{j=2}^n \sum_{i=1}^{j-1} E\left(X_i X_j; \bigcup\nolimits_{k=j}^{n+1} A_k\right). \end{aligned}$$

Man beachte nun, dass

$$\bigcup\nolimits_{k=j}^{n+1} A_k = \{|S_1| < \lambda, \ldots, |S_{j-1}| < \lambda\},$$

ist also in der von $X_1, \ldots, X_{j-1}$ erzeugten σ-Algebra. Somit ist für $i < j$

$$E\left(X_i X_j; \bigcup\nolimits_{k=j}^{n+1} A_k\right) = E\left(X_i; \bigcup\nolimits_{k=j}^{n+1} A_k\right) E(X_j) = 0.$$

Zusammen erhalten wir

$$P\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) \leq \lambda^{-2} \sum_{j=1}^{n} E\left(X_j^2\right).$$

■

Bemerkung 2.53
Der entscheidende Punkt ist, dass man $\max_{1\leq k\leq n} |S_k|$ praktisch gleich gut abschätzen kann, wie $|S_n|$. Die Tschebyscheff-Ungleichung ergibt ja

$$P(|S_n| \geq \lambda) \leq \lambda^{-2} \operatorname{var}(S_n) = \lambda^{-2} \sum_{j=1}^{n} E\left(X_j^2\right).$$

Eine sehr naive Abschätzung besteht darin, dass die Wahrscheinlichkeit der Vereinigung

$$\{\max\nolimits_{1\leq k\leq n} |S_k| \geq \lambda\} = \bigcup\nolimits_{k=1}^{n} \{|S_k| \geq \lambda\}$$

durch die Summe abgeschätzt wird:

$$P(\max\nolimits_{1\leq k\leq n} |S_k| \geq \lambda) \leq \sum_{k=1}^{n} P(|S_k| \geq \lambda),$$

und dann weiter mit Tschebyscheff $P(|S_k| \geq \lambda) \leq \lambda^{-2} \sum_{j=1}^{k} E\left(X_j^2\right)$, was schlussendlich die folgende Abschätzung ergibt

$$\begin{aligned} P(\max\nolimits_{1\leq k\leq n} |S_k| \geq \lambda) &\leq \lambda^{-2} \sum_{k=1}^{n} \sum_{j=1}^{k} E\left(X_j^2\right) \\ &= \lambda^{-2} \sum_{j=1}^{n} (n-j+1) E\left(X_j^2\right). \end{aligned}$$

Die im Vergleich zur Kolmogoroff-Ungleichung zusätzlichen Faktoren $n-j+1$ machen diese Abschätzung aber für fast alle Zwecke unbrauchbar.

Bemerkung 2.54
Wir können in der obigen Proposition auch $n = \infty$ einsetzen: Ist $\{X_i\}_{i \geq 1}$ eine Folge von Zufallsgrössen mit denselben Bedingugngen wie oben, so gilt

$$P\left(\sup_{k \in \mathbb{N}} |S_k| \geq \lambda\right) \leq \lambda^{-2} \sum_{k=1}^{\infty} \operatorname{var}(X_k).$$

Um dies zu zeigen, wählen wir erst $\lambda' \in (0, \lambda)$. Dann gilt

$$\left\{\max_{1 \leq k \leq n} |S_k| > \lambda'\right\} \uparrow \left\{\sup_{1 \leq k} |S_k| > \lambda'\right\},$$

also

$$\begin{aligned} P\left(\sup_{1 \leq k} |S_k| > \lambda'\right) &= \lim_{n \to \infty} P\left(\max_{1 \leq k \leq n} |S_k| > \lambda'\right) \leq \lim_{n \to \infty} P\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda'\right) \\ &\leq \lambda'^{-2} \lim_{n \to \infty} \sum_{k=1}^{n} \operatorname{var}(X_k) = \lambda'^{-2} \sum_{k=1}^{\infty} \operatorname{var}(X_k). \end{aligned}$$

Mit $\lambda' \uparrow \lambda,\ \lambda' \in \mathbb{Q}$ folgt die Behauptung.

Wir benötigen ferner das folgende sogenannte „Kronecker Lemma“ über reelle Zahlenfolge.

Lemma 2.55 (Kronecker Lemma)
Sei $\{x_k\}_{k \geq 1}$ eine Folge reeller Zahlen, deren Reihe $\sum_k x_k$ konvergiert. Sei ferner $\{a_k\}_{k \geq 1}$ eine Folge postiver reeller Zahlen mit $a_k \leq a_{k+1}$, $\forall k$, und $a_k \uparrow \infty$. Dann gilt

$$\lim_{n \to \infty} \frac{1}{a_n} \sum_{k=1}^{n} a_k x_k = 0.$$

Beweis. Übungsaufgabe in Analysis. ■
Beweis von Satz 2.51.
Wir beweisen zunächst die folgende Aussage:
Ist $\{Z_n\}$ eine Folge von unabhängigen Zufallsgrössen in $\mathcal{L}_2$ mit $EZ_n = 0$, $\sum_n \operatorname{var}(Z_n) < \infty$, so existiert

$$\lim_{N \to \infty} \sum_{n=1}^{N} Z_n \tag{2.5}$$

f.s. in $\mathbb{R}$. Aus dieser Aussage folgt der Satz sofort mit $Z_n := X_n / a_n$ und dem Kronecker Lemma: Aus der Voraussetzung (2.4) folgt nämlich $\sum_n \operatorname{var}(Z_n) < \infty$, also ergibt sich, sofern (2.5) bewiesen ist, dass für fast alle $\omega \in \Omega$ die Reihe $\sum_{n=1}^{\infty} Z_n(\omega) = \sum_{n=1}^{\infty} X_n(\omega) / a_n$ konvergiert. Mit dem Kronecker Lemma folgt daraus

$$\lim_{n \to \infty} a_n^{-1} \sum_{k=1}^{n} a_k (X_k / a_k) = \lim_{n \to \infty} a_n^{-1} \sum_{k=1}^{n} X_k = 0,$$

fast sicher.

Wir beweisen nun (2.5). Wir setzen $S_n; = \sum_{i=1}^n Z_i$. Aus der Voraussetzung folgt die Existenz einer Folge $\{n_k\}_{k\geq 1}$ für die

$$\sum_{j=n_k}^{\infty} \operatorname{var}(Z_j) \leq k^{-4}$$

gilt. Wir definieren

$$\begin{aligned} A_k \quad : \quad &= \left\{ \sup_{n_k \leq i < j} |S_i - S_j| \geq k^{-1} \right\} \subset \left\{ \sup_{n_k < i} |S_i - S_{n_k}| \geq \frac{1}{2k} \right\} \\ &= \left\{ \sup_{n_k < i} \left| \sum\nolimits_{r=n_k+1}^{i} Z_r \right| \geq \frac{1}{2k} \right\}. \end{aligned}$$

Wir wenden nun die Kolomogoroff-Ungleichung auf die unabhängige Folge $Z_{n_k+1}, Z_{n_k+2}, \ldots$ an und erhalten

$$P(A_k) \leq 4k^2 \sum_{i=n_k+1}^{\infty} \operatorname{var}(Z_i) \leq \frac{4}{k^2}.$$

Somit gilt $\sum_k P(A_k) < \infty$, und nach dem Borel-Cantelli Lemma folgt $P(\limsup_k A_k) = 0$. Für $\omega \notin \limsup_k A_k$ ist die reelle Zahlenfolge $\{S_n(\omega)\}_{n\geq 1}$ eine Cauchy-Folge, die somit konvergiert. Damit ist (2.5) gezeigt. ■

Angewandt auf identisch verteilte Zufallsgrössen und $a_n = n$ ergibt sich daraus die Aussage von Satz 2.50, allerdings unter der Voraussetzung $X_n \in \mathcal{L}_2$. Man kann in Satz 2.51 im Fall identisch verteilter Zufallsgrössen auch $a_n = n^{\delta+\frac{1}{2}}$ oder $a_n = n^{1/2}(\log n)^{\delta+\frac{1}{2}}$ für $\delta > 0$ wählen, nicht jedoch $a_n = n^{1/2}(\log n)^{1/2}$ (da dann $\sum a_n^{-2}$ divergiert). Tatsächlich ist jedoch die Aussage von Satz 2.51 noch richtig für diese letzte Folge.

Man wird sich also fragen, ob das Gesetz der grossen Zahlen die Konvergenzgeschwindigkeit von S_n/n gegen EX_1 richtig beschreibt. Nehmen wir der Einfachheit halber an, EX_1 sei 0. Die Frage präzise gestellt lautet: Für welche Folgen $\{a_n\}_{n\in\mathbb{N}}$ reeller positiver Zahlen mit $a_n \uparrow \infty$ gilt

$$\lim_{n\to\infty} \frac{S_n}{a_n} = 0 \quad P\text{–f. s.?}$$

Nach Satz 2.48 existieren $\overline{c}, \underline{c} \in [-\infty, \infty]$ mit

$$P\left(\limsup_{n\to\infty} \frac{S_n}{a_n} = \overline{c}\right) = 1, \quad P\left(\liminf_{n\to\infty} \frac{S_n}{a_n} = \underline{c}\right) = 1.$$

Die Frage ist also, ob $\underline{c} = \overline{c} = 0$ gilt. ($\underline{c}$, $\overline{c}$ hängen natürlich von der Folge $\{a_n\}$ und der Verteilung der X_i ab.) Nehmen wir zum Beispiel $a_n = \sqrt{n}$. Nach Beispiel 2.49 c) ergibt sich $\overline{c} = \infty$, $\underline{c} = -\infty$, sofern $X_i \in \mathcal{L}_2$ ist. Anderseits besagt das Gesetz der grossen Zahlen, dass $\underline{c} = \overline{c} = 0$ für $a_n = n$, $X_i \in \mathcal{L}_1$ gilt. Eine natürliche Frage ist also: Gibt es Folgen $\{a_n\}_{n\in\mathbb{N}}$ mit $-\infty < \underline{c} < \overline{c} < \infty$? Die Antwort ist: Ja (zumindest für $X_i \in \mathcal{L}_2$):

Satz 2.56 (Gesetz vom iterierten Logarithmus)
Es seien X_n für $n \in \mathbb{N}$ unabhängige Zufallsgrössen in $\mathcal{L}_2$, alle mit derselben Verteilung, und es gelte $EX_n = 0$. Dann gelten

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = \sigma \text{ fast sicher und}$$
$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = -\sigma \text{ fast sicher,}$$

wobei $\sigma^2 = \mathrm{var}(X_n)$ ist.

Dies ist das berühmte Hartman-Wintner Gesetz vom iterierten Logarithmus. Wir wollen es hier nicht beweisen. Obwohl Mathematikern natürlich geläufig ist, dass $\log(\log n)$ für $n \to \infty$ gegen ∞ strebt, ist diese Funktion "für praktische Zwecke" beschränkt. Dazu das folgende numerische Gedankenexperiment: Jemand hat zum Zeitpunkt x mit einem Münzwurfexperiment begonnen, das er nun abbricht. Er habe in jeder Sekunde die Münze einmal geworfen. n bezeichne die Anzahl der Würfe.

x	$\log(\log n)$
vor einer Stunde	2.103
beim Tode Cäsars	3.214
Geburt des Universums	3.706

Zum Schluss noch eine partielle Umkehrung des 1. Borel-Cantelli-Lemmas.

Lemma 2.57 (2. Borel-Cantelli-Lemma)
Es seien A_n für $n \in \mathbb{N}$ unabhängige Ereignisse mit $\sum_{n=1}^{\infty} P(A_n) = \infty$. Dann gilt $P(\limsup_{n\to\infty} A_n) = 1$.

Beweis. Für $n \in \mathbb{N}$ gilt

$$\begin{aligned} P\left(\bigcup\nolimits_{m=n}^{\infty} A_m\right) &= 1 - P\left(\bigcap\nolimits_{m=n}^{\infty} A_m^c\right) = 1 - \lim_{k\to\infty} P\left(\bigcap\nolimits_{m=n}^{k} A_m^c\right) \\ &= 1 - \prod_{m=n}^{\infty} P(A_m^c) = 1 - \prod_{m=n}^{\infty} (1 - P(A_m)). \end{aligned}$$

Ein einfaches Lemma aus der Analysis besagt, dass wenn $\sum_{n=1}^{\infty} a_n$ divergiert für eine Folge von Zahlen $a_n \in [0,1]$, das unendliche Produkt $\prod_{n=1}^{\infty}(1-a_n) = 0$ ist. (Falls nicht bekannt: Übungsaufgabe.) Somit ist $P(\bigcup_{m=n}^{\infty} A_m) = 1$ für alle $n \in \mathbb{N}$, das heisst

$$P(\limsup_{n\to\infty} A_n) = P\left(\bigcap\nolimits_{n=1}^{\infty} \bigcup\nolimits_{m=n}^{\infty} A_m\right) = \lim_{n\to\infty} P\left(\bigcup\nolimits_{m=n}^{\infty} A_m\right) = 1.$$

∎

Bemerkung 2.58
Man kann auf die Voraussetzung der Unabhängigkeit im obigen Lemma nicht vollständig verzichten. Zum Beispiel gilt mit $A_n = A$ für alle $n \in \mathbb{N}$ und $0 < P(A) < 1$ natürlich $\sum_n P(A_n) = \infty$, aber $P(\limsup_{n\to\infty} A_n) = P(A) < 1$. Es gibt jedoch Verallgemeinerungen, die mit stark abgeschwächten Voraussetzungen auskommen.

Das Lemma spielt bei vielen feineren Untersuchungen über Folgen von unabhängigen Zufallsgrössen eine grosse Rolle (zum Beispiel beim Beweis vom Gesetz vom iterierten Logarithmus).

2.7 Verteilungskonvergenz, Zentraler Grenzwertsatz

In vielen Fällen interessiert nicht so sehr, ob eine Folge $\{X_n\}$ von Zufallsgrössen in einem der obigen Sinne gegen eine Zufallsgrösse X konvergiert, sondern nur, ob die Verteilung von X durch die Verteilungen der X_n für $n \to \infty$ approximiert wird.

Definition 2.59

a) *Seien $\mu_n,\ n \in \mathbb{N},\ \mu$ Wahrscheinlichkeitsmasse auf $(\mathbb{R}, \mathcal{B})$. Wir sagen, dass $\{\mu_n\}$ schwach gegen μ konvergiert, wenn für jede stetige und beschränkte Funktion $f : \mathbb{R} \to \mathbb{R}$ die Gleichung*
$$\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$$
gilt. Wir schreiben dafür auch
$$\text{w-}\lim_{n\to\infty} \mu_n = \mu.$$

b) *Seien $X_n,\ n \in \mathbb{N},\ X$ Zufallsgrössen. Wir sagen, dass die Folge $\{X_n\}$ schwach oder in Verteilung gegen das Wahrscheinlichkeitsmass μ konvergiert, wenn*
$$\text{w-}\lim_{n\to\infty} P X_n^{-1} = \mu$$
gilt. Dies bedeutet also, dass für jede stetige und beschränkte Funktion f, die Gleichung
$$\lim_{n\to\infty} Ef(X_n) = \lim_{n\to\infty} \int f d\left(P X_n^{-1}\right) = \int f d\mu$$
gilt.

Lemma 2.60
Konvergiert $\{X_n\}$ in Wahrscheinlichkeit gegen X, so gilt
$$\text{w-}\lim_{n\to\infty} P X_n^{-1} = P X^{-1},$$
d.h. es gilt
$$\lim_{n\to\infty} Ef(X_n) = Ef(X)$$
für jede beschränkte stetige Funktion $f : \mathbb{R} \to \mathbb{R}$.

Beweis. Sie f eine stetige und beschränkte Funktion. $C; = \sup_x |f(x)|$. Sei $\varepsilon > 0$. Es existiert $K > 0$ mit $P(|X| \geq K) \leq \varepsilon$. Aus der Konvergenz in Wahrscheinlichkeit folgt insbesondere

$$\lim_{n\to\infty} P(|X_n - X| \geq 1) = 0.$$

Es existiert also N mit

$$P(|X_n| \geq K + 1) \leq \varepsilon$$

für $n \geq N$.

f ist gleichmässig stetig auf dem kompakten Intervall $I := [-K-1, K+1]$. Es existiert somit ein $\delta > 0$ mit

$$x, y \in I,\ |x - y| \leq \delta \Longrightarrow |f(x) - f(y)| \leq \varepsilon. \tag{2.6}$$

Aus

$$\begin{aligned} E(f(X) - f(X_n)) &= E\left((f(X) - f(X_n))\, 1_{\{X\in I, X_n\in I, |X_n - X| \leq \delta\}}\right) \\ &\quad + E\left((f(X) - f(X_n))\, 1_{\{X\in I, X_n\in I, |X_n - X| \leq \delta\}^c}\right) \end{aligned}$$

folgt für $n \geq N$

$$\begin{aligned} |E(f(X) - f(X_n))| &\leq E\left(|f(X) - f(X_n)|\, 1_{\{X\in I, X_n\in I, |X_n - X| \leq \delta\}}\right) \\ &\quad + 2C\left(P(X \notin I) + P(X_n \notin I) + P(|X_n - X| > \delta)\right) \\ &\leq E\left(|f(X) - f(X_n)|\, 1_{\{X\in I, X_n\in I, |X_n - X| \leq \delta\}}\right) \\ &\quad + 4C\varepsilon + 2CP(|X_n - X| > \delta) \\ &\leq \varepsilon + 4C\varepsilon + 2CP(|X_n - X| > \delta). \end{aligned}$$

Die erste Ungleichung wegen Lemma 2.10 c) und $|f(X)| \leq C,\ |f(X_n)| \leq C$, die zweite Ungleichung nach (2.6). Wegen der Konvergenz in Wahrscheinlichkeit folgt

$$\limsup_{n\to\infty} |E(f(X) - f(X_n))| \leq \varepsilon + 4C\varepsilon.$$

Da $\varepsilon > 0$ beliebig ist, ergibt sich

$$\lim_{n\to\infty} Ef(X_n) = Ef(X).$$

■

Bemerkung 2.61
Es ist wichtig zu bemerken, dass die Umkehrung der obigen Aussage nicht gilt. Haben z.B. X_n und X alle die gleiche Verteilung, so gilt trivialerweise

$$\mathrm{w-}\lim_{n\to\infty} PX_n^{-1} = PX^{-1}$$

aber natürlich müssen die X_n nicht in Wahrscheinlichkeit gegen X konvergieren, z.B. wenn sie alle unabhängig sind.

Wir können Verteilungskonvergenz mit Hilfe der Verteilungsfunktionen charakterisieren. Seien also μ_n, μ Wahrscheinlichkeitsmasse auf $(\mathbb{R}, \mathcal{B})$ mit Verteilungsfunktionen F_n bzw. F.

Lemma 2.62
Für jede Verteilungsfunktion F ist die Menge der Unstetigkeitsstellen abzählbar. Insbesondere liegt die Menge der Stetigkeitspunkte von F dicht in $\mathbb{R}$.

Beweis. Da F monoton ist, existiert

$$F(t-) := \lim_{s \uparrow t} F(s)$$

und erfüllt $F(t-) \leq F(t)$. Die Menge der Unstetigkeitsstellen ist

$$\{t : F(t-) < F(t)\} = \bigcup\nolimits_{n \in \mathbb{N}} \{t : F(t) - F(t-) \geq 1/n\}.$$

Nun gilt offensichtlich $|\{t : F(t) - F(t-) \geq 1/n\}| \leq n$. Daraus folgt die Behauptung. ■

Proposition 2.63
$\mathrm{w-}\lim\limits_{n\to\infty}\mu_n = \mu$ *gilt genau dann, wenn für jeden Stetigkeitspunkt t von F die Gleichung*

$$\lim_{n\to\infty} F_n(t) = F(t) \tag{2.7}$$

gilt.

Beweis. (I): Wir nehmen zunächst an, dass $\mathrm{w-}\lim\limits_{n\to\infty}\mu_n = \mu$. Für $s < t$ betrachten wir die stetige und beschränkte Funktion $f_{s,t} : \mathbb{R} \to [0,1]$ die $f(x) = 1$ für $x \leq s$, $f(x) = 0$ für $x \geq t$ erfüllt und die linear im Intervall $[s,t]$ verläuft. Sei t eine Stetigkeitsstelle von F. Für jedes $\delta > 0$ gilt

$$F(t-\delta) \leq \int f_{t-\delta,t}(x)\,\mu(dx) \leq F(t) \leq \int f_{t,t+\delta}(x)\,\mu(dx) \leq F(t+\delta)$$

und die gleichen Ungleichungen gelten für F_n, μ_n. Demzufolge ist

$$\begin{aligned}
\limsup_{n\to\infty} F_n(t) &\leq \limsup_{n\to\infty} \int f_{t,t+\delta}(x)\,\mu_n(dx) = \int f_{t,t+\delta}(x)\,\mu(dx) \leq F(t+\delta), \\
\liminf_{n\to\infty} F_n(t) &\geq \liminf_{n\to\infty} \int f_{t-\delta,t}(x)\,\mu_n(dx) = \int f_{t-\delta,t}(x)\,\mu(dx) \leq F(t-\delta).
\end{aligned}$$

Wegen $\lim_{\delta\to 0} F(t+\delta) = \lim_{\delta\to 0} F(t-\delta) = F(t)$ folgt

$$\lim_{n\to\infty} F_n(t) = F(t).$$

(II): Wir nehmen an, dass (2.7) für jede Stetigkeitsstelle von F gilt. Zunächst bemerken wir, dass wenn $s < t$ Stetigkeitsstellen von F sind, die Gleichung

$$\begin{aligned}
\lim_{n\to\infty} \mu_n((s,t]) &= \lim_{n\to\infty} (F_n(t) - F_n(s)) \\
&= F(s) - F(t) = \mu((s,t])
\end{aligned}$$

gilt.

Sei $f : \mathbb{R} \to \mathbb{R}$ eine stetige und beschränkte Funktion. Wir wählen $a > 0,\ b \in \mathbb{R}$ so dass $g := af + b$ nur Werte in $[0,1]$ annimmt. Sofern wir

$$\lim_{n\to\infty} \int g\ d\mu_n = \int g\ d\mu$$

zeigen, so folgt die entsprechende Gleichung auch für f. Wir können daher o.B.d.A. annehmen, dass f nur Werte in $[0,1]$ annimmt.

Sei nun $\varepsilon > 0$ beliebig. Wir können Stetigkeitsstellen $A < 0 < B$ von F so wählen, dass $F(A) \le \varepsilon,\ 1 - F(B) \le \varepsilon$ gelten. f ist gleichmässig stetig auf dem Intervall $[A, B]$. Zu ε können wir daher ein $\delta > 0$ so finden, dass in jedem Teilintervall I von $[A, B]$ der Länge $\le \delta$ die Ungleichung

$$\sup_{s,t\in I} |f(s) - f(t)| \le \varepsilon$$

gilt. Wir wählen $K \in \mathbb{N}$ und $A = t_0 < t_1 < t_2 < \ldots < t_{K-1} < t_K = B$ mit $t_i - t_{i-1} \le \delta$, t_i Stetigkeitsstellen von F. Eine derartige Wahl ist wegen Lemma 2.62 möglich. Wir setzen $I_0 := (-\infty, t_0],\ I_k := (t_{k-1}, t_k]$ für $1 \le k \le K,\ I_{K+1} := (t_K, \infty)$. Da die t_k Stetigkeitsstellen von F sind folgt aus (2.7)

$$\lim_{n\to\infty} \mu_n(I_k) = \mu(I_k),\ 0 \le k \le K + 1. \tag{2.8}$$

Ferner gilt für $\nu = \mu_n$ oder $\nu = \mu$

$$\int f\ d\nu = \sum_{k=0}^{K+1} \int_{I_k} f\ d\nu$$

und somit

$$\begin{aligned}
\int f\ d\nu &\le \sum_{k=1}^{K} \sup_{t\in I_k} f(t)\,\nu(I_k) + \nu(I_0) + \nu(I_{K+1}) \\
\int f\ d\nu &\ge \sum_{k=1}^{K} \inf_{t\in I_k} f(t)\,\nu(I_k).
\end{aligned}$$

Anwedung dieser Ungleichungen sowohl auf $\nu = \mu_n$, wie $\nu = \mu$ ergibt zusammen mit (2.8)

$$\begin{aligned}
\limsup_{n\to\infty} \int f\ d\mu_n &\le \lim_{n\to\infty} \left(\sum_{k=1}^{K} \sup_{t\in I_k} f(t)\,\mu_n(I_k) + \mu_n(I_0) + \mu_n(I_{K+1}) \right) \\
&= \sum_{k=1}^{K} \mu(I_k) \sup_{t\in I_k} f(t) + \mu(I_0) + \mu(I_{K+1}) \\
&\le \sum_{k=1}^{K} \mu(I_k) \left(\inf_{t\in I_k} f(t) + \varepsilon \right) + 2\varepsilon \\
&\le \int f\ d\mu + 3\varepsilon.
\end{aligned}$$

$$
\begin{aligned}
\liminf_{n\to\infty} \int f \, d\mu_n &\geq \liminf_{n\to\infty} \sum_{k=1}^{K} \inf_{t\in I_k} f(t)\, \mu_n(I_k) \\
&= \sum_{k=1}^{K} \inf_{t\in I_k} f(t)\, \mu(I_k) \geq \sum_{k=1}^{K} \sup_{t\in I_k} f(t)\, \mu(I_k) + \mu(I_0) + \mu(I_{K+1}) - 3\varepsilon \\
&\geq \int f \, d\mu - 3\varepsilon.
\end{aligned}
$$

Da ε beliebig ist, folgt

$$
\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu.
$$

■

Eine sehr grundlegender Askpekt der Verteilungskonvergenz ist die Eigenschaft, dass Folgen von Wahrscheinlichkeitsmassen unter sehr schwachen Bedingungen konvergente Teilfolgen besitzen.

Definition 2.64
Sei $\{\mu_n\}_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmassen auf $(\mathbb{R}, \mathcal{B})$. Diese heisst ***straff*** *(engl. tight), wenn zu jedem $\varepsilon > 0$ ein $K \in \mathbb{N}$ existiert mit $\mu_n([-K, K]^c) \leq \varepsilon$ für alle n.*

Satz 2.65
Sei $\{\mu_n\}_{n\in\mathbb{N}}$ eine straffe Folge von Wahrscheinlichkeitsmassen auf $(\mathbb{R}, \mathcal{B})$. Dann existiert eine Teilfolge $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmass μ mit

$$
\mathrm{w-}\lim_{k\to\infty} \mu_{n_k} = \mu.
$$

Beweis. Wir argumentieren mit Verteilungsfunktionen. Sei als F_n die Verteilungsfunktion von μ_n. Wir wählen eine Abzählung der rationnalen Zahlen $\{t_i\}_{i\in\mathbb{N}} = \mathbb{Q}$. Da $\{F_n(t_1)\}$ eine Zahlenfolge im Intervall $[0,1]$ ist existiert eine Teilfolge $\{n_k^1\}_{k\in\mathbb{N}} \subset \mathbb{N}$ sodass $\hat{F}(t_1) = \lim_{k\to\infty} F_{n_k^1}(t_1)$ existiert. Weiter existiert eine Teilfolge $\{n_k^2\}$ der Folge $\{n_k^1\}$ mit der Eigenschaft, dass $\hat{F}(t_2) = \lim_{k\to\infty} F_{n_k^1}(t_2)$ existiert. Da diese zweite Folge eine Teilfolge der ersten ist, gilt natürlich auch $\hat{F}(t_1) = \lim_{k\to\infty} F_{n_k^2}(t_1)$. In dieser Weise konstruieren wir eine Folge von Folgen $\{n_k^m\}_{k,m\in\mathbb{N}}$ mit den folgenden Eigenschaften:

1. Für jedes m ist $\{n_k^{m+1}\}_{k\in\mathbb{N}}$ eine Teilfolge von $\{n_k^m\}_{k\in\mathbb{N}}$.
2. $$\hat{F}(t_m) := \lim_{k\to\infty} F_{n_k^m}(t_m)$$
   existiert für alle m.

Wegen der Monotonie der F_n ist natürlich $\hat{F}$, das im Moment nur auf $\{t_i\}_{i\in\mathbb{N}} = \mathbb{Q}$ definiert ist, eine monotone Funktion, d.h. für $t_i < t_j$ gilt $\hat{F}(t_i) \leq \hat{F}(t_j)$. Wir benötigen eine weitere Eigenschaft von $\hat{F}$, die aus der Straffheit folgt. Zu jedem $\varepsilon > 0$ existiert nach

Voraussetzung ein $K(\varepsilon) \in \mathbb{N}$ mit $\mu_n([-K, K]^c) \leq \varepsilon,\ \forall n$. Dies impliziert $F_n(t) \geq 1 - \varepsilon$, für $t \geq K(\varepsilon)$ und alle n, sowie $F_n(t) \leq \varepsilon$ für $t \leq -K(\varepsilon)$, $\forall n$. Dies impliziert $\hat{F}(t) \geq 1-\varepsilon$, für $t \geq K(\varepsilon)$, $\hat{F}(t) \leq \varepsilon$ für $t \leq -K(\varepsilon)$.

Wir wählen nun die „Diagonalfolge" $\{n_k^k\}_{k\in\mathbb{N}}$. Offensichtlich gilt

$$\hat{F}(t_m) = \lim_{k\to\infty} F_{n_k^k}(t_m)$$

für alle m.

$\hat{F}$ ist noch nicht ganz, was wir brauchen, denn es ist nur auf $\mathbb{Q}$ definiert und ausserdem braucht es nicht rechtsseitig stetig zu sein. Wir erweitern und modifizieren es, damit es die gewünschte Eigenschaft hat. Für $t \in \mathbb{R}$ setzen wir

$$F(t) := \lim_{q\downarrow t,\ q\in\mathbb{Q},\ q>t} \hat{F}(q).$$

Der Limes existiert wegen der Monotonie von $\hat{F}$. Beachte, dass selbst für $t \in \mathbb{Q}$ nicht notwendigerweise $F(t) = \hat{F}(t)$ gilt. F ist aber per Konstruktion rechtsseitig stetig und natürlich gilt $F(t) \geq 1 - \varepsilon$ für $t \geq K(\varepsilon)$ und $F(t) \leq \varepsilon$ zumindest für $t < -K(\varepsilon)$. F ist somit eine Verteilungsfunktion. Nach Satz 1.25 gibt es genau ein Wahrscheinlichkeitsmass μ, dessen Verteilungsfunktion F ist.

Wir zeigen nun, dass

$$\lim_{k\to\infty} F_{n_k^m}(t) = F(t)$$

für jeden Stetigkeitspunkt von F gilt. Damit ist nach Proposition 2.63 der Satz bewiesen.

Sei also $t \in \mathbb{R}$ ein Stetigkeitspunkt von F. Zu $\varepsilon > 0$ existiert $\delta > 0$ mit

$$\begin{aligned} F(t) - \varepsilon &\leq F(t-\delta) \leq F(s_1) \leq F(t) \\ &\leq F(s_2) \leq F(t+\delta) \leq F(t) + \varepsilon \end{aligned}$$

für $t - \delta < s_1 < t < s_2 < t + \delta$. Nach Konstruktion von F folgt dann auch

$$\begin{aligned} F(t) - \varepsilon &\leq F(t-\delta) \leq \hat{F}(s_1) \leq F(t) \\ &\leq \hat{F}(s_2) \leq F(t+\delta) \leq F(t) + \varepsilon \end{aligned}$$

sofern zusäztlich $s_1, s_2 \in \mathbb{Q}$ gilt. Somit erhalten wir

$$\begin{aligned} \limsup_{k\to\infty} F_{n_k^k}(t) &\leq \limsup_{k\to\infty} F_{n_k^k}(s_2) = \hat{F}(s_2) \leq F(t) + \varepsilon, \\ \liminf_{k\to\infty} F_{n_k^m}(t) &\geq \liminf_{k\to\infty} F_{n_k^k}(s_1) = \hat{F}(s_1) \geq F(t) - \varepsilon. \end{aligned}$$

Da $\varepsilon > 0$ beliebig ist, folgt die Behauptung. ■

Da sin und cos stetige und beschränkte Funktionen sind folgt sofort

$$\text{w-}\lim_{n\to\infty} \mu_n = \mu \Longrightarrow \lim_{n\to\infty} \hat{\mu}_n(t) = \hat{\mu}(t),\ \forall t \in \mathbb{R}, \tag{2.9}$$

wobei $\hat{\mu}(t)$ die charakteristische Funktion von μ ist. Von dieser Aussage gilt auch die Umkehrung:

Satz 2.66
μ_n, μ seien Wahrscheinlichkeitsmasse auf $(\mathbb{R}, \mathcal{B})$. Dann gilt $\underset{n\to\infty}{\text{w}-\lim}\mu_n = \mu$ genau dann, wenn $\lim_{n\to\infty} \hat{\mu}_n(t) = \hat{\mu}(t)$ für alle $t \in \mathbb{R}$ gilt.

Der Beweis der Straffheit aus der Konvergenz der charakteristischen Funktionen ist nicht ganz einfach. Da sich diese in der von uns ins Auge gefassten Anwendung quasi gratis ergibt, begnügen wir uns mit dem folgenden billigeren Teilresultat:

Lemma 2.67
μ_n, μ seien Wahrscheinlichkeitsmasse auf $(\mathbb{R}, \mathcal{B})$, $n \in \mathbb{N}$. Ist $\{\mu_n\}$ straff und gilt $\lim_{n\to\infty} \hat{\mu}_n(t) = \hat{\mu}(t)$, $\forall t$, so folgt $\underset{n\to\infty}{\text{w}-\lim}\mu_n = \mu$.

Beweis. Wir müssen zeigen, dass für jede stetige und beschränkte Funktion $f : \mathbb{R} \to \mathbb{R}$ die Gleichung $\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$ gilt. $\int f d\mu_n$ ist eine beschränkte Zahlenfolge. Falls die Behauptung nicht gilt, so existiert eine Teilfolge $\{n_k\}$ sodass $\alpha := \lim_{k\to\infty} \int f d\mu_{n_k}$ existiert, aber $\alpha \neq \int f d\mu$ ist. Aus der Definition der Straffheit folgt sofort, dass auch $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ straff ist. Nach Satz 2.65 folgt die Existenz einer konvergenten Teilfolge $\left\{\mu_{n_{k_l}}\right\}_{l\in\mathbb{N}}$, d.h.

$$\nu := \underset{l\to\infty}{\text{w}-\lim}\mu_{n_{k_l}}$$

existiert. Nach (2.9) gilt $\hat{\nu}(t) = \lim_{l\to\infty} \hat{\mu}_{n_{k_l}}(t)$ $\forall t$. Daraus folgt jedoch $\hat{\nu}(t) = \hat{\mu}(t)$, $\forall t$. Nach Satz 2.21 folgt $\nu = \mu$ im Widerspruch zu $\alpha \neq \int f d\mu$. ∎

Wir kommen nun zum Zentralen Grenzwertsatz, von dem wir eine Version schon in der Stochastik bewiesen hatten.

Ausgangspunkt ist eine Folge $\{X_n\}$ von Zufallsgrössen. Wir nehmen an, dass die X_n unabhängig und $\in \mathcal{L}_2$ sind. Wir wollen jedoch nicht annehmen, dass alle die gleiche Verteilung haben. Wir setzen

$$\alpha_n := EX_n,\ \sigma_n^2 := \operatorname{var} X_n,$$

$$S_n := \sum_{i=1}^n X_i.$$

Dann gelten

$$\begin{aligned} ES_n &= \sum_{i=1}^n \alpha_i, \\ \operatorname{var} S_n &= s_n^2 := \sum_{i=1}^n \sigma_n^2. \end{aligned}$$

Im Spezialfall identisch verteilter X_i sind die α_i, σ_i^2 natürlich alle gleich und man erhält $ES_n = n\alpha$, $s_n^2 = n\sigma^2$. Wir zeigen nun, dass unter einer ziemlich allgemeinen Bedingung, die Folge

$$\hat{S}_n := \frac{S_n - ES_n}{s_n},\ s_n := \sqrt{s_n^2},$$

asymptotisch normalverteilt ist. Die Bedingung, die wir benötigen, besagt einfach, dass keine der einzelnen Variablen einen substantiellen Beitrag zur Gesamtsumme gibt. Die genaue Formulierung stammt von Lindeberg und heisst deshalb die *Lindeberg-Bedingung.*

Condition 2.68 (Lindeberg-Bedingung)
$\{X_n\}$ erfüllt die Lindeberg-Bedingung, falls die X_n eine endliche Varianz besitzen, und falls für jedes $\delta > 0$

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} E\left((X_i - \alpha_i)^2 ; |X_i - \alpha_i| \geq \delta s_n\right) = 0.$$

Wir können die α_k direkt von den X_k substrahieren und können deshalb ohne Einschränkung der Allgemeinheit annehmen, dass $\alpha_k = 0$ gilt. Die Lindeberg-Bedingung ist klarerweise im Falle indentisch verteilter Zufallsgrössen mit $\sigma^2 > 0$ erfüllt, denn in diesem Fall ist $s_n^2 = n\sigma^2$, und somit

$$\begin{aligned} \frac{1}{s_n^2} \sum_{i=1}^{n} E\left(X_i^2; |X_i| \geq \delta s_n\right) &= \frac{1}{n\sigma^2} nE\left(X_1^2; |X_1| \geq \delta\sigma\sqrt{n}\right) \\ &= \frac{E\left(X_1^2; |X_1| \geq \delta\sigma\sqrt{n}\right)}{\sigma^2} \end{aligned}$$

und der Zähler geht nach dem Satz von Lebesgue gegen 0 für $n \to 0$.

Aus der Lindeberg-Bedingung folgt wegen

$$\begin{aligned} \frac{1}{s_n^2} \sum_{i=1}^{n} E\left(X_i^2; |X_i| \geq \delta s_n\right) &\geq \frac{1}{s_n^2} \sup_{i\leq n} E\left(X_i^2; |X_i| \geq \delta s_n\right) \\ &= \frac{1}{s_n^2} \sup_{i\leq n} \left[\sigma_i^2 - E\left(X_i^2; |X_i| < \delta s_n\right)\right] \\ &\geq \frac{1}{s_n^2} \sup_{i\leq n} \sigma_i^2 - \delta \end{aligned}$$

die Abschätzung

$$\limsup_{n\to\infty} \frac{1}{s_n^2} \sup_{k\leq n} \sigma_k^2 \leq \delta,$$

für jedes $\delta > 0$, also

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sup_{k\leq n} \sigma_k^2 = 0. \tag{2.10}$$

Satz 2.69
Sei $\{X_n\}$ eine Folge unabhängiger Zufallsgrössen, welche die Lindeberg-Bedingung erfüllt. Dann konvergiert die Folge der Verteilungen von $\hat{S}_n$ schwach gegen die Standardnormalverteilung.

Beweis. Wie schon oben erwähnt, können wir annehmen, dass $\alpha_i = 0$ gilt. Die Straffheit der Folge $\mu_n := P\hat{S}_n^{-1}$ ergibt sich unmittelbar aus der Tschebyscheff-Ungleichung: Für $K > 0$ gilt

$$P\left(\left|\hat{S}_n\right| \geq K\right) \leq \frac{\operatorname{var}\hat{S}_n}{K^2} = \frac{1}{K^2}.$$

Wir zeigen nun, dass für jedes $t \in \mathbb{R}$

$$\lim_{n\to\infty} \hat{\mu}_n(t) = \mathrm{e}^{-t^2/2}$$

gilt. Die rechte Seite ist die charakteristische Funktion der Standardnormalverteilung. Nach dem Lemma 2.67 folgt die Behauptung des Satzes.

$$\begin{aligned}
\hat{\mu}_n(t) &= E\left(\exp\left[it\hat{S}_n\right]\right) = E\left(\exp\left[\frac{it}{s_n}\sum\nolimits_{k=1}^n X_k\right]\right) \\
&= \prod\nolimits_{k=1}^n E\left(\exp\left[\frac{itX_k}{s_n}\right]\right).
\end{aligned}$$

Da $s_n \to \infty$ gilt, wird man versuchen, die Faktoren nach Taylor zu entwickeln:

$$\begin{aligned}
E\left(\exp\left[\frac{it}{s_n}X_k\right]\right) &\approx 1 + \frac{it}{s_n}EX_k + \frac{1}{2}\left(\frac{it}{s_n}\right)^2 E\left(X_k\right)^2 \\
&= 1 - \frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2
\end{aligned}$$

und somit

$$\begin{aligned}
\hat{\mu}_n(t) &\approx \prod\nolimits_{k=1}^n \left(1 - \frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2\right) \approx \exp\left[\sum\nolimits_{k=1}^n \log\left(1 - \frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2\right)\right] \\
&\approx \exp\left[-\frac{1}{2}\sum\nolimits_{k=1}^n \frac{t^2}{s_n^2}\sigma_k^2\right] = \mathrm{e}^{-t^2/2}.
\end{aligned}$$

Die Lindeberg-Bedingung wird benötigt, um die Fehlerterme in diesen Approximationen genau abzuschätzen.

Wir verwenden die Taylor-Formel mit Restterm

$$\mathrm{e}^{itx} = 1 + itx - \frac{t^2x^2}{2} - t^2\int_0^x \left(\mathrm{e}^{itu} - 1\right)(x-u)\,du,$$

also

$$E\left(\exp\left[\frac{it}{s_n}X_k\right]\right) = 1 - \frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2 - t^2E\theta\left(t, \frac{X_k}{s_n}\right)$$

mit

$$\begin{aligned}
\theta(t,x) &= \int_0^x \left(\mathrm{e}^{itu} - 1\right)(x-u)\,du = x^2\int_0^1 \left(\mathrm{e}^{itux} - 1\right)(1-u)\,du \\
|\theta(t,x)| &\leq \frac{x^2}{2}\sup_{0\leq v\leq 1}\left|\mathrm{e}^{itvx} - 1\right|.
\end{aligned}$$

Man beachte, dass für $\varphi \in \mathbb{R}$

$$\left|\mathrm{e}^{i\varphi} - 1\right| \leq |\varphi| \tag{2.11}$$

gilt.

$$\begin{aligned}
\left|E\theta\left(t, \frac{X_k}{s_n}\right)\right| &\leq E\left|\theta\left(t, \frac{X_k}{s_n}\right)\right| \\
&\leq \frac{1}{2s_n^2}E\left(X_k^2 \sup_{0\leq v\leq 1}\left|\mathrm{e}^{itvX_k/s_n} - 1\right|\right) \\
&= \frac{1}{2s_n^2}E\left(X_k^2 \sup_{0\leq v\leq 1}\left|\mathrm{e}^{itvX_k/s_n} - 1\right|; |X_k| < \delta s_n\right) \\
&\quad + \frac{1}{2s_n^2}E\left(X_k^2 \sup_{0\leq v\leq 1}\left|\mathrm{e}^{itvX_k/s_n} - 1\right|; |X_k| \geq \delta s_n\right) \\
&\leq \frac{\delta\,|t|\,EX_k^2}{2s_n^2} + \frac{1}{s_n^2}E\left(X_k^2; |X_k| \geq \delta s_n\right),
\end{aligned}$$

wobei wir die letzten Abschätzung für den ersten Summanden aus der Ungleichung (2.11) folgt. Nun verwenden wir die folgende Ungleichung für zwei Folgen $a_1, \ldots, a_n, b_1, \ldots, b_n$ von komplexen Zahlen vom Betrag ≤ 1, die dem Leser als Übungsaufgabe überlassen sei.

$$\left|\prod\nolimits_{k=1}^n a_k - \prod\nolimits_{k=1}^n b_k\right| \leq \sum_{k=1}^n |a_k - b_k|\,. \tag{2.12}$$

Demzufolge gilt

$$\begin{aligned}
&\left|E\left(\exp\left[it\hat{S}_n\right]\right) - \prod\nolimits_{k=1}^n \left(1 - \frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2\right)\right| \\
\leq\ & \sum_{k=1}^n \left(\frac{\delta\,|t|\,EX_k^2}{2s_n^2} + \frac{1}{s_n^2}E\left(X_k^2; |X_k| \geq \delta s_n\right)\right) \qquad (2.13) \\
=\ & \frac{\delta\,|t|}{2} + \frac{1}{s_n^2}\sum_{k=1}^n E\left(X_k^2; |X_k| \geq \delta s_n\right).
\end{aligned}$$

Man beachte nun

$$\mathrm{e}^{-t^2/2} = \exp\left[-\sum\nolimits_{k=1}^n \frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2\right] = \prod\nolimits_{k=1}^n \exp\left[-\frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2\right].$$

Wir verwenden, dass für $x > 0$

$$1 - x \leq \mathrm{e}^{-x} \leq 1 - x + x^2/2$$

gilt. Unter Verwendung von (2.12) ergibt sich

$$
\begin{aligned}
& \left| \mathrm{e}^{-t^2/2} - \prod\nolimits_{k=1}^{n} \left(1 - \frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2\right)\right| \\
= & \left| \prod\nolimits_{k=1}^{n} \exp\left[-\frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2\right] - \prod\nolimits_{k=1}^{n} \left(1 - \frac{1}{2}\frac{t^2}{s_n^2}\sigma_k^2\right)\right| \\
\leq & \frac{t^4}{8}\sum_{k=1}^{n}\frac{\sigma_k^4}{s_n^4} \leq \frac{t^4}{8}\sup_{k\leq n}\frac{\sigma_k^2}{s_n^2}\sum_{k=1}^{n}\frac{\sigma_k^2}{s_n^2} = \frac{t^4}{8}\sup_{k\leq n}\frac{\sigma_k^2}{s_n^2}.
\end{aligned}
$$

Zusammen mit (2.13) erhalten wir

$$\left| E\left(\exp\left[it\hat{S}_n\right]\right) - \mathrm{e}^{-t^2/2}\right| \leq \frac{\delta\,|t|}{2} + \frac{1}{s_n^2}\sum_{k=1}^{n} E\left(X_k^2; |X_k| \geq \delta s_n\right) + \frac{t^4}{8}\sup_{k\leq n}\frac{\sigma_k^2}{s_n^2}.$$

Für festes δ konvergieren der zweite und der dritte Summand für $n \to \infty$ gegen 0, der zweite wegen der Lindeberg-Bedingung und der dritte nach (2.10). Da $\delta > 0$ beliebig ist, folgt

$$\lim_{n\to\infty}\left| E\left(\exp\left[\frac{it}{s_n}S_n\right]\right) - \mathrm{e}^{-t^2/2}\right| = 0$$

■

3 Bedingte Wahrscheinlichkeiten und Erwartungswerte

Definition 3.1
Sind $A, B \in \mathcal{F}$ mit $P(B) > 0$, so ist die bedingte Wahrscheinlichkeit von A gegeben B durch $P(A \,|\, B) := P(A \cap B)/P(B)$ definiert.

Diese elementare Definition ist aus folgendem Grund nicht ganz befriegend. Oft möchte man die bedingte Wahrscheinlichkeit von A nicht nur bedingt auf ein Ereignis B mit $P(B) > 0$ betrachten, sondern auf die Werte einer Zufallsgrösse X. Z.B. möchte man $P(A \,|\, X = x)$ für x im Wertebereich von X kennen. Falls die Verteilungsfunktion von X jedoch stetig ist, so gilt $P(X = x) = 0$ für alle x. Demzufolge ist $P(A \,|\, X = x)$ für *kein* x definiert. Der Ausweg besteht darin, dass $P(A \,|\, X = x)$ nicht einzeln für jedes x definiert wird, sondern als Funktion von x. Die Bedingung, die diese Funktion erfüllen muss, ergibt sich aus einer natürlichen Übertragung des Satzes über die totale Wahrscheinlichkeit.

Es ist formal bequemer, statt mit Zufallsgrössen X, gleich mit Teil-σ-Algebren $\mathcal{G} \subset \mathcal{F}$ zu arbeiten. Wir kommen dann danach wieder auf die Situation mit Zufallsgrössen zurück, wobei dann $\mathcal{G} := \sigma(X) = X^{-1}(\mathcal{B})$ genommen wird.

Wir beginnen mit einer einfachen Situation. Es sei $\{B_1, \ldots, B_n\}$ eine Zerlegung von Ω in paarweise disjunkte messbare Teilmengen. Für ein feste $A \in \mathcal{F}$ können wir die Folge $P(A \,|\, B_i)$, $i = 1, \ldots, n$ definieren, sofern alle $P(B_i)$ grösser als 0 sind. Nach dem Satz über die totale Wahrscheinlichkeit gilt für jedes Teilmenge $J \subset \{1, \ldots, n\}$:

$$P\left(A \cap \bigcup_{j \in J} B_j\right) = \sum_{j \in J} P(A \,|\, B_j) P(B_j) \,. \tag{3.1}$$

Wir können das noch etwas komplizierter interpretieren und $\{P(A \,|\, B_j)\}$ als Funktion auf Ω auffassen, nennen wir sie ψ_A :

$$\psi_A(\omega) := P(A \,|\, B_j), \text{ für } \omega \in B_j.$$

Wir bezeichnen mit $\mathcal{G}$ die von den B_j erzeugte σ-Algebra. $\mathcal{G}$ enthält genau die Mengen, die sich als $\bigcup_{j \in J} B_j$ mit $J \subset \{1, \ldots, n\}$ darstellen lassen. Da ψ_A konstant auf den B_j ist, ist es eine $\mathcal{G}$-m.b. Funktion. Um dies zu betonen, schreiben wir $\psi_{A,\mathcal{G}}$. Wir können diese Eigenschaften wie folgt zusammenfassen:

Lemma 3.2
In der obigen Situation gilt:

a) $\psi_{A,\mathcal{G}} : \Omega \to [0,1]$ *ist* $\mathcal{G}$*-messbar.*

b) Für alle $B \in \mathcal{G}$ *gilt*
$$P(A \cap B) = \int_B \psi_{A,\mathcal{G}}(\omega) \; P(d\omega) \,.$$

Beweis. a) ist offensichtlich nach der Konstruktion von $\psi_{A,\mathcal{G}}$. b) ergibt sich unmittelbar aus (3.1). ■

Diese beiden Eigenschaften charakterisieren $\psi_{A,\mathcal{G}}$ eindeutig. Der (sehr einfache) Beweis sei dem Leser überlassen. Die Diskussion legt eine weitgehende Verallgemeinerung von bedingten Wahrscheinlichkeiten nahe. Wir betrachten eine *beliebige* Teil-σ-Algebra $\mathcal{G}$ von $\mathcal{F}$.

Definition 3.3
Sei $A \in \mathcal{F}$*. Eine Abbildung* $\psi_{A,\mathcal{G}} : \Omega \to [0,1]$ *heisst* ***bedingte Wahrscheinlichkeit von*** A ***gegeben*** $\mathcal{G}$*, wenn a) und b) von Lemma 3.2 erfüllt sind.*

Satz 3.4
Sei $A \in \mathcal{F}$ *und* $\mathcal{G}$ *eine Teil-*σ*-Algebra von* $\mathcal{F}$*. Dann existiert eine bis auf f.s.-Gleichheit eindeutige bedinge Wahrscheinlichkeit von* A *gegeben* $\mathcal{G}$*.*

Beweis. Wir betrachten die Abbildung $\mathcal{G} \ni B \to \mu(B) := P(A \cap B)$. Sie definiert ein Mass auf $\mathcal{G}$ mit $\mu \ll P$. Nach dem Satz von Radon-Nikodym existiert eine bis auf P-f.s.-Gleichheit eindeutige $\mathcal{G}$-m.b. Funktion $\psi_{A,\mathcal{G}} : \Omega \to \mathbb{R}^+$ mit

$$\int_B \psi_{A,\mathcal{G}} \; dP = \mu(B) = P(A \cap B) \,,$$

für alle $B \in \mathcal{G}$. $\psi_{A,\mathcal{G}}$ ist die gewünschte bedingte Wahrscheinlichkeit. Die Eindeutigkeit bis auf f.s.-Gleichheit ergibt sich aus der Eindeutigkeit der Radon-Nikodym-Ableitung.

Wir müssen noch nachweisen müssen, dass $\psi_{A,\mathcal{G}}$ so gewählt werden kann, dass $0 \leq \psi_{A,\mathcal{G}}(\omega) \leq 1$ für alle $\omega \in \Omega$ gilt. Erfüllt $\psi_{A,\mathcal{G}} : \Omega \to \mathbb{R}^+$ die obige Gleichung, so setzen wir $C := \{\omega : \psi_{A,\mathcal{G}}(\omega) > 1\} \in \mathcal{G}$. Dann folgt

$$\int_C \psi_{A,\mathcal{G}} dP = P(A \cap C) \leq P(C) = \int_C 1 dP,$$

$$\int_C \left(1 - \psi_{A,\mathcal{G}}\right) dP \geq 0.$$

Wegen $\psi_{A,\mathcal{G}} > 1$ auf C folgt $P(C) = 0$. Wir können daher $\psi_{A,\mathcal{G}}$ durch $\min\left(\psi_{A,\mathcal{G}}, 1\right)$ ersetzen. Diese Funktion stimmt mit $\psi_{A,\mathcal{G}}$ f.s. überein und nimmt nur Werte in $[0, 1]$ an.
■

In Zukunft schreiben wir $P(A \mid \mathcal{G})$ anstelle von $\psi_{A,\mathcal{G}}$. Man muss sich klar darüber sein, dass bedingte Wahrscheinlichkeiten im obigen Sinn nur bis auf f.s.-Gleichheit eindeutig gegeben sind. Man spricht dann oft von einer *Version* der bedingten Wahrscheinlichkeit, wenn eine bestimmte $\mathcal{G}$-m.b. Funktion gewählt wird. Dies hat eine unangenehme Konsequenz, die wir kurz diskutieren.

Bisher hatten wir die bedingten Wahrscheinlichkeiten $P(A \mid \mathcal{G})$ nur für ein festes $A \in \mathcal{F}$ diskutiert. Eine naheliegende Frage ist, ob das tatsächlich eine Wahrscheinlichkeit als Funktion von A ist. Präziser ausgedrückt: Ist es möglich, für jedes $A \in \mathcal{F}$, eine Version von $P(A \mid \mathcal{G})$ so zu wählen, dass für alle $\omega \in \Omega$ die Abbildung $A \to P(A \mid \mathcal{G})(\omega)$ ein Wahrscheinlichkeitsmass auf $\mathcal{F}$ ist?

Definition 3.5
Es sei $\mathcal{G}$ eine Teil-σ-Algebra von $\mathcal{F}$. Eine ***reguläre bedingte Wahrscheinlichkeit*** *gegeben $\mathcal{G}$ ist eine Abbildung $Q : \Omega \times \mathcal{F} \to [0, 1]$ mit den folgenden Eigenschaften*

1. *Für jedes $A \in \mathcal{F}$ ist die Abbildung $\Omega \ni \omega \to Q(\omega, A)$ $\mathcal{G}$-m.b.*
2. *Für jedes $\omega \in \Omega$ ist die Abbildung $\mathcal{F} \ni A \to Q(\omega, A)$ ein Wahrscheinlichkeitsmass auf $\mathcal{F}$.*
3. *$\omega \to Q(\omega, A)$ ist für jedes $A \in \mathcal{F}$ eine Version der bedingten Wahrscheinlichkeit von A gegeben $\mathcal{G}$ ist, d.h. es gilt*

$$P(A \cap B) = \int_B Q(\omega, A) P(d\omega),$$

für alle $B \in \mathcal{G}$. (Wir schreiben in Zukunft natürlich wieder $P(A \mid \mathcal{G})(\omega)$ anstelle von $Q(\omega, A)$ auch für eine reguläre bedingte Wahrscheinlichkeiten).

Unglücklicherweise existieren reguläre bedingte Wahrscheinlichkeiten nicht immer. Es gilt jedoch:

Satz 3.6
Ist Ω ein vollständiger, separabler metrischer Raum und $\mathcal{F} = \mathcal{B}_\Omega$ die Borel-σ-Algebra, das heisst die von den offenen Mengen in Ω erzeugte σ-Algebra, so existiert für jede Teil-σ-Algebra $\mathcal{G}$ eine reguläre bedingte Wahrscheinlichkeit.

Der Beweis ist nicht ganz einfach und kann hier nicht ausgeführt werden. $\mathbb{R}$ und $\mathbb{R}^n$ sind natürlich vollständige, separabler metrische Räume (mit der üblichen Metrik). Obwohl also die Bedingungen des Satzes oft erfüllt sind, ist die Tatsache, dass reguläre bedingte Wahrscheinlichkeiten nicht immer existieren, natürlich etwas enttäuschend.

Nun zu bedingten Erwartungswerten. Sei X eine auf $(\Omega, \mathcal{F}, P)$ definierte Zufallsgrösse und $\mathcal{G}$ eine Teil-σ-Algebra von $\mathcal{F}$. Falls eine reguläre bedingte Wahrscheinlichkeit gegeben $\mathcal{G}$ existiert, so wird man den bedingten Erwartungswert einfach durch

$$E(X \mid \mathcal{G})(\omega) := \int X(\omega') \, P(d\omega' \mid \mathcal{G})(\omega)$$

definieren, sofern $X \in \mathcal{L}_1(\Omega, \mathcal{F}, P(\cdot \mid \mathcal{G})(\omega))$ für alle $\omega \in \Omega$ ist. Wie schon oben ewähnt, exisitieren jedoch reguläre bedingte Wahrscheinlichkeiten nicht immer und man konstruiert daher die bedingten Erwartungswerte besser direkt über den Satz von Radon-Nikodym:

Satz 3.7
Sei $X \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ und $\mathcal{G}$ eine Teil-σ-Algebra von $\mathcal{F}$. Dann existiert eine bis auf P-f.s.-Gleichheit eindeutig definierte integrierbare und $\mathcal{G}$-m.b. Zufallsgrösse $E(X \mid \mathcal{G})$ mit

$$\int_B E(X \mid \mathcal{G}) \, dP = \int_B X \, dP \tag{3.2}$$

für alle $B \in \mathcal{G}$.

Beweis. Wir nehmen zunächst an, dass $X \geq 0$ gilt. Dann ist $\mu(B) := \int_B X \, dP,\ B \in \mathcal{G}$, ein Mass auf $(\Omega, \mathcal{G})$, das absolutstetig bezüglich P ist. Aus dem Satz von Radon-Nikodym folgt daher die Existenz einer $\mathcal{G}$-m.b. nicht negativen Zufallsgrösse $E(X \mid \mathcal{G})$, die die Eigenschaft (3.2) erfüllt. Ist X integrierbar, so gilt

$$\int_\Omega E(X \mid \mathcal{G}) \, dP = \int_\Omega X \, dP < \infty.$$

Daher ist $E(X \mid \mathcal{G})$ integrierbar und insbesondere fast überall endlich. Ist $X \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$, so zerlegen wir X in Positiv- und Negativteil $X = X^+ - X^-$ und setzen

$$E(X \mid \mathcal{G}) = E(X^+ \mid \mathcal{G}) - E(X^- \mid \mathcal{G}).$$

Diese Zufallsgrösse erfüllt offensichtlich (3.2) und ist integrierbar.

Eindeutigkeit: Sind f, g zwei $\mathcal{G}$-m.b. integrierbare Zufallsgrössen mit $\int_B f \, dP = \int_B g \, dP$ für alle $B \in \mathcal{G}$, so folgt sehr einfach $f = g$ P-f.ü.. ■

Man beachte, dass bedingte Wahrscheinlichkeiten einfach Spezialfälle von bedingten Erwartungswerten sind:

$$P(A \mid \mathcal{G}) = E(1_A \mid \mathcal{G}).$$

Der folgende Satz ist eine Auflistung der wichtigsten Eigenschaften bedingter Erwartungswerte.

Satz 3.8
Es seien $X, X_1, X_2 \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$, und $\mathcal{G}$ sowie $\mathcal{G}'$ seien Teil-σ-Algebren von $\mathcal{F}$. Dann gelten:

a) *Ist $\mathcal{G}$ die triviale σ-Algebra, das heisst $\mathcal{G} = \{\emptyset, \Omega\}$, so gilt $E(X \,|\, \mathcal{G}) = E(X)$ f.s.*

b) *Ist X $\mathcal{G}$-messbar, so gilt $E(X \,|\, \mathcal{G}) = X$ f.s..*

c) *Sind $a_1, a_2 \in R$, so gilt $E(a_1 X_1 + a_2 X_2 \,|\, \mathcal{G}) = a_1 E(X_1 \,|\, \mathcal{G}) + a_2 E(X_2 \,|\, \mathcal{G})$ f.s.*

d) *Ist $X \geq 0$ f.s, so ist $E(X \,|\, \mathcal{G}) \geq 0$ f.s..*

e) *Es gilt $|E(X \,|\, \mathcal{G})| \leq E(|X| \,|\, \mathcal{G})$ f.s.*

f) *Ist $\mathcal{G}' \subset \mathcal{G}$, so gilt $E(X \,|\, \mathcal{G}') = E(E(X \,|\, \mathcal{G}) \,|\, \mathcal{G}')$ f.s.*

g) *Sind X und $\mathcal{G}$ unabhängig, so gilt $E(X \,|\, \mathcal{G}) = E(X)$ f.s.*[2]

Beweis. a) Da $E(X \,|\, \mathcal{G})$ $\mathcal{G}$-m.b. ist, so diese Funktion konstant. Wegen $\int_\Omega E(X \,|\, \mathcal{G}) dP = \int_\Omega X \, dP$ folgt die Behauptung.

b) Folgt unmittelbar aus der Definition des bedingten Erwartungswertes.

c) Sei Y die rechte Seite der behaupteten Gleichung. Dann gilt für alle $D \in \mathcal{G}$:

$$\begin{aligned}\int_D Y \, dP &= a_1 \int_D E(X_1 \,|\, \mathcal{G}) \, dP + a_2 \int_D E(X_2 \,|\, \mathcal{G}) \, dP \\ &= a_1 \int_D X_1 \, dP + a_2 \int_D X_2 \, dP = \int_D (a_1 X_1 + a_2 X_2) \, dP.\end{aligned}$$

Ausserdem ist Y natürlich $\mathcal{G}$-messbar, erfüllt also die definierenden Eigenschaften für einen bedingten Erwartungswert von $a_1 X_1 + a_2 X_2$ gegeben $\mathcal{G}$.

d) Für alle $D \in \mathcal{G}$ gilt $\int_D E(X \,|\, \mathcal{G}) \, dP = \int_D X \, dP \geq 0$. Ist $D = \{\, E(X \,|\, \mathcal{G}) < 0 \,\}$, so folgt $P(D) = 0$.

e) Wegen $E(X \,|\, \mathcal{G}) = E(X^+ \,|\, \mathcal{G}) - E(X^- \,|\, \mathcal{G})$ folgt

$$|E(X \,|\, \mathcal{G})| \leq E(X^+ \,|\, \mathcal{G}) + E(X^- \,|\, \mathcal{G}) = E(|X| \,|\, \mathcal{G}).$$

f) Sei wieder Y die rechte Seite der behaupteten Gleichung, und sei $D' \in \mathcal{G}'$. Dann gilt $\int_{D'} Y \, dP = \int_{D'} E(X \,|\, \mathcal{G}) \, dP = \int_{D'} X \, dP$, da $D' \in \mathcal{G}$ ist. Da Y $\mathcal{G}'$-messbar ist, folgt die Aussage.

g) Die konstante Abbildung $\omega \mapsto E(X)$ ist natürlich $\mathcal{G}$-messbar. Für $D \in \mathcal{G}$ gilt wegen der Unabhängigkeit

$$\int_D X \, dP = E(1_D X) = P(D) E(X) = \int_D E(X) \, dP.$$

[2] Für mathematische Puristen sind die obigen Formulierungen ganz leicht anrüchig: Der Satz von Radon-Nidkodym liefert die entsprechenden Objekte ohnehin nur bis auf f.s.-Gleichheit. Genau genommen sollte man daher die bedingten Erwartungswerte als Äquivalenzklassen unter f.s.-Gleichheit auffassen. Dann sind die obigen Gleichungen einfach Gleichungen zwischen Äquivalenzklassen und man kann sich die ständige Wiederholung von „fast sicher“ ersparen.

$\blacksquare$

Aus d) und c) folgt sofort, dass $X_1 \leq X_2$ f.s. die Ungleichung $E(X_1 \,|\, \mathcal{G}) \leq E(X_2 \,|\, \mathcal{G})$ impliziert.

Für bedingte Erwartungswerte gelten die üblichen Konvergenzsätze:

Satz 3.9
Seien X_n, $n \in \mathbb{N}$, und X integrierbare Zufallsgrössen und $\mathcal{G}$ sei eine Teil-σ-Algebra von $\mathcal{F}$.

a) *Gilt $X_n \geq 0$ und $X_n \uparrow X$ f.s., so folgt $E(X_n \,|\, \mathcal{G}) \uparrow E(X \,|\, \mathcal{G})$ f.s.*

b) *Gilt $X_n \geq 0$ und $X_n \to X$ f.s., so gilt das Lemma von Fatou:*

$$E(X \,|\, \mathcal{G}) \leq \liminf_{n\to\infty} E(X_n \,|\, \mathcal{G}) \text{ f.s.}$$

c) *Existiert $Y \in \mathcal{L}_1$ mit $|X_n| \leq Y$ f.s. für alle $n \in N$, und $X_n \to X$ f.s., so gilt:*

$$E(X \,|\, \mathcal{G}) = \lim_{n\to\infty} E(X_n \,|\, \mathcal{G}) \text{ f.s.}$$

Beweis. Wir beweisen nur a). b) und c) folgen dann in gleicher Weise wie das Lemma von Fatou und der Satz von Lebesgue.

Aus $X_n \leq X_{n+1} \leq X$ f.s. für alle $n \in \mathbb{N}$ folgt $E(X_n \,|\, \mathcal{G}) \leq E(X_{n+1} \,|\, \mathcal{G}) \leq E(X \,|\, \mathcal{G})$ f.s. für alle $n \in \mathbb{N}$. Die Folge $(E(X_n \,|\, \mathcal{G}))_{n\in\mathbb{N}}$ konvergiert also fast sicher gegen eine $\mathcal{G}$-messbare Zufallsgrösse Y mit $Y \leq E(X \,|\, \mathcal{G})$. Für alle $D \in \mathcal{G}$ gilt

$$\begin{aligned}
\int_D (E(X \,|\, \mathcal{G}) - Y)\, dP &= \int_D X\, dP - \int_D Y\, dP \\
&= \int_D X\, dP - \int_D \lim_{n\to\infty} E(X_n \,|\, \mathcal{G})\, dP \\
&= \int_D X\, dP - \lim_{n\to\infty} \int_D E(X_n \,|\, \mathcal{G})\, dP \\
&= \int_D X\, dP - \lim_{n\to\infty} \int_D X_n\, dP = 0.
\end{aligned}$$

Somit folgt $Y = E(X \,|\, \mathcal{G})$ fast sicher. $\blacksquare$

Satz 3.10
Seien $X, Y \in \mathcal{L}_1$ und $X \cdot Y \in \mathcal{L}_1$. Ist Y $\mathcal{G}$-messbar, so gilt $E(YX \,|\, \mathcal{G}) = YE(X \,|\, \mathcal{G})$ fast sicher.

Beweis. Sind $X \geq 0$ und $Y = 1_D$ mit $D \in \mathcal{G}$, so folgt für jedes $D' \in \mathcal{G}$

$$\int_{D'} YE(X \,|\, \mathcal{G})\, dP = \int_{D\cap D'} E(X \,|\, \mathcal{G})\, dP = \int_{D\cap D'} X\, dP = \int_{D'} 1_D X\, dP.$$

Somit erfüllt $YE(X \,|\, \mathcal{G})$ die definierenden Eigenschaften eines bedingten Erwartungswertes von YX gegeben $\mathcal{G}$, das heisst, es gilt $YE(X \,|\, \mathcal{G}) = E(XY \,|\, \mathcal{G})$ fast sicher.

Der allgemeine Fall folgt dann auf die übliche Weise mit einem montonen Klassenarguement. ■

Wir wollen noch den wichtigen Spezialfall betrachten, wo $\mathcal{G}$ die von einer Zufallsgrösse Z erzeugt σ-Algebra ist. Z nehme Werte in einem allgemeinen messbaren Raum $(S, \mathcal{S})$ an. Ist X eine integrierbare Zufallsgrösse (reellwertig), so schreibt man meist $E(X \mid Z)$ anstelle von $E(X \mid \sigma(Z))$. $E(X \mid Z)$ ist also eine auf Ω definierte, integrierbare, $\sigma(Z)$ messbare Zufallsgrösse. Wir verwenden nun das folgende Resultat:

Lemma 3.11
Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $(S, \mathcal{S})$ ein messbarer Raum und $F : \Omega \to S$ eine messbare Abbildung. Eine (reellwertige) Funktion $f : \Omega \to \mathbb{R}$ ist genau dann $\sigma(F) - \mathcal{B}$-m.b., wenn eine $\mathcal{S} - \mathcal{B}$-m.b. Abbildung $\varphi : S \to \mathbb{R}$ existiert mit $f = \varphi \circ F$. φ ist eindeutig durch F bestimmt, bis auf PF^{-1}-f.s.-Gleichheit.

Beweis. Übungsaufgabe. ■

Unter Verwendung dieses Lemmas sehen wir, dass wir $E(X \mid Z)$ über $(S, \mathcal{S})$ faktorisieren können, d.h. es gibt eine messbare Abbildung, nennen wir sie ad hoc wieder $\varphi : S \to \mathbb{R}$, mit

$$E(X \mid Z)(\omega) = \varphi(Z(\omega)),$$

und φ ist eindeutig bis auf PF^{-1}-f.s.-Gleichheit. Man schreibt diese Abbildung meist als $E(X \mid Z = \cdot)$, oder unter Betonung des Arguments $E(X \mid Z = z)$, $z \in S$. Mit dem Transformationssatz 1.50 sehen wir, dass diese Funktion auf S durch die Eigenschaft

$$\int_A E(X \mid Z = z)\, PZ^{-1}(dz) = \int_{Z^{-1}(A)} X dP, \ \forall A \in \mathcal{S}$$

charakterisiert ist.

Wenn Z eine Zufallsgrösse ist, die nur abzählbar viele Werte annimmt, mit $P(Z = z) > 0$ für alle z im Wertebereich von Z, so ist $E(X \mid Z = z)$ natürlich einfach der elementar definierte bedingte Erwartungswert, d.h. der Erwartungswert von X bezüglich des Wahrscheinlichkeitsmasses $A \to P(A \mid Z = z)$, das wegen $P(Z = z) > 0$ elementar definiert ist. Der Leser möge sich das als Übungsaufgabe überlegen.

Hier noch einige Beispiele und Ergänzungen.

Beispiel 3.12
Sei $(\Omega, \mathcal{F}, P) = (\mathbb{R}, \mathcal{B}, \mu)$, wobei μ ein symmetrisches Wahrscheinlichkeitsmass auf $\mathcal{B}$ ist. Symmetrie von μ bedeutet $\mu(-A) = \mu(A)$, $A \in \mathcal{B}$, wobei $-A := \{-x : x \in A\}$ ist. Sei ferner $\mathcal{B}_{\mathrm{sym}}$ die σ-Algebra der symmetrischen Mengen:

$$\mathcal{B}_{\mathrm{sym}} := \{A \in \mathcal{B} : -A = A\}.$$

Ist $X : \mathbb{R} \to \mathbb{R}$ irgend eine μ-integrierbare Zufallsgrösse, so ist

$$\hat{X}(\omega) := \frac{X(\omega) + X(-\omega)}{2}$$

eine Version des bedingten Erwartungswertes $E(X|\mathcal{B}_{\text{sym}})$. Zunächst ist $\hat{X}$ offensichtlich $\mathcal{B}_{\text{sym}}$-m.b. Wir setzen $\tau(x) := -x$. Für $A \in \mathcal{B}_{\text{sym}}$ gilt:

$$\begin{aligned}
\int_A \hat{X} d\mu &= \frac{1}{2}\int_A X d\mu + \frac{1}{2}\int_A (X \circ \tau)\, d\mu \\
&= \frac{1}{2}\int_A X d\mu + \frac{1}{2}\int_A X d\left(\mu\tau^{-1}\right) \\
&= \frac{1}{2}\int_A X d\mu + \frac{1}{2}\int_A X d\mu = \int_A X d\mu.
\end{aligned}$$

In der zweiten Gleichung haben wir den Transformationssatz und die Symmetrie von A benützt, in der dritten Gleichung die Symmetrie von μ.

Falls μ nicht symmetrisch ist, so ist die Berechnung von $E(X|\mathcal{B}_{\text{sym}})$ etwas schwieriger. Wir nehmen an, dass μ eine überall positive Dichte f bezüglich des Lebesgue-Masses hat. Dann ist

$$\hat{X}(\omega) := \frac{f(\omega)X(\omega) + f(-\omega)X(-\omega)}{f(\omega) + f(-\omega)}$$

eine Version des bedingten Erwartungswertes.

Beispiel 3.13

Seien $X_1, X_2, \ldots$ unabhängige integrierbare Zufallsgrössen. Wir setzen $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$ und $S_n := \sum_{i=1}^n X_i$. S_n ist natürlich $\mathcal{F}_n$-m.b. und X_{n+1} ist unabhängig von $\mathcal{F}_n$. Demzufolge ist

$$\begin{aligned}
E(S_{n+1}|\mathcal{F}_n) &= E(S_n + X_{n+1}|\mathcal{F}_n) = E(S_n|\mathcal{F}_n) + E(X_{n+1}|\mathcal{F}_n) \\
&= S_n + E(X_{n+1}).
\end{aligned}$$

In der zweiten Gleichung haben wir c) von Satz 3.8 verwendet, und in der dritten b) und g).

Wir können auch $E\left(S_{n+1}^2\middle|\mathcal{F}_n\right)$ berechnen, unter der Voraussetzung, dass die X_i quadratisch integrierbar sind:

$$\begin{aligned}
E\left(S_{n+1}^2\middle|\mathcal{F}_n\right) &= E\left((S_n + X_{n+1})^2\middle|\mathcal{F}_n\right) = E\left(S_n^2 + 2S_nX_{n+1} + X_{n+1}^2\middle|\mathcal{F}_n\right) \\
&= S_n^2 + 2E(S_nX_{n+1}|\mathcal{F}_n) + E\left(X_{n+1}^2\right) \\
&= S_n^2 + 2S_nE(X_{n+1}) + E\left(X_{n+1}^2\right).
\end{aligned}$$

In der letzten Gleichung haben wir noch Satz 3.10 benützt und dann nochmals $E(X_{n+1}|\mathcal{F}_n) = E(X_{n+1})$ wegen der Unabhängigkeit.

Ist (X, Y) ein zweidimensionaler Zufallsvektor, so ist in der Regel die Berechnung von $E(Y|X)$ nicht ganz einfach. Es gibt jedoch einen wichtigen Spezialfall, den wir nun diskutieren. Zunächst benötigen wir das folgende Ergebnis:

Proposition 3.14

Sei $(X_1, \ldots, X_m, X_{m+1}, \ldots, X_n)$ ein n-dimensionaler normalverteilter Zufallsvektor. Gilt $\operatorname{cov}(X_i, X_j) = 0$ für $1 \le i \le m,\ m+1 \le j \le n$, so sind $(X_1, \ldots, X_m), (X_{m+1}, \ldots, X_n)$ unabhängig.

Beweis. Durch Subtraktion der Erwartungswerte können wir annehmen, dass $EX_i = 0$ für alle i gilt. Sei μ die Verteilung von $(X_1, \ldots, X_m, X_{m+1}, \ldots, X_n)$ auf $(\mathbb{R}^n, \mathcal{B}_n)$, ν_1 die Verteilung von $(X_1, \ldots, X_m)$ sowie ν_2 die Verteilung von $(X_{m+1}, \ldots, X_n)$. Gemäss Proposition 2.45 ist die behauptete Unabhängigkeite äquivalent mit

$$\mu = \nu_1 \otimes \nu_2.$$

Wir zeigen diese Gleichung, indem wir Gleichheit der charakteristischen Funktionen nachweisen:

$$\hat{\mu} = \widehat{\nu_1 \otimes \nu_2}.$$

Mit Hilfe des Satzes von Fubini folgt sofort

$$\widehat{\nu_1 \otimes \nu_2}\,(t_1, \ldots, t_n) = \widehat{\nu}_1\,(t_1, \ldots, t_m)\,\widehat{\nu}_2\,(t_{m+1}, \ldots, t_n)$$

Nun sind sowohl μ wie ν_1, ν_2 Normalverteilungen. μ mit Kovarianzmatrix

$$\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$$

wobie Σ_1, Σ_2 die Kovarianzmatrizen von ν_1, ν_2. 0 steht für die $m \times (n-m)$- bzw. $(n-m) \times m$-Nullmatrix. Die angegebene Form ergibt sich aus der Voraussetzung, dass Kovarianzen zwischen der ersten Gruppe und der zweiten verschwinden. Wir erhalten somit

$$\hat{\mu}\,(t) = \exp\left[-\frac{1}{2}\langle t, \Sigma t\rangle\right] = \exp\left[-\frac{1}{2}\left\langle t^{(1)}, \Sigma_1 t^{(1)}\right\rangle\right]\exp\left[-\frac{1}{2}\left\langle t^{(2)}, \Sigma_2 t^{(2)}\right\rangle\right],$$

mit $t = (t_1, \ldots, t_n)$, $t^{(1)} = (t_1, \ldots, t_m)$, $t^{(2)} = (t_{m+1}, \ldots, t_2)$. Die beiden Faktoren auf der rechten Seite sind die charakteristischen Funktionen von ν_1 bzw. ν_2. ■

Wir wollen dieses Ergebnis nun benützen, um für einen n-dimensionalen normalverteilten Zufallsvektor $(X_1, \ldots, X_n)$ den bedingten Erwartungswert

$$E\left(X_n \mid \mathcal{F}_{n-1}\right)$$

zu berechnen, wobei $\mathcal{F}_{n-1} := \sigma\,(X_1, \ldots, X_{n-1})$ ist. Der Einfachheit halben nehmen wir wieder an, dass $EX_i = 0$ gilt. (Der allgemeine Fall lässt sich sofort darauf zurückführen).

Wir betrachten die Vektorraum L_2^0 der Äquivalenzklassen quadratisch integrierbaren Zufallsgrössen mit Erwartungswert 0. Die Äquivalenzrelation ist dabei durch die f.s.-Gleichheit gegeben. Wir unterscheiden nachfolgend nicht sorgfältig zwischen einer Zufallsgrösse und ihrer Äquivalenzklasse, um die Notation nicht zu überlasten. Auf diesem Vektorraum betrachten wir das Skalarprodukt

$$\langle X, Y\rangle = E\,(XY) = \operatorname{cov}(X, Y).$$

L_2^0 versehen mit diesem Sklarprodukt ist ein reeller Hilbertraum. U sei der lineare Unterraum von L_2^0, der von $X_1, \ldots, X_{n-1}$ aufgespannt wird. Die orthogonale Projektion von L_2^0 auf U bezeichnen wir mit π_U. $\pi_U(X_n)$ ist in jedem Fall eine Linearkombination der $X_1, \ldots, X_{n-1}$:

$$\pi_U(X_n) = \sum_{i=1}^{n-1} a_i X_i, \; a_i \in \mathbb{R}.$$

Diese Darstellung ist natürlich eindeutig, sofern die $X_1, \ldots, X_{n-1}$ unabhängig im Sinne der Linearen Algebra sind. Aus Sätzen der Linearen Algebra ergibt sich, dass $X_1, \ldots, X_{n-1}$ genau dann unabhängig im Sinne der LA sind, wenn die Kovarianzmatrix

$$\Sigma := (\langle X_i, X_j \rangle)_{1 \le i,j \le n-1} = (\operatorname{cov}(X_i, X_j))_{1 \le i,j \le n-1}$$

regulär ist. In diesem Fall erhalten wir die $a_1, \ldots, a_{n-1}$ durch

$$\begin{pmatrix} a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = \Sigma^{-1} \begin{pmatrix} \operatorname{cov}(X_1, X_n) \\ \vdots \\ \operatorname{cov}(X_{n-1}, X_n) \end{pmatrix}.$$

Proposition 3.15
Ist $(X_1, \ldots, X_n)$ ein normalverteilter Zufallsvektor mit $EX_i = 0, \; \forall i$, so gilt

$$E(X_n | \mathcal{F}_{n-1}) = \pi_U(X_n)$$

Beweis.

$$\begin{aligned} E(X_n | \mathcal{F}_{n-1}) &= E(X_n - \pi_U(X_n) + \pi_U(X_n) | \mathcal{F}_{n-1}) \\ &= E(X_n - \pi_U(X_n) + \pi_U(X_n) | \mathcal{F}_{n-1}) \\ &= E(X_n - \pi_U(X_n) | \mathcal{F}_{n-1}) + E(\pi_U(X_n) | \mathcal{F}_{n-1}). \end{aligned}$$

$\pi_U(X_n)$ ist jedoch $\mathcal{F}_{n-1}$-m.b. Daher gilt nach Satz 3.8 b) $E(\pi_U(X_n) | \mathcal{F}_{n-1}) = \pi_U(X_n)$.
Zur Auswertung des ersten Summanden bemerken wir, dass

$$(X_1, \ldots, X_{n-1}, X_n - \pi_U(X_n))$$

nach Korollar 2.24 ebenfalls normalverteilt ist. Für $i = 1, \ldots, n-1$ gilt

$$E(X_i X_n) = \langle X_i, X_n \rangle = \langle X_i, \pi_U(X_n) \rangle,$$

d.h.

$$\operatorname{cov}(X_i, X_n - \pi_U(X_n)) = 0.$$

Nach der vorangegangen Proposition 3.14 sind daher $X_n - \pi_U(X_n)$ und $\mathcal{F}_{n-1}$ unabhängig, woraus sich nach Satz 3.8 g) die Gleichung

$$E(X_n - \pi_U(X_n) | \mathcal{F}_{n-1}) = E(X_n - \pi_U(X_n)) = 0.$$

ergibt. Daraus folgt

$$E(X_n | \mathcal{F}_{n-1}) = \pi_U(X_n).$$

■

Es muss betont werden, dass wir in massiver Weise ausgenützt haben, dass $(X_1, \ldots, X_n)$ normalverteilt ist, weil nur unter dieser Voraussetzung aus der Unkorreliertheit die Unabhängigkeit folgt.

4 Masse auf unendlichen Produkträumen

Wir stellen in diesem Kapitel die zwei wichtigsten Sätze über die Existenz von Massen auf unendlichen Produkträumen vor: Den Satz von Kolmogoroff und den Satz von Ionescu-Tulcea. Aus Zeitgründen werden die Sätze nicht bewiesen. Wir leiten jedoch den Rahmen für die Sätze präzise her.

4.1 Stochastische Prozesse und ihre endlichdimensionalen Verteilungen

Wir hatten in Abschnitt 1.6 Produkte von messbaren Räumen und Produkte von Massen eingeführt. Wir erweitern das nun auf abzählbar unendlich viele Faktoren.

Ist $(S, \mathcal{S})$ ein messbarer Raum, so ist $S^{\mathbb{N}}$ die Menge der unendlich langen Folgen $\mathbf{s} = (s_1, s_2, \ldots)$ mit $s_i \in S$, oder anders ausgedrückt, die Menge der Abbildungen $\mathbb{N} \to S$. Mit π_n bezeichnen wir die Abbildung $S^{\mathbb{N}} \to S$, $\pi_n(\mathbf{s}) := s_n$. Mit $\pi^{(n)} : S^{\mathbb{N}} \to S^n$ bezeichnen wir die Abbildung $\pi^{(n)}(\mathbf{s}) := (s_1, \ldots, s_n)$. Mit $\mathcal{S}^{\otimes\mathbb{N}}$ bezeichnen wir die sogenannte Produkt-σ-Algebra auf $S^{\mathbb{N}}$, die von den π_n erzeugt wird:

$$\mathcal{S}^{\otimes\mathbb{N}} := \sigma\left(\left\{\pi_n^{-1}(A) : n \in \mathbb{N},\ A \in \mathcal{S}\right\}\right).$$

$\left\{\pi_n^{-1}(A) : n \in \mathbb{N},\ A \in \mathcal{S}\right\}$ ist also ein Erzeugendensystem von $\mathcal{S}^{\otimes\mathbb{N}}$. Dieses ist offensichtlich nicht durchschnittstabil. Ein durchschnittstabiles Erzeugendensystem von $\mathcal{S}^{\otimes\mathbb{N}}$ ist

$$\mathcal{D} := \left\{\pi^{(n)-1}(A) : n \in \mathbb{N},\ A \in \mathcal{S}^{\otimes n}\right\}.$$

Der Einfachheit halber schreiben wir $\mathcal{S}^{\mathbb{N}}$ und $\mathcal{S}^n$ anstelle von $\mathcal{S}^{\otimes\mathbb{N}}$ und $\mathcal{S}^{\otimes n}$. Man muss sich jedoch klar darüber sein, dass dies nicht die Menge der Abbildung $\mathbb{N} \to \mathcal{S}$ bzw. $\{1, \ldots, n\} \to \mathcal{S}$ ist.

Definition 4.1
Es sei $(S, \mathcal{S})$ ein messbarer Raum. Eine Folge $\{X_n\}_{n\in\mathbb{N}}$ von $(S, \mathcal{S})$-wertigen Zufallsgrössen, die auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert sind, heisst (diskreter) $(S, \mathcal{S})$-wertiger ***stochastischer Prozess****. Im Falle $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B})$ sprechen wir einfach von einem stochastischen Prozess.*

Eine Folge von Abbildungen $X_n : \Omega \to S$ können wir natürlich einfach als Abbildung $X : \Omega \to S^{\mathbb{N}}$ auffassen. Die X_n sind dann die Kompositionen $\pi_n \circ X$.

Satz 4.2
Es sei $X = \{X_n\}_{n\in\mathbb{N}}$ eine Folge von Abbildungen $X_n : \Omega \to S$. Dann ist $X : \Omega \to S^{\mathbb{N}}$ genau dann $\mathcal{F}$-$\mathcal{S}^{\mathbb{N}}$-messbar, wenn für jedes $n \in \mathbb{N}$ die Abbildung X_n $\mathcal{F}$-S-messbar ist.

Beweis. Nach Definition von $\mathcal{S}^{\mathbb{N}}$ sind alle Projektionen $\pi_n : S^{\mathbb{N}} \to S$ $\mathcal{S}^{\mathbb{N}}$-$\mathcal{S}$-messbar. Ist X $\mathcal{F}$-$\mathcal{S}^{\mathbb{N}}$-messbar, so ist somit $X_n = \pi_n \circ X$ für jedes $n \in \mathbb{N}$ eine $\mathcal{F}$-$\mathcal{S}$-messbare Abbildung. Sind umgekehrt alle X_n $\mathcal{F}$-S-messbar, so gilt $X^{-1}(\pi_n^{-1}(A)) = X_n^{-1}(A) \in \mathcal{F}$

für alle $n \in \mathbb{N}$ und alle $A \in \mathcal{S}$. Da per Definition $\{ \pi_n^{-1}(A) : n \in \mathbb{N}, \; A \in \mathcal{S} \}$ ein Erzeugendensystem von $\mathcal{S}^{\mathbb{N}}$ ist, folgt die Messbarkeit von X. ■

Die Verteilung eines stochastischen Prozesses X ist einfach seine Verteilung als $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}})$-wertige Zufallsgrösse, d.h. das Wahrscheinlichkeitsmass PX^{-1} auf $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}})$. Die meisten der uns interessierenden Fragen hängen nur von der Verteilung des stochastischen Prozesses ab. Ist X ein stochastischer Prozess, so ist die Folge $\{\pi_n\}_{n\in\mathbb{N}}$ der Projektionen $\pi_n : S^{\mathbb{N}} \to S$ ein auf $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, PX^{-1})$ definierter stochastischer Prozess, der dieselbe Verteilung wie X hat. Es ist daher meist keine Einschränkung anzunehmen, dass der zugrunde liegende Wahrscheinlichkeitsraum von der Form $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, P)$ ist und die X_n die Projektionen von $S^{\mathbb{N}}$ auf S sind.

Die Beschreibung von Verteilungen (d.h. Wahrscheinlichkeitsmassen) auf einem unendlichen Produktraum ist nicht ganz einfach. Sie geschieht in der Regel über die sogenannten endlichdimensionalen Verteilungen.

Definition 4.3
Ist Q ein Wahrscheinlichkeitsmass auf $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}})$, so ist

$$Q^{(n)} := Q(\pi^{(n)})^{-1} \tag{4.1}$$

*ein Wahrscheinlichkeitsmass auf $(S^n, \mathcal{S}^n)$. Die Masse $Q^{(n)}$ heissen die **endlichdimensionalen Verteilungen** von Q.*

Wir wollen im folgenden zwei Fragen nachgehen:

1. Legen die endlichdimensionalen Verteilungen $\{Q^{(n)}\}_{n\in\mathbb{N}}$ die Verteilung Q eindeutig fest?

2. Gibt es zu einer vorgegebenen Folge $\{Q^{(n)}\}_{n\in\mathbb{N}}$ von Wahrscheinlichkeitsmassen auf den Räumen $(S^n, \mathcal{S}^n)$, $n \in \mathbb{N}$, stets ein Wahrscheinlichkeitsmass Q auf $(S^{\mathbb{N}}$,$\mathrm{S}^{\mathbb{N}})$, dessen endlichdimensionale Verteilungen die $Q^{(n)}$ sind?

Die Antwort auf die erste Frage lautet uneingeschränkt *Ja.*

Satz 4.4
Die Folge der endlichdimensionalen Verteilungen $\{Q^{(n)}\}_{n\in\mathbb{N}}$ bestimmt das Wahrscheinlichkeitsmass Q eindeutig.

Beweis. Durch die endlichdimensionalen Verteilungen ist das Mass Q auf dem durchschnittstabilen Erzeugendensystem $\mathcal{D}$. Daraus folgt die Behauptung. ■

Im Spezialfall $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B})$ ist die folgende Aussage nützlich.

Proposition 4.5
Sind $X = \{X_n\}_{n\in\mathbb{N}}$ und $X' = \{X'_n\}_{n\in\mathbb{N}}$ zwei $(\mathbb{R}, \mathcal{B})$-wertige stochastische Prozesse auf den Wahrscheinlichkeitsräumen $(\Omega, \mathcal{F}, P)$ beziehungsweise $(\Omega', \mathcal{F}', P')$, so gilt $PX^{-1} = P'X'^{-1}$ genau dann, wenn

$$P(X_1 \le t_1, \dots, X_n \le t_n) = P'(X'_1 \le t_1, \dots, X'_n \le t_n)$$

für alle $n \in \mathbb{N}$ und alle $t_1, \dots, t_n \in \mathbb{R}$ gilt.

Beweis. Die Mengen der Form $\{ x \in \mathbb{R}^{\mathbb{N}} : x_1 \leq t_1, x_2 \leq t_2, \ldots, x_n \leq t_n \}$ $n \in \mathbb{N}$, $t_1, \ldots, t_n \in \mathbb{R}$ bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}^{\mathbb{N}}$. Nach Voraussetzung stimmen PX^{-1} und $P'X'^{-1}$ auf diesen Mengen überein. Somit folgt $PX^{-1} = P'X'^{-1}$.. ■

Die Antwort auf die zweite Frage ist viel schwieriger. Es ist jedoch offensichtlich, dass die $Q^{(n)}$ einer *Verträglichkeitsbedingung* genügen müssen, damit überhaupt eine Chance besteht, ein Wahrscheinlichkeitsmass Q auf $(S^{\mathbb{N}}, \mathrm{S}^{\mathbb{N}})$ zu finden mit

$$Q^{(n)} = Q(\pi^{(n)})^{-1}.$$

Ist nämlich $\varphi_n : S^n \to S^{n-1}$ die Projektion auf die ersten $n-1$ Koordinaten von S^n, so gilt $\pi^{(n-1)} = \varphi_n \circ \pi^{(n)}$. Daher folgt für jede Menge $A \in \mathrm{S}^{n-1}$:

$$\left(\pi^{(n-1)}\right)^{-1}(A) = \left(\pi^{(n)}\right)^{-1}\left(\varphi_n^{-1}(A)\right)$$

und somit

$$Q^{(n-1)} = Q^{(n)}\varphi_n^{-1}, \quad n \geq 2, \tag{4.2}$$

falls ein Q auf $(S^{\mathbb{N}}, \mathrm{S}^{\mathbb{N}})$ existiert, dessen endlichdimensionalen Verteilungen die $Q^{(n)}$.

Definition 4.6
*Für jedes $n \in \mathbb{N}$ sei $Q^{(n)}$ ein Wahrscheinlichkeitsmass auf $(S^n, \mathcal{S}^n)$. Die Folge $\{Q^{(n)}\}_{n \in \mathbb{N}}$ heisst **verträglich**, wenn (4.2) gilt.*

Wie wir gesehen haben, ist diese Bedingung *notwendig* für die Existenz von Q. Die Frage muss also dahin präzisiert werden, ob zu jeder *verträglichen* Folge ein Wahrscheinlichkeitsmass Q auf $(S^{\mathbb{N}}, \mathrm{S}^{\mathbb{N}})$ existiert, deren endlichdimensionale Verteilungen die $Q^{(n)}$ sind. Die Antwort ist leider "Nein". Es gibt jedoch wichtige Spezialfälle, in denen die Antwort "Ja" lautet.

Satz 4.7 (Kolmogoroff)
Ist S ein vollständiger metrischer Raum und $\mathcal{S}$ die Borel-σ-Algebra von S, so existiert für jede verträgliche Familie $\{Q^{(n)}\}$ eindeutig ein Wahrscheinlichkeitsmass auf $\left(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}\right)$.

Die Satz lässt sich insbesondere auf $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B})$ anwenden. Wir wollen ihn hier nicht beweisen. Er lässt sich auf den Satz von Ionescu-Tulcea zurückführen, den wir nun diskutieren wollen. (Die Herleitung des Satzes von Kolmogoroff aus dem Satz von Ionescu-Tulcea ist jedoch nicht völlig trivial).

4.2 Markoff Kerne und der Satz von Ionescu-Tulcea

Definition 4.8
*Es seien $(S_1, \mathcal{S}_1)$, $(S_2, \mathcal{S}_2)$ zwei messbare Räume. Ein **Markoffkern** K von $(S_1, \mathcal{S}_1)$ nach $(S_2, \mathcal{S}_2)$ ist eine Abbildung $K : S_1 \times \mathcal{S}_2 \to [0, 1]$ mit den folgenden zwei Eigenschaften:*

1. *Für alle $x \in S_1$ ist $K(x, \cdot)$ ein Wahrscheinlichkeitsmass auf $(S_2, \mathcal{S}_2)$.*

2. Für alle $A \in \mathcal{S}_2$ ist $K(\cdot, A)$ eine $\mathcal{S}_1$-messbare Funktion auf S_1.

Beispiel 4.9

a) $K(x, A) := \mu(A)$ für ein Wahrscheinlichkeitsmass μ auf $(S_2, \mathcal{S}_2)$. Hier hängt also der Kern gar nicht von x ab.

b) $K(x, A) := 1_A(f(x))$ für eine messbare Abbildung f von $(S_1, \mathcal{S}_1)$ nach $(S_2, \mathcal{S}_2)$.

c) Sei I eine höchstens abzählbare Menge, $(p_{ij})_{i,j \in I}$ eine stochastische Matrix und $(S_1, \mathcal{S}_1) = (S_2, \mathcal{S}_2) = (I, \mathcal{P}(I))$. Für $i \in I$, $A \subset I$ wird ein Markoffkern K durch

$$K(i, A) := \sum_{j \in A} p_{ij}$$

definiert.

d) Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $\mathcal{G}$ eine Teil-σ-Algebra. Eine reguläre bedingte Wahrscheinlichkeit von P gegeben $\mathcal{G}$ definiert einen Markoff-Kern von $(\Omega, \mathcal{G})$ nach $(\Omega, \mathcal{F})$. (siehe Definition 3.5).

Man stellt sich einen Kern am besten als eine Art "fuzzy" oder "verrauschte" Abbildung vor. Wir werden auch die folgende Notation verwenden, die diesen Aspekt betont: Ist K ein Markoffkern von $(S_1, \mathcal{S}_1)$ nach $(S_2, \mathcal{S}_2)$, so schreiben wir $K : (S_1, \mathcal{S}_1) \rightsquigarrow (S_2, \mathcal{S}_2)$ oder kurz $K : S_1 \rightsquigarrow S_2$. (Vorsicht: Diese Notation ist in der wahrscheinlichkeitstheoretischen Literatur nicht üblich). Wichtig ist, dass sich mit Kernen der Begriff des Produktmasses verallgemeinern lässt:

Definition 4.10

Sei μ ein Wahrscheinlichkeitsmass auf $(S_1, \mathcal{S}_1)$ und $K : (S_1, \mathcal{S}_1) \rightsquigarrow (S_2, \mathcal{S}_2)$. Dann ist das Wahrscheinlichkeitsmass $\mu \otimes K$ auf $\mathcal{S}_1 \otimes \mathcal{S}_2$ wie folgt definiert:

$$(\mu \otimes K)(A) = \int \mu(d\omega_1) K(\omega_1, A_{\omega_1}) \quad \text{für } A \in \mathcal{S}_1 \otimes \mathcal{S}_2, \tag{4.3}$$

wobei $A_{\omega_1} = \{\, \omega_2 \in S_2 : (\omega_1, \omega_2) \in A \,\}$ der ω_1-Schnitt der Menge A ist.

Damit dies eine sinnvolle Definition ist, müssen einige Punkte geprüft werden:

- Nach Satz 1.52 ist $A_{\omega_1} \in \mathcal{S}_2$, und demzufolge ist $K(\omega_1, A_{\omega_1})$ für jedes $\omega_1 \in S_1$ definiert.

- Nun muss nachgewiesen werden, dass $\omega_1 \mapsto K(\omega_1, A_{\omega_1})$ eine $\mathcal{S}_1$-messbare Funktion ist. Dies sieht man wie folgt ein:

$$\mathcal{D} = \{\, A \in \mathcal{S}_1 \otimes \mathcal{S}_2 : \omega_1 \mapsto K(\omega_1, A_{\omega_1}) \text{ ist } \mathcal{S}_1\text{−messbar} \,\}$$

ist ein Dynkinsystem (einfache Übungsaufgabe) und enthält die Mengen der Form $A_1 \times A_2$ mit $A_i \in \mathrm{S}_i$ für $i = 1, 2$, denn es gilt

$$K(\omega_1, (A_1 \times A_2)_{\omega_1}) = 1_{A_1}(\omega_1) K(\omega_1, A_2).$$

Diese Produktmengen bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{S}_1 \otimes \mathcal{S}_2$. Nach Satz 1.9 folgt $\mathcal{D} = \mathcal{S}_1 \otimes \mathcal{S}_2$. Damit ist das Integral auf der rechten Seite von (4.3) eindeutig definiert und damit ist $(\mu \otimes K)(A)$ für jedes $A \in \mathcal{S}_1 \otimes \mathcal{S}_2$ definiert.

- Mit Hilfe des Satzes von der majorisierten Konvergenz (Satz von Lebesgue) zeigt man sofort, dass $\mu \otimes K$ auch σ-additiv und somit ein Wahrscheinlichkeitsmass ist.

Ist $K : (S_1, \mathcal{S}_1) \rightsquigarrow (S_2, \mathcal{S}_2)$ gemäss Beispiel 4.9 a) durch ein Wahrscheinlichkeitsmass ν auf $(S_2, \mathcal{S}_2)$ gegeben, so ist $\mu \otimes K$ natürlich einfach das Produktmass $\mu \otimes \nu$.

Wir betrachten die beiden Randverteilungen von $\mu \otimes K$ auf $(S_1, \mathcal{S}_1)$ bzw. $(S_2, \mathcal{S}_2)$. Sind $\pi_i : S_1 \times S_2 \to S_i$ die Projektionen, so ist

$$\left((\mu \otimes K)\, \pi_1^{-1}\right)(A) = (\mu \otimes K)\,(A \times S_2) = \mu\,(A)\,,\ A \in \mathcal{S}_1,$$

d.h.

$$(\mu \otimes K)\, \pi_1^{-1} = \mu. \tag{4.4}$$

Die zweiten Randverteilung ist gegeben durch

$$\left((\mu \otimes K)\, \pi_2^{-1}\right)(A) = (\mu \otimes K)\,(S_1 \times A) = \int \mu\,(dx)\, K\,(x, A)\,,\ A \in \mathcal{S}_2. \tag{4.5}$$

Diese Randverteilung auf $(S_2, \mathcal{S}_2)$ bezeichnet man meist mit μK.

Wir benötigen noch eine Verallgemeinerung des Satzes von Fubini-Tonelli:

Satz 4.11
Sei μ ein Wahrscheinlichkeitsmass auf $(S_1, \mathcal{S}_1)$ und $K : (S_1, \mathcal{S}_1) \rightsquigarrow (S_2, \mathcal{S}_2)$. Ferner sei $f : S_1 \times S_2 \to \mathbb{R}$ eine messbare Funktion.

a) *Ist $f \geq 0$ so gilt*

$$\int f\, d\,(\mu \otimes K) = \int \left(\int f\,(x,y)\, K\,(x, dy) \right) \mu\,(dx)\,. \tag{4.6}$$

b) *Ist $f \in \mathcal{L}_1\,(\mu \otimes K)$, so ist für μ-fast alle $x \in S_1$ die Abbildung $S_2 \ni y \to f\,(x,y)$ integrierbar bezüglich $K\,(x,\cdot)$. Ferner ist $\int f\,(x,y)\, K\,(x,dy)$ als Funktion von x μ-integrierbar, und es gilt (4.6).*

Beweis. Der Beweis folgt nach dem üblichen Schema: Für $f = 1_A$, $A \in \mathcal{S}_1 \otimes \mathcal{S}_2$, ergibt sich (4.6) aus der Definition und a) folgt dann gemäss Satz 1.37. b) folgt durch Zerlegung in Positiv- und Negativteil. Die Details seien dem Leser überlassen. ■

Nun zurück zu den endlichdimensionalen Verteilungen und dem Satz von Ionescu-Tulcea. Sei $(S, \mathcal{S})$ ein beliebiger messbarer Raum, und seien μ ein Wahrscheinlichkeitsmass auf $(S, \mathcal{S})$ (das "Startmass") und für jedes $n \in \mathbb{N}$ sei $K_n : (S^n, \mathcal{S}^n) \rightsquigarrow (S, \mathcal{S})$. Durch

$$\begin{aligned} Q^{(1)} &= \mu, \qquad (4.7)\\ Q^{(n)} &= Q^{(n-1)} \otimes K_{n-1} \text{ für } n \geq 2 \end{aligned}$$

erhalten wir ein Wahrscheinlichkeitsmass $Q^{(n)}$ auf $(S^n, \mathcal{S}^n)$ für jedes $n \in \mathbb{N}$. Nach (4.4) ist die Folge $\{Q^{(n)}\}_{n\in\mathbb{N}}$ dieser Wahrscheinlichkeitsmasse verträglich, d.h. sie erfüllt (4.2).

Satz 4.12 (Satz von Ionescu-Tulcea)
In der oben beschriebenen Situation existiert stets ein Wahrscheinlichkeitsmass Q auf $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}})$, das (4.1) erfüllt.

5 Stationäre Prozesse, Ergodensatz

Definition 5.1
*$(\Omega, \mathcal{F}, P)$ sei ein Wahrscheinlichkeitsraum. Eine $\mathcal{F}$-$\mathcal{F}$-messbare Abbildung $T : \Omega \to \Omega$ heisst **masserhaltende Transformation**, wenn $PT^{-1} = P$ gilt. Wir sagen dafür auch, dass P T-invariant (oder invariant unter T) ist.*

Masserhaltende Transformationen hängens eng mit stationären stochastischen Prozessen zusammen.

Definition 5.2
*Ein stochastischer Prozess $X = (X_1, X_2, \ldots)$ heisst **stationär,** wenn X dieselbe Verteilung wie der Prozess $(X_2, X_3, \ldots)$ hat.*

Lemma 5.3
Es seien Y eine Zufallsgrösse und T eine masserhaltende Transformation auf $(\Omega, \mathcal{F}, P)$. Definiere $X_1 := Y$ und $X_{n+1} = X_n \circ T$ für $n \in \mathbb{N}$. Dann ist der Prozess $\{X_n\}_{n\in\mathbb{N}}$ stationär.

Beweis. Seien $n \in \mathbb{N}$ und $t_1, \ldots, t_n \in \mathbb{R}$. Dann gilt:

$$\begin{aligned}
P(X_1 \le t_1, \ldots, X_n \le t_n) &= PT^{-1}(X_1 \le t_1, \ldots, X_n \le t_n) \\
&= P(X_1 \circ T \le t_1, \ldots, X_n \circ T \le t_n) \\
&= P(X_2 \le t_1, \ldots, X_{n+1} \le t_n).
\end{aligned}$$

Das Lemma folgt nun aus Proposition 4.5. ■

Es gilt auch, dass umgekehrt jeder stationäre Prozess in gewisser Weise so beschrieben werden kann. Genauer: Zu jedem auf $(\Omega, \mathcal{F}, P)$ definierten, stationären Prozess $X = \{X_n\}_{n\in\mathbb{N}}$ existiert ein mit Hilfe einer masserhaltenden Transformation beschriebener Prozess, der dieselbe Verteilung hat. Dies sieht man wie folgt: Wie wir im letzten Kapitel gesehen haben, ist ein stochastischer Prozess X nichts anderes als eine $\mathcal{F}$-$\mathcal{B}^{\mathbb{N}}$-messbare Abbildung $\Omega \to \mathbb{R}^{\mathbb{N}}$. Sei $P' = PX^{-1}$ die Verteilung von X. Die sogenannte Verschiebungsabbildung $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, definiert durch $T((x_1, x_2, \ldots)) = (x_2, x_3, \ldots)$, ist masserhaltend auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$, wenn X stationär ist: Für $A \in \mathcal{B}^{\mathbb{N}}$ gilt nämlich $P'(A) = P(X \in A) = P(T \circ X \in A) = PX^{-1}T^{-1}(A) = P'T^{-1}(A)$ wegen der Stationarität von X. Natürlich ist P' auch die Verteilung der Folge der Projektionen $\pi_1, \pi_2, \dots : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, denn $\Pi = \{\pi_n\}_{n\in\mathbb{N}}$ ist die identische Abbildung auf $\mathbb{R}^{\mathbb{N}}$. Offensichtlich gilt $\pi_n = \pi_{n-1} \circ T$ für $n \ge 2$. Der Prozess $\{\pi_n\}_{n\in\mathbb{N}}$ auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ ist also von der in Lemma 5.3 definierten Form und besitzt dieselbe Verteilung wie X.

Wir werden die in Definition beschriebene Situation zugrundelegen, was bequemer ist. Alle gewonnenen Resultate, insbesondere der Ergodensatz (Satz **??**), sind unmittelbar auf stationäre Folgen von Zufallsgrössen übertragbar.

Bemerkung 5.4
Eine messbare Abbildung $T : \Omega \to \Omega$ ist genau dann eine masserhaltende Transformation, wenn $PT^{-1}(A) = P(A)$ für alle A aus einem durchschnittstabilen Erzeugendensystem von $\mathcal{F}$ gilt. Dies folgt sofort aus Satz 1.15.

Beispiel 5.5

1. Eine Folge von unabhängigen identisch verteilten Zufallsgrössen ist ein stationärer Prozess. Anders ausgedrückt: Ist μ ein Wahrscheinlichkeitsmass auf $(\mathbb{R}, \mathcal{B})$, so ist das unendliche Produktmass $\mu^{\mathbb{N}}$ invariant unter der Verschiebungsabbildung $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$.

2. Sei $\Omega = D := \{ z \in \mathbb{C} : |z| = 1 \}$. Wir können D als Kreis mit Radius 1 in $\mathbb{R}^2$ auffassen. Überträgt man von dem Intervall $[0, 2\pi)$ die Borel-σ-Algebra und die gleichförmige Verteilung (=Lebesgue-Mass/2π) auf D, so erhält man einen Wahrscheinlichkeitsraum $(D, \mathcal{B}_D, \lambda_D)$. Für $c \in D$ sei der Multiplikations- oder auch Drehoperator $T_c : D \to D$ definiert durch $T_c(\omega) := c\omega$. Ist $A \subset D$ ein Intervall, so gilt offensichtlich $\lambda_D T_c^{-1}(A) = \lambda_D(A)$. Aus Bemerkung 5.4 folgt, dass T_c masserhaltend ist.

3. Auf $([0,1), \mathcal{B}_{[0,1)}, \lambda)$ betrachten wir die Transformation $T : [0,1) \to [0,1)$, die gegeben ist durch
$$T(\omega) := \begin{cases} 2\omega & \text{für } \omega \in [0, 1/2), \\ 2\omega - 1 & \text{für } \omega \in [1/2, 1). \end{cases}$$
Für $0 \le a < b \le 1$ ist $T^{-1}([a,b)) = [\frac{a}{2}, \frac{b}{2}) \cup [\frac{a+1}{2}, \frac{b+1}{2})$, also folgt $PT^{-1}([a,b)) = b - a = P([a,b))$. Die Mengen der Form $[a,b)$ mit $0 \le a < b \le 1$ bilden zusammen mit $\emptyset$ ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}_{[0,1)}$. Es folgt daher $PT^{-1} = P$.

Definition 5.6

1. *Für eine Abbildung $T : \Omega \to \Omega$ heisst eine Teilmenge A von Ω T-**invariant** (oder invariant unter T), wenn $T^{-1}(A) = A$ gilt.*

2. *Eine masserhaltende Abbildung T auf $(\Omega, \mathcal{F}, P)$ heisst **ergodisch**, wenn für jede T-invariante Menge $A \in \mathcal{F}$ gilt: $P(A) \in \{0, 1\}$.*

Bemerkung 5.7
Die Familie der T-invarianten Mengen aus $\mathcal{F}$ ist offensichtlich eine Teil-σ-Algebra von $\mathcal{F}$. Diese wird meist mit $\mathcal{J}$ bezeichnet. Ist T ergodisch, so sagt man auch, dass $\mathcal{J}$ *P-trivial* ist. Nach Lemma 2.39 ist in diesem Fall jede $\mathcal{J}$-messbare Zufallsgrösse fast sicher konstant.

Für das folgende Lemma erinnern wir an die Diskussion nach Lemma 5.3:

Lemma 5.8
Es sei $X = \{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsgrössen auf $(\Omega, \mathcal{F}, P)$ mit Verteilung $P' := PX^{-1}$. Dann ist die Verschiebungsabbildung $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, $\{x_n\}_{n\in\mathbb{N}} \mapsto \{x_{n+1}\}_{n\in\mathbb{N}}$, ergodisch auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$.

Beweis. Für eine T-invariante Menge $A \in \mathcal{B}^{\mathbb{N}}$ gilt für alle $n \in \mathbb{N}$ auch $A = T^{-n}(A) := (T^n)^{-1}(A) = \{\, \{x_k\}_{k\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid \{x_{n+k}\}_{k\in\mathbb{N}} \in A \,\}$. Demzufolge liegt die Menge $X^{-1}(A) = \{\, \omega : (X_{n+1}(\omega), X_{n+2}(\omega), \dots) \in A \,\}$ in $\check{\mathcal{F}}_{n+1} := \bigvee_{k=n+1}^{\infty} X_k^{-1}(\mathcal{B})$. Da dies für alle $n \in \mathbb{N}$ gilt, ist $X^{-1}(A) \in \mathcal{T}_\infty$ in der Notation von Satz 2.38, und aus diesem Satz folgt $P'(A) = P(X^{-1}(A)) \in \{0, 1\}$. ■

Bemerkung 5.9
Der obige Beweisgang zeigt folgendes: Die σ-Algebra $\mathcal{J}$ der verschiebungsinvarianten messbaren Mengen in $\mathbb{R}^{\mathbb{N}}$ ist eine Teil-σ-Algebra der terminalen σ-Algebra $\mathcal{T}_\infty = \bigcap_{n=1}^{\infty} \sigma(\pi_k : k \geq n)$. Es gilt jedoch keinesfalls $\mathcal{J} = \mathcal{T}_\infty$, und ohne Beweis sei bemerkt, dass es ergodische Masse auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ gibt, bezüglich denen $\mathcal{T}_\infty$ nicht trivial ist.

Wir untersuchen nun den Multiplikations- beziehungsweise Drehoperator T_c aus Beispiel) auf Ergodizität:

Satz 5.10
Es seien $D = \{\, z \in \mathbb{C} : |z| = 1 \,\}$, $\mathcal{B}_D$ die Borel-σ-Algebra auf D und λ_D das normierte Lebesgue-Mass auf $(D, \mathcal{B}_D)$ sowie $T_c : D \to D$ definiert durch $T_c(\omega) := c\omega$ für $c, \omega \in D$. Dann ist T_c genau dann ergodisch auf $(D, \mathcal{B}_D, \lambda_D)$, wenn c keine Einheitswurzel ist.

Beweis. Wir schreiben $T = T_c$. Ist c eine Einheitswurzel, das heisst existiert ein $n \in \mathbb{N}$ mit $c^n = 1$, so ist T^n die Identität. Somit ist für jedes $A \in \mathcal{B}_D$ die Menge $A \cup T^{-1}A \cup \dots \cup T^{-n+1}A$ invariant unter T. Wir wählen ein $A \in \mathcal{B}_D$ mit $0 < \lambda_D(A) < 1/n$ und erhalten

$$0 < \lambda_D(A \cup \dots \cup T^{-n+1}A) \leq \sum_{k=0}^{n-1} \lambda_D(T^{-k}A) = n\lambda_D(A) < 1.$$

Somit ist T nicht ergodisch. Der Beweis der Umkehrung erfordert einige Vorbereitungen:

Lemma 5.11
Sei $c \in D$ keine Einheitswurzel. Dann liegt $\{\, c^n : n \in \mathbb{N}_0 \,\}$ dicht in D.

Da die c^n, $n \in \mathbb{N}$, alle verschieden sind, besitzt diese Folge mindestens einen Häufungspunkt $\omega_0 \in D$. Seien $\varepsilon > 0$ und $m > n$ zwei natürliche Zahlen mit $|c^n - \omega_0| < \varepsilon$ und $|c^m - \omega_0| < \varepsilon$. Dann gilt $|c^{m-n} - 1| \in (0, 2\varepsilon)$. Daraus folgt, dass für jedes $\omega \in D$ ein $k \in \mathbb{N}$ existiert mit $|\omega - c^{k(m-n)}| < 2\varepsilon$. Da $\varepsilon > 0$ beliebig ist, folgt die Behauptung. ■

Wir benötigen noch das folgende masstheoretische Resultat (Dies war eine Übungsaufgabe. Siehe auch Satz 5.7 in dem Buch von H. Bauer):

Lemma 5.12
Es sei $(\Omega, \mathcal{F}, P)$ ein beliebiger Wahrscheinlichkeitsraum und $\mathcal{A}$ eine Algebra mit $\mathcal{F} = \sigma(\mathcal{A})$. Zu $\varepsilon > 0$ und $A \in \mathcal{F}$ existiert $B \in \mathcal{A}$ mit $P(A \triangle B) < \varepsilon$.

Fortsetzung des Beweises von 5.10. Sei c keine Einheitswurzel. Sei $\mathcal{A}$ die Familie der endlichen Vereinigungen paarweise disjunkter Intervalle in D, dann ist $\mathcal{A}$ eine Algebra. Ist A invariant unter T_c mit $\lambda_D(A) > 0$ und $\varepsilon \in (0, \frac{1}{2})$, so existiert nach Lemma **??** eine endliche Vereinigung $B = \bigcup_{i=1}^n I_i$ disjunkter Intervalle $I_1, \ldots, I_n \subset D$ mit $\lambda_D(A \triangle B) \leq \varepsilon \lambda_D(A)$. Wir können annehmen, dass $\lambda_D(I_i) \leq \varepsilon$ ist für $i = 1, \ldots, n$.

Wir haben

$$\lambda_D(A \triangle B) \leq \varepsilon \lambda_D(A) \leq 2\varepsilon(1-\varepsilon)\lambda_D(A) \leq 2\varepsilon(\lambda_D(A) - \lambda_D(A \triangle B)) \leq 2\varepsilon \lambda_D(B),$$

und daraus folgt

$$\sum_{i=1}^n \lambda_D(A \cap I_i) = \lambda_D(A \cap B) \geq \lambda_D(B) - \lambda_D(A \triangle B) \geq (1-2\varepsilon)\lambda_D(B) = (1-2\varepsilon)\sum_{i=1}^n \lambda_D(I_i).$$

Mindestens eines der I_i — kurz mit I bezeichnet — erfüllt also die Ungleichung $\lambda_D(A \cap I) \geq (1-2\varepsilon)\lambda_D(I)$. Nach Lemma **??** existieren ein $k \in \mathbb{N}$ und $n_1, \ldots, n_k \in \mathbb{N}$ derart, dass die Intervalle $T^{-n_1}I, T^{-n_2}I, \ldots, T^{-n_k}I$ paarweise disjunkt sind und D bis auf eine Menge von kleinerem Mass als 2ε ausfüllen. Wegen der T-Invarianz von A und λ_D gilt für $j = 1, \ldots, k$:

$$\begin{aligned} \lambda_D(A \cap T^{-n_j}I) &= \lambda_D(T^{-n_j}A \cap T^{-n_j}I) = \lambda_D T^{-n_j}(A \cap I) \\ &= \lambda_D(A \cap I) \geq (1-2\varepsilon)\lambda_D(I) = (1-2\varepsilon)\lambda_D(T^{-n_j}I), \end{aligned}$$

und dies führt zu

$$\lambda_D(A) \geq \sum_{j=1}^k \lambda_D(A \cap T^{-n_j}I) \geq (1-2\varepsilon)\sum_{j=1}^k \lambda_D(T^{-n_j}I) \geq (1-2\varepsilon)^2.$$

Daraus folgt $\lambda_D(A) = 1$. ■

Die Transformation aus Beispiel) ist ebenfalls ergodisch. Um das einzusehen, führen wir den folgenden Begriff ein:

Definition 5.13
Eine masserhaltende Transformation T auf $(\Omega, \mathcal{F}, P)$ heisst mischend, wenn für alle $A, B \in \mathcal{F}$ gilt:

$$\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A)P(B).$$

Lemma 5.14
Jede mischende Transformation ist ergodisch.

Beweis. Ist $A \in \mathcal{F}$ eine T-invariante Menge, so gilt

$$P(A) = P(A \cap A) = P(A \cap T^{-n}A) \to P(A)P(A) \quad \text{für } n \to \infty.$$

Daraus folgt sofort $P(A) = P(A)^2$, das heisst $P(A) \in \{0, 1\}$. ■

Lemma 5.15
Es sei $\mathcal{A}$ eine Algebra, die $\mathcal{F}$ erzeugt. Falls die Gleichung für alle $A, B \in \mathcal{A}$ gilt, so ist T mischend.

Beweis. Seien $A, B \in \mathcal{F}$ und $\varepsilon > 0$. Nach Lemma 5.12 existieren $A_0, B_0 \in \mathcal{A}$ mit $P(A \triangle A_0) < \varepsilon$ und $P(B \triangle B_0) < \varepsilon$. Daraus ergibt sich

$$\begin{aligned} |P(A \cap T^{-n}B) - P(A_0 \cap T^{-n}B_0)| &\leq P(A \triangle A_0) + P(T^{-n}B \triangle T^{-n}B_0) \\ &= P(A \triangle A_0) + P(B \triangle B_0) < 2\varepsilon. \end{aligned}$$

Die Folge $\{P(A_0 \cap T^{-n}B_0)\}_{n\in\mathbb{N}}$ konvergiert gegen $P(A_0)P(B_0)$, und es gilt

$$\begin{array}{rcl} |P(A_0) - P(A)| & \leq & P(A \triangle A_0) < \varepsilon, \\ |P(B_0) - P(B)| & < & \varepsilon. \end{array}$$

Da $\varepsilon > 0$ beliebig ist, folgt $\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A)P(B)$. ■

Nicht jede ergodische Transformation ist mischend. Man sieht leicht ein, dass der Drehoperator T_c auf D (siehe Beispiel) für kein c mischend ist.

Satz 5.16
Die in Beispiel 5.5 c) definierte Transformation T auf $([0,1), \mathcal{B}_{[0,1)}, \lambda)$ ist mischend.

Beweis. Für jede Menge $A \subset [0,1)$ sind $T^{-1}A \cap [0, \frac{1}{2})$ und $T^{-1}A \cap [\frac{1}{2}, 1)$ nur um 1/2 gegeneinander verschobene Mengen. Daraus folgt:

$$\begin{aligned} P\left(T^{-1}A \cap [1/2, 1)\right) = P\left(T^{-1}A \cap [0, 1/2)\right) &= \frac{1}{2} P\left(T^{-1}A\right) \\ &= P(A)P\left([0, 1/2)\right). \end{aligned}$$

Völlig analog folgt für $n \in \mathbb{N}$ und $k = 0, 1, \ldots, 2^n - 1$:

$$P(T^{-n}A \cap [k2^{-n}, (k+1)2^{-n})) = P(A)P([k2^{-n}, (k+1)2^{-n})).$$

Ist $\mathcal{A}_0$ die Familie der Intervalle der Form $[k2^{-n}, (k+1)2^{-n})$ mit $n \in \mathbb{N}$ und $0 \leq k \leq 2^n - 1$, so gilt also

$$\lim_{m\to\infty} P(T^{-m}A \cap I) = P(A)P(I), \ \forall I \in \mathcal{A}_0.$$

Diese Gleichung bleibt auch richtig, wenn I eine endliche Vereinigung paarweise disjunkter Intervalle dieser Form ist. Diese Intervallfiguren bilden jedoch eine Algebra, die die Borel-σ-Algebra auf $[0,1)$ erzeugt. Der Satz folgt somit aus Lemma 5.15. ■

Wir wollen nun nachweisen, dass eine stationäre, irreduzible, positiv rekurrente und aperiodische Markoffkette mischend ist. Sei I eine abzählbare Menge und $\mathbb{P} = (p_{ij})_{i,j\in I}$ eine irreduzible, positiv rekurrente und aperiodische stochastische Matrix. Es existiert dann eine eindeutige stationäre Wahrscheinlichkeitsverteilung auf I, das heisst ein Wahrscheinlichkeitsmass π mit $\sum_{i\in I} \pi(i)p_{ij} = \pi(j)$ für alle j. (Siehe Einführung in die Stochastik). Aus Satz 4.12 folgt, dass ein Wahrscheinlichkeitsmass P auf $(I^{\mathbb{N}}, \mathcal{P}(I)^{\mathbb{N}})$ existiert, sodass die Folge der Projektionen $X_n : I^{\mathbb{N}} \to I$ eine Markoff-Kette ist, die π als Startverteilung und $\mathbb{P}$ als Matrix der Übergangswahrscheinlichkeiten hat. Das π stationär ist, folgt sofort, dass P invariant unter dem Verscheibungsoperator $T : I^{\mathbb{N}} \to I^{\mathbb{N}}$ ist.

Satz 5.17
Es sei I eine abzählbare Menge und P die Verteilung einer stationären irreduziblen, positiv rekurrenten und aperiodischen Markoffkette auf $(I^{\mathbb{N}}, \mathcal{P}(I)^{\mathbb{N}})$. Dann ist $T : I^{\mathbb{N}} \to I^{\mathbb{N}}$, definiert durch $T(\{x_n\}_{n\in\mathbb{N}}) := \{x_{n+1}\}_{n\in\mathbb{N}}$, mischend.

Beweis. Für $k \in \mathbb{N}$ sei $X_k : I^{\mathbb{N}} \to I$ die k-te Projektion, definiert durch $X_k(\{i_n\}_{n\in\mathbb{N}}) := i_k$. Sei $\mathbb{A}$ die Algebra $\bigcup_{m=1}^{\infty} \sigma(X_1, \ldots, X_m)$. Beachte, dass gilt: $\sigma(X_1, \ldots, X_m) = X^{(m)-1}(\mathcal{P}(I^m)) = \{\, X^{(m)-1}(C) : C \subset I^m \,\}$, wobei $X^{(m)} = (X_1, \ldots, X_m) : I^{\mathbb{N}} \to I^m$ die Projektion auf die ersten m Faktoren bezeichnet. Für jedes $A \in \mathcal{A}$ existiert also ein $m \in \mathbb{N}$ und ein $C \subset I^m$ mit $A = X^{(m)-1}(C) = \{\, i \in I^{\mathbb{N}} : X^{(m)}(i) \in C \,\}$.

Da $\sigma(\mathcal{A}) = \mathcal{P}(I)^{\mathbb{N}}$ ist, genügt es nach Lemma 5.15 nachzuweisen, dass für $A, B \in \mathcal{A}$ gilt:

$$\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A)P(B).$$

Dies ist äquivalent zu

$$\lim_{n\to\infty} P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D) = P(X^{(m)} \in C)P(X^{(k)} \in D) \tag{5.1}$$

für $k, m \in \mathbb{N}$ und $C \subset I^m$ sowie $D \subset I^k$. Wir überlegen uns zunächst, dass es genügt, die Gleichung für endliche Mengen C und D nachzuweisen. Ist nämlich D eine beliebige Teilmenge von I^k, so existiert zu $\varepsilon > 0$ eine endliche Teilmenge D' von D mit

$$|P(X^{(k)} \in D) - P(X^{(k)} \in D')| \le \varepsilon.$$

Wegen der T-Invarianz von P folgt

$$\begin{aligned}
&|P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D) - P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D')| \\
&\quad \le |P(T^n(X^{(k)}) \in D) - P(T^n(X^{(k)}) \in D')| \\
&\quad \le |P(X^{(k)} \in D) - P(X^{(k)} \in D')| \le \varepsilon.
\end{aligned}$$

Da $\varepsilon > 0$ beliebig ist, genügt es also, die Gleichung (5.1) für endliches D zu zeigen. Für C ist das Argument dasselbe. Seien also $k, m \in \mathbb{N}$ und $C \subset I^m$, $D \subset I^k$ endliche Mengen. Für $n > m$ gilt:

$$\begin{aligned}
&P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D) \\
&\quad = P((X_1, \ldots, X_m) \in C,\ (X_{n+1}, \ldots, X_{n+k}) \in D) \\
&\quad = \sum_{(i_1,\ldots,i_m)\in C} \sum_{(j_1,\ldots,j_k)\in D} P(X_1 = i_1, \ldots, X_m = i_m, X_{n+1} = j_1, \ldots,\ X_{n+k} = j_k) \\
&\quad = \sum_{(i_1,\ldots,i_m)\in C} \sum_{(j_1,\ldots,j_k)\in D} \pi(i_1) p_{i_1 i_2} \cdots p_{i_{m-1} i_m} p^{(n-m)}_{i_m j_1} p_{j_1 j_2} \cdots p_{j_{k-1} j_k}.
\end{aligned}$$

Nach Satz 8.52 der „Einführung in die Stochastik“ gilt $\lim_{n\to\infty} p^{(n-m)}_{i_m j_1} = \pi(j_1)$. Da die obigen Summen endlich sind, können wir die Summation mit dem Limes $n \to \infty$

vertauschen und erhalten

$$\begin{aligned}
&\lim_{n\to\infty} P(X^{(m)} \in C,\ T^n(X^{(k)}) \in D) \\
&= \sum_{(i_1,\dots,i_m)\in C} \pi(i_1)p_{i_1 i_2}\cdots p_{i_{m-1}i_m} \sum_{(j_1,\dots,j_k)\in D} \pi(j_1)p_{j_1 j_2}\cdots p_{j_{k-1}j_k} \\
&= P(X^{(m)} \in C)P(X^{(k)} \in D).
\end{aligned}$$

■

Satz 5.18 (Ergodensatz)
Sei T eine masserhaltende Transformation auf $(\Omega, \mathcal{F}, P)$, und es sei $X \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$. Dann konvergiert $S_n := \frac{1}{n}\sum_{j=0}^{n-1} X \circ T^j$ fast sicher für $n \to \infty$ gegen $E\,(X|\,\mathcal{J})$ eine $\mathcal{J}$-messbare Zufallsgrösse Y, für die $\int X\,dP = \int Y\,dP$ gilt.

Bemerkung 5.19

1. Ist T ergodisch, so folgt $E\,(X|\,\mathcal{J}) = EX$ fast sicher. Jede $\mathcal{J}$-m.b. Zufallsgrösse ist in diesem Fall wegen Lemma 2.39 f.s. konstant. Wegen
$$\int E\,(X|\,\mathcal{J})\,dP = \int X dP = EX$$
folgt daher die Behauptung.
2. Es gilt auch Konvergenz im 1. Mittel. Dies soll hier nicht bewiesen werden. Ein einfacher funktionalanalytischer Beweis für die Konvergenz im 1. Mittel findet sich etwa in dem Buch 'Ergodic Theory and Information' von P˙ Billingsley.

Beweis von Satz 5.18. Wir zeigen zunächste, die folgende etwas bescheidenere Aussage

$$Y \stackrel{\text{def}}{=} \lim_{n\to\infty} S_n$$

existiert f.s., ist $\mathcal{J}$-m.b. und derfüllt

$$\int X\,dP = \int Y\,dP \tag{5.2}$$

Wir zeigen anschliessend, dass daraus die Behauptung des Satzes folgt.

Wegen der Zerlegung $X = X^+ - X^-$ reicht es aus, diese Behauptung für nichtnegative Zufallsgrössen zu beweisen. Sei also $X \geq 0$.

Man definiert $\overline{X}(\omega) := \limsup_{n\to\infty} S_n(\omega)$ sowie $\underline{X}(\omega) := \liminf_{n\to\infty} S_n(\omega)$. Offenbar gelten $\underline{X} \circ T = \underline{X}$ und $\overline{X} \circ T = \overline{X}$. Daraus folgt sofort, dass $\underline{X}$ und $\overline{X}$ $\mathcal{J}$-messbar sind.

Man zeigt nun, dass

$$\int \overline{X}\,dP \leq \int X\,dP \leq \int \underline{X}\,dP \tag{5.3}$$

gilt. Wegen $\underline{X} \leq \overline{X}$ folgt daraus $\underline{X} = \overline{X}$ fast sicher, das heisst, $(S_n)_n$ konvergiert fast sicher gegen eine $\mathcal{J}$-messbare Zufallsgrösse mit den gewünschten Eigenschaften.

Wir beweisen nun die 1. Ungleichung in (5.3). Die Grundidee des Beweises ist sehr einfach: Man schaut auf die „Zeitpunkte“ $0 = n_0 < n_1 < n_2 < n_3 < \dots$, für die der durchschnittliche Zuwachs $\frac{1}{n_{j+1}-n_j}[X(T^{n_j}\omega) + \dots + X(T^{n_{j+1}-1}\omega)]$ dem $\limsup$ „nahe“ kommt. Es ergeben sich zwei technische Probleme: Erstens kann man nicht von vorneherein ausschliessen, dass $\overline{X} = \infty$ ist. Um diese Schwierigkeit zu umgehen, schneidet man $\overline{X}$ ab. Für $M \in (0,\infty)$ sei $\overline{X}_M = \min(\overline{X}, M)$. Sei $\varepsilon > 0$ beliebig. Für $\omega \in \Omega$ sei $n(\omega) \in \mathbb{N}$ die kleinste Zahl $k \in \mathbb{N}$ mit $S_k(\omega) \geq \overline{X}_M(\omega) - \varepsilon$. Die Abbildung $n : \Omega \to \mathbb{N}$ ist $\mathcal{F}$-$\mathcal{P}(\mathbb{N})$-messbar, denn für $k \in \mathbb{N}$ gilt

$$\{\,\omega : n(\omega) = k\,\} = \bigcap_{j=1}^{k-1} \{\, S_j < \overline{X}_M - \varepsilon \,\} \cap \{\, S_k \geq \overline{X}_M - \varepsilon \,\}.$$

Die zweite technische Schwierigkeit besteht darin, dass n als Funktion von ω nicht beschränkt zu sein braucht. Allerdings gilt natürlich $\lim_{N\to\infty} P(\{\,\omega : n(\omega) > N\,\}) = 0$. Es existiert also ein $N_\varepsilon \in \mathbb{N}$ mit

$$P(\{\,\omega : n(\omega) > N_\varepsilon\,\}) \leq \varepsilon. \tag{5.4}$$

Wir stutzen nun auch $n(\omega)$ und X zurecht: Mit $A := \{\,\omega : n(\omega) \leq N_\varepsilon\,\}$ seien

$$\begin{aligned} \widetilde{X}(\omega) \quad &: = \begin{cases} X(\omega) & \omega \in A, \\ M & \omega \notin A \end{cases} \\ \widetilde{n}(\omega) \quad &: = \begin{cases} n(\omega) & \omega \in A, \\ 1 & \omega \notin A. \end{cases} \end{aligned}$$

Gilt $X(\omega) > M$, so ist $n(\omega) = 1$ (wegen $S_1(\omega) = X(\omega)$), das heisst, es gilt $\omega \in A$. Es folgt also

$$X(\omega) \leq \widetilde{X}(\omega) \quad \text{für alle } \omega. \tag{5.5}$$

Für $\omega \notin A$ gilt wegen $\widetilde{n}(\omega) = 1$ und $\widetilde{X}(\omega) = M$ die Abschätzung

$$\frac{1}{\widetilde{n}(\omega)} \sum_{j=0}^{\widetilde{n}(\omega)-1} \widetilde{X}(T^j\omega) \geq \overline{X}_M(\omega) - \varepsilon. \tag{5.6}$$

Diese Ungleichung gilt auch für $\omega \in A$, denn dann ist $\widetilde{n}(\omega) = n(\omega)$, und wegen (5.5) ist die linke Seite von (5.6) nicht kleiner als $S_{n(\omega)}(\omega)$, also auch nicht kleiner als $\overline{X}_M(\omega) - \varepsilon$ nach der Definition von $n(\omega)$.

Wegen (5.4) folgt

$$\int \widetilde{X}\, dP = \int_A \widetilde{X}\, dP + \int_{A^c} \widetilde{X}\, dP \leq \int X\, dP + M\varepsilon. \tag{5.7}$$

Man definiert nun rekursiv:

$$\begin{aligned} &n_0(\omega) := 0; \\ &n_1(\omega) := \widetilde{n}(\omega); \\ &n_k(\omega) = n_{k-1}(\omega) + \widetilde{n}(T^{n_{k-1}(\omega)}(\omega)) \quad \text{für } k \in \mathbb{N} \setminus \{1\}. \end{aligned}$$

Für $m \in \mathbb{N}$ sei $K(\omega)$ die grösste Zahl $k \in \mathbb{N}$ mit $n_k(\omega) \leq m$. (Der Index $K(\omega)$ hängt natürlich von m ab.) Wegen $\widetilde{n}(\omega) \leq N_\varepsilon$ folgt $m - n_{K(\omega)}(\omega) \leq N_\varepsilon$.

Es folgt

$$\begin{aligned}\sum_{j=0}^{m-1} \widetilde{X}(T^j\omega) &\geq \sum_{j=0}^{n_{K(\omega)}(\omega)-1} \widetilde{X}(T^j\omega)\\ &= \sum_{j=0}^{n_1(\omega)-1} \widetilde{X}(T^j\omega) + \sum_{j=n_1(\omega)}^{n_2(\omega)-1} \widetilde{X}(T^j\omega) + \cdots + \sum_{j=n_{K(\omega)-1}(\omega)}^{n_{K(\omega)}(\omega)-1} \widetilde{X}(T^j\omega).\end{aligned}$$

Die Ungleichung (5.6), angewandt auf ω, $T^{n_1(\omega)}(\omega)$, $T^{n_2(\omega)}(\omega), \ldots, T^{n_{K(\omega)-1}(\omega)}(\omega)$, ergibt für die rechte Seite die Abschätzung

$$\begin{aligned}&\geq n_1(\omega)(\overline{X}_M(\omega) - \varepsilon) + (n_2(\omega) - n_1(\omega))(\overline{X}_M(T^{n_1(\omega)}(\omega)) - \varepsilon)\\ &\quad + \cdots + (n_{K(\omega)}(\omega) - n_{K(\omega)-1}(\omega))(\overline{X}_M(T^{n_{K(\omega)-1}(\omega)}(\omega)) - \varepsilon).\end{aligned}$$

Für alle $\omega \in \Omega$ und $j \in \mathbb{N}$ ist $\overline{X}_M(T^j\omega) = \overline{X}_M(\omega)$. Somit ist dieser Ausdruck

$$\begin{aligned}= n_{K(\omega)}(\omega)\overline{X}_M(\omega) - n_{K(\omega)}(\omega)\varepsilon &\geq m\overline{X}_M(\omega) + (n_{K(\omega)}(\omega) - m)\overline{X}_M(\omega) - m\varepsilon\\ &\geq m\overline{X}_M(\omega) - N_\varepsilon \cdot M - m\varepsilon.\end{aligned}$$

Dividiert man durch m und beachtet $\int \widetilde{X}(T^j\omega)P(d\omega) = \int \widetilde{X}\, dP$ (denn T ist masserhaltend), so folgt:

$$\int \widetilde{X}\, dP \geq \int \overline{X}_M\, dP - \frac{N_\varepsilon\, M}{m} - \varepsilon.$$

Somit gilt wegen (5.7):

$$\int X\, dP \geq \int \overline{X}_M\, dP - \frac{N_\varepsilon\, M}{m} - \varepsilon - M\varepsilon.$$

Da dies für alle $m \in \mathbb{N}$ und $\varepsilon > 0$ gilt, folgt $\int X dP \geq \int \overline{X}_M dP$ für alle $M > 0$.

Nach dem monotonen Konvergenzsatz folgt $\int X dP \geq \int \overline{X} dP$. Damit ist die 1. Ungleichung von (5.3) bewiesen; die zweite folgt völlig analog und ist sogar noch etwas einfacher, da $\underline{X}$ nicht nach unten abgeschnitten werden muss.

Damit ist die f.s.-Existenz von $Y = \lim_{n\to\infty} S_n$ bewiesen, und Y ist $\mathcal{Y}$-m.b. und erfüllt (5.2). Wir zeigen nun noch, dass $Y = E(X|\,\mathcal{J})$ ist. Sei eine beliegibe Menge $B \in \mathcal{J}$. Wir wenden nun das oben Gezeigte auf $1_B X$ anstelle von X an. Da B invariant ist, folgt $(1_B X) \circ T = 1_B\,(X \circ T)$ und

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (1_B X) \circ T^k = 1_B \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X \circ T^k = 1_B Y, \text{ f.s..}$$

Wegen (5.2), angewandt auf $1_B X$ folgt

$$\int_B X dP = \int_B Y dP.$$

Dies gilt für jede Menge $B \in \mathcal{J}$. Da Y $\mathcal{J}$-m.b. ist fogt $Y = E(X|\,\mathcal{J})$ f.s. ■

Bemerkung 5.20
Sei $X = \{X_n\}_{n\in\mathbb{N}}$ ein stationärer Prozess. Die Verschiebungstransformation $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ ist masserhaltend für $P' := PX^{-1}$. Sei $\pi_1 : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ die Projektion auf den 1. Faktor. Die Zufallsgrösse X_1 liegt genau dann in $\mathcal{L}_1(P)$, wenn π_1 in $\mathcal{L}_1(P')$ liegt. In diesem Fall konvergiert nach dem Ergodensatz die Folge $(\frac{1}{n}\sum_{j=0}^{n-1}\pi_1 \circ T^j)_n$ fast sicher gegen $\psi := E\left(\pi_1 | \mathcal{J}\right)$, was eine auf $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P')$ definierte T-invariante Zufallsgrösse ist. Im ergodischen Fall ist $\psi = \int \pi_1 \, dP'$. Somit folgt die f.s.-Konvergenz von $\{\frac{1}{n}\sum_{j=1}^{n} X_j\}_{n\in\mathbb{N}}$ gegen $\psi \circ X$ im ergodischen Fall also gegen EX_1. Als Spezialfall erhalten wir das starke Gesetz der grossen Zahlen, Satz 2.50, denn wie wir gesehen hatten (Lemma 5.8) ist der zugehörige Verschiebungsoperator ergodisch.

Als weiteres Korollar des Ergodensatzes erhalten wir zusammen mit Satz 5.17:

Satz 5.21
Sei $\{X_i\}_{i\in\mathbb{N}}$ eine irreduzible, aperiodische, positiv rekurrente Markoff-Kette mit stationärer Startverteilung π auf I. Sei ferner $f : I \to \mathbb{R}$ eine Abbildung, mit

$$\sum_{i\in I} \pi(i) \left| f(i) \right| < \infty.$$

Dann gilt

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} f(X_j) = \int f d\pi = \sum_{j\in I} f(j)\pi(j), \text{ f.s.}$$

Bemerkung 5.22
Auf die Aperiodizität kann man mit einer Zusatzüberlegung verzichten, was wir jedoch hier nicht ausführen wollen.

Ebenfalls kann man in der Annahme darauf verzichten, dass die Kette mit der invarianten Startverteilung π startet. Dies erfordert ebenfalls eine Zusatzüberlegung, die dem Leser als Übungsaufgabe überlassen sei. (Die Annahme, dass die stochastische Matrix irreduzibel und aperiodisch ist, muss man jedoch beibehalten). Als Beispiel können wir etwas $f := 1_C$, $C \subset I$, nehmen. Dann besagt der Satz, dass

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} 1_C(X_j) = \lim_{n\to\infty} \frac{1}{n}\#\left\{k \le n : X_k \in C\right\} = \pi(C), \text{ f.s.}$$

gilt.

6 Brownsche Bewegung

6.1 Stochastische Prozesse

Die Brownsche Bewegung ist ein stochastischer Prozess mit Zeitparameter $t \in \mathbb{R}^+$. Wir hatten bisher derartige Prozesse nicht betrachtet; ihre Definition macht jedoch keinerlei Schwierigkeiten.

Definition 6.1

a) *Ein* ***stochastischer Prozess mit stetigem Zeitparameter*** $X = \{X_t\}_{t\in\mathbb{R}^+}$ *ist eine auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, P)$ *definierte Familie von* $(\mathbb{R}, \mathcal{B})$*-wertigen Zufallsgrössen. Die reellwertigen Funktionen* $\mathbb{R}^+ \ni t \to X_t(\omega)$ *heissen die* ***Pfade*** *des stochastischen Prozesses.*

b) X *hat* **stetige Pfade** *(kurz: ist ein stetiger stochastischer Prozess), wenn für jedes* $\omega \in \Omega$ *die Abbildung* $\mathbb{R}^+ \ni t \to X_t(\omega)$ *stetig ist.*

Analog wie in Kapitel 4 betrachten wir die *endlichdimensionalen Verteilungen* (kurz e.d. Verteilungen) eines stochastischen Prozesses $X = \{X_t\}_{t\in\mathbb{R}^+}$.

Für $0 \le t_1 < t_2 < \cdots < t_n$, sei

$$\mu_{t_1,\ldots,t_n} := P(X_{t_1}, X_{t_2}, \ldots, X_{t_n})^{-1}.$$

Dies ist ein Wahrscheinlichkeitsmass auf $(\mathbb{R}^n, \mathcal{B}^n)$. Die Familie dieser Wahrscheinlichkeitsmasse heisst die **Familie der e.d. Verteilungen** von X. Für $a_i, b_i \in \mathbb{R}$, $1 \le i \le n$, $a_i < b_i$ ist also etwa

$$\begin{aligned} & \mu_{t_1,\ldots,t_n}((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n]) \\ = \; & P(a_1 < X_{t_1} \le b_1, a_2 < X_{t_2} \le b_2, \ldots, a_n < X_{t_n} \le b_n) \end{aligned}$$

Diese Familie von Verteilungen besitzt offenbar die folgende **Verträglichkeitseigenschaft**: Für $t_1 < t_2 < \cdots < t_n$ und $j \in \{1, \ldots, n\}$ ist

$$\mu_{t_1,\ldots,t_{j-1},t_{j+1},\ldots,t_n} = \mu_{t_1,\ldots,t_n}\varphi_j^{-1},$$

wobei φ_j die Projektion $\mathbb{R}^n \to \mathbb{R}^{n-1}$,

$$(x_1, \ldots, x_n) \to (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)$$

ist.

Definition 6.2

Eine Familie $\{\, \mu_{t_1,\ldots,t_n} : n \in \mathbb{N},\ t_1 < \cdots < t_n \,\}$ *von endlichdimensionalen Verteilungen, die diese Bedingung erfüllt, nennen wir* ***verträglich****.*

Der folgende Satz stammt von Kolmogoroff. Wir werden ihn jedoch nicht explizit benützen

Satz 6.3

Zu jeder verträglichen Familie von e.d. Verteilungen existiert ein Wahrscheinlichkeitsraum und ein darauf definiertet stochastischer Prozess mit diesen e.d. Verteilungen.

Im Zusammenhang mit diesem Satz muss jedoch betont werden, dass damit keinesfalls die Frage geklärt ist, ob ein stochastischer Prozess mit speziellen Pfadeigenschaften existiert, z.B. ob der stochastische Prozess stetige Pfade hat. Für den Fall der Brownschen Bewegung werden wir den Prozess mit stetigen Pfaden konstruieren ohne auf den obigen Satz Bezug zu nehmen.

Definition 6.4

a) *Ein stochastischer Prozess* $\{X_t\}_{t\in\mathbb{R}^+}$ *heisst* ***Gauss-Prozess****, wenn alle e.d. Verteilungen Normalverteilungen sind.*

b) *Ein Gauss-Prozess heisst* ***zentriert****, wenn* $EX_t = 0$ *für alle* $t \in \mathbb{R}^+$ *gilt.*

Eine n-dimensionale Normalverteilung μ auf $(\mathbb{R}^n, \mathcal{B}^n)$ ist vollständig durch die Angabe des Vektors der Erwartungswerte $m = (m_1, \ldots, m_n)$

$$m_i := \int x_i \, \mu(dx)$$

und durch die (symmetrische, positiv semidefinite) Kovarianzmatrix $\Sigma = (\sigma_{ij})$

$$\sigma_{ij} = \int (x_i - m_i)(x_j - m_j) \, \mu(dx)$$

festgelegt (Satz 2.23).

Ist $X = \{X_t\}_{t\in T}$ ein Gauss-Prozess, so sei die Funktion $m : \mathbb{R}^+ \to \mathbb{R}$ definiert durch

$$m(t) = EX_t, \tag{6.1}$$

und die Kovarianzfunktion $\sigma : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ durch

$$\sigma(t,s) = \operatorname{cov}(X_t, X_s). \tag{6.2}$$

Die obigen Überlegungen führen direkt auf den untenstehenden Satzes.

Satz 6.5

a) *Ist* X *ein Gauss-Prozess, so sind die e.d. Verteilungen festgelegt durch die Funktionen* $m : \mathbb{R}^+ \to \mathbb{R}$, $\sigma : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$.

b) σ *hat die folgenden Eigenschaften:*

- $\sigma(t,s) = \sigma(s,t)$, $\forall\, s,t \in \mathbb{R}^+$.
- *Für* $t_1, \ldots, t_n \in \mathbb{R}^+$ *is die symmetrische Matrix* $\{\sigma(t_i,t_j)\}_{1\leq i,j\leq n}$ *positiv semidefinit.*

Aus Satz 6.3 folgt leicht, dass zu Funktionen m und σ, die die obigen Bedingungen erfüllen, ein zugehöriger Gauss-Prozess existiert.

Wir interessieren uns hier nur für *stetige* stochastische Prozesse. Ein derartiger Prozess X definiert offenbar eine Abbildung $\Omega \to C(\mathbb{R}^+)$, wobei $C(\mathbb{R}^+)$ die Menge der stetigen reellwertigen Funktionen auf $\mathbb{R}^+$ ist. Wir schreiben nachfolgend einfach C für $C(\mathbb{R}^+)$. Die Abbildung ist gegeben durch $\omega \to X_.(\omega) \in C$, die wir der Einfachheit halber auch mit X bezeichnen. Auf C definieren wir die σ-Algebra $\mathcal{C}$, die von den Auswertungsabbildungen $\pi_t : C \to \mathbb{R}$, $C \ni f \to \pi_t(f) := f(t) \in \mathbb{R}$, erzeugt wird. Es ist offensichtlich, dass X eine messbare Abbildung $(X, \mathcal{F}) \to (C, \mathcal{C})$ ist. Die **Verteilung** des stetigen stochastischen Prozesses X ist dann das induzierte Wahrscheinlichkeitsmass PX^{-1} auf $(C, \mathcal{C})$.

Lemma 6.6
Die Verteilung eines stetigen stochastischen Prozesses auf $(C,\mathcal{C})$ wird durch seine e.d. Verteilungen eindeutig festgelegt.

Beweis. Das ist das übliche Argument: Die Mengen der Form $\pi_{t_1}^{-1}(A_1)\cap\ldots\cap\pi_{t_k}^{-1}(A_k)$, $k\in\mathbb{N}$, $t_1,\ldots,t_k\in\mathbb{R}^+$, $A_1,\ldots,A_k\in\mathcal{B}$ bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{C}$. Die e.d. Verteilungen von X legen PX^{-1} offensichtlich auf den Mengen dieser Form fest. ■

Wie schon in Kapitel 4 im Falle von stochastischen Prozessen mit diskretem Zeitparameter diskutiert, ist das Problem, einen stetigen stochastischen Prozess zu konstruieren einfach das Problem, ein Wahrscheinlichkeitsmass auf $(C,\mathcal{C})$ zu konstruieren, sodass die Projektionen $\{\pi_t\}_{t\in\mathbb{R}^+}$ die richtigen e.d. Verteilungen haben.

Beispiel 6.7
Sei $\sigma:\mathbb{R}^+\times\mathbb{R}^+\to\mathbb{R}$ definiert durch $\sigma(t,s):=\min(t,s)$. Dann erfüllt σ die Bedingungen b) von Satz 6.5.

Beweis. Die Symmetrie ist klar. Wir müssen nur den zweiten Teil zeigen. Seien $0\le t_1<\cdots<t_n$. Wir müssen nachweisen, dass für einen beliebigen Vektor $\lambda=(\lambda_1,\ldots,\lambda_n)\in\mathbb{R}^n$ die Ungleichung

$$\sum_{i,j=1}^{n}\sigma(t_i,t_j)\lambda_i\lambda_j\ge 0$$

gilt.

Wir setzen $t_0:=0$. Eine elementare Umformung ergibt sofort:

$$\sum_{i,j=1}^{n}\min(t_i,t_j)\lambda_i\lambda_j=\sum_{k=1}^{n}(t_k-t_{k-1})\left(\sum_{i=k}^{n}\lambda_i\right)^2\ge 0.$$

■

Definition 6.8
Ein zentrierter Gauss Prozess $X=\{X_t\}_{t\in\mathbb{R}^+}$ mit $EX_tX_s=t\wedge s:=\min(t,s)$ für alle $t,s\ge 0$ heisst ***Brownsche Bewegung****. Hat der Prozess stetige Pfade, so nennt man ihn eine* ***stetige Brownsche Bewegung.***

6.2 Die Lévy–Ciesielski Konstruktion der Brownschen Bewegung

Wir konstruieren in diesem Abschnitt eine stetige Brownsche Bewegung. Wir sind zunächst bescheiden und schränken uns auf das Zeitintervall $[0,1]$ ein. Wir brauchen einige Fakten über (reelle) Hilberträume, die als bekannt vorausgesetzt werden. Hier eine Zusammenstellung:

Ein reeller **Hilbertraum** H ist ein $\mathbb{R}$-Vektorraum, versehen mit einem positive definiten Skalarprodukt $\langle\cdot,\cdot\rangle$ und zugehöriger Norm $\|x\|:=\sqrt{\langle x,x\rangle}$, der bezüglich dieser Norm vollständig ist. Der Hilbertraum heisst **separabel**, falls eine abzählbare dichte Teilmenge in H existiert. Eine abzählbare Folge $\{h_n\}_{n\in\mathbb{N}}$ heisst **vollständige Orthonormalbasis** von H, falls $\langle h_i,h_j\rangle=\delta_{ij},\ i,j\in\mathbb{N}$ gilt, und falls die Menge

$\left\{\sum_{n=1}^{N} a_n h_n : N \in \mathbb{N},\ a_1, \ldots, a_N \in \mathbb{R}\right\}$ dicht in H ist. (Vorsicht: Das ist keine Basis im Sinne der Linearen Algebra: Dort würde man verlangen, dass diese Menge gleich H ist). Hier die Fakten, die wir benötigen:

- $L^2[0,1]$, die Menge der (Äquivalenzklassen von) reellwertigen quadratintegrierbaren Funktionen auf $[0,1]$, versehen mit

$$\langle f, g \rangle = \int_0^1 f(t)\, g(t)\, dt$$

  ist ein separabler Hilbertraum.

- Jeder separable Hilbertraum besitzt eine vollständige Orthonormalbasis.

- Ist H ein separabler Hilbertraum und ist $\{h_n\}_{n\in\mathbb{N}}$ eine Folge mit

$$\langle h_i, h_j \rangle = \delta_{ij},\ i, j \in \mathbb{N}, \tag{6.3}$$

  so ist diese Folge genau dann eine vollständige Orthonormalbasis wenn

$$\{x \in H : \langle x, h_n \rangle = 0,\ \forall n \in \mathbb{N}\} = \{0\} \tag{6.4}$$

  gilt.

- Ist $\{h_n\}_{n\in\mathbb{N}}$ eine vollständige Orthonormalbasis, so gilt für $x, y \in H$

$$\langle x, y \rangle = \sum_{n=1}^{\infty} \langle x, h_n \rangle \langle y, h_n \rangle. \tag{6.5}$$

Ausgangspunkt der Lévy-Ciesielski-Konstruktion ist eine spezielle vollständige Orthonormalbasis von $L^2[0,1]$, die sogenannte Haar Basis. Sie ist definiert durch:

$$f_0(t) \equiv 1,$$

und für $n \in \mathbb{N}$, $1 \leq k \leq 2^{n-1}$:

$$f_{n,k}(t) := \begin{cases} 2^{(n-1)/2} & \text{für } t \in [(2k-2)2^{-n}, (2k-1)2^{-n}) \\ -2^{(n-1)/2} & \text{für } t \in [(2k-1)2^{-n}, 2k2^{-n}) \\ 0 & \text{sonst.} \end{cases}$$

Lemma 6.9
$F := \{f_0\} \cup \{f_{n,k} : n \in \mathbb{N}, 1 \leq k \leq 2^{n-1}\}$ ist eine vollständige Orthonormalbasis von $L^2[0,1]$.

Beweis. (6.3) ist eine einfache Rechnung. Wir beweisen (6.4). Sei $h \in L^2[0,1]$ mit

$$\langle h, f_{n,k} \rangle = 0 \text{ für alle } n, k \tag{6.6}$$

h ist natürlich auch integrierbar. Wir zeigen, dass für $0 \le a < b \le 1$ die Gleichung

$$\int_a^b h \, dt = 0 \tag{6.7}$$

gilt. Eine Lebesgue-integrierbare Funktion auf $[0,1]$, die diese Eigenschaft hat, ist gleich 0, fast überall. (Falls nicht bekannt: Einfache Übungsaufgabe zur Masstheorie).

Zunächst folgt aus (6.6), $\langle h, 1 \rangle = 0$. Wir zeigen nun mit Induktion nach $n \in \mathbb{N}$, dass

$$\int_{(k-1)2^{-n}}^{k2^{-n}} h \, dt = 0, \quad n \in \mathbb{N}, \, 1 \le k \le 2^n, \tag{6.8}$$

gilt. Zunächst $n = 1$. Aus (6.6) folgt

$$0 = \langle h, f_{1,1} \rangle = \int_0^{1/2} h \, dt - \int_{1/2}^1 h \, dt.$$

Zusammen mit $\int_0^1 h \, dt = 0$ folgt (6.8) für $n = 1$.

Nun der Induktionsschluss: Sei $n \ge 2$ und $1 \le k \le 2^{n-1}$. Nach Induktionsvoraussetzung gilt

$$\int_{(k-1)2^{-n+1}}^{k2^{-n+1}} h \, dt = 0,$$

und aus (6.6) folgt

$$0 = 2^{(n-1)/2} \int_{(2k-2)2^{-n}}^{(2k-1)2^{-n}} h \, dt - 2^{(n-1)/2} \int_{(2k-1)2^{-n}}^{2k2^{-n}} h \, dt.$$

Daraus folgt

$$\int_{(2k-2)2^{-n}}^{(2k-1)2^{-n}} h \, dt = \int_{(2k-1)2^{-n}}^{2k2^{-n}} h \, dt = 0.$$

Damit ist (6.8) bewiesen.

Aus (6.8) folgt, dass für $0 \le a < b \le 1$, $a, b \in D := \{ k2^{-n} : n \in \mathbb{N}, 0 \le k \le 2^n \}$, die Gleichung $\int_a^b h \, dt = 0$ gilt. Da D dicht in $[0,1]$ liegt, folgt mit einer einfachen Anwendung des Satzes von Lebesgue (6.7). ∎

Aus dem Lemma 6.9 folgt die Parcevalsche Gleichung (6.5), in unserem Spezialfall:

$$\langle h_1, h_2 \rangle = \langle h_1, f_0 \rangle \langle f_0, h_2 \rangle + \sum_{n=1}^{\infty} \sum_{k=1}^{2^{n-1}} \langle h_1, f_{n,k} \rangle \langle f_{n,k}, h_2 \rangle. \tag{6.9}$$

Wir setzen $F_0(t) := \int_0^t f_0(s) \, ds$ und für $n \in \mathbb{N}$, $1 \le k \le 2^{n-1}$, $F_{n,k}(t) := \int_0^t f_{n,k}(s) \, ds$.

Es seien nun ξ_0, $\xi_{n,k}$, $n \in \mathbb{N}$, $1 \leq k \leq 2^{n-1}$, unabhängige standard-normalverteilte Zufallsgrössen, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. (Die Existenz eines derartigen Wahrscheinlichkeitsraums und einer solchen Folge wurde im Kapitel 4 diskutiert. (Sie folgt z.B. aus dem Satz von Ionescu-Tulcea).

Wir definieren für $N \in \mathbb{N}$, $t \in [0,1]$, $\omega \in \Omega$:

$$B_t^{(N)}(\omega) := F_0(t)\xi_0(\omega) + \sum_{n=1}^{N} \sum_{k=1}^{2^{n-1}} F_{n,k}(t)\xi_{n,k}(\omega). \tag{6.10}$$

Zunächst zwei triviale Beobachtungen:

- Für alle $N \in \mathbb{N}$, $\omega \in \Omega$ ist $B_t^{(N)}(\omega)$ stetig in $t \in [0,1]$ mit $B_0^{(N)}(\omega) = 0$. $B^{(N)} = \{B_t^{(N)}\}_{t \in [0,1]}$ ist daher ein stetiger stochastischer Prozess.
- Für $0 \leq t_1 < \cdots < t_k \leq 1$ ist der Zufallsvektor $(B_{t_1}^{(N)}, \ldots, B_{t_k}^{(N)})$ normalverteilt und zentriert. $\mathbf{B}^{(N)}$ ist daher für jedes N ein zentrierter stetiger Gauss-Prozess.

Wir wollen nun nachweisen, dass die Folge der Prozesse $B^{(N)}$ in einem zu präzisierenden Sinn gegen eine Brownsche Bewegung konvergiert. Der Beweis dieser Aussage besteht aus einem einfachen Teil, und einem schwierigeren. Der einfache Teil ist zu zeigen, dass

$$\lim_{N \to \infty} E\left(B_s^{(N)} B_t^{(N)}\right) = \min(s,t)$$

gilt. Das hat nichts mit der speziell gewählten Haar-Basis zu tun hat. Die Aussage ist richtig mit jedem vollständigen Orthonormalsystem:

Lemma 6.10
Sei $\{h_n\}_{n \in \mathbb{N}}$ ein vollständiges Orthonormalsystem in $L_2[0,1]$, seien $H_n(t) := \int_0^t h_n(s)\, ds$, und seien unabhängige, standard Normalverteilte Zufallsgrössen ξ_n, $n \in \mathbb{N}$ gegeben. Dann gilt

$$\lim_{N \to \infty} E\left(\sum\nolimits_{n=1}^{N} \xi_n H_n(t) \sum\nolimits_{n=1}^{N} \xi_n H_n(s)\right) = \min(s,t)$$

für alle $s, t \in [0,1]$.

Beweis. Das ist einfach ein Spezialfall der Parcevalschen Gleichung: Für $N \to \infty$ gilt

$$\begin{aligned} E\left(\sum_{n=1}^{N} \xi_n H_n(t) \sum_{n=1}^{N} \xi_n H_n(s)\right) &= \sum_{n=1}^{N} H_n(t) H_n(s) \\ &= \sum\nolimits_{n=1}^{N} \left\langle 1_{[0,t]} h_n \right\rangle \left\langle 1_{[0,s]} h_n \right\rangle \\ &\to \left\langle 1_{[0,t]} 1_{[0,s]} \right\rangle = \min(s,t). \end{aligned}$$

■

Dieses Lemma legt nahe, dass $\lim_{N\to\infty}\sum_{n=1}^{N}\xi_n H_n(\cdot)$ eine Brownsche Bewegung ist, gleichgültig, ob wir die Haar-Basis verwendet haben oder eine andere vollständige Orthonormalbasis. Es gibt jedoch noch zwei Probleme. Als Erstes müssen wir präzisieren, in welchem Sinn der Limes überhaupt existiert. Für dieses Problem ist die spezielle Haar-Basis sehr nützlich. Im untenstehenden Satz nehmen wir also an, dass $B_t^{(N)}$ wie oben in (6.10) definiert ist.

Lemma 6.11
Es existiert $\Omega_0 \in \mathcal{F}$ mit $P(\Omega_0) = 1$, so dass für alle $\omega \in \Omega_0$ die Folge von stetigen Funktionen $[0,1] \ni t \to B_t^{(N)}(\omega)$ gleichmässig auf $[0,1]$ gegen eine Funktion $t \to B_t(\omega)$ konvergiert.

Beweis. Wir betrachten

$$D_t^{(N)}(\omega) := B_t^{(N)}(\omega) - B_t^{(N-1)}(\omega) = \sum_{k=1}^{2^{N-1}} F_{N,k}(t)\xi_{N,k}(\omega).$$

Die Funktionen $F_{n,k}$ haben die folgenden Eigenschaften:

- $$F_{n,k}(t) \geq 0,\ \forall t. \tag{6.11}$$

- $$\{t : F_{n,k}(t) > 0\} = ((k-1)2^{-n+1}, k2^{-n+1}). \tag{6.12}$$

- $$\max_t F_{n,k}(t) = 2^{-(n+1)/2} \tag{6.13}$$

Aus (6.12) und (6.13) folgt

$$\sup_t |D_t^{(N)}(\omega)| \leq 2^{-(N+1)/2}\bar{\xi}_N(\omega), \tag{6.14}$$

wobei $\bar{\xi}_N := \max_{1\leq k\leq 2^{N-1}} |\xi_{N,k}|$ ist. Demzufolge ist für jedes $x > 0$:

$$\begin{aligned} P\left(\sup_t |D_t^{(N)}| \geq x\right) &\leq P(\bar{\xi}_N \geq 2^{(N+1)/2}x) \\ &\leq 2^{N-1}P(|\xi_{N,1}| \geq 2^{(N+1)/2}x) \\ &= 2^N\left(1 - \Phi(2^{(N+1)/2}x)\right), \end{aligned} \tag{6.15}$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist:

$$\Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-s^2/2}\, ds.$$

Wir verwenden nun die Ungleichung

$$1 - \Phi(y) \leq \frac{1}{\sqrt{2\pi} y} \mathrm{e}^{-y^2/2},$$

für $y > 0$ (Übungsaufgabe), und setzen in (6.15) $x := \frac{N}{2} 2^{-N/2}$ ein. Dann ergibt sich

$$\begin{aligned} P\left(\sup_t |D_t^{(N)}| \geq \frac{N}{2} 2^{-N/2}\right) &\leq 2^N \left(1 - \Phi(N/\sqrt{2})\right) \\ &\leq 2^N \frac{1}{\sqrt{2\pi}} \frac{\sqrt{2}}{N} \exp\left[-\frac{N^2}{4}\right]. \end{aligned}$$

Eine elementare Rechnung ergibt, dass die Reihe

$$\sum_N 2^N \frac{1}{N} \exp\left[-\frac{N^2}{4}\right]$$

konvergiert. Wir setzen

$$\Omega_0^c = \bigcap_m \bigcup_{N \geq m} \left\{\sup_t |D_t^{(N)}| \geq \frac{N}{2} 2^{-N/2}\right\}.$$

Aus dem Borel–Cantelli Lemma folgt, dass $P(\Omega_0^c) = 0$ und daher $P(\Omega_0) = 1$ gilt. Für $\omega \in \Omega_0$ existiert $m(\omega)$, so dass für alle $N \geq m(\omega)$

$$\sup_t |D_t^N(\omega)| \leq \frac{N}{2} 2^{-N/2}$$

gilt. Da $\sum_N N 2^{-N/2} < \infty$ ist, folgt, dass für $\omega \in \Omega_0$ die Funktionenfolge $(B_\cdot^{(N)}(\omega))_{N \in \mathbb{N}}$ gleichmässig konvergiert. Wir setzen $B_t(\omega) = \lim_{N \to \infty} B_t^{(N)}(\omega)$, $\omega \in \Omega_0$, was wegen der gleichmässigen Konvergenz stetig in t ist. Ferner gilt natürlich $B_0(\omega) = 0$. ■

Wir schränken nun die Definition von $B = (B_t)$ auf Ω_0 ein. Wegen $\Omega_0 \in \mathcal{F}$ können wir natürlich auch die σ-Algebra einschränken: $\mathcal{F}_0 := \{A \in \mathcal{F} : A \subset \Omega_0\}$, und P auf $\mathcal{F}_0$ einschränken, wofür wir P_0 schreiben. Wegen $P(\Omega_0) = 1$ ist natürlich P_0 ein Wahrscheinlichkeitsmass. B ist daher ein auf $(\Omega_0, \mathcal{F}_0, P_0)$ definierter stetiger stochastischer Prozess. Der Einfachheit halber lassen wir den Index 0 gleich wieder weg.

Satz 6.12
Der so konstruierte Prozess $B = (B_t)_{t \in [0,1]}$ ist eine stetige Brownsche Bewegung.

Beweis. Wir haben schon fast alles gezeigt. Wir wissen aus Lemma 6.11, dass B ein stetiger stochastischer Prozess ist und aus Lemma 6.10, dass

$$\lim_{N \to \infty} E\left(B_t^{(N)} B_s^{(N)}\right) = \min(s, t)$$

ist. Wir haben jedoch zwei Dinge noch nicht gezeigt: Erstens, dass der Limes mit dem Erwartungswert vertauscht, d.h. dass

$$E\left(B_t B_s\right) = \min\left(s, t\right)$$

gilt und zweitens, dass B überhaupt ein Gauss-Prozess ist. Für jedes N ist $\left(B_t^{(N)}\right)_{t\in[0,1]}$ natürlich ein Gauss-Prozess. Was uns noch fehlt ist das nachfolgende Lemma, das wir etwas allgemeiner formulieren, als wir es hier benötigen würden. ■

Lemma 6.13
Sei $X^{(N)} = (X_1^{(N)}, \ldots, X_d^{(N)})$, $N \in \mathbb{N}$, eine Folge von normalverteilten Zufallsvektoren, die in Wahrscheinlichkeit gegen einen Zufallsvektor $X = (X_1, \ldots, X_d)$ konvergiert. Dann ist X ebenfalls normalverteilt, und es gelten

$$\begin{aligned} EX_i &= \lim_{N\to\infty} EX_i^{(N)}, \\ \operatorname{cov}(X_i, X_j) &= \lim_{N\to\infty} \operatorname{cov}(X_i^{(N)}, X_j^{(N)}). \end{aligned}$$

Beweis. Wir verwenden charakteristische Funktionen. Für $t \in \mathbb{R}^d$ gilt

$$\begin{aligned} E\left(\exp\left[i\left\langle t, X\right\rangle\right]\right) &= \lim_{N\to\infty} E\left(\exp\left[i\left\langle t, X^{(N)}\right\rangle\right]\right) \\ &= \lim_{N\to\infty} \exp\left[i\left\langle t, a^{(N)}\right\rangle - \frac{1}{2}\left\langle t, \Sigma^{(N)} t\right\rangle\right], \end{aligned} \tag{6.16}$$

wobei $a^{(N)}$ der Vektor der Erwartungswerte ist: $a^{(N)} = \left(EX_1^{(N)}, \ldots, EX_d^{(N)}\right)$ und $\Sigma^{(N)}$ die positive semidefinite Kovarianzmatrix von $X^{(N)}$. Hier haben wir die folgende Verallgemeinerung des Satzes von Lebesgue verwendet: Falls die Folge $X^{(N)}$ in Wahrscheinlichkeit gegen X konvergiert und falls $f : \mathbb{R}^d \to \mathbb{R}$ eine stetige und beschränkte Funktion ist, so gilt $\lim_{N\to\infty} Ef\left(X^{(N)}\right) = Ef\left(X\right)$. Der Leser möge sich dies als Übungsaufgabe überlegen.

Aus der Existenz des Limes auf der rechten Seite von (6.16), für jedes $t \in \mathbb{R}^d$, folgt jedoch sofort die Existenz der Limiten $a := \lim_{N\to\infty} a^{(N)}$ und $\Sigma = \lim_{N\to\infty} \Sigma^{(N)}$ (einfache Übungsaufgabe zu Diff-Int) und dass Σ positiv semidefinit ist. Damit ergibt sich

$$E\left(\exp\left[i\left\langle t, X\right\rangle\right]\right) = \exp\left[i\left\langle t, a\right\rangle - \frac{1}{2}\left\langle t, \Sigma t\right\rangle\right].$$

Somit ist X normalverteilt mit Mittel a und Kovarianzmatrix Σ. Damit ist das Lemma bewiesen. ■

Wir haben somit die Brownsche Bewegung auf dem Zeitintervall $[0, 1]$ konstruiert. Es verbleibt noch, eine stetige Brownsche Bewegung mit Zeitachse $\mathbb{R}^+$ zu konstruieren. Dazu verwenden wir die oben konstruierte Brownsche Bewegung $\{B_t\}_{t\in[0,1]}$ und definieren für $0 \le t < \infty$

$$\bar{B}_t := (1 + t)B_{(1+t)^{-1}} - B_1.$$

Satz 6.14
$\{\bar{B}_t\}_{t\geq 0}$ ist eine stetige Brownsche Bewegung.

Beweis. $\{\bar{B}_t\}_{t\geq 0}$ ist evidenterweise ein stetiger zentrierter Gauss Prozess mit $B_0 = 0$. Es bleibt daher nur noch die Aufgabe, die Kovarianzfunktion auszurechnen.

$$\begin{aligned} E(\bar{B}_t\bar{B}_s) &= (1+t)(1+s)E\left(B\left(\frac{1}{1+t}\right)B\left(\frac{1}{1+s}\right)\right) \\ &\quad -(1+t)EB\frac{1}{1+t}B_1 - (1+s)EB_1B\frac{1}{1+s} + E\left(B_1^2\right) \\ &= (1+t)\wedge(1+s) - 1 = t\wedge s. \end{aligned}$$

■

Bemerkung 6.15
Eine stetige Brownsche Bewegung ist in gewissem Sinne nicht eindeutig, denn für die Wahl des Wahrscheinlichkeitsraumes und der Abbildungen $B_t : \Omega \to \mathbb{R}$ gibt es viele Möglichkeiten. Eindeutig ist hingegen die Verteilung $W := PB^{-1}$ auf $(C, \mathcal{C})$. W nennt man auch das **Wiener Mass** (nach Norbert Wiener, der dieses Mass zuerst mathematisch präzise konstruiert hat).

6.3 Einfache Eigenschaften der Brownschen Bewegung

Die Brownsche Bewegung hat einige einfache aber wichtige Skalierungseigenschaften.

Satz 6.16
Sei $\{B_t\}_{t\geq 0}$ eine stetige Brownsche Bewegung.

a) Der Prozess $\{-B_t\}_{t\geq 0}$ ist eine stetige Brownsche Bewegung.

b) Für jedes $\lambda > 0$ ist $\{\sqrt{\lambda}B_{t/\lambda}\}_{t\geq 0}$ eine stetige Brownsche Bewegung.

c) Für jedes $u \geq 0$ ist $\{B_{t+u} - B_u\}_{t\geq 0}$ eine stetige Brownsche Bewegung.

d) Der Prozess $\{tB_{1/t}\}_{t\geq 0}$ mit $0B_{1/0} := 0$ ist eine Brownsche Bewegung. Es existiert $\Omega_0 \in \mathcal{F}$ mit $P(\Omega_0) = 1$ und

$$\lim_{t\downarrow 0} tB_{1/t}(\omega) = 0$$

für $\omega \in \Omega_0$. $\{tB_{1/t}\}_{t\geq 0}$, eingeschränkt auf Ω_0, ist somit eine stetige Brownsche Bewegung.

Beweis. Die Stetigkeit in a), b), c) und in d) auf $(0, \infty)$ ist klar, ebenso, dass die Prozesse zentrierte Gauss-Prozesse sind. Um nachzuweisen, dass es Brownsche Bewegungen sind, müssen wir also nur die Kovarianzen ausrechnen. Seien $s, t \geq 0$

a) $E((-B_s)(-B_t)) = E(B_sB_t) = s \wedge t$.

b) $E(\sqrt{\lambda}B_{s/\lambda}\sqrt{\lambda}B_{t/\lambda}) = \lambda(\frac{s}{\lambda}) \wedge (\frac{t}{\lambda}) = s \wedge t$.

c) $E((B_{s+u}-B_u)(B_{t+u}-B_u)) = (s+u)\wedge(t+u)-u = s\wedge t.$

d) Hier setzen wir $s,t>0$ voraus: $E(sB_{1/s}\,tB_{1/t}) = st(\frac{1}{t}\wedge\frac{1}{s}) = s\wedge t$. Mit unserer Festsetzung von $0B_{1/0}:=0$ gilt dies jedoch auch für s oder $t=0$.

Es bleibt noch die letzte Behauptung in d) nachzuweisen. Dies ist nun ziemlich evident: Da $X_t := tB_{1/t}$ schon stetig auf $(0,\infty)$ ist, ist die Stetigkeit in 0 ein messbares Ereignis:

$$\Omega_0 := \{\omega : \lim_{t\downarrow 0} X_t(\omega)=0\} = \bigcap_{m\in\mathbb{N}}\bigcup_{n\in\mathbb{N}}\bigcap_{s\in\mathbb{Q}\cap(0,1/n]}\{|X_s|\le 1/m\}\in\mathcal{F}.$$

Offensichtlich ergibt sich $P(\Omega_0)$ aus der Kenntnis der e.d. Verteilungen. Wir wissen jedoch schon, dass eine stetige Brownsche Bewegung existiert. Somit ist $P(\Omega_0)$ gleich dem Wert, den dieser Ausdruck für eine stetige Brownsche Bewegung hätte, also gleich 1. ■

Wir weisen nun nach, dass die Pfade der Brownschen Bewegung fast sicher nirgends differenzierbar sind. Genauer:

Satz 6.17
$B=\{B_t\}_{t\ge 0}$ sei eine stetige Brownsche Bewegung. Dann existiert $\Omega_0\in\mathcal{F}$ mit $P(\Omega_0)=1$, so dass für alle $\omega\in\Omega_0$ die Funktion $t\to B_t(\omega)$ in keinem Punkt differenzierbar ist.

Beweis. Es genügt zu zeigen, dass für jedes $N>0$ die Pfade der Brownschen Bewegung fast sicher nirgends differenzierbar auf dem Zeitintervall $[0,N]$ sind. Der notationellen Einfachheit halber nehmen wir $N=1$. Sei $f:[0,\infty)\to\mathbb{R}$ eine stetige Funktion, die differenzierbar in mindestens einem Punkt $s\in[0,1]$ ist. Per Definition existiert dann $\lim_{t\to s}(f(t)-f(s))/(t-s)\in\mathbb{R}$. Dies impliziert insbesondere, dass $\varepsilon>0$ und $l\in\mathbb{N}$ existieren mit $|f(t)-f(s)|\le l(t-s)$ für $s\le t\le s+\varepsilon$. Ist $n\ge m:=[\frac{4}{\varepsilon}]+1$, so gilt mit $i:=[ns]+1$:

$$s<\frac{i}{n}<\cdots<\frac{i+3}{n}\le s+\varepsilon,$$

und demzufolge für $j=i+1,i+2,i+3$:

$$\begin{aligned}\left|f\left(\frac{j}{n}\right)-f\left(\frac{j-1}{n}\right)\right| &\le \left|f\left(\frac{j}{n}\right)-f(s)\right|+\left|f\left(\frac{j-1}{n}\right)-f(s)\right|\\ &\le l\left(\frac{j}{n}-s\right)+l\left(\frac{j-1}{n}-s\right)\le\frac{7l}{n}\end{aligned}$$

Wir haben somit gezeigt, dass, wenn f in mindestens einem Punkt $\in[0,1]$ differenzierbar ist, natürliche Zahlen m, l existieren, so dass für alle $n\ge m$ eine Zahl $i\in\{1,\ldots,n+1\}$ existiert, sodass für die drei Zahlen $j=i+1,i+2,i+3$ die folgende Ungleichung gilt:

$$\left|f\left(\frac{j}{n}\right)-f\left(\frac{j-1}{n}\right)\right|\le\frac{7l}{n}.$$

Demzufolge ist für

$$\omega \notin N := \bigcup_{l\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \bigcup_{i=1}^{n+1} \bigcap_{j=i+1}^{i+3} \left\{ |B_{j/n} - B_{(j-1)/n}| < \frac{7l}{n} \right\}$$

die Funktion $t \to B_t(\omega)$ nirgends auf $[0,1]$ differenzierbar. Offensichtlich ist $N \in \mathcal{F}$. Es bleibt also zu zeigen, dass $P(N) = 0$ ist. Dafür müssen wir einfach für jedes l, m

$$P\left(\bigcap\nolimits_{n\geq m} \bigcup\nolimits_{i=1}^{n+1} \bigcap\nolimits_{j=i+1}^{i+3} \left\{ |B_{j/n} - B_{(j-1)/n}| < \frac{7l}{n} \right\} \right) = 0 \tag{6.17}$$

nachweisen. Die Zufallsgrössen $B_{j/n} - B_{(j-1)/n}, i+1 \leq j \leq i+3$, sind drei unabhängige, normalverteilte Zufallsgrössen mit Mittel 0 und Varianz $1/n$, was man sofort aus den Kovarianzen der Brownschen Bewegung ablesen kann. Demzufolge ist die linke Seite von (6.17)

$$\begin{aligned}
&\leq \liminf_{n\to\infty} P\left(\bigcup\nolimits_{i=1}^{n+1} \bigcap\nolimits_{j=i+1}^{i+3} \left\{ |B_{j/n} - B_{(j-1)/n}| < \frac{7l}{n} \right\} \right) \\
&\leq \liminf_{n\to\infty} n \max_{1\leq i\leq n+1} P\left(\bigcap\nolimits_{j=i+1}^{i+3} \left\{ |B_{j/n} - B_{(j-1)/n}| < \frac{7l}{n} \right\} \right) \\
&= \liminf_{n\to\infty} n \left\{ \int_{-7l/n}^{7l/n} \sqrt{\frac{n}{2\pi}} \exp\left[-\frac{n}{2} x^2 \right] dx \right\}^3 \leq \liminf_{n\to\infty} n \left(\frac{n}{2\pi} \right)^{3/2} \left(\frac{14l}{n} \right)^3 = 0.
\end{aligned}$$

■

6.4 Prozesse mit unabhängigen Zuwächsen, Markoffeigenschaft

Die Brownsche Bewegung hat, wie man sagt, *unabhängige Zuwächse* (was wir im Beweis von (6.17) implizit schon ausgenutzt haben). Wir wollen das nun etwas eingehender untersuchen.

Definition 6.18
Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum.

a) *Sei $X = \{X_t\}_{t\in\mathbb{R}^+}$ ein auf $(\Omega, \mathcal{F}, P)$ definierter stochastischer Prozess. Für $t \in \mathbb{R}^+$ sei $\mathcal{F}_t^X := \sigma(X_s : s \in \mathbb{R}^+, s \leq t)$. Die Familie $\{\mathcal{F}_t^X\}_{t\in\mathbb{R}^+}$ heisst die zu X gehörende* ***Filtrierung****.*

b) *Allgemeiner heisst eine Familie $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$ von Teil-σ-Algebren eine* ***Filtrierung*** *von $\mathcal{F}$, wenn $\mathcal{F}_s \subset \mathcal{F}_t$ für alle $s, t \in \mathbb{R}^+$ mit $s \leq t$ gilt.*

c) *Ein stochastischer Prozess $\{X_t\}_{t\in\mathbb{R}^+}$ heisst* ***angepasst*** *(oder adaptiert) an eine Filtrierung $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$, wenn für jedes $t \in \mathbb{R}^+$ die Zufallsgrösse X_t bezüglich $\mathcal{F}_t$-messbar ist.*

Offenbar ist $X = \{X_t\}_{t\in\mathbb{R}^+}$ genau dann $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$-angepasst, wenn $\mathcal{F}_t^X \subset \mathcal{F}_t$ für jedes $t \in \mathbb{R}^+$ gilt. Natürlich ist X stets $\{\mathcal{F}_t^X\}_{t\in\mathbb{R}^+}$-angepasst.

Definition 6.19
*Ein stochastischer Prozess, der angepasst an eine Filtrierung $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$ ist, hat **unabhängige Zuwächse** bezüglich $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$, falls für jedes $t \in \mathbb{R}^+$ der stochastische Prozess $\{X_s - X_t\}_{s\in[t,\infty)}$ und $\mathcal{F}_t$ unabhängig sind. Man sagt einfach, X habe unabhängige Zuwächse, falls er unabhängige Zuwächse bezüglich $\{\mathcal{F}_t^X\}_{t\in\mathbb{R}^+}$ hat.*

Im Moment mag es etwas unklar sein, weshalb wir in der Definition 6.19 eine allgemeinere Filtrierung als $\{\mathcal{F}_t^X\}_{t\in\mathbb{R}^+}$ zulassen. Es gibt aber dafür eine Reihe von Gründen, die später klar werden.

Lemma 6.20
Seien $X = (X_1, \ldots, X_m)$ und $Y = (Y_1, \ldots, Y_n)$ zwei Zufallsvektoren. $\chi_{(X,Y)} : \mathbb{R}^{n+m} \to \mathbb{C}$ sei die charakteristische Funktion des Vektors $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ und χ_X, χ_Y seien die charakteristischen Funktion von X bzw. Y. Dann sind X und Y genau dann unabhängig, wenn für alle $(s_1, \ldots, s_m, t_1, \ldots, t_n) \in \mathbb{R}^{n+m}$ die Gleichung

$$\chi_{(X,Y)}(s_1, \ldots, s_m, t_1, \ldots, t_n) = \chi_X(s_1, \ldots, s_m)\, \chi_Y(t_1, \ldots, t_n).$$

Beweis. Aus der Unabhängigkeit folgt sofort diese Faktorisierung. Wir zeigen die Umkehrung. Die rechte Seite der obigen Gleichung ist offensichtlich die charakteristische Funktion des Produktmasses $\mu_X \otimes \mu_Y$, wobei μ_X und μ_Y die Verteilungen von X bzw. Y sind. Da charakteristische Funktionen die Verteilungen eindeutig charakterisieren, folgt, dass $\mu_X \otimes \mu_Y$ die Verteilung von $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ ist. Daraus folgt die Unabhängigkeit von X und Y. ■

Lemma 6.21
Seien $(X_1, \ldots, X_m)$ und $(Y_1, \ldots, Y_n)$ zwei normalverteilte Zufallsvektoren. Sie sind genau dann unabhängig, wenn $\operatorname{cov}(X_i, Y_j) = 0$ für $1 \le i \le m,\ 1 \le j \le n$ gilt.

Beweis. Lemma 6.20. ■

Satz 6.22
Eine Brownsche Bewegung $B = \{B_t\}_{t\ge 0}$ hat unabhängige Zuwächse.

Beweis. Wir müssen nachweisen, dass für jedes t die σ-Algebren $\mathcal{G} := \mathcal{F}_t^B$ und $\mathcal{H} := \sigma((B_s - B_t) : s \ge t)$ unabhängig sind. Es reicht dafür aus, die Unabhängigkeit von zwei durchschnittstabilen Erzeugendensystemen von $\mathcal{G}$ bzw. $\mathcal{H}$ zu zeigen. Ein durchschnittstabiles Erzeugendensystem von $\mathcal{G}$ ist

$$\bigcup_{m,\, 0\le t_1<\cdots<t_m\le t} \sigma(B_{t_1}, \ldots, B_{t_m})$$

und von $\mathcal{H}$:

$$\bigcup_{m,\, t\le s_1<\cdots<s_n} \sigma(B_{s_1} - B_t, \ldots, B_{s_n} - B_t).$$

Es genügt also einfach zu zeigen, dass für $0 \leq t_1 < \cdots < t_m \leq t$ und $t \leq s_1 < \cdots < s_n$ die Zufallsvektoren $(B_{t_1}, \ldots, B_{t_m})$ und $(B_{s_1} - B_t, \ldots, B_{s_n} - B_t)$ unabhängig sind. Beide Vektoren sind gemeinsam normalverteilt. Nach Lemma 6.21 müssen wir nur die Kovarianzen ausrechnen:

$$\mathrm{cov}(B_{t_i}, B_{s_j} - B_t) = E(B_{t_i}(B_{s_j} - B_t)) = t_i \wedge s_j - t_i \wedge t = 0$$

für $1 \leq i \leq m$ und $1 \leq j \leq n$. ∎

Prozesse mit unabhängigen Zuwächsen sind Beispiele von Markoffprozessen.

Definition 6.23
*Es sei $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$ eine Filtrierung. Ein $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$-angepasster stochastischer Prozess $X = \{X_t\}_{t\in\mathbb{R}^+}$ heisst $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$-**Markoffprozess**, wenn für alle $t \in \mathbb{R}^+$ und $A \in \sigma(X_s : s \in \mathbb{R}^+, s \geq t)$*

$$P(A \mid \mathcal{F}_t) = P(A \mid X_t) \quad P\text{-fast sicher} \tag{6.18}$$

*gilt. Gilt (6.18) mit $\mathcal{F}_t = \mathcal{F}_t^X$ für alle $t \in \mathbb{R}^+$, so nennen wir X einfach **Markoffprozess**.*

Bemerkung 6.24

1.

   a) Da X_t bezüglich $\mathcal{F}_t$-messbar ist, ist $P(A \mid X_t)$ natürlich $\mathcal{F}_t$-messbar. Nach der Definition der bedingten Wahrscheinlichkeit ist die Bedingung (6.18) also gleichbedeutend damit, dass für alle $B \in \mathcal{F}_t$

   $$P(A \cap B) = \int_B P(A \mid X_t)\, dP$$

   gilt.

   b) Gleichung (6.18) impliziert auf die übliche Weise, dass für jede $\sigma(X_s : s \geq t,\ s \in \mathbb{R}^+)$-messbare und P-integrierbare Funktion $\varphi : \Omega \to \mathbb{R}$

   $$E(\varphi \mid \mathcal{F}_t) = E(\varphi \mid X_t) \quad P\text{-fast sicher}$$

   gilt: Ist φ eine Indikatorfunktion, so ist die Behauptung nur eine Umformulierung von (6.18). Aus der Linearität des bedingten Erwartungswertes folgt die Behauptung für einfache φ. Der Satz von Lebesgue in der Version für bedingte Erwartungswerte (Satz 3.9 c)) überträgt das Resultat auf nichtnegative, messbare φ, woraus sich schliesslich die Behauptung für integrierbare φ ergibt.

Satz 6.25
Ein $\mathbb{R}$-wertiger stochastischer Prozess X, der bezüglich einer Filtrierung $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$ unabhängige Zuwächse hat, ist ein $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$-Markoffprozess.

Beweis. Sei $t \in \mathbb{R}^+$. Es genügt, die Bedingung (6.18) für A's aus einem durchschnittstabilen Erzeugendensystem von $\sigma(X_s : s \in \mathbb{R}^+, s \geq t)$ zu beweisen. Sei $k \in \mathbb{N}$, und sei $\psi : \mathbb{R}^k \to \mathbb{R}$ eine beschränkte messbare Funktion. Seien $s_1, \ldots, s_k \in \mathbb{R}^+$ mit $t \leq s_1 < \cdots < s_k$. Wir definieren eine beschränkte messbare Funktion $\varphi : \mathbb{R} \to \mathbb{R}$ durch

$$\varphi(x) := E\left(\psi(x + (X_{s_1} - X_t), \ldots, x + (X_{s_k} - X_t))\right).$$

Nach Voraussetzung sind $(X_{s_1} - X_t, \ldots, X_{s_k} - X_t)$ und $\mathcal{F}_t$ unabhängig. Nach dem Satz von Fubini folgt für jedes $B \in \mathcal{F}_t$

$$\begin{aligned} \int_B \varphi(X_t)\, dP &= \int_B \psi(X_t + (X_{s_1} - X_t), \ldots, X_t + (X_{s_k} - X_t))\, dP \\ &= \int_B \psi(X_{s_1}, \ldots, X_{s_k})\, dP = \int_B E(\psi(X_{s_1}, \ldots, X_{s_k}) \mid \mathcal{F}_t)\, dP. \end{aligned}$$

Demzufolge ist

$$\varphi(X_t) = E(\psi(X_{s_1}, \ldots, X_{s_k}) \mid \mathcal{F}_t) \quad P\text{-fast sicher}$$

was impliziert, dass die rechte Seite dieser Gleichung $\sigma(X_t)$-messbar ist, also gilt

$$E(\psi(X_{s_1}, \ldots, X_{s_k}) \mid \mathcal{F}_t) = E(\psi(X_{s_1}, \ldots, X_{s_k}) \mid X_t) \quad P\text{-fast sicher.}$$

Hieraus folgt (6.18) für alle $A \in \sigma(X_{s_1}, \ldots, X_{s_k})$. Da für $t \geq 0$

$$\bigcup \{\sigma(X_{s_1}, \ldots, X_{s_k}) : k \in \mathbb{N},\ t \leq s_1 < \cdots < s_k\}$$

ein durchsnittstabiles Erzeugendensystem von $\sigma(X_s : s \in \mathbb{R}^+, s \geq t)$ ist, folgt die Behauptung. ■

Aus verschiedenen Gründen ist es manchmal bequem, wenn die verwendete Filtrierung, wie man sagt, rechtsstetig ist. Sei $\{\mathcal{F}_t\}_{t \in [0,\infty)}$ eine Filtrierung. Wir definieren

$$\mathcal{F}_{t+} := \bigcap_{s: s > t} \mathcal{F}_s = \bigcap_{m=1}^{\infty} \mathcal{F}_{t+1/m}.$$

Definition 6.26
Eine Filtrierung heisst ***rechtsstetig*** *falls $\mathcal{F}_{t+} = \mathcal{F}_t$ für alle $t \geq 0$ gilt.*

Bemerkung 6.27

1.

   a) Ist $\{\mathcal{F}_t\}_{t \geq 0}$ eine beliebige Filtrierung, so sieht man ganz leicht, dass $\{\mathcal{F}_{t+}\}_{t \geq 0}$ eine rechtsstetige Filtrierung ist.

   b) Die durch einen stochastischen Prozess X induzierte Filtrierung $\{\mathcal{F}_t^X\}_{t \geq 0}$ ist in der Regel nicht *rechtsstetig*, selbst wenn X stetig ist. Ist zum Beispiel $\Omega = C[0, \infty)$ und ist $\Pi = \{\pi_t\}_{t \geq 0}$ der Prozess der Auswertungsabbildungen, so ist $\{\mathcal{F}_t^\Pi\}_{t \geq 0}$ nicht rechtsstetig. So ist zum Beispiel für jedes $t \in (0, \infty)$ die Teilmenge $\{ f \in C : f$ ist differenzierbar in $t \}$ in $\mathcal{F}_{t+}^\Pi$ aber nicht in $\mathcal{F}_t^\Pi$, was der Leser sich als Übungsaufgabe überlegen möge.

Es ist wichtig, dass für eine stetige Brownsche Bewegung B zwischen $\{\mathcal{F}_t^B\}_{t\geq 0}$ und $\{\mathcal{F}_{t+}^B\}_{t\geq 0}$ kein "grosser" Unterschied ist, wie wir nun nachweisen werden.

Satz 6.28
Eine stetige Brownsche Bewegung $B = \{B_t\}_{t\in[0,\infty)}$ hat unabhängige Zuwächse bezüglich $\{\mathcal{F}_{t+}^{\mathbf{B}}\}_{t\in[0,\infty)}$.

Beweis. Es genügt zu zeigen, dass für jede Wahl von $s_1, \ldots, s_k \in (t, \infty)$ die σ-Algebra $\mathcal{F}_{t+}$ und der Vektor $(B_{s_1} - B_t, \ldots, B_{s_k} - B_t)$ der Zuwächse unabhängig sind. Um dies nachzuweisen, wollen wir zeigen, dass für jedes $A \in \mathcal{F}_{t+}^{\mathbf{B}}$ und jede beschränkte stetige Funktion $\varphi : \mathbb{R}^k \to \mathbb{R}$

$$E(1_A\varphi(X_{s_1} - X_t, \ldots, X_{s_k} - X_t)) = P(A)E(\varphi(X_{s_1} - X_t, \ldots, X_{s_k} - X_t)) \qquad (6.19)$$

gilt. Hieraus ergibt sich der Satz wie folgt: Ist $C \subset \mathbb{R}^k$ abgeschlossen, so ist durch $\varphi_n(x) = \max\{0, 1 - n\,\mathrm{dist}(x, C)\}$ für alle $x \in \mathbb{R}^k$ und $n \in \mathbb{N}$ eine monoton fallende Folge stetiger beschränkter Funktionen mit $\varphi_n \downarrow 1_C$ für $n \to \infty$ definiert, und der Satz von der majorisierten Konvergenz liefert Gleichung (6.19) für $\varphi = 1_C$. Da $\{\, C \in \mathcal{B}^k \mid (6.19) \text{ gilt für } C \,\}$ ein Dynkinsystem ist, das den durchschnittstabilen Erzeuger $\{\, C \in \mathcal{B}^k \mid C \text{ abgeschlossen} \,\}$ von $\mathcal{B}^k$ enthält, folgt (6.19) für alle $\varphi = 1_C$ mit $C \in \mathcal{B}^k$. Also sind das Ereignis A und der Vektor $(B_{s_1} - B_t, \ldots, B_{s_k} - B_t)$ unabhängig.

Für den Beweis von (6.19) sei $m_0 \in \mathbb{N}$ so gross gewählt, dass $t + 1/m_0 \leq s_j$ für alle $j \in \{1, \ldots, k\}$ gilt. Für jedes $m \geq m_0$ ist $A \in \mathcal{F}_{t+1/m}^B$, und aus Satz 6.22 folgt

$$\begin{aligned} & E(1_A\varphi(B_{s_1} - B_{t+1/m}, \ldots, B_{s_k} - B_{t+1/m})) \\ = \; & P(A)E(\varphi(B_{s_1} - B_{t+1/m}, \ldots, B_{s_k} - B_{t+1/m})). \end{aligned}$$

Da die Pfade des Prozesses B insbesondere rechtsstetig in t sind, konvergiert der Vektor $(B_{s_1} - B_{t+1/m}, \ldots, B_{s_k} - B_{t+1/m})$ gegen $(B_{s_1} - B_t, \ldots, B_{s_k} - B_t)$ für $m \to \infty$. Da φ als beschränkte stetige Funktion gewählt war, folgt (6.19) aus dieser Gleichung mit Hilfe des Satzes von der majorisierten Konvergenz. ■

Korollar 6.29
Eine stetige Brownsche Bewegung B ist ein $\{\mathcal{F}_{t+}^B\}_{t\geq 0}$-Markoffprozess.

Beweis. Folgt unmittelbar aus dem vorangegangen Satz und Satz 6.25. ■

Korollar 6.30 (Blumenthal 0-1-Gesetz)
Sei $B = \{B_t\}_{t\geq 0}$ eine stetige Brownsche Bewegung. Für jedes $A \in \mathcal{F}_{0+}^B$ gilt $P(A) \in \{0, 1\}$.

Beweis. Ist $A \in \mathcal{F}_{0+}^{\mathbf{B}}$, so folgt aus Korollar 6.29

$$1_A = P(A|\mathcal{F}_{0+}^B) = P(A|B_0) = P(A) \quad P\text{-fast sicher},$$

da $B_0 \equiv 0$ ist. Multiplikation mit 1_A liefert $1_A = 1_A 1_A = P(A)1_A$ P-fast sicher, also gilt $P(A) = E(1_A) = E(P(A)1_A) = P(A)^2$, woraus $P(A) \in \{0, 1\}$ folgt. ■

Das Blumenthalsche 0-1-Gesetz hat einige interessante Folgerungen. Hier ist ein Beispiel:

Korollar 6.31
Sei $\{B_t\}_{t\geq 0}$ eine stetige Brownsche Bewegung und sei $\varepsilon > 0$. Für P-fast alle $\omega \in \Omega$ wechselt der Pfad $[0,\infty) \ni t \mapsto B_t(\omega)$ im Intervall $[0,\varepsilon]$ unendlich oft das Vorzeichen, das heisst, mit Wahrscheinlichkeit 1 existiert eine (von ω abhängige) Folge $t_1 > t_2 > \cdots > 0$ mit $B_{t_{2k-1}}(\omega) > 0$ und $B_{t_{2k}}(\omega) < 0$ für alle $k \in \mathbb{N}$.

Beweis. Für jedes ω aus dem Komplement des im Korollar beschriebenen Ereignisses existiert eine natürliche Zahl n mit $\omega \in A_n^+ := \{B_t \geq 0 \text{ für alle } t \in [0,1/n]\}$ oder $\omega \in A_n^- := \{B_t \leq 0 \text{ für alle } t \in [0,1/n]\}$. Es gelten $A_n^+ \subset A_{n+1}^+$ und $A_n^- \subset A_{n+1}^-$ sowie

$$A_n^+ = \bigcap_{t\in\mathbb{Q}\cap[0,1/n]} \{B_t \geq 0\} \in \mathcal{F}_{1/n}^B$$

und $A_n^- \in \mathcal{F}_{1/n}^B$ für alle $n \in \mathbb{N}$. Folglich sind $A^+ := \bigcup_{n\in\mathbb{N}} A_n^+$ und $A^- := \bigcup_{n\in\mathbb{N}} A_n^-$ in $\mathcal{F}_{0+}^B$. Da nach Satz 6.16 auch $\{-B_t\}_{t\geq 0}$ eine stetige Brownsche Bewegung ist, gilt $P(A^+) = P(A^-)$. Wäre $P(A^+) = 1$, so wäre auch $P(A^-) = 1$ und demzufolge $P(A^+ \cap A^-) = 1$. Da aber $A^+ \cap A^- \subset \bigcup_{n\in\mathbb{N}}\{B_{1/n} = 0\}$ ist und $P(B_{1/n} = 0) = 0$ für jedes $n \in \mathbb{N}$ gilt, muss $P(A^+ \cap A^-) = 0$ sein. Nach Korollar 6.30 verbleibt also nur die Möglichkeit $P(A^+) = 0$, woraus $P(A^+ \cup A^-) = 0$ folgt. ■

6.5 Die starke Markoff-Eigenschaft

Es sei $B = \{B_t\}_{t\geq 0}$ eine stetige Brownsche Bewegung. Wir wollen in diesem Abschnitt die Aussage von Satz 6.28 verallgemeinern, der besagt, dass für jedes $t \in [0,\infty)$ der Prozess $\{B_{t+s} - B_t\}_{s\geq 0}$ eine von $\mathcal{F}_{t+}^B$ unabhängige Brownsche Bewegung ist. Es wird sich herausstellen, dass diese Aussage (nach einigen Präzisierungen) richtig bleibt, wenn t durch eine zufällige Zeit ersetzt wird.

Definition 6.32
*$(\Omega, \mathcal{F}, P)$ sei ein Wahrscheinlichkeitsraum und $\{\mathcal{F}_t\}_{t\geq 0}$ sei eine Filtrierung. Eine $\mathcal{F}$-messbare Abbildung $\tau : \Omega \to [0,\infty] := [0,\infty) \cup \{\infty\}$ heisst $\{\mathcal{F}_t\}$-**Stoppzeit**, wenn für alle $t \in [0,\infty)$*

$$\{\tau \leq t\} \in \mathcal{F}_t$$

gilt.

Das folgende Lemma gibt einige einfache Eigenschaften von Stoppzeiten.

Lemma 6.33

1. *τ, σ seien zwei $\{\mathcal{F}_t\}$-Stoppzeiten. Dann sind $\min(\sigma,\tau)$, $\max(\sigma,\tau)$ und $\sigma+\tau$ Stoppzeiten.*

2. *τ_n, $n \in \mathbb{N}$, seien $\{\mathcal{F}_t\}$-Stoppzeiten mit $\tau_n \uparrow \tau$ für $n \to \infty$. Dann ist τ eine Stoppzeit.*

3. *τ_n, $n \in \mathbb{N}$, seien $\{\mathcal{F}_t\}$-Stoppzeiten und $\{\mathcal{F}_t\}$ sei eine rechtsstetige Filtrierung. Dann sind $\liminf_{n\to\infty} \tau_n$, $\limsup_{n\to\infty} \tau_n$ $\{\mathcal{F}_t\}$-Stoppzeiten.*

Beweis. (a) sei dem Leser überlassen.

(b):

$$\{\tau \le t\} = \bigcap_n \{\tau_n \le t\} \in \mathcal{F}_t.$$

(c)

$$\begin{aligned}
\left\{\limsup_{n\to\infty} \tau_n \le t\right\} &= \left\{\inf_n \sup_{m\ge n} \tau_m \le t\right\} \\
&= \bigcap_{k\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} \bigcap_{m:m\ge n} \left\{\tau_m \le t + \frac{1}{k}\right\} \in \mathcal{F}_{t+} = \mathcal{F}_t.
\end{aligned}$$

Der Beweis für $\liminf_{n\to\infty} \tau_n$ verläuft analog. ■

Die wichtigsten Beispiele für Stoppzeiten sind Ersteintrittszeiten von stochastischen Prozessen. Sei $X = \{X_t\}_{t\ge 0}$ ein Prozess, der an eine Filtrierung $\{\mathcal{F}_t\}$ angepasst ist. Leider ist im allgemeinen nicht richtig, dass für jede Borel-Menge $A \in \mathcal{B}$ die Zufallszeit

$$\tau_A := \inf\{\, t : X_t \in A \,\}$$

($\inf \emptyset := \infty$) eine $\{\mathcal{F}_t\}$-Stoppzeit ist, selbst wenn X stetig und $\{\mathcal{F}_t\}$ rechtsstetig sind. Um zu gewährleisten, dass τ_A für jede Borel-Menge A eine Stoppzeit ist, muss die Filtrierung in geeigneter Weise erweitert werden. Dies führt zu lästigen masstheoretischen Diskussionen, die wir uns hier ersparen wollen. Für spezielle Mengen ist nämlich die Sache sehr viel einfacher zu beweisen.

Lemma 6.34
X sei $\{\mathcal{F}_t\}$-angepasst und habe rechtsstetige Pfade und $\{\mathcal{F}_t\}$ sei rechtsstetig.

1. *Ist A offen, so ist τ_A eine Stoppzeit.*

2. *Ist A abgeschlossen und X stetig, so ist τ_A eine Stoppzeit.*

Beweis. (a) Sei A offen. Sei $t \in [0,\infty)$, $\omega \in \Omega$.

$$\begin{aligned}
\tau_A(\omega) \le t \quad &\Leftrightarrow \quad \forall\, \varepsilon > 0\ \exists\, s < t+\varepsilon \text{ mit } X_s(\omega) \in A \\
&\Leftrightarrow \quad \forall\, \varepsilon > 0\ \exists\, q \in \mathbb{Q},\ q < t+\varepsilon \text{ mit } X_q(\omega) \in A,
\end{aligned}$$

da A offen ist und $t \to X_t\,(\omega)$ rechtsstetig ist. Somit ist

$$\{\tau_A \le t\} = \bigcap_m \underbrace{\bigcup_{q\in\mathbb{Q}\cap[0,t+1/m]} \{X_q \in A\}}_{\in\mathcal{F}_{t+1/m}} \in \mathcal{F}_{t+} = \mathcal{F}_t.$$

b) Sei nun A abgeschlossen. Für $n \in \mathbb{N}$ sei $A^{(n)}$ die offene $\frac{1}{n}$-Umgebung von A:

$$A^{(n)} := \left\{ x \in \mathbb{R} : \exists\, y \in A \text{ mit } |x-y| < \frac{1}{n} \right\}.$$

Nach dem vorher schon Bewiesenen sind die $\tau_{A^{(n)}}$ Stoppzeiten. Es gilt $\tau_{A^{(n)}} \leq \tau_A$, $\forall\, n$, und die Folge der $\tau_{A^{(n)}}$ steigt monoton an. Sei $\tau' := \lim_{n\to\infty} \tau_{A^{(n)}}$. Ist $\tau'(\omega) = \infty$, so ist auch $\tau_A(\omega) = \infty$. Ist $\tau'(\omega) < \infty$, so ist wegen der Stetigkeit der Pfade und der Abgeschlossenheit von A $X_{\tau'(\omega)}(\omega) = \lim_{n\to\infty} X_{\tau^{(n)}(\omega)}(\omega) \in A$, das heisst, es gilt $\tau_A(\omega) \leq \tau'(\omega)$. Somit ist $\tau_A = \tau'$ gezeigt, und aus Lemma 6.33 b) folgt, dass τ_A eine Stoppzeit ist. ∎

Wir definieren die σ-Algebra der prä-τ-Ereignisse wie folgt:

Definition 6.35
Sei τ eine Stoppzeit bezüglich der Filtrierung $\{\mathcal{F}_t\}_{t\geq 0}$. Dann ist

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t,\ \forall\, t\}\,.$$

Es folgt sehr einfach, dass $\mathcal{F}_\tau$ eine σ-Algebra ist. Ferner ist $\mathcal{F}_\tau = \mathcal{F}_t$ falls $\tau \equiv t$ ist.

Wir werden des öftern das folgende Resultat benutzen:

Lemma 6.36

a) σ, τ seien zwei Stoppzeiten mit $\sigma(\omega) \leq \tau(\omega)$, $\forall\, \omega$. Dann gilt $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

b) $\{\mathcal{F}_t\}$ sei rechtsstetig. Dann ist $\mathcal{F}_\tau = \bigcap_m \mathcal{F}_{\tau+1/m}$.

Beweis. a) Sei $A \in \mathcal{F}_\sigma$. Dann gilt für alle $t \geq 0$:

$$A \cap \{\tau \leq t\} = A \cap \{\sigma \leq t\} \cap \{\tau \leq t\} \in \mathcal{F}_t,$$

da $A \cap \{\sigma \leq t\} \in \mathcal{F}_t$ und $\{\tau \leq t\} \in \mathcal{F}_t$ sind.

b) $\tau + 1/m$ ist eine Stoppzeit, wie man leicht nachprüft. Wegen a) folgt $\mathcal{F}_\tau \subset \bigcap_m \mathcal{F}_{\tau+1/m}$. Sei $A \in \bigcap_m \mathcal{F}_{\tau+1/m}$. Dann ist

$$A \cap \{\tau \leq t\} = A \cap \left\{\tau + \frac{1}{m} \leq t + \frac{1}{m}\right\} \in \mathcal{F}_{t+1/m}$$

$\forall\, m$, das heisst $A \cap \{\tau \leq t\} \in \mathcal{F}_{t+} = \mathcal{F}_t$. ∎

Sei $X = \{X_t\}_{t\geq 0}$ ein $\{\mathcal{F}_t\}$-angepasster Prozess und τ eine Stoppzeit. Wir wollen nun den Prozess X zum zufälligen Zeitpunkt τ betrachten: $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$. X_τ sollte natürlicherweise $\mathcal{F}_\tau$-messbar sein. Zunächst ergibt sich die Schwierigkeit, dass $\tau(\omega)$ durchaus ∞ sein kann, und $X_\infty(\omega)$ nicht definiert ist.

Lemma 6.37
Sei X rechtsstetig und angepasst an eine rechtsstetige Filtrierung $\{\mathcal{F}_t\}_{t\geq 0}$. Dann ist X_τ $\mathcal{F}_\tau$-messbar auf $\{\tau < \infty\}$, das heisst, für jede Borelmenge $A \in \mathcal{B}$ gilt

$$\{X_\tau \in A\} \cap \{\tau < \infty\} \in \mathcal{F}_\tau.$$

Beweis. Zunächst ist $\{\tau < \infty\} \in \mathcal{F}_\tau$, denn $\{\tau < \infty\} \cap \{\tau \leq t\} = \{\tau \leq t\} \in \mathcal{F}_t$, $\forall\, t$.

Wir approximieren nun τ von rechts durch eine Folge von Zufallsgrössen $\tau^{(n)}$, die nur abzählbare Wertebereiche haben. Dazu definieren wir

$$\tau^{(n)}(\omega) := \begin{cases} k2^{-n} & \text{für } \tau(\omega) \in \left(\frac{k-1}{2^n}, \frac{k}{2^n}\right], k \in \mathbb{N}_0, \\ \infty & \text{für } \tau(\omega) = \infty. \end{cases}$$

Die $\tau^{(n)}$ sind Stoppzeiten, denn für $t \geq 0$ gilt

$$\{\tau^{(n)} \leq t\} = \{\tau \leq 2^{-n}[2^n t]\} \in \mathcal{F}_{2^{-n}[2^n t]} \subset \mathcal{F}_t.$$

Für jedes $\omega \in \Omega$ fällt $\tau^{(n)}(\omega)$ monoton gegen $\tau(\omega)$.

Da X als rechtsstetig vorausgesetzt ist, folgt

$$X_\tau = \lim_{n\to\infty} X_{\tau_n}$$

auf $\{\tau < \infty\} = \{\tau_n < \infty\}$.

Da die Stoppzeit τ_n nur abzählbar viele Werte annimmt, folgt leicht, dass X_{τ_n} auf $\{\tau_n < \infty\}$ $\mathcal{F}_{\tau_n}$-messbar ist: Für $A \in \mathcal{B}$ und $t \in [0,\infty)$ ist

$$\{X_{\tau_n} \in A\} \cap \{\tau_n \leq t\} = \bigcup_{k: k2^{-n} \leq t} \{X_{k2^{-n}} \in A, \tau_n = k2^{-n}\} \in \mathcal{F}_t,$$

und somit ist $\{X_{\tau_n} \in A\} \cap \{\tau_n < \infty\} \in \mathcal{F}_{\tau_n}$, das heisst, X_{τ_n} ist auf $\{\tau_n < \infty\}$ $\mathcal{F}_{\tau_n}$-messbar. Für jedes $\varepsilon > 0$ gilt $\tau_n \leq \tau + \varepsilon$, sofern $2^{-n} \leq \varepsilon$ ist. $\tau + \varepsilon$ ist ebenfalls eine Stoppzeit, und es gilt $\mathcal{F}_{\tau_n} \subset \mathcal{F}_{\tau+\varepsilon}$ für $2^{-n} \leq \varepsilon$ (Lemma 6.36 (a)). Demzufolge ist $X_\tau = \lim_{n\to\infty} X_{\tau_n}$ für jedes $\varepsilon > 0$ $\mathcal{F}_{\tau+\varepsilon}$-messbar auf $\{\tau < \infty\}$, und somit auch $\mathcal{F}_{\tau+} := \bigcap_m \mathcal{F}_{\tau+1/m}$-messbar.

Da $\{\mathcal{F}_t\}$ als rechtsstetig vorausgesetzt war, gilt $\mathcal{F}_{\tau+} = \mathcal{F}_\tau$ (Lemma 6.36 (b)). Wir haben somit gezeigt, dass X_τ $\mathcal{F}_\tau$-messbar auf $\{\tau < \infty\}$ ist. ∎

Wir kehren nun zu einer stetigen Brownschen Bewegung $B = \{B_t\}_{t\geq 0}$ zurück. Die uns hier interessierende Filtrierung ist $\mathcal{F}_t = \mathcal{F}^B_{t+}$, $t \geq 0$. Wir werden für den Rest des Kapitels stillschweigend stets mit dieser Filtrierung arbeiten.

τ sei eine $\{\mathcal{F}_t\}$-Stoppzeit, und wir setzen voraus, dass $P(\tau < \infty) > 0$ ist. Wir definieren $(\Omega', \mathcal{F}', P')$ als die Einschränkung von $(\Omega, \mathcal{F}, P)$ auf $\Omega' := \{\tau < \infty\}$, $\mathcal{F}' := \{\, A' \in \mathcal{F} : A' \subset \{\tau < \infty\}\,\}$, $P'(A') := P(A' | \tau < \infty)$. B_τ ist eine auf $(\Omega', \mathcal{F}', P')$ definierte Zufallsgrösse und $\{B_{\tau+s} - B_\tau\}_{s\geq 0}$ ein stetiger stochastischer Prozess.

Satz 6.38
$\mathbf{B}^{(\tau)} := \{B_{\tau+s} - B_\tau\}_{s\geq 0}$ *ist eine Brownsche Bewegung, die unabhängig von* $\mathcal{F}_\tau$ *ist (bezüglich* P'*).*

Beweis. Wir fassen $B^{(\tau)}$ als messbare Abbildung $\Omega' = \{\tau < \infty\} \to C[0,\infty)$ auf.

Sei $\psi : C[0,\infty) \to \mathbb{R}$ eine beschränkte messbare Abbildung. Wir beweisen, dass für jedes derartige ψ und für jede beschränkte $\mathcal{F}_\tau$-messbare Funktion $\xi : \Omega \to \mathbb{R}$ die

Gleichung $E'(\psi(B^{(\tau)})\xi) = E(\psi(B))E'(\xi)$ gilt. Daraus folgt im Spezialfall $\xi = 1$, dass $E'(\psi(B^{(\tau)})) = E(\psi(B))$ gilt, also dass $B^{(\tau)}$ unter P' eine Brownsche Bewegung ist, und damit wiederum, dass $E'(\psi(B^{(\tau)})\xi) = E(\psi(B))E'(\xi) = E'(\psi(B^{(\tau)}))E'(\xi)$ gilt, was äquivalent zur Unabhängigkeit ist. Wegen $P'(\cdot) = P'(\,\cdot\,|\tau < \infty)$ ist das gleichbedeutend mit

$$E\left(\psi(B^{(\tau)})\xi; \tau < \infty\right) = E(\psi(B))E(\xi; \tau < \infty). \tag{6.20}$$

In dieser Formulierung brauchen wir $P(\tau < \infty) > 0$ nicht vorauszusetzen, da für $P(\tau < \infty) = 0$ die Gleichung trivial ist. Wie üblich genügt es, die Gleichung (6.20) für spezielle Funktionen ψ zu zeigen, nämlich für stetige und beschränkte Funktionen $\psi : C[0,\infty) \to \mathbb{R}$, die nur von endlich vielen Stellen der Elemente $f \in C[0,\infty)$ abhängen. Wir betrachten also ψ's der Form

$$\psi(f) = \varphi(f(s_1), f(s_2), \dots, f(s_k)),$$

wobei $\varphi : \mathbb{R}^k \to \mathbb{R}$ stetig und beschränkt ist. Ferner reicht es, $\xi = 1_A$, $A \in \mathcal{F}_\tau$ zu betrachten. Die nachzuweisende Gleichung (6.20) hat dann die folgende Form: Sei $X_t := B_{\tau+t} - B_\tau$.

$$\int_{\{A,\tau<\infty\}} \varphi(X_{s_1}, \dots, X_{s_k})\, dP = E(\varphi(B_{s_1}, \dots, B_{s_k}))P(A, \tau < \infty). \tag{6.21}$$

Wir approximieren nun die Stoppzeit τ auf dieselbe Weise wie im Beweis von Lemma 6.36 durch die Folge der Stoppzeiten τ_n, die Werte in $2^{-n}\mathbb{N}_0$ annehmen, und für die $\tau_n(\omega) \downarrow \tau(\omega)$, $n \to \infty$ gilt. Sei

$$X_t^{(n)} := B_{\tau_{n+t}} - B_{\tau_n}.$$

Für alle $A \in \mathcal{F}_{\tau_n}$ gilt

$$\begin{aligned}
&\int_{\{A,\tau_n<\infty\}} \varphi(X_{s_1}^{(n)}, \dots, X_{s_k}^{(n)})\, dP \\
= &\sum_{j=0}^{\infty} \int_{\{A,\tau_n=j2^{-n}\}} \varphi(B_{j2^{-n}+s_1} - B_{j2^{-n}}, \dots, B_{j2^{-n}+s_k} - B_{js^{-n}})\, dP \\
= &\sum_{j=0}^{\infty} P(A, \tau_n = j2^{-n})E(\varphi(B_{s_1}, \dots, B_{s_k})) = P(A, \tau_n < \infty)E\varphi((B_{s_1}, \dots, B_{s_k})).
\end{aligned}$$

Die zweitletzte Gleichung folgt wegen $A \cap \{\tau_n = j2^{-n}\} \in \mathcal{F}_{js^{-n}}$ (da $A \in \mathcal{F}_{\tau_n}$) und Satz 6.22.

Sei nun $A \in \mathcal{F}_\tau$. Wegen $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$ folgt $A \in \mathcal{F}_{\tau_n}, \forall n$. Man beachte ferner $\{\tau < \infty\} = \{\tau_n < \infty\}$. Lässt man daher in der obigen Folge von Gleichungen n gegen ∞ streben, so bleibt die rechte Seite unabhängig von n, und

$$\lim_{n\to\infty} \int_{\{A,\tau_n<\infty\}} \varphi\left(X_{s_1}^{(n)}, \dots, X_{s_k}^{(n)}\right) dP = \int_{\{A,\tau<\infty\}} \varphi(X_{s_1}, \dots, X_{s_k})\, dP$$

nach dem Satz von Lebesgue (da φ stetig ist). Damit ist (6.21) bewiesen. ■

Der Satz 6.38 impliziert eine Version der sogenannten **starken Markoffeigenschaft**:

Korollar 6.39 (Starke Markoff-Eigenschaft)
τ sei wieder eine Stoppzeit. Sei ferner $\mathcal{Z}_\tau := \sigma\left(B_{\tau+s} : s \geq 0\right)$. Dann gilt für $A \in \mathcal{Z}_\tau$:

$$P\left(A \mid \mathcal{F}_\tau\right) = P\left(A \mid B_\tau\right), \ P\text{−f.s. auf } \{\tau < \infty\}.$$

Der Beweis geht analog zum Beweis der einfachen Markoff-Eigenschaft.

Die Restriktion auf $\{\tau < \infty\}$ in den Formulierungen ist natürlich etwas lästig. In vielen Büchern geht man dem aus dem Weg, indem man den Wertebereich des Prozesses - bei uns $\mathbb{R}$ - durch ein sogenanntes “Grab” Δ ergänzt. Man setzt einfach $B_\infty := \Delta$ und B_τ ist dann auf ganz Ω definiert.

Wir diskutieren nun einige wichtige Anwendungen. Zunächst benötigen wir ein analytisches Lemma über Laplacetransformationen, das wir nicht beweisen.

Ist $f : \mathbb{R}^+ := [0, \infty) \to \mathbb{R}$ eine integrierbare oder beschränkte, messbare Funktion, so ist die **Laplacetransformation** von f definiert durch

$$L_f(\lambda) = \int_{\mathbb{R}^+} \mathrm{e}^{-\lambda x} f(x)\, dx, \quad \lambda > 0.$$

Nach dem Satz von Lebesgue ist L_f stetig, und es gelten

$$\lim_{\lambda\to\infty} L_f(\lambda) = 0, \qquad \lim_{\lambda\to 0} L_f(\lambda) = \int_{\mathbb{R}^+} f(x)\, dx, \tag{6.22}$$

letzteres, sofern f integrierbar ist (was natürlich nicht aus der Beschränktheit folgt). Das nachfolgende analytische Result soll hier nicht bewiesen werden.

Lemma 6.40
Zwei beschränkte messbare oder integrierbare Funktionen, deren Laplacetransformierte auf $(0, \infty)$ übereinstimmen, sind Lebesgue-fast überall gleich.

Ist μ ein endliches Mass auf $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$, so ist die **Laplacetransformierte** von μ definiert durch

$$L_\mu(\lambda) := \int_{\mathbb{R}^+} \mathrm{e}^{-\lambda x} \mu(dx), \quad \lambda \geq 0.$$

Hat μ eine Dichte f bezüglich des Lebesgue Masses, so gilt natürlich $L_\mu = L_f$. $L_\mu(\lambda)$ ist ebenfalls stetig in $\lambda \in [0, \infty)$, und es gilt

$$\lim_{\lambda\to\infty} L_\mu(\lambda) = \mu(\{0\}), \qquad \lim_{\lambda\to 0} L_\mu(\lambda) = \mu(\mathbb{R}^+). \tag{6.23}$$

Lemma 6.41
Seien μ, ν zwei endliche Masse. Falls $L_\mu(\lambda) = L_\nu(\lambda)$, $\lambda \geq 0$, so folgt $\mu = \nu$.

Beweis. Wir benützen Lemma 6.40. Sei $\lambda > 0$

$$\begin{aligned}
\int_0^\infty \mathrm{e}^{-\lambda x} \mu([0, x])\, dx &= \int_0^\infty \mathrm{e}^{-\lambda x} \left(\int 1_{[0,x]}(y)\, \mu(dy)\right) dx \\
&= \int_{[0,\infty)} \mu(dy) \int_y^\infty \mathrm{e}^{-\lambda x}\, dx = \frac{1}{\lambda} L_\mu(\lambda) = \frac{1}{\lambda} L_\nu(\lambda) \\
&= \int_0^\infty \mathrm{e}^{-\lambda x} \nu([0, x])\, dx.
\end{aligned}$$

Aus Lemma 6.40 folgt $\mu([0,x]) = \nu([0,x])$ für Lebesgue-fast alle $x \in [0,\infty)$. Da die beiden Funktionen rechtsstetig in x sind, folgt die Gleichheit für alle $x \geq 0$. Daraus ergibt sich $\mu = \nu$. ■

Als erste Anwendung der starken Markoff-Eigenschaft berechnen wir die Verteilung von einfachen *Ersteintrittszeiten.* Sei $a > 0$ und

$$\tau_a := \inf\{\, t \geq 0 : B_t = a \,\}.$$

Nach Lemma 6.34 ist τ_a eine Stoppzeit. Im Moment wissen wir noch nicht, ob $P(\tau_a < \infty) = 1$ gilt (es wird sich gleich herausstellen, dass das richtig ist), wir können jedoch dennoch von der Verteilung von τ_a sprechen. Wir wollen diese nun berechnen. Zu diesem Zweck berechnen wir die Laplacetransformierte von $P(\tau_a \leq x)$:

Wir wenden nun (6.20) auf die folgende Funktion $\psi : C[0,\infty) \to \mathbb{R}$ an. Für $\lambda > 0$ sei

$$\psi(f) := \int_0^\infty \mathrm{e}^{-\lambda s} 1_{f(s) \geq 0}\, ds.$$

ψ ist messbar, was der Leser selbst nachweisen möge. ψ ist natürlich beschränkt (für $\lambda > 0$). Wir wählen ferner $\xi := e^{-\lambda \tau_a}$. Es folgt:

$$E\left(\mathrm{e}^{-\lambda\tau_a} \int_0^\infty \mathrm{e}^{-\lambda s} 1_{B_{\tau_a+s} - B_{\tau_a} \geq 0}\, ds\right) = E\left(\mathrm{e}^{-\lambda\tau_a}\right) E \int_0^\infty \mathrm{e}^{-\lambda s} 1_{B_s \geq 0}\, ds. \tag{6.24}$$

(Auf den Einschluss von $\{\tau_a < \infty\}$ in der Gleichung können wir verzichten, wenn wir $\mathrm{e}^{-\lambda\infty}$ als 0 interpretieren). Wir werten nun beiden Seiten der Gleichung weiter aus: Zunächst die rechte Seite: Nach Fubini ist

$$E \int_0^\infty \mathrm{e}^{-\lambda s} 1_{B_s \geq 0}\, ds = \int_0^\infty \mathrm{e}^{-\lambda s} P(B_s \geq 0)\, ds = \int_0^\infty \mathrm{e}^{-\lambda s} \frac{1}{2}\, ds = \frac{1}{2\lambda}.$$

Ferner ist

$$E\left(\mathrm{e}^{-\lambda\tau_a}\right) = \lambda E \int_{\tau_a}^\infty \mathrm{e}^{-\lambda s}\, ds = \lambda E \int_0^\infty 1_{\{\tau_a \leq s\}} \mathrm{e}^{-\lambda s}\, ds = \lambda \int_0^\infty \mathrm{e}^{-\lambda s} P(\tau_a \leq s)\, ds.$$

In der linken Seite von (6.24) substituieren wir $t = \tau_a + s$ und beachten, dass $B_{\tau_a} = a$ auf $\{\tau_a < \infty\}$ ist:

$$\begin{aligned} E\left(\mathrm{e}^{-\lambda\tau_a} \int_0^\infty \mathrm{e}^{-\lambda s} 1_{B_{\tau_a+s} - B_{\tau_a} \geq 0}\right) &= E \int_{\tau_a}^\infty \mathrm{e}^{-\lambda t} 1_{B_t \geq a}\, dt = E \int_0^\infty \mathrm{e}^{-\lambda t} 1_{B_t \geq a}\, dt \\ &= \int_0^\infty \mathrm{e}^{-\lambda t} P(B_t \geq a)\, dt. \end{aligned}$$

Setzen wir diese Umformungen in (6.24) ein, so ergibt sich

$$\int_0^\infty \mathrm{e}^{-\lambda t} P(\tau_a \leq t)\, dt = 2 \int_0^\infty \mathrm{e}^{-\lambda t} P(B_t \geq a)\, dt.$$

$P(\tau_a \leq t)$ ist rechtsstetig in t, und $P(B_t \geq a) = \int_a^\infty \frac{1}{\sqrt{2\pi t}} \mathrm{e}^{-x^2/2t}\, dx$ ist stetig. Aus Lemma 6.40 folgt daher

Satz 6.42
Für alle $a > 0$, $t \geq 0$ gilt

$$P(\tau_a \leq t) = 2P(B_t \geq a) = 2\int_a^\infty \frac{1}{\sqrt{2\pi t}} \mathrm{e}^{-x^2/2t}\, dx.$$

Beweis. $P(\tau_a < \infty) = \lim_{t\to\infty} P(\tau_a \leq t) = \lim_{t\to\infty} 2P(B_t \geq a) = 1.$ ∎

Als weitere Anwendung der starken Markoff-Eigenschaft betrachten wir die Nullstellenmenge $N(\omega) := \{\, t \geq 0 : B_t(\omega) = 0 \,\}$. Für eine stetige Brownsche Bewegung ist $N(\omega)$ natürlich abgeschlossen. Aus Korollar 6.31 wissen wir, dass mit Wahrscheinschlichkeit 1 der Punkt $t = 0$, der natürlich in $N(\omega)$ ist, Häufungspunkt von anderen Nullstellen ist.

Eine abgeschlossene nicht leere Teilmenge $A \subset \mathbb{R}$ heisst **perfekt***,* falls jeder Punkt $p \in A$ Häufungspunkt von anderen Punkten aus A ist, das heisst, wenn für alle $p \in A$ die Menge A gleich dem Abschluss von $A \setminus \{p\}$ ist.

Es ist bekannt, dass jede nicht leere perfekte abgeschlossene Teilmenge von $\mathbb{R}$ überabzählbar ist. Dies soll hier nicht bewiesen werden.

Satz 6.43
Die Nullstellenmenge $N(\omega)$ einer stetigen Brownschen Bewegung ist für fast alle ω eine perfekte Menge von Lebesgue Mass 0.

Beweis. Wir wollen nicht nachweisen, dass $\{\,\omega : N(\omega) \text{ ist perfekt}\}$ messbar ist. Wir zeigen nur: Es existiert $\Omega_0 \in \mathcal{F}$ mit $P(\Omega_0) = 1$, so dass $N(\omega)$ perfekt ist für alle $\omega \in \Omega_0$.

Ist eine abgeschlossene, nicht leere Teilmenge $N \subset [0, \infty)$ nicht perfekt, so existiert ein Interval $[q_1, q_2)$, $q_1, q_2 \in \mathbb{Q}$, so dass $N \cap [q_1, q_2)$ genau einen Punkt enthält. Wir definieren daher für $q_1 < q_2$, $q_1, q_2 \in \mathbb{Q}$:

$$A_{q_1,q_2} := \{\omega : t \to B_t(\omega) \text{ hat genau eine Nullstelle in } [q_1, q_2)\,\}.$$

Offensichtlich ist $N(\omega)$ perfekt für $\omega \in (\bigcup_{q_1<q_2} A_{q_1,q_2})^c$. Es genügt daher zu zeigen, dass $P(A_{q_1,q_2}) = 0$ für alle $q_1 < q_2$ ist. Wir halten q_1,q_2 fest und definieren

$$\tau := \inf\{\, t \in [q_1, q_2) : B_t = 0 \,\}.$$

τ ist eine Stoppzeit, wie man leicht nachprüft. Nun ist offensichtlich

$$\begin{aligned} A_{q_1,q_2} \quad &\subset \quad \bigcup_m \left(\left\{ B_{\tau+s} < 0, 0 < s \leq \frac{1}{m} \right\} \cap \{\tau < \infty\} \right) \\ &\quad \cup \bigcup_m \left(\left\{ B_{\tau+s} > 0, 0 < s \leq \frac{1}{m} \right\} \cap \{\tau < \infty\} \right). \end{aligned}$$

Eine Anwendung des Satzes 6.38 und von Korollar 6.31 ergibt

$$P\left(B_{\tau+s} < 0, 0 < s \leq \frac{1}{m}, \tau < \infty \right) = P(\tau < \infty) P\left(B_s < 0, 0 < s \leq \frac{1}{m} \right) = 0.$$

Analog natürlich

$$P\left(B_{\tau+s} > 0, 0 < s \leq \frac{1}{m}, \tau < \infty\right) = 0.$$

Somit folgt $P(\bigcup_{q_1<q_2\in\mathbb{Q}} A_{q_1,q_2}) = 0$.

Es bleibt noch zu zeigen, dass für fast alle ω die Nullstellenmenge $N(\omega)$ Lebesgue Mass 0 hat. Sei λ das Lebesgue Mass auf $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$. Dann gilt nach Fubini:

$$\int \lambda(N(\omega))\, P(d\omega) = \int \int_{\mathbb{R}^+} 1_{\{B_t(\omega)=0\}}\, dt\, P(d\omega) = \int_{\mathbb{R}^+} P(B_t = 0)\, dt = 0.$$

Wegen $\lambda(N(\omega)) \geq 0$ folgt $\lambda(N(\omega)) = 0$ für P-fast alle $\omega \in \Omega$. $\blacksquare$