Wahrscheinlichkeitstheorie

Matthias Löwe

1 Einleitung

Glück, Zufall und Schicksal haben seit jeher im Zentrum des menschlichen Interesses gestanden. Spätestens seit der Erfindung des Glücksspiels und seiner systematischen Durchführung um größere Geldbeträge gibt es den Wunsch nach einem systematischen, wissenschaftlichen Studium des Zufalls und seiner Gesetze. Diese Erforschung geschieht in der Mathematik durch die Disziplin der Stochastik.
Schon die Erfahrung lehrt uns, dass es selbst in chaotischen und rein zufälligen Prozessen Regelmäßigkeiten gibt: Selbst wenn man nicht weiß, was der nächste Wurf einer fairen Münze bringen wird, so "weiß" man doch, dass auf lange Sicht gesehen in etwa 50% aller Fälle "Kopf" und in den anderen 50% "Zahl" fallen wird. Dieses Wissen hat Jakob Bernoulli 1713 (posthum) in seiner "Doctrine of Chance" bewiesen und veröffentlicht. Es ist die erste Formulierung eines "Gesetzes der großen Zahlen". Wenig später analysierte der französische Mathematiker de Moivre die Fluktuationen um dieses Gesetz der großen Zahlen und erhielt die berühmte Gaußsche Glockenkurve (die früher auf dem 10 DM-Schein abgedruckt war). Dies ist die erste Version dessen, was heute allgemeiner als der zentrale Grenzwertsatz bekannt ist.
Nichtsdestotrotz war es für diese Mathematiker und ihre Nachfolger nicht einfach, ihre Resultate mathematisch exakt zu formuieren. Dies lag in erster Linie daran, dass man, bevor man daran geht zu formulieren, was die Regelmäßigkeit in einem zufälligen Experiment ist, zunächst beschreiben muss, was denn ein zufälligen Experiment, was denn Zufall, was denn Wahrscheinlichkeit ist.
Eine intuitive Herangehensweise würde sich das obige "Wissen" (gemeint ist, das Gesetz der großen Zahlen, dass ja ohne eine Fundierung der Wahrscheinlichkeitstheorie noch nicht auf sicherem Boden steht) zunutze machen und es als Axiom postulieren. Die Wahrscheinlichkeit eines Ereignisses würde dann definiert als der Limes der relativen Häufigkeiten seines Eintretens in einer sehr langen Versuchsreihe. Das Problem in dieser Definition liegt darin, dass dieser Limes der relativen Häufigkeiten nur typischerweise gegen die zugrunde liegende Wahrscheinlichkeit konvergieren wird: Dass man in einem Bernoulli-Experiment der Länge n (wobei wir uns n sehr groß vorstellen) mit Erfolgswahrscheinlichkeit $p = \frac{1}{2}$ nur Erfolge beobachten ist durchaus möglich, es ist nur extrem unwahrscheinlich. Um dieses typischerweise zu mathematisieren bräuchte man also am besten schon den Begriff der Wahrscheinlichkeit, der ja erst erklärt werden soll. Unabhängig von dieser Schwierigkeit stoßen wir

bei einer solchen Modellierung von Wahrscheinlichkeit als Limes der relativen Häufigkeiten auf ein zweites, technisches Problem: wollen wir mit einer solchen Definition ein Theorem beweisen, müssen wir etwas über alle typischen Versuchsausgänge in einer sehr langen Reihe von Experimenten aussagen; wie man sich leicht vorstellen kann, ist dies nicht gerade eine einfache Aufgabe, so dass selbst relativ "übersichtliche" mathematische Sachverhalte einen schwierigen Beweis haben können.
Aus diesen Gründen formulierte Hilbert bei seinem erinnerungswürdigen Vortrag 1900 auf dem Weltkongress in Paris die axiomatische Begründung der Wagrscheinlichkeitstheorie (und Physik (!)) als das sechste seiner berühmt gewordenen 23 Probleme. Der entsprechende Passus aus dem Hilbertschen Vortrag zeigt auch heute noch die enge Verwandschaft von Stochastik und mathematischer Physik: "Durch die Untersuchungen über die Grundlagen der Geometrie wird uns die Aufgabe nahegelegt, nach diesem Vorbilde diejenigen physikalischen Disciplinen axiomatisch zu behandeln, in denen schon heute die Mathematik eine hervorragende Rolle spielt; dies sind in erster Linie die Wahrscheinlichkeitsrechnung und die Mechanik.
Was die Axiome der Wahrscheinlichkeitsrechnung (Vgl. Bohlmann, Ueber Versicherungsmathematik 2te Vorlesung aus Klein und Riecke, Ueber angewandte Mathematik und Physik, Leipzig und Berlin 1900) angeht, so scheint es mir wünschenswert, daß mit der logischen Untersuchung derselben zugleich eine strenge und befriedigende Entwickelung der Methode der mittleren Werte in der mathematischen Physik, speciell in der kinetischen Gastheorie Hand in Hand gehe."
Darauf folgt ein längerer Abschnitt über weitere Probleme der Physik, die einer Mathematisierung bedürfen.

Das Problem der Axiomatisierung der Wahrscheinlichkeitstheorie wurde 1933 durch A. N. Kolmogorov gelöst. Er verband die Wahrscheinlichkeitstheorie mit dem damals jungen Feld der Maßtheorie. Diese Verbindung löst das Problem der Axiomatisierung der Wahrscheinlichkeitstheorie auf beinahe geniale Art und Weise: Eine Wahrscheinlichkeit ist nach Kolmogorov nichts anderes als ein Maß auf der Menge aller möglichen Versuchsausgänge eines zufälligen Experiments. Um auch die relativen Haäufigkeiten bei mehrfacher Ausführung des Experiments zu einem solchen Maß zu machen, fordert man weiter, dass die Gesamtmasse des Maßes eins sein soll.

Von diesem Ausgangspunkt entwickelte sich die gesamte moderne Wahrscheinlichkeitstheorie: Angefangen von den Gesetzen der großen Zahlen und dem Zentralen Grenzwertsatz, die zumindest in Spezialfällen schon lange bekannt waren, über die Theorie stochastischer Prozesse, bis hin zum sehr jungen Gebiet der Finanzmathematik (und vieles, vieles mehr).

Dieser Kurs soll die wesentlichen Konzepte der Wahrscheinlichkeitstheorie aufzeigen. Zunächst sollen die dafür notwendigen Grundlagen der Maßtheorie wiederholt werden und typische wahrscheinlichkeitstheoretische Konzepte der Maßtheorie neu vorgestellt werden.
Auf dieser Grundlage wollen wir dann die zentralen Sätze der Wahrscheinlichkeitstheorie entwickeln: Gesetze der großen Zahlen, Zentrale Grenzwertsätze, Prinizipien großer Abweichungen und Ergodensätze. Dies geschieht natürlicherweise zunächst für den Fall unabhänginger Zufallsgrößen. Abhängige Zufallsvariablen kommen in der Wahrscheinlichkeitstheorie häufig in der Form sogenannter stochastischer Prozesse vor. Diese sollen im zweiten Teil der Vorlesung etwas genauer studiert werden. Insbesondere wollen wir dort den stochastischen Prozess, die Brownsche Bewegung, kennenlernen, und sehen, wie er aus bekannten Prozeesen als Limesprozess gewonnen werden kann.

2 Grundlagen der Maßtheorie

2.1 Mengensysteme

Wie schon in der Einleitung erwähnt, basiert die Modellierung eines stochastischen Phänomens auf der Angabe eines Tripels $(\Omega, \mathcal{F}, \mathbb{P})$. Wir nennen $(\Omega, \mathcal{F}, \mathbb{P})$ einen Wahrscheinlichkeitsraum, falls die folgenden Bedingungen erfüllt sind.

Definition 2.1 *Ein Wahrscheinlichkeitsraum ist ein Tripel* $(\Omega, \mathcal{F}, \mathbb{P})$*, so dass*

- Ω *eine Menge ist*
- $\mathcal{F}$ *eine* σ*–Algebra über* Ω *ist und*
- $\mathbb{P}$ *ein Maß auf* $\mathcal{F}$ *mit* $\mathbb{P}(\Omega) = 1$ *ist.*

Um diese Definition mit Leben zu füllen, müssen wir allererst definieren, was eine σ–Algebra und was ein Maß über einer σ–Algebra sein sollen und was ihre Eigenschaften sind. Die Idee der Einführung einer σ–Algebra besteht darin, dass sie sämtliche Mengen versammelt, die wir messen wollen

(bzw. mit obiger Interpretation, denen wir eine Wahrscheinlichkeit zuerkennen wollen). Wünschenswert wäre es natürlich $\mathcal{F} = \mathcal{P}(\Omega)$, also die σ–Algebra gleich der Potenzmenge der zugrunde liegenden Menge zu wählen. Leider ist es nicht immer möglich, Maße mit gewissen zusätzlichen Eigenschaften auf der Potenzmenge von Ω zu definieren (wir werden dies an anderer Stelle sehen). Eine σ-Algebra ist gewissermaßen eine Minimalforderung, die wir an die Menge der Teilmengen von Ω stellen wollen, deren Maß wir bestimmen können wollen.
Natürlich wollen wir stets Ω messen können. Außerdem wollen wir mit einer Menge $A \subset \Omega$ immer auch ihr Komplement A^c messen können. Schlußenlich wollen wir mit einer Folge von $(A_n)_{n\in\mathbb{N}}$, $A_n \subset \Omega$, auch ihre Vereinigung $\bigcup_{n\in\mathbb{N}} A_n$ messen können.
Dies führt zu folgender Definition:

Definition 2.2 *Ein Mengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt σ–Algebra über Ω, falls gilt*

$$\Omega \in \mathcal{A} \tag{2.1}$$

$$A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A} \tag{2.2}$$

$$\textit{Falls } A_n \in \mathcal{A} \textit{ für } n = 1, 2, \ldots \textit{ dann ist auch } \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}. \tag{2.3}$$

Beispiele für σ–Algebren sind

Beispiel 2.3 *1. $\mathcal{P}(\Omega)$ ist eine σ-Algebra.*

2. Ist $\mathcal{A}$ eine σ–Algebra über Ω und $\Omega' \subseteq \Omega$, dann ist

$$\mathcal{A}' := \{\Omega' \cap A : A \in \mathcal{A}\}$$

eine σ–Algebra über Ω' (die sogenannten Spur-σ-Algebra).

3. Sind Ω, Ω' Mengen und ist $\mathcal{A}$ eine σ–Algebra über Ω, ist ferner

$$T : \Omega \to \Omega'$$

eine Abbildung. Dannn ist

$$\mathcal{A}' := \left\{A' \subset \Omega' : T^{-1}[A'] \in \mathcal{A}\right\}$$

eine σ–Algebra über Ω'.

Bemerkung 2.4 *Die Vereinigung zweier σ–Algebren (über derselben Menge Ω), d.h.*

$$\mathcal{A}_1 \cup \mathcal{A}_2 := \{A \in \mathcal{A}_1 \text{ oder } A \in \mathcal{A}_2\}$$

is i.a. keine σ–Algebra.

Dagegen ist für eine beliebige Indexmenge I und σ–Algebren $\mathcal{A}_i, i \in I$ über derselben Menge Ω

$$\bigcap_{i \in I} \mathcal{A}_i$$

stets eine σ–Algebra.
Dies impliziert, dass für ein Mengensystem $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ eine kleinste σ–Algebra $\sigma(\mathcal{E})$ existiert, die $\mathcal{E}$ enthält. In der Tat muss man nur

$$\mathcal{S} := \{\mathcal{A} \text{ is a } \sigma\text{-algebra, } \mathcal{E} \subset \mathcal{A}\}$$

betrachten. Dann ist

$$\sigma(\mathcal{E}) = \bigcap_{\mathcal{A} \in \mathcal{S}} \mathcal{A}$$

eine σ–Algebra. Offensichtlich gilt $\mathcal{E} \subseteq \sigma(\mathcal{E})$ und $\sigma(\mathcal{E})$ ist minimal mit dieser Eigenschaft. $\mathcal{E}$ heißt auch der Erzeuger von $\sigma(\mathcal{E})$.

Oftmals wird ein Erzeuger auch selbst schon eine Struktur tragen. Diese bekommen eigene Namen:

Definition 2.5 *Ein Mengensystem $\mathcal{R} \subset \mathcal{P}(\Omega)$ heißt Ring, falls es den folgenden Anforderungen genügt*

$$\emptyset \in \mathcal{R} \tag{2.4}$$

$$A, B \in \mathcal{R} \Longrightarrow A \backslash B \in \mathcal{R} \tag{2.5}$$

$$A, B \in \mathcal{R} \Longrightarrow A \cup B \in \mathcal{R} \tag{2.6}$$

Falls zusätzlich

$$\Omega \in \mathcal{R} \tag{2.7}$$

gilt, dann heißt $\mathcal{R}$ eine Algebra.

Es bliebe zu bemerken, dass für einen Ring $\mathcal{R}$ unnd $A, B \in \mathcal{R}$ auch

$$A \cap B = A \backslash (A \backslash B) \in \mathcal{R}$$

ist.
Manchmal ist es nicht einfach, festzustellen ob ein Mengensystem eine σ–Algebra ist oder nicht. Der folgende Begriff geht auf Dynkin zurück und hilft uns dieses Problem zu lösen.

Definition 2.6 *Ein System $\mathcal{D} \subset \mathcal{P}(\Omega)$ heißt Dynkin–System, falls es den folgenden Kriterien genügt*

$$\Omega \in \mathcal{D} \tag{2.8}$$

$$D \in \mathcal{D} \quad \Longrightarrow D^c \in \mathcal{D} \tag{2.9}$$

Für jede Folge $(D_n)_{n\in\mathbb{N}}$ paarweise disjunkter Mengen in $\mathcal{D}$ ist ihre Vereinigung $\bigcup_{n\in\mathbb{N}} \mathcal{D}_n$ auch in $\mathcal{D}$. (2.10)

Beispiel 2.7

1. *Jede σ–Algebra ist ein Dynkin–System.*
2. *Sei $|\Omega|$ endlich und $|\Omega| = 2n, n \in \mathbb{N}$.*
   Dann ist
   $$\mathcal{D} = \{D \subset \Omega, |D| \text{ ist gerade}\}$$
   ein Dynkin–System, für $n > 1$ aber keine Algebra, also auch keine σ–Algebra.

Der Zusammenhang zwischen Dynkin–Systemen und σ–Algebren wird im folgenden Satz geklärt

Theorem 2.8 *Ein Dynkin–System $\mathcal{D}$ ist eine σ–Algebra genau dann, wenn es durchschnittsstabil ist, d.h. wenn mit $A, B \in \mathcal{D}$ auch*

$$A \cap B \in \mathcal{D} \tag{2.11}$$

gilt.

Beweis. Der Beweis findet sich beispielsweise im Buch von Bauer [**?**], Satz 2.3. ■
Ähnlich wie σ-Algebren gibt es zu jedem Mengensystem $\mathcal{E} \subset \mathcal{P}(\Omega)$ ein kleinstes Dynkin–System $\mathcal{D}(\mathcal{E})$, dass von $\mathcal{E}$ erzeugt wird und es enthält. Hinter den meisten der zentralen Anwendungen von Dynkin–Systemen steckt das

Theorem 2.9 *Für jedes Mengensystem $\mathcal{E}$ mit*

$$A, B \in \mathcal{E} \Longrightarrow A \cap B \in \mathcal{E}$$

gilt

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

Beweis. Da jede σ–Algebra ein Dynkin–System ist, $\mathcal{D}(\mathcal{E}) \subseteq \sigma(\mathcal{E})$. Wüssten wir umgekehrt, dass $\mathcal{D}(\mathcal{E})$ eine σ–Algebra ist, so folgte auch die umgekehrte $\sigma(\mathcal{E}) \subseteq \mathcal{D}(\mathcal{E})$. Dazu müssen wir nach Theorem 2.8 nur zeigen, dass $\mathcal{D}(\mathcal{E})$ $\cap$-stabil ist. Dazu sei für $D \in \mathcal{D}(\mathcal{E})$

$$\mathcal{D}_D = \{Q \in \mathcal{P}(\Omega) : Q \cap D \in \mathcal{D}(\mathcal{E})\} \tag{2.12}$$

Man rechnet nach, dass $\mathcal{D}_{\mathrm{D}}$ ein Dynkin–System ist. Für jedes $E \in \mathcal{E}$ wissen wir laut Voraussetzung, dass $\mathcal{E} \subset \mathcal{D}_E$ und somit $\mathcal{D}(\mathcal{E}) \subseteq \mathcal{D}_E$. Dies aber zeigt, dass für jedes $D \in \mathcal{D}(\mathcal{E})$ und jedes $E \in \mathcal{E}$ gilt: $E \cap D \in \mathcal{D}(\mathcal{E})$. Daher folgt $\mathcal{E} \subseteq \mathcal{D}_D$ und somit auch $\mathcal{D}(\mathcal{E}) \subseteq \mathcal{D}_D$ für alle D $\in \mathcal{D}_D$. Übersetzt man dies zurück, ergibt sich

$$E \cap D \in \mathcal{D}(\mathcal{E}) \qquad \text{für alle E,D} \in \mathcal{D}(\mathcal{E})$$

also genau das, was man für die Anwendung von Theorem 2.8 benötigt. ■

2.2 Volumen, Inhalt, Prämaß, Maß

Oft hat man, wenn man ein Maß über einem Mengensystem konstruieren will, eine a priori Vorstellung, was das Maß gewisser Mengen sein sollte. Beispielsweise hat man für Teilmengen von $\mathbb{R}^d$ den intuitiven (und korrekten) Eindruck, dass ein Maß, das die Rechtecken $[a, b[= [a_1, b_1[\times \ldots [a_d, b_d[$ mit ihrem geometrischen Volumen , d.h. mit $\prod_{i=1}^{d} (b_i - a_i)$ misst, interessant sein könnte. Die Frage, ob man dann auch andere Mengen als Rechtecke messen kann, stellt sich natürlicherweise. Kann man beispielsweise auch einen Kreis

messen? Schon seit Archimedes wissen wir, dass eine Möglichkeit darin besteht, den Kreis (oder die Kugel) durch Rechtecke zu approximieren (von innen oder von außen). Dies hängt natürlich sehr davon ab, dass die Klasse der Rechtecke sehr reich ist, also reich genug, um beispielsweise Kugeln zu approximieren. Eigentlich ist natürlich an dem Spezialfall $\Omega = \mathbb{R}^d$ und einem Volumen, das von Rechtecken herkommt, nichts Besonderes, obschon wir diesen Fall mit besonderer Genauigkeit anschauen wollen.
In diesem Abschnitt wollen wir allgemein das Konzept von Volumina (Inhalten) und Maßen, die von Inhalten stammen, diskutieren. Hierzu ist es hilfreich, diese Begriffe zunächst zu definieren.

Definition 2.10 *Sei $\mathcal{R}$ ein Ring. Eine Mengenfunktion*

$$\mu : \mathcal{R} \to [0, \infty] \tag{2.13}$$

heißt Volumen, falls

$$\mu(\emptyset) = 0 \tag{2.14}$$

und

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i) \tag{2.15}$$

für alle paarweis disjunkten Mengen $A_1, \ldots, A_n \in \mathcal{R}$ und alle $n \in \mathbb{N}$. Ein Volumen heißt Prämaß, falls

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{2.16}$$

gilt für alle paarweise disjunkten Folgen $(A_i)_{i\in\mathbb{N}} \in \mathcal{R}$.
Wir nennen (2.15) endliche Additivität und (2.16) σ-Additivität.
Schließlich nennen wir ein Prämaß ein Maß, falls der zugrunde liegende Ring schon eine σ-Algebra ist. Falls $\mu(\Omega) < \infty$, heißt das Maß μ endlich; falls es eine Folge $\Omega_n \in \mathcal{A}$ mit $\Omega_n \uparrow \Omega$ und $\mu(\Omega_n) < \infty$ gibt, heißt μ σ–endlich.

Beispiel 2.11 *Sei $\mathcal{R}$ ein Ring über Ω und für $\omega \in \Omega$ sei*

$$\delta_\omega(A) = \begin{cases} 1 & \omega \in A \\ 0 & \text{sonst} \end{cases}$$

für $A \in \mathcal{R}$. Dann ist $\delta_\omega(\cdot)$ ein Prämaß. Ist $\mathcal{R}$ eine σ-Algebra , dann ist δ_ω sogar ein Maß, das sogenannte Dirac-Maß.

Beispiel 2.12 *Sei Ω eine abzählbar unendliche Menge und $\mathcal{A}$ die Algebra*

$$\mathcal{A} := \{A \subseteq \Omega : A \text{ finite or } A^c \text{ finite}\}.$$

Definiere

$$\mu(A) = \begin{cases} 0 & \text{falls } A \text{ endlich ist} \\ 1 & \text{falls } A^c \text{ endlich ist} \end{cases}$$

für $A \in \mathcal{A}$. Dann ist μ ein Volumen, aber kein Prämaß (Übung).

Einige leicht nachzuprüfende Folgerungen für Volumina sind:

Lemma 2.13 *Sei $\mathcal{R}$ ein Ring und $A, B, A_1, A_2, \ldots \in \mathcal{R}$. Sei μ ein Volumen über $\mathcal{R}$. Dann gilt:*

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B) \tag{2.17}$$

$$A \subseteq B \Longrightarrow \mu(A) \leq \mu(B) \tag{2.18}$$

$$A \subseteq B, \mu(A) < \infty \Longrightarrow \mu(B \backslash A) = \mu(B) - \mu(A) \tag{2.19}$$

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) \leq \sum_{i=1}^{n} \mu(A_i) \tag{2.20}$$

und falls die $(A_i)_{i\in\mathbb{N}}$ paarweise disjunkt sind und $\bigcup_{n=1}^{\infty} A_i \in \mathcal{R}$

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu\left(\bigcup_{i=1}^{\infty} A_n\right). \tag{2.21}$$

Beweis. Den Beweis führt der interessierte Leser entweder selbst oder liest ihn in einem beliebigen Maßtheoriebuch nach.

■

Falls μ ein Prämaß ist, erhalten wir für $A_1, A_2, .. \in \mathcal{R}$ mit $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$, dass

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_k). \tag{2.22}$$

gilt.

Eine wichtige Konsequenz der σ–Additivität ist eine Stetigkeit Eigenschaft der Prämaß. Um die Notation zu vereinfachen, schreiben wir $E_n \uparrow E$, falls $E_1 \subset E_2 \subset \ldots$ und $E = \bigcup_{n=1}^{\infty} E_n$ und $E_n \downarrow E$, falls $E_1 \supset E_2 \supset \ldots$ und $E = \bigcap_{n=1}^{\infty} E_n$.

Theorem 2.14 *Sei $\mathcal{R}$ ein Ring und μ ein Volume über $\mathcal{R}$. Betrachte*

(a) μ ist ein Prämaß.

(b) Für $(A_n)_n, A_n \in \mathcal{R}, A_n \uparrow A \in \mathcal{R}$ gilt

$$\lim_{n\to\infty} \mu(A_n) = \mu(A)$$

(c) Für $(A_n)_n, A_n \in \mathcal{R}, A_n \downarrow A \in \mathcal{R}$ und $\mu(A_n) < \infty$ gilt

$$\lim_{n\to\infty} \mu_n(A_n) = \mu(A)$$

(d) Für alle $(A_n)_n, A_n \in \mathcal{R}$ mit $\mu(A_n) < \infty$ und $A_n \downarrow \emptyset$ gilt

$$\lim_{n\to\infty} \mu(A_n) = 0.$$

Dann gilt

$$(a) \Leftrightarrow (b) \Rightarrow (c) \Leftrightarrow (d)$$

Falls μ endlich ist, sind $(a) - (d)$ sogar äquivalent.

2.3 Der Satz von Carathéodory

Eines der Schüsselprobleme der gesamten Maßtheorie besteht darin zu klären, unter welchen Voraussetzungen sich ein ein Volumen μ über einem Ring $\mathcal{R}$ auf eine größere σ-Algebra ausgeweitet werden kann, d.h. wann gibt es eine σ-Algebra $\mathcal{A} \supseteq \mathcal{R}$ und ein Maß $\tilde{\mu}$ auf $\mathcal{A}$, so dass $\tilde{\mu} \mid \mathcal{R} = \mu$. Eine notwendige Bedingung dafür haben wir schon gesehen: μ muss ein Prämaß (weil ein Maß σ-additiv ist). Wir werden nun die erstaunliche Tatsache kennenlernen, dass diese Eigenschaft auch schon hinreichend ist (was den Namen Prämaß rechtfertigt).

Theorem 2.15 *(Carathéodory) Für jedes Prämaß μ über einem Ring $\mathcal{R}$ über Ω gibt es mindestens eine Art μ zu einem Maß auf die σ-Algebra $\sigma(\mathcal{R})$ fortzusetzen.*

Beweis. Im ersten Schritt benutzen wir die geometrische Idee der Überdeckungen. Wir wollen eine gegebene Menge so genau wie möglich mit σ-Algebra Elementen approximieren. Für ein gegebenes $Q \subseteq \Omega$ sei $\mathcal{C}(Q)$ die Menge aller Folgen $(A_n)_n; A_n \in \mathcal{R}$ mit $Q \subseteq \bigcup_{n=1}^{\infty} A_n$. Definiere μ^* auf $\mathcal{P}(\Omega)$ vermöge

$$\mu^*(Q) := \begin{cases} \inf\left\{\sum_{n=1}^{\infty} \mu(A_n), (A_n)_n \in \mathcal{C}(Q)\right\} & \text{,falls } \mathcal{C}(Q) \neq \emptyset \\ \infty & \text{sonst} \end{cases} \tag{2.23}$$

Diese Funktion hat die folgenden Eigenschaften

$$\mu^*(\emptyset) = 0 \tag{2.24}$$

$$\mu^*(Q_1) \leq \mu^*(Q_2) \qquad \text{falls } Q_1 \subseteq Q_2 \tag{2.25}$$

$$\mu^*\left(\bigcup_{n=1}^{\infty} Q_n\right) \leq \sum_{n=1}^{\infty} \mu^*(Q_n) \tag{2.26}$$

für alle Folgen $(Q_n)_n, Q_n \in \mathcal{P}(\Omega)$. Dies ist eine Übungsaufgabe. Bemerke nun, dass für alle $A \in \mathcal{R}$ und $Q \in \mathcal{P}(\Omega)$ gilt

$$\mu^*(Q) \geq \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \tag{2.27}$$

und

$$\mu^*(A) = \mu(A). \tag{2.28}$$

Für den Beweis von (2.27) können wir natürlich $\mu^*(Q) < \infty$ annehmen und somit $\mathcal{C}(Q) \neq \emptyset$. Aus der endlichen Additivität von Prämaßen erhalten wir

$$\sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n \cap A) + \sum_{n=1}^{\infty} \mu(A_n \cap A^c)$$

für alle $(A_n)_n \in \mathcal{C}(Q)$. Darüber hinaus gilt $(A_n \cap A)_n \in \mathcal{C}(Q \cap A)$ und $(A_n \backslash A)_n \in \mathcal{C}(Q \backslash A)$. Daher ist $\sum_{n=1}^{\infty} \mu(A_n) \geq \mu^*(Q \cap A) + \mu^*(Q \backslash A)$. Hieraus folgt (2.27).
(2.28) folgt, da $(A, \emptyset, \emptyset, \ldots) \in \mathcal{C}(A)$ wegen $\mu(A) \leq \mu^*(A)$.

Die Bedeutung der obigen Beobachtung liegt in der Tatsache, dass wir zeigen werden, dass das System $\mathcal{A}^*$ aller Mengen, die (2.27) gehorchen, eine σ-Algebra ist und dass $\mu^* \mid \mathcal{A}^*$ ein Maß ist.

(2.27) zeigt, dass $\mathcal{R} \subset \mathcal{A}^*$, also $\sigma(\mathcal{R}) \subseteq \mathcal{A}^*$. (2.28) schließlich zeigt, dass $\mu^* \mid \mathcal{R} = \mu$, also dass μ^* eine Fortsetzung von μ ist und danach hatten wir ja gesucht.
Der Beweis wird durch die folgende Definitition 2.17 und Theorem 2.18 abgeschlossen. ■

Übung 2.16 *Man beweise 2.24, 2.25 und 2.26.*
Tipp für 2.26: Für jedes $\varepsilon > 0, n \in \mathbb{N}$ können wir $(A_{m,n})_m \in \mathcal{C}(Q_n)$ finden, so dass

$$\left|\sum_{m=1}^{\infty} \mu(A_{m,n}) - \mu^*(Q_n)\right| < \varepsilon 2^{-n}$$

Dann ist

$$(A_{m,n})_{n,m\in\mathbb{N}} \in \mathcal{C}\left(\bigcup_{m=1}^{\infty} Q_m\right).$$

Definition 2.17 *Eine Funktion μ^* auf $\mathcal{P}(\Omega)$ mit (2.24) - (2.26) heißt äußeres Maß auf Ω. $A \subseteq \Omega$ heißt μ^*- messbar, falls (2.27) gilt für alle $Q \subseteq \Omega$.*

Theorem 2.18 *Sei μ^* ein äußeres Maß über Ω. Das System $\mathcal{A}^*$ der μ^* - messbaren Mengen ist eine σ - Algebra. $\mu^* \mid \mathcal{A}^*$ ist ein Maß.*

Beweis. Bemerke dass (2.27) äquivalent ist zu

$$\mu^*(Q) = \mu^*(Q \cap A) + \mu^*(Q \backslash A) \text{ für alle } Q \in \mathcal{P}(\Omega)). \tag{2.29}$$

In der Tat: wenden wir (2.26) auf die Folge

$$Q \cap A, Q\backslash A, \emptyset, \emptyset, ... \tag{2.30}$$

an, erhalten wir sofort

$$\mu^*(Q) \leq \mu^*(Q \cap A) + \mu^*(Q \backslash A) \qquad \text{für alle } Q \in \mathcal{P}(\Omega).$$

(2.29) impliziert, dass $\Omega \in \mathcal{A}^*$ und dass mit $A \in \mathcal{A}^*$ auch $A^c \in \mathcal{A}^*$ gilt.
Als nächstes zeigen wir, dass $\mathcal{A}^*$ eine Algebra ist. Seien also $A, B \in \mathcal{A}^*$. Wendet man die definierende Eigenschaft (2.29) auf B (und $Q = Q \cap A$ und $Q = Q \cap A^c$) an, erhält man

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

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

Da auch $A \in \mathcal{A}^*$ gilt, bekommen wir

$$\begin{aligned} \mu^*(Q) &= \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \\ &= \mu^*(Q \cap A \cap B) + \mu^*(Q \cap A \cap B^c) \\ &+ \mu^*(Q \cap A^c \cap B) + \mu^*(Q \cap A^c \cap B^c). \end{aligned} \tag{2.31}$$

Da dies für alle $Q \in \mathcal{P}(\Omega)$ gilt, können wir auch Q durch $Q \cap (A \cup B)$ ersetzen. Somit folgt

$$\mu^*(Q \cap (A \cup B)) = \mu^*(Q \cap A \cap B) + \mu^*(Q \cap A \cap B^c) + \mu^*(Q \cap A^c \cap B) \tag{2.32}$$

für alle $Q \in \mathcal{P}(\Omega)$. (2.31) zusammen mit (2.32) ergibt

$$\mu^*(Q) = \mu^*(Q \cap (A \cup B)) + \mu^*(Q \backslash (A \cup B))$$

für alle $Q \in \mathcal{P}(\Omega)$. Dies zeigt, dass $A \cup B \in \mathcal{A}^*$.

In den folgenden beiden Schritten zeigen wir, dass die Algebra $\mathcal{A}^*$ ein $\cap$–stabiles Dynkin–System, somit eine σ–Algebra.
Sei also $(A_n)_n$ eine Folge paarweise disjunkter Mengen in $\mathcal{A}^*$ und setze $A := \bigcup_{n=1}^{\infty} A_n$. (2.32) ergibt durch Induktion:

$$\mu^*\left(Q \cap \bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu^*(Q \cap A_i)$$

für alle $n \in \mathbb{N}, Q \in \mathcal{P}(\Omega)$. Bedenkt man, dass wir aus obigem schon wissen, dass $B_n := \bigcup_{i=1}^{n} A_i \in \mathcal{A}^*$ und dass $Q \backslash B_n \supseteq Q \backslash A$ gilt und daher

$$\mu^*(Q \backslash B_n) \geq \mu^*(Q \backslash A)$$

folgt, erhalten wir

$$\mu^*(Q) = \mu^*(Q \cap B_n) + \mu^*(Q \backslash B_n) \geq \sum_{i=1}^{n} \mu^*(Q \cap A_i) + \mu^*(Q \backslash A).$$

Unter Zuhilfenahme von (2.26) ergibt sich

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

Nach dem, was wir eingangs bemerkt hatten, bedeutet dies aber schon, dass

$$\mu^*(Q) = \mu^*(Q \cap A) + \mu^*(Q \backslash A) = \sum_{i=1}^{\infty} \mu^*(Q \cap A_i) + \mu^*(Q \backslash A). \qquad (2.33)$$

gilt. Also ist $A \in \mathcal{A}^*$. Wir haben somit gezeigt, dass $\mathcal{A}^*$ ein Dynkin–System ist.
Außerdem ist $\mathcal{A}^*$ eine Algebra. Aber ein Dynkin–System, das eine Algebra ist, ist $\cap$ - stabil (weil $A \cap B = (A^c \cup B^c)^c$. Also ist $\mathcal{A}^*$ eine σ - algebra.

Wählt man nun $A = Q$ in (2.33), erhält man

$$\mu^*(A) = \sum_{i=1}^{\infty} \mu^*(A_i).$$

Also ist μ^* eingeschränkt auf $\mathcal{A}^*$ ein Maß. ■

Natürlich wäre es noch schöner zu wissen, dass μ nicht nur auf $\mathcal{A}^*$ fortgesetzt werden kann, sondern dass diese Fortsetzung auch eindeutig ist. Dies ist in vielen Fällen in der Tat wahr. Der Beweis benutzt auf instuktive Art die Technik der Dynkin–Systeme.

Theorem 2.19 *Sei $\mathcal{E}$ ein $\cap$–stabiler Erzeuger einer σ–Algebra $\mathcal{A}$ über Ω. Es existiere eine Folge $(E_n)_n$, $E_n \in \mathcal{E}$ mit $\bigcup_{i=1}^{\infty} E_i = \Omega$. Sind dann μ_1, μ_2 zwei Maße über $\mathcal{A}$ mit*

$$\mu_1(E) = \mu_2(E) \qquad \text{für alle } E \in \mathcal{E} \qquad (2.34)$$

und

$$\mu_1(E_n) < \infty \qquad \text{für alle } n \in \mathbb{N}. \qquad (2.35)$$

Dann gilt $\mu_1 = \mu_2$.

Beweis. Sei $\hat{\mathcal{E}}$ das System aller $E \in \mathcal{E}$ mit $\mu_1(E) = \mu_2(E) < \infty$. Für ein beliebiges $E \in \hat{\mathcal{E}}$ betrachte

$$\mathcal{D}_E := \{D \in \mathcal{A} : \mu_1(E \cap D) = \mu_2(E \cap D)\}.$$

Man rechnet nach, dass $\mathcal{D}_E$ ein Dynkin–System ist (Übung). Da $\mathcal{E}$ $\cap$–stabil ist, folgt $\mathcal{E} \subset \mathcal{D}_E$, aufgrund von (2.34) und der Definition von $\mathcal{D}_E$. Daher gilt $\mathcal{D}(\mathcal{E}) \subseteq \mathcal{D}_E$. Andererseits impliziert die $\cap$–Stabilität von $\mathcal{E}$, dass $\mathcal{A} = \mathcal{D}(\mathcal{E}) = \sigma(\mathcal{E})$ und daher (da $\mathcal{D}_E \subset \mathcal{A}$), dass $\mathcal{D}_E = \mathcal{A}$. Somit folgt

$$\mu_1(E \cap A) = \mu_2(E \cap A) \tag{2.36}$$

für alle $E \in \hat{\mathcal{E}}$ und $A \in \mathcal{A}$. Wegen (2.35) heißt dies insbesondere

$$\mu_1(E_n \cap A) = \mu_2(E_n \cap A)$$

für alle $A \in \mathcal{A}, n \in \mathbb{N}$. Der Rest des Beweises besteht darin, A in Stücke zu zerschneiden. Wir definieren

$$F_1 := E_1 \quad \text{and} \quad F_n := E_n \backslash \left(\bigcup_{i=1}^{n-1} E_i \right) \quad n \in \mathbb{N}.$$

Dann sind die (F_n) paarweise disjunkt mit $F_n \subset E_n$ und $\bigcup_{n=1}^{\infty} F_n = \bigcup_{n=1}^{\infty} E_n = \Omega$. Da $F_n \cap A \in \mathcal{A}$ erhalten wir mit (2.36):

$$\mu_1(F_n \cap A) = \mu_1(E_n \cap F_n \cap A) = \mu_2(E_n \cap F_n \cap A) = \mu_2(F_n \cap A).$$

für alle $A \in \mathcal{A}$ und $n \in \mathbb{N}$. Da

$$A = \bigcup_{n=1}^{\infty} (F_n \cap A)$$

folgt aus der σ–Additivität von μ_1 und μ_2

$$\mu_1(A) = \sum_{n=1}^{\infty} \mu_1(F_n \cap A) = \sum_{n=1}^{\infty} \mu_2(F_n \cap A) = \mu_2(A) \text{ für alle } A \in \mathcal{A}$$

was dasselbe ist wie $\mu_1 = \mu_2$. ■

Schon die Konstruktion im Beweis des Satzes von Carathéodory legt nahe, dass man das Maß von $A \in \mathcal{A}^*$ durch Maße von Ringelemente approximiert werden kann. Dies ist auch der Inhalt des folgenden

Theorem 2.20 *Sei μ ein endliches Maß über einer σ–Algebra $\mathcal{A}$ über Ω, die von einer Algebra $\mathcal{A}_0$ über Ω erzeugt wird. Dann gibt es zu $A \in \mathcal{A}$ eine Folge $(C_n)_{n\in\mathbb{N}}, C_n \in \mathcal{A}_0$, so dass*

$$\mu(A \Delta C_n) \to 0 \tag{2.37}$$

wenn $n \to \infty$. Hier schreiben wir für zwei Mengen $A, B \subseteq \Omega$

$$A \Delta B := A \backslash B \cup B \backslash A.$$

Beweis. Den Beweis findet man beispielsweise im Buch von Bauer [?], Satz 5.7. ■

2.4 Das Lebesgue–Maß

Wendet man die oben eingeführten Techniken auf den Fall $\mathbb{R}^d$, kommen wir auf ein besonders wichtiges Maß, das Lebesgue-Maß. Wie schon oben weiter erwähnt haben wir in diesem Fall ein intuitives Gefühl, dass man einfachen geometrischen Objekten, z.B. Quadern, ihr geometrisches Volumen zuordnen sollte. Wir wollen dieses Maß auf subtilere Mengen ausweiten.

Definition 2.21 *Seien $a, b \in \mathbb{R}^d$. Mit dem Quader $[a, b[$ meinen wir*

$$[a, b[:= \left\{x \in \mathbb{R}^d : a_i \leq x_i < b_i, i = 1...d\right\}$$

Ähnlich definieren wir $]a, b[,]a, b]$, and $[a, b]$.
Weiter sei

$$\mathcal{J}^d := \left\{]a, b] : \ a, b \in \mathbb{R}^d\right\}$$

und

$$\mathcal{F}^d := \left\{\bigcup_{i=1}^{n} \mathcal{J}_i, n \in \mathbb{N}, \mathcal{J}_i \in \mathcal{J}^d\right\}.$$

Übung 2.22 *$\mathcal{F}^d$ ist ein Ring über $\mathbb{R}^d$.*

Mit diesen Vorbereitungen können wir die Methoden des vorigen Abschnitts anwenden. Das zugehörige Volumen auf $\mathcal{F}^d$ ist das geometrische Volumen.

Definition 2.23 *Sei $I \in \mathcal{J}^d$, $I =]a, b]$. Wir definieren*

$$\overline{\lambda}(I) = \begin{cases} \prod_{i=1}^{d} (b_i - a_i) & \text{falls } I \neq \emptyset \\ 0 & \text{sonst} \end{cases}$$

Theorem 2.24 *Es gibt ein eindeutiges Volumen λ auf $\mathcal{F}^d$, so dass λ eine Erweiterung von $\overline{\lambda}$ auf $\mathcal{J}^d$ ist. λ ist ein Prämaß.*

Beweis. Man kontrolliert, dass $F \in \mathcal{F}^d$ als $F = \bigcup_{i=1}^{n} I_i$ mit paarweise disjunkten $I_i \in \mathcal{F}^d$ geschrieben werden kann. Schlussrichtig definiert man $\lambda(F)$ als

$$\lambda(F) = \sum_{i=1}^{n} \overline{\lambda}(I_i).$$

Durch Übergang zur gemeinsamen Verfeinerung zweier Darstellungen von F sehen wir, dass dies wohldefiniert ist.

Um zu sehen, λ auch ein Prämaß ist, müssen wir überprüfen, dass λ $\emptyset$-stetig ist. Sei also $(F_n)_{n\in\mathbb{N}}$ eine fallende Folge aus $\mathcal{F}^d$. Wir zeigen, dass

$$\delta := \lim_{n\to\infty} \lambda(F_n) = \inf_{n\to\infty} \lambda(F_n) > 0$$

impliziert, dass $\bigcap_{n=1}^{\infty} F_n \neq \emptyset$.
Da jedes F_n eine endliche Vereinigung disjunkter Elemente aus $\mathcal{J}^d$ ist, können wir $G_n \in \mathcal{F}^d$ finden, so dass

$$G_n \subset \overline{G}_n \subset F_n \quad \text{und}$$

$$|\lambda(G_n) - \lambda(F_n)| \leq 2^{-n}\delta.$$

Sei $H_n := \bigcap_{i=1}^{n} G_i$, dann ist $H_n \in \mathcal{F}^d$ und $H_n \supseteq H_{n+1}$ ebenso wie $\overline{H}_n \subseteq \overline{G}_n \subseteq F_n$. F_n ist beschränkt. Daher ist $\left(\overline{H}_n\right)_n$ eine Folge beschränkter und daher kompakter Teilmengen des $\mathbb{R}^d$ mit $F_n \supseteq \overline{H}_{n+1}$. Daher gilt $\bigcap_{n=0}^{\infty} \overline{H}_n \neq \emptyset$ (und daher auch $\bigcap_{n=1}^{\infty} F_n \neq \emptyset$), wenn nur $H_n \neq \emptyset$ für jedes n. Dazu zeigen wir

$$\lambda(H_n) \geq \lambda(F_n) - \delta\left(1 - 2^{-n}\right) \geq \delta 2^{-n} \tag{2.38}$$

wobei nur die erste Ungleichung zu zeigen ist. Dies zeigt man mit Induktion über n. Für $n = 1$ rechnet man (2.38) nach. Ist die Hypothese für n wahr, so folgt

$$\lambda(H_n) \geq \lambda(F_n) - \delta\left(1 - 2^{-n}\right)$$

und

$$\lambda(G_{n+1}) \geq \lambda(F_{n+1}) - \delta 2^{-(n+1)}$$

und $G_{n+1} \cup H_n \subseteq F_{n+1} \cup F_n = F_n$. Fügt man dies zusammen, ergibt sich

$$\begin{aligned} \lambda(H_{n+1}) &\geq \lambda(F_{n+1}) - \delta 2^{-(n+1)} - \delta\left(1 - 2^{-n}\right) \\ &\geq \lambda(F_{n+1}) - \delta\left(1 - 2^{-(n+1)}\right). \end{aligned}$$

Dies zeigt, dass $H_n \neq \emptyset$ für alle n und daher das Theorem. ■
Daher wissen wir, dass λ ein σ-endliches Prämaß auf $\mathcal{F}^d$ ist. Aus dem letzten Abschnitt erhalten wir sofort

Korollar 2.25 *Das Prämaß λ auf $\mathcal{F}^d$ kann eindeutig zu einem Maß λ auf $\sigma\left(\mathcal{F}^d\right)$ erweitert werden. λ (für das wir manchmal λ^d schreiben, um die Dimensionsabhängigkeit anzudeuten) heißt das Lebesgue-Maß, $\sigma\left(\mathcal{F}^d\right)$ die Borelsche σ–Algebra; sie kürzt man $\mathcal{B}^d$ ab.*

Von einem topologischen Gesichtspunkt ist das folgende Resultat sehr zufriedenstellend.

Theorem 2.26 *Seien $\mathcal{O}^d, \mathcal{C}^d$, und $\mathcal{K}^d$ die Systeme aller offenen, geschlossenen und kompakten Teilmengen des $\mathbb{R}^d$. Dann gilt*

$$\mathcal{B}^d = \sigma\left(\mathcal{O}^d\right) = \sigma\left(\mathcal{C}^d\right) = \sigma\left(\mathcal{K}^d\right) \tag{2.39}$$

Beweis. Wegen $\mathcal{K}^d \subseteq \mathcal{C}^d$ gilt $\sigma\left(\mathcal{K}^d\right) \subseteq \sigma\left(\mathcal{C}^d\right)$. Andererseits ist jedes $C \in \mathcal{C}^d$ die abzählbare Vereinigung von Mengen in $C_n \in \mathcal{K}^d$.Man wählt z.B.

$$K_n := \left\{x \in \mathbb{R}^d : ||x|| \leq n\right\}$$

dann ist $C = \bigcup_{n=1}^{\infty}(C \cap K_n)$. Daher $\mathcal{C}^d \subseteq \sigma\left(\mathcal{K}^d\right)$, was insgesamt zeigt, dass $\sigma\left(\mathcal{C}^d\right) = \sigma\left(\mathcal{K}^d\right)$. Andererseits ist das Komplement einer offenen Menge eine geschlossen Menge. Daher

$$\sigma\left(\mathcal{O}^d\right) = \sigma\left(\mathcal{K}^d\right) = \sigma\left(\mathcal{C}^d\right).$$

Schließlich kann man $]a,b] \in \mathcal{J}^d$ als abzählbaren Durchschnitt von $]a,b^{(n)}[$ schreiben, wo

$$b^{(n)} = \left(a_1 + \frac{1}{n}, .., a_d + \frac{1}{n}\right).$$

Daher gilt $\mathcal{B}^d = \sigma\left(\mathcal{J}^d\right) \subseteq \sigma\left(\mathcal{O}^d\right)$. Andererseits ist $]a,b[\in \mathcal{O}^d$ die Vereinigung von $]a,\tilde{b}^{(n)}] \in \mathcal{J}^d$, wobei

$$\tilde{b}^{(n)} = \left(a_1 - \frac{1}{n}, .., a_d - \frac{1}{n}\right).$$

Schließlich ist jede offene Menge $G \in \mathcal{O}^d$ die abzählbare Vereinigung von $]a,b[\in \mathcal{O}^d$ (z.B. denen mit rationalen Koordinaten). Dies zeigt $\sigma\left(\mathcal{O}^d\right) \subseteq \mathcal{B}^d$, daher $\sigma\left(\mathcal{O}^d\right) = \mathcal{B}^d$ und dies zeigt den Satz. ■

Das Lebesgue–Maß ist der Prototyp eines Borel–Maßes, d.h. eines Maßes auf $\left(\mathbb{R}^d, \mathcal{B}^d\right)$. Ein genauerer Blick zeigt, dass es der Startpunkt in Dimension eins ist, jedem Intervall $]a,b]$ seine Länge $b-a$ als Maß zuzuweisen. Dies ist geometrisch vernünftig, aber i.a., besonders wenn wir an Wahrscheinlichkeiten denken, ist dies natürlich nicht notwendig. Man könnte beispielsweise eine beliebige steigende Funktion $F : \mathbb{R} \to \mathbb{R}$, die überdies rechtsseitig-stetig ist und dem Intervall $[a,b[$ das Maß $F(b) - F(a)$ zuzuweisen. Hierbei ist es notwendig, dass F steigt, da es sonst Intervalle negativen Maßes gäbe und die Rechtsstetigkeit benötigen wir wegen der Stetigkeit von Maßen.

Definition 2.27 *Eine Funktion F heißt maßerzeugend, falls F steigend und links-stetig ist. Ist $\lim_{x\to\infty} F(x) - \lim_{y\to-\infty} F(y) = 1$, so heißt F Verteilungsfunktion.*

Der Beweis des folgenden Theorems ist größtenteils sehr ähnlich zur Konstruktion des Lebesguemaßes.

Theorem 2.28 *Sei F maßerzeugend. Dann gibt es genau ein Maß μ_F auf B^1 mit*

$$\mu_F\left(]a,b]\right) = F(b) - F(a).$$

Darüberhinaus gilt: falls G eine andere maßerzeugende Funktion auf $\mathbb{R}$ mit $\mu_F = \mu_G$ ist, dann gilt $F = G + c$ für ein c.

Um die Eigenschaften des Lebesguemaßes weiter zu untersuchen, müssen wir einen kleinen Einschub über den Zusammenhang von Maßen und Abbildungen bringen. Dies geschieht im nächsten Abschnitt.

2.5 Messbare Abbildungen und Bildmaße

Sei Ω eine Menge und $\mathcal{A}$ eine σ–Algebra über Ω (wir nennen $(\Omega, \mathcal{A})$ einen messbaren Raum). Darüber hinaus sei μ ein Maß auf $\mathcal{A}$. In diesem Abschnitt diskutierebn wir, wie man ein Maß μ auf einen anderen messbaren Raum (Ω', A') mit Hilfe einer Abbildung "teleportiert".

Definition 2.29 *Seien $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ messbare Räume. Eine Abbildung $T : \Omega \to \Omega'$ heißt $\mathcal{A} - A'$–messbar, falls gilt*

$$T^{-1}(A') \in \mathcal{A} \quad \textit{für alle } A' \in \mathcal{A}.' \tag{2.40}$$

Beispiel 2.30 *Jede konstante Abbildung ist messbar, denn*

$$T^{-1}(A') \in \{\emptyset, \Omega\} \quad \textit{für alle } A' \in \mathcal{A}'.$$

Im Prinzip kann man allerlei Eigenschaften messbarer Abbildung ähnlich diskutieren wie die Eigenschaften stetiger Abbildungen in Topologie. Wir untersuchen hier nur die o.g. "Teleportations"-Eigenschaften messbarer Abbildungen sowie eine Proposition, die uns später nützlich sein wird.

Theorem 2.31 *Sei $T : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ messbar. Dann definiert für jedes Maß μ auf $(\Omega, \mathcal{A})$*

$$T(\mu)(A') := \mu\left(T^{-1}(A')\right), \qquad A' \in \mathcal{A}'$$

ein Maß auf $(\Omega', \mathcal{A}')$.

Beweis. Dies folgt unmittelbar aus der Eigenschaft inverser Abbildungen, dass für paarweise disjunkte $(A'_n)_{n\in\mathbb{N}}$ in $\mathcal{A}'$ auch $(T^{-1}(A'_n))_{n\in\mathbb{N}}$ paarweise disjunkt sind und dass

$$T^{-1}\left(\bigcup_{n=1}^{\infty} A'_n\right) = \bigcup_{n=1}^{\infty} T^{-1}(A'_n).$$

■

Definition 2.32 *Das Maß $T(\mu)$ in Theorem 2.31 heißt Bildmaß von μ unter der Abbildung T.*

Proposition 2.33 *Es seien $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$ Maßräume und $\mathcal{E}'$ ein Erzeuger von $\mathcal{F}'$. Eine Abbildung $T : \Omega \to \Omega'$ ist genau dann messbar, wenn gilt*

$$T^{-1}(A') \in \mathcal{F} \qquad \text{für alle } A' \in \mathcal{F}' \tag{2.41}$$

Beweis. Sei

$$\mathcal{S}' := \{S' \subseteq \Omega' : T^{-1}(S') \in \mathcal{F}\}.$$

$\mathcal{S}'$ ist eine σ-Algebra . Folglich ist $\mathcal{F}' \subseteq \mathcal{S}'$ genau dann wenn $\mathcal{E}' \subseteq \mathcal{S}'$. Umgekehrt ist $\mathcal{F}' \subseteq \mathcal{S}'$ gerade die Messbarkeit von T, während $\mathcal{E}' \subseteq \mathcal{S}'$ dasselbe ist wie (2.41). ■

Ein sehr wichtiges Beispiel, nämlich $\left(\mathbb{R}^d, \mathcal{B}^d\right)$ und Abbildungen T, die Translationen sind, diskutieren wir als nächstes.

2.6 Eigenschaften des Lebesgue–Maßes

Betrachten wir also Translationen T, d.h. eine Translation um $a \in \mathbb{R}^d$ also eine Abbildung $T_a : \mathbb{R}^d \to \mathbb{R}^d$ mit $T_a(x) = a + x$ für alle $x \in \mathbb{R}^d$.

Proposition 2.34 *Das Lebesgue-Maß ist translationsinvariant, d.h. $T_a\left(\lambda^d\right) = \lambda^d$ für alle $a \in \mathbb{R}^d$.*

Beweis. Dies folgt sofort, da für jedes $I \in \mathcal{J}^d$ gilt

$$\lambda^d(I) = T_a\left(\lambda^d\right)(I).$$

■

Die Translarionsinvarianz ist sogar charakteristisch für das Lebesgue–Maß. Sei $W := [0,1[$ der Einheitswürfel in $\mathbb{R}^d$ (0 und 1 seien d-dimensionale Vektoren). Dann gilt

Theorem 2.35 *Sei μ ein Maß auf $\left(\mathbb{R}^d, \mathcal{B}^d\right)$ mit $T_a(\mu) = \mu$ für alle $a \in \mathbb{R}^d$ und*

$$\alpha := \mu(W) < \infty. \tag{2.42}$$

Dann gilt

$$\mu = \alpha\, \lambda^d. \tag{2.43}$$

d.h. bis auf Skalierung ist das Lebesgue–Maß das einzige translationsinvariante Maß auf $(\mathbb{R}^d, \mathcal{B}^d)$.

Beweis. Der Trick ist, dass $\mu(W)$ auch $\mu(W_n)$ festlegt, wo

$$W_n = [0, \frac{1}{n}[\in \mathcal{J}^d$$

und dass jedes $I \in \mathcal{J}^d$ beliebig gut approximiert werden kann durch $W_{n's}$. All dies folgt aus der Translationsinvarianz. Genaueres findet man bei Bauer [?], Satz 8.1. ■

Tatsächlich gilt sogar viel mehr: Sei W die Menge

$$\mathcal{W}\left(\mathbb{R}^d\right) := \left\{T : \mathbb{R}^d \to \mathbb{R}^d : d\left(T(x), T(y)\right) = d(x,y) \quad \text{for all } x, y \in \mathbb{R}\right\},$$

d.h. die Menge aller Abbildungen, die sich als orthogonale, lineare Abbildung plus einen Shift schreiben lassen. Es stellt sich heraus, dass das Lebesgue–Maß invariant ist unter allen $T \in \mathcal{W}$.

Theorem 2.36 *Für jedes $T \in \mathcal{W}$ gilt $T(\lambda^d) = \lambda^d$.*

Beweis. $T \in \mathcal{W}$ lässt sich darstellen als

$$T(x) = T_0(x) + a, \qquad , x \in \mathbb{R}$$

mit $a \in \mathbb{R}^d$ und einer orthogonalen linearen Abbildung T_0. Wir wissen schon, dass das Lebesgue–Maß translationsinvariant ist. Daher genügt es den Fall, dass T orthogonal ist, also insbesondere $T(0) = 0$ gilt, zu betrachten. Für so ein T, $a \in \mathbb{R}^d$ und $b = T^{-1}(a)$ gilt

$$T_a \circ T = T \circ T_b \tag{2.44}$$

wo $T_c(x) = x + c$ (Übung 2.37 unten). Unter Ausnutzung der Translationsinvarianz von λ^d erhalten wir also

$$T_a(T(\lambda^d)) = T(T_b(\lambda^d)) = T(\lambda^d)$$

für alle $a \in \mathbb{R}^d$. Dies aber heißt, dass $\mu := T(\lambda^d)$ translationsinvariant ist. Also folgt $T(\lambda^d) = \alpha\lambda^d$ mit $\alpha = \mu(W)$, und $W = [0, 1[$. Betrachte

$$B_1(0) := \{x \in \mathbb{R}^d : ||x|| \leq 1\}.$$

Mit T ist auch T^{-1} orthogonal und somit gilt $T^{-1}[B_1(0)] := B_1(0)$. Daher folgt aus $T(\lambda^d) = \alpha\lambda^d$

$$\lambda^d(B_1(0)) = \lambda^d(T^{-1}[B_1(0)]) = T(\lambda^d)(B_1(0)) = \alpha\lambda^d(B_1(0)).$$

Da $\lambda^d(B_1(0)) \notin \{0, \infty\}$ impliziert dies $\alpha = 1$. ■

Übung 2.37 *Man zeige (2.44).*

Es gilt sogar noch mehr:

Theorem 2.38 *Für jedes $T \in GL(d, \mathbb{R}^d)$ gilt*

$$T(\lambda^d) = \frac{1}{|detT|}\lambda^d. \tag{2.45}$$

Beweis. Sei $T \in GL(d, \mathbb{R}^d)$ eine invertierbare, lineare Abbildung. Wie im Beweis von Theorem 2.36 zeigt man, dass $T(\lambda^d)$ translationsinvariant ist. Da

$$\Phi(T) := T(\lambda^d)(W) < \infty \qquad (W := [0, 1[)$$

gilt, erhalten wir

$$T(\lambda^d) = \Phi(T)(\lambda^d). \tag{2.46}$$

Um $\Phi(T)$ zu brechnen, bemerken wir, dass für $S \in GL(d, \mathbb{R}^d)$

$$\Phi(S \cdot T) = \Phi(S)\Phi(T) \tag{2.47}$$

(siehe Übung 2.39 unten). Das heißt Φ ist ein Homorphismus von $GL(d,\mathbb{R})$ in die multiplikative Gruppe $\mathbb{R}\setminus\{0\}$. Aus der linearen Algebra wissen wir, dass es einen Homomorphismus

$$\phi : \mathbb{R}\setminus\{0\} \to \mathbb{R}\setminus\{0\}$$

gibt, so dass

$$\Phi(T) = \phi(\det T).$$

Falls also $\det T = 1$, wissen wir, dass $\Phi(T) = 1$. Für beliebiges T sei $\gamma := \det T \neq 0$. Für $\alpha \neq 0$ sei

$$D_\alpha : (x_1,\dots,x_d) \mapsto (x_1,\dots x_{d-1}, \alpha x_d).$$

Offensichtlich gilt $\det D_\alpha = \alpha$. Definiert man $S := T \circ D_{1/\gamma}$, so ergibt sich $\det S = 1$ mit $T = S \circ D_\gamma$. Da D_γ translationsinvariant in den ersten $d-1$ Koordinaten ist und die d'te Koordinate um einen Faktor γ gestreckt wird, sehen wir dass

$$D_\gamma(\lambda^d) = \frac{1}{|\gamma|}\lambda^d.$$

Dies kann man sofort für Intervalle kontrollieren und dann mit Standardargumenten auf beliebige Mengen fortsetzen. Da Φ ein Homorphismus mit $\Phi(S) = 1$ ist, erhalten wir

$$\Phi(T) = \Phi(S\circ D_\gamma) = \Phi(S)\Phi(D_\gamma) = \Phi(D_\gamma) = \frac{1}{|\gamma|} = \frac{1}{|\det T|}.$$

■

Übung 2.39 *Man zeige, dass Φ aus (2.46) der Gleichung (2.47) genügt.*

Die folgende Übung zeigt, dass das obige Theorem in gewissem Sinne auch für Abbildungen mit verschwindender Determinante gilt.

Übung 2.40 *Sei $T : \mathbb{R}^d \to \mathbb{R}^d$ eine lineare Abbildung mit $\det T = 0$. Man zeige, dass es für alle $A \in \mathcal{B}^d$ eine messbare Menge $N \subset \mathbb{R}^d$ mit $\lambda^d(N) = 0$ gibt, so dass $T(A) \subseteq N$. In diesem Sinne gilt (2.45) auch hier.*

Die Translationsinvarianz des Lebesgue–Maßes begründet auch zwei seiner wichtigsten Eigenschaften (die gewissermaßen "Nicht-Eigenschaften" sind, denn sie sagen etwas über das Nichtbestehen gewisser Qualitäten des Lebesgue–Maßes aus). Die erste davon sagt i.w.: "Das Lebesgue–Maß gibt es nicht auf $\mathbb{R}^\infty$". Diese Aussage ist analog (und folgt im wesentlichen den gleichen Ideen) wie die Aussage, dass es keine Laplaceverteilung auf einer unendlichen Menge gibt.

Theorem 2.41 *Auf* $(\mathbb{R}^\infty, \mathcal{B}^\infty)$ *gibt es kein translationsinvariantes Maß* λ, *das beschränkten Mengen* W *mit nichtleerem Inneren eine positive, endliche Masse zuweist, d.h.*

$$0 < \lambda(W) < \infty.$$

Beweis. Angenommen obiges λ existiert. Wir nehmen eine abgeschlossene Kugel mit Radius $0 < \varepsilon < \frac{\sqrt{2}}{2}$. Sei

$$B_\varepsilon(y) := \{x \in \mathbb{R}^\infty : \|x - y\| < \varepsilon\}.$$

$B_\varepsilon(0)$ hätte ein positives Maß, da es einen kleinen Würfel enthält. Andererseits sind die $(B_\varepsilon(e_i))_{i=1}^\infty$ disjunkt (wobei die e_i die i'ten Einheitsvektoren bezeichnen) und

$$\left(\bigcup_{i=1}^{\infty} B_\varepsilon(e_i)\right) \subseteq B_3(0). \tag{2.48}$$

(2.49) impliziert, dass $\lambda\left(\bigcup_{i=1}^\infty B_\varepsilon(e_i)\right)$ endlich ist. Aber wegen der Translationsinvarianz haben alle $B_\varepsilon(e_i)$ das gleiche Maß. Dann hat aber $\bigcup_{i=1}^\infty B_\varepsilon(e_i)$ ein unendliches Maß im Widerrspruch zu dessen Beschränktheit. ■

Die Translationsinvarianz des Lebesgue–Maßes ist es auch, die verhindert, dass alle Teilmengen $\mathbb{R}^d$ messbar sind:

Theorem 2.42 $\mathcal{B}^d \neq \mathcal{P}\left(\mathbb{R}^d\right)$, *d.h. es gibt nicht–messbare Teilmengen des* $\mathbb{R}^d$.

Beweis. Es genügt den Fall $d = 1$ zu betrachten. Wir studieren die folgende Äquivalenzrelation $\sim$ auf $\mathbb{R}$:

$$x \sim y :\Longleftrightarrow x - y \in \mathbb{Q} \tag{2.49}$$

Da es für jedes $x \in \mathbb{R}$, ein $n \in \mathbb{Q}$ mit $n \leq x < n + 1$ gibt, können wir für jede Äquivalenzklasse bzgl. $\sim$ einen Repräsentanten in $[0, 1[$ wählen. Somit gibt es eine Menge $K \subset [0, 1[$, die genau ein Element jeder Äquivalenzklasse enthält. (Bemerke, dass wir hier vom Auswahlaxiom Gebrauch machen.) K genügt

$$\bigcup_{y \in \mathbb{Q}} (y + K) = \mathbb{R} \tag{2.50}$$

und

$$y_1 \neq y_2 \Rightarrow (y_1 + K) \cap (y_2 + K) = \emptyset. \tag{2.51}$$

Angenommen K ist messbar. Dann gilt aufgrund der oben bemerkten Disjunktheit

$$\lambda\left(\bigcup_{y\in\mathbb{Q}}(y+K)\right)=\lambda(\mathbb{R})=\infty$$

Aus der Translationsinvarianz von λ bekommen wir:

$$\lambda(y+K)=\lambda(K)\qquad \text{für alle } y\in\mathbb{Q}$$

Da $\mathbb{Q}$ abzählbar ist, kann $\lambda(K)$ nicht Null sein. Andererseits ist

$$\bigcup_{y\in\mathbb{Q}\cap[0,1]}(y+K)\subseteq[0,2[.$$

Daher kann $\lambda(K)$ auch nicht positiv sein. Dies ist ein Widerspruch, also ist K nicht messbar. ■

3 Eine Einführung in die Integrationstheorie

In diesem Kapitel konstruieren wir zu vorgelegtem Maß ein Integral.
Aus gutem Grund (z.B. weil das Supremum einer Folge von $\mathbb{R}$–wertigen Funktionen auch den Wert unendlich annehmen kann) wollen wir Funktionen mit Werten in $\overline{\mathbb{R}}:=\mathbb{R}\cup\{\pm\infty\}$ integrieren. Die zugehörige Borel σ–Algebra ist definiert als

$$\overline{\mathcal{B}}^1:=\{B,B\cup\{\infty\},B\cup\{-\infty\},B\cup\{\infty\}\cup\{-\infty\}|B\in\mathcal{B}^1\}.$$

Offensichtlich ist $\overline{\mathcal{B}}^1$ eine σ–Algebra mit

$$\mathbb{R}\cap\overline{\mathcal{B}}^1:=\mathcal{B}^1.$$

Eine Funktion $f:\Omega\to\overline{\mathbb{R}}$ heißt numerische Funktion, falls sie $\mathcal{A}-\overline{\mathcal{B}}^1$-messbar ist.

Beispiel 3.1 *Beispielsweise sind alle Indikatoren*

$$1_A(\omega):=\begin{cases}1 & \omega\in A\\ 0 & \omega\notin A\end{cases}$$

für $A\in\mathcal{A}$ numerische Funktionen.

Fakt 3.2 *Für Indikatoren gelten die folgenden Rechenregeln*

- $A \subseteq B \Rightarrow 1_A \leq 1_B \quad A, B \in \mathcal{A}.$
- $1_{\bigcup_{i=1}^{\infty} A_i} = \sup_i 1_{A_i} \quad A_i \in \mathcal{A}.$
- $1_{\bigcap_{i=1}^{\infty} A_i} = \prod_{i=1}^{\infty} 1_{A_i} \quad A_i \in \mathcal{A}.$

Ob eine Funktion eine numerische Funktion ist, lässt sich schnell feststellen.

Proposition 3.3 *Sei* $(\Omega, \mathcal{A})$ *ein messbarer Raum. Dann ist* $f : \Omega \to \overline{\mathbb{R}}$ $\mathcal{A} - \overline{\mathcal{B}}^1$*-messbar, dann und nur dann, wenn*

$$\{f \geq \alpha\} \in \mathcal{A} \qquad \text{für alle } \alpha \in \mathbb{R} \tag{3.1}$$

Beweis. Wegen Proposition 2.33 ist lediglich zu zeigen, dass $\overline{\mathcal{J}} := \{[\alpha, \infty], \alpha \in \mathbb{R}\}$ die σ-Algebra $\overline{\mathcal{B}}^1$ erzeugt. Nun ist $[a, b[= [a, \infty] \setminus [b, \infty]$. und damit $\sigma(\mathcal{J}) = \mathcal{B}^1 \subseteq \overline{\mathcal{J}}$. Weiter enthält $\overline{\mathcal{J}}$ die Punkte $\{\infty\} := \bigcap_{n=1}^{\infty} [n, \infty]$ und $\{-\infty\} := \overline{\mathbb{R}} \setminus \bigcup_{n=1}^{\infty} [-n, \infty]$. Dies genügt um die Behauptung zu zeigen. ∎

Die folgende Tatsache ist leicht zu zeigen

Fakt 3.4 *Äquivalent sind:*

1. f *ist* $\mathcal{A} - \overline{\mathcal{B}}^1$*-messbar.*
2. $\{f \geq \alpha\} \in \mathcal{A}$ *für alle* $\alpha \in \mathbb{R}$.
3. $\{f > \alpha\} \in \mathcal{A}$ *für alle* $\alpha \in \mathbb{R}$.
4. $\{f < \alpha\} \in \mathcal{A}$ *für alle* $\alpha \in \mathbb{R}$.
5. $\{f \leq \alpha\} \in \mathcal{A}$ *für alle* $\alpha \in \mathbb{R}$.

Neben der Messbarkeit von f und g interessiert und natürlich auch die Messbarkeit aus f und g zusammengesetzter Funktionen. Dies wird im folgenden vorbereitet.

Proposition 3.5 *Seien* f, g *numerische Funktionen. Dann sind* $\{f \leq g\}$, $\{f < g\}$, $\{f = g\}$ *und* $\{f \neq g\}$ *messbar.*

Beweis. Da $\mathbb{Q}$ abzählbar ist, folgt die Behauptung aus

$$\{f < g\} = \bigcap_{q \in \mathbb{Q}} \{f < q\} \cap \{q < g\}$$

und

$$\{f \leq g\} = \{g < f\}^c, \{f = g\} = \{g \leq f\} \cap \{f \leq g\}, \{f \neq g\} = \{f = g\}^c.$$

■

Proposition 3.6 *Seien f, g numerische Funktionen. Dann sind auch $f \pm g$ und $f \cdot g$ – falls definiert – messbar.*

Beweis. Mit g ist auch $-g$ messbar, denn $\{-g \leq \alpha\} = \{g \geq -\alpha\}$. Also ist mit $f+g$ auch $f-g$ messbar. Desweiteren ist mit g auch $g+t$, $t \in \mathbb{R}$ messbar, denn $\{g + t \leq a\} = \{g \leq \alpha - t\}$. Für alle messbaren reellen Funktionen f, g ist nun

$$\{f + g \leq \alpha\} = \{f \leq -g + \alpha\},$$

was zusammen mit dem oben Bemerkten und Proposition 3.5 die Messbarkeit von $f \pm g$ beweist. Weiter ist

$$f \cdot g = \frac{1}{4} \left[(f+g)^2 - (f-g)^2 \right].$$

Nun ist mit f auch f^2 messbar, denn

$$\{f^2 \leq \alpha\} = \{f \leq \sqrt{\alpha}\} \cap \{f \geq -\sqrt{\alpha}\}.$$

Dies zeigt die Behauptung für relle Funktionen. Die Erweiterung auf numerische Funktionen folgt, denn die Mengen $\{f \pm g = \pm\infty\}$ und $\{f \cdot g = \pm\infty\}$ sind messbar, wenn die entsprechenden Operationen definiert sind. ■
Der folgende Satz ist essentiell für das weitere Vorgehen in der Integrationstheorie

Theorem 3.7 *Sei $(f_n)_{n \in \mathbb{N}}$ eine Folge numerischer Funktionen auf einem messbaren Raum $(\Omega, \mathcal{A})$. Dann sind auch die folgenden Funktionen messbar*

$$\sup_n f_n, \inf_n f_n, \limsup f_n, \liminf_n f_n$$

Beweis. $\sup f_n$ ist messbar, weil

$$\{\sup_n f_n \le \alpha\} = \bigcap_{n=1}^{\infty} \{f_n \le \alpha\}.$$

Weiter ist $\inf_n f_n = -\sup_n -f_n$, $\limsup f_n = \inf_n \sup_{m \ge n} f_m$ und $\liminf_n f_n = -\limsup -f_n$. Daher sind alle obigen Funktionen messbar. ■

Korollar 3.8 *Seien $f, (f_n)_{n \in \mathbb{N}}$ messbare Funktionen auf einem messbaren Raum $(\Omega, \mathcal{A})$. Dann ist für alle n*

$$\inf_{m=1\dots n} f_m \quad \text{und} \quad \sup_{m=1\dots n} f_m \tag{3.2}$$

und (falls existent)

$$\lim_n f_n \tag{3.3}$$

und $|f|$ messbar.

Beweis. (3.2) folgt aus dem vorhergehenden Satz, wenn man $f_k = f_n$ für alle $k \ge n$ setzt. Für (3.3) rufe man sich ins Gedächtnis dass, falls $\lim f_n$ existiert, man $\lim f_n = \limsup f_n$ hat. Schließlich ist

$$|f| = f^+ - f^-,$$

wo $f^+ = \max(f, 0)$ und $f^- = (-f)^+$ messbar sind. ■

3.1 Konstruktion des Integrals

Die Grundidee bei der Konstruktion des Maßintegrals ist nicht sehr verschieden von der, die zur Konstruktion des Riemann–Integrals: Wenn $(\Omega, \mathcal{A})$ ein messbarer Raum mit einem Maß μ ist und $A \in \mathcal{A}$, dann beschreibt die Indikatorfunktion 1_A geometrisch ein verallgemeinertes "Rechteck" in $\Omega \times \mathbb{R}$ mit Kantenlängen $\mu(A)$ und 1. Daher sollte man

$$\int 1_A d\mu$$

gleich $\mu(A)$ setzen. Trotzdem gibt es essentielle Unterschiede zwischen dem Maßintegral und dem Riemannintegral, die wesentlich daher kommen, dass Riemann von den Mengen A forderte, dass sie die Gestalt von Intervallen haben müssten.
Die elementarsten Funktionen, die man integrieren möchte, nennen wir wieder Treppenfunktionen (oder Elementarfunktionen):

Definition 3.9 *Sei* $(\Omega, \mathcal{A})$ *ein messbarer Raum. Eine Treppenfunktion oder Elementarfunktion ist eine Abbildung der Form*

$$f(\omega) = \sum_{i=1}^{n} \alpha_i 1_{A_i}(\omega).$$

Hier sind die $A_i \in \mathcal{A}$, $i = 1, \dots, n$ *paarweise disjunkt mit* $\bigcup A_i = \Omega$ *und* $\alpha_i \geq 0$. *Die Menge der Elementarfunktionen auf* $(\Omega, \mathcal{A})$ *kürzen wir mit*

$$E := E(\Omega, \mathcal{A}) \tag{3.4}$$

ab.

Definition 3.9 und die Betrachtungen im vorigen Abchnitt implizieren

Lemma 3.10 *Es seien* $u, v \in E$ *und* $\alpha \in \mathbb{R}^+$. *Dann gilt auch* $\alpha u, u + v, \max\{u, v\}, \min\{u, v\} \in E$.

Beweis. Offensichtlich. ■

Das folgende Lemma bereitet die Integration von Treppenfunktionen vor

Lemma 3.11 *Sei* $u \in E$*:*

$$u = \sum_{i=1}^{m} \alpha_i 1_{A_i} = \sum_{i=1}^{n} \beta_i 1_{B_i}.$$

Dann gilt

$$\sum_{i=1}^{m} \alpha_i \mu(A_i) = \sum_{i=1}^{n} \beta_i \mu(B_i).$$

Beweis. Der Beweis beruht auf einem Standardtrick, nämlich von den (A_i) und (B_i) zu einer gemeinsamen Verfeinerung $(A_i \cap B_j)_{i,j}$ überzugehen. Wegen $A_i = \bigcup_{j=1}^{n} (A_i \cap B_j)$ und $B_j = \bigcup_{i=1}^{m} (A_i \cap B_j)$ folgt wegen der Additivität von μ

$$\mu(A_i) = \sum_{j=1}^{n} \mu(A_i \cap B_j) \quad \text{ und } \mu(B_j) = \sum_{i=1}^{m} \mu(A_i \cap B_j),$$

was wegen $\alpha_i = \beta_j$ auf $(A_i \cap B_j)$ die Behauptung impliziert:

$$\sum_{i=1}^{m} \alpha_i \mu(A_i) = \sum_{i,j} \alpha_i \mu(A_i \cap B_j) = \sum_{i,j} \beta_i \mu(A_i \cap B_j) = \sum_{i=1}^{n} \beta_i \mu(B_i)$$

■

Somit ist die folgende Definition wohldefiniert:

Definition 3.12 *Sei $u \in E$ mit der Darstellung*

$$u = \sum_{i=1}^{n} \alpha_i 1_{A_i}. \tag{3.5}$$

Dann ist

$$\int u d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i)$$

das (μ)-Integral von u. Hier ist μ ein Maß auf $(\Omega, \mathcal{A})$. $\int u d\mu$ ist unabhängig von der speziellen Darstellung (7.2).

Die folgenden Eigenschaften des Integrals von Elementarfunktionen prüft man schnell nach

Proposition 3.13 *Seien $u, v \in E, \alpha \in \mathbb{R}^+$ und $A \in \mathcal{A}$. Dann gilt*

$$\int 1_A d\mu = \mu(A), \tag{3.6}$$

$$\int \alpha u d\mu = \alpha \int u d\mu, \tag{3.7}$$

$$\int (u+v) d\mu = \int u d\mu + \int v d\mu, \tag{3.8}$$

$$u \leq v \Rightarrow \int u d\mu \leq \int v d\mu \tag{3.9}$$

Beweis. (3.6) und (3.7) sind offensichtlich.
Für (3.8) seien

$$u = \sum_{i=1}^{m} \alpha_i 1_{A_i} \qquad v = \sum_{j=1}^{n} \beta_j 1_{B_j}. \tag{3.10}$$

Dann ist $u + v$ eine Elementarfunktion mit

$$u + v = \sum_{i,j} (\alpha_i + \beta_j) 1_{A_i \cap B_j}.$$

Da $A_i = \bigcup_{j=1}^n (A_i \cap B_j)$ und $B_j = \bigcup_{i=1}^m (A_i \cap B_j)$ ebenso eine Partition ist wie $(A_i)_i$ und $(B_j)_j$, folgt

$$\begin{aligned} \int (u+v) d\mu &= \sum_{i,j} (\alpha_i + \beta_j) \mu(A_i \cap B_j) \\ &= \sum_{i,j} \alpha_i \mu(A_i \cap B_j) + \sum_{i,j} \beta_j \mu(A_i \cap B_j) \\ &= \int u d\mu + \int v d\mu. \end{aligned}$$

(3.9) folgt wieder per Übergang zu einer gemeinsamen Verfeinerung der Darstellungen

$$u = \sum_{i,j} \alpha_i 1_{A_i \cap B_j} \text{ und } v = \sum_{i,j} \beta_j 1_{A_i \cap B_j}.$$

$u \leq v$ impliziert, dass $\alpha_i \leq \beta_j$ auf $A_i \cap B_j \neq \emptyset$ und hieraus folgt die Behauptung. ■

Die letzte Proposition heißt insbesondere, dass für $u \in E$ und $A_i \in \mathcal{A}$, $\alpha_i \in \mathbb{R}^+$, so dass

$$u = \sum_{i=1}^n \alpha_i 1_{A_i}$$

(wobei wir nicht mehr fordern, dass die $(A_i)_i$ eine Partition bilden) immer noch $\int u d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$ gilt.

Wir werden nun das Integral für numerische Funktionen definieren. Dies geschieht zunächst für positive Funktionen, später werden wir dann beliebige Funktionen als Differenz zweier positiver Funktionen schreiben. Die Schlüsselidee ist es eine numerische Funktionen durchTreppenfunktionen zu approximieren. Der wesentliche Unterschied zur Konstruktion des Riemannintegrals besteht darin, dass die Klasse der Elementarfunktionen bei uns i.a. größer ist als die Klasse der Treppenfunktionen bei der Konstruktion des Riemannintegrals.

Das folgende Lemma spielt eine Schlüsselrolle bei unseren weiteren Überlegungen:

Lemma 3.14 *Seien $u, (u_n)_n \in E$. Weiter sei die Folge (u_n) steigend. Dann gilt*

$$u \leq \sup_n u_n \Rightarrow \int u d\mu \leq \sup_n \int u_n d\mu.$$

Beweis. u habe die folgende Darstellung

$$u = \sum_{i=1}^{m} \alpha_i 1_{A_i}$$

mit $A_i \in \mathcal{A}$ und $\alpha_i \in \mathbb{R}^+$. Sei $\alpha \in (0,1)$. Da u_n messbar ist, folgt

$$B_n := \{u_n \geq \alpha u\} \in \mathcal{A}.$$

Definitionsgemäß gilt auf B_n: $u_n \geq \alpha u 1_{B_n}$, also

$$\int u_n d\mu \geq \alpha \int u 1_{B_n} d\mu.$$

Nun ist (u_n) wachsend und $u \leq \sup_n u_n$. Daher folgt $B_n \uparrow \Omega$ und daher $A_i \cap B_n \uparrow A_i$ für alle i. Da μ von unten stetig ist, erhalten wir

$$\int u d\mu \quad = \quad \sum_{i=1}^{m} \alpha_i \mu(A_i) = \lim_{n\to\infty} \sum_{i=1}^{m} \alpha_i \mu(A_i \cap B_n) = \lim_{n\to\infty} \int u 1_{B_n} d\mu.$$

Also

$$\sup_n \int u_n d\mu \quad \geq \quad \alpha \sup_n \int u 1_{B_n} d\mu = \alpha \lim_{n\to\infty} \int u 1_{B_n} d\mu = \alpha \int u d\mu.$$

Da $\alpha \in (0,1)$ beliebig war, folgt die Behauptung des Lemmas. ■

Korollar 3.15 *Für zwei wachsende Folgen (u_n) und (v_n) in E gilt*

$$\sup u_n = \sup v_n \Rightarrow \sup_n \int u_n d\mu = \sup_n \int v_n d\mu \tag{3.11}$$

Beweis. Es gilt $u_n \leq \sup v_m$ und $v_m \leq \sup u_n$ für alle n, m. Damit folgt die Behauptung aus dem vorhergehenden Lemma. ■

Definition 3.16 *$E^*(\Omega, \mathcal{A}) = E^*$ bezeichnet die Menge aller numerischer Funktionen, für die es eine wachsende Folge (u_n) in E gibt, so dass*

$$f = \sup u_n.$$

Der Punkt ist, dass Korollar 3.15 uns sagt, dass $\sup_n \int u_n d\mu$ unabhängig ist von der speziellen Wahl der Folge (u_n), mit der wir f approximieren. Daher definieren wir

Definition 3.17 *Sei $f \in E^*$. Wir definieren das Integral von f bezüglich μ durch*

$$\int f d\mu = \sup_n \int u_n d\mu,$$

wobei (u_n) eine wachsende Folge in E ist, die

$$f = \sup u_n$$

genügt.

Bemerke, dass $E \subseteq E^*$. Ebenso überprüft man, dass mit $f, g \in E^*$ und $\alpha \in \mathbb{R}^+$ gilt:

$$\alpha f, f+g, f \cdot g, \min f, g, \max f, g \in E^*$$

$$\int \alpha f d\mu = \alpha \int f d\mu,$$

$$\int (f+g) d\mu = \int f d\mu + \int g d\mu,$$

$$f \leq g \Rightarrow \int f d\mu \leq \int g d\mu$$

Somit ist das Integral eine positive, steigende Linearform auf der Menge aller inetgrierbarer Funktionen.
Es mag nicht zu sehr überraschen, dass Lemma 3.14 von $f \in E^*$ geerbt wird. Dies ist trotzdem einer der wesentlichen Vorteile des neuen Integralbegriffs.

Theorem 3.18 *(Levi, Satz von der monotonen Konvergenz):*
Sei $(f_n)_n$ eine wachsende Folge in E^*. Dann gilt $\sup f_n \in E^*$ und

$$\int \sup f_n d\mu = \sup \int f_n d\mu. \tag{3.12}$$

Beweis. Definiere $f := \sup f_n$. Für $f_n \in E^*$ gibt es eine Folge $(u_{m,n})_m$ in E, die gegen f_n aufsteigt. Es ist

$$v_m := \max(u_{m,1}, ...u_{m,m}) \in E.$$

Da die $(u_{m,n})$ wachsend sind, sind dies auch die $(v_m)_m$. Offensichtlich $v_m \leq f_m$, also $v_m \leq f$. Andererseits gilt für $m \leq n$ dass $u_{m,n} \leq v_m$ und somit

$$\sup_{m \in \mathbb{N}} u_{mn} = f_n \leq \sup_{m \in \mathbb{N}} v_m.$$

Daher gilt $\sup v_m = f$, was $f \in E^*$ impliziert. Aber dann gilt definitionsgemäß $\int f d\mu = \sup \int v_n d\mu$. Nun folgt aus $v_n \leq f_n$ dass $\int v_n d\mu \leq \int f_n d\mu$ gilt. Dies zeigt, dass

$$\int f d\mu \leq \int f_n d\mu.$$

Die umgekehrte Ungleichung $\int f d\mu \geq \int f_n d\mu$ ist offensichtlich, da $f_n \leq f$ für alle $n \in \mathbb{N}$. ■

Korollar 3.19 *Sei $(f_n)_n$ eine Folge in E^*, dann gilt*

$$\sum_{n=1}^{\infty} f_n \in E^*$$

und

$$\int \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

Beweis. Man wende den Satz von der monotonen Konvergenz auf $(f_1 + ... + f_n)_n$ an. ■

Es stellt sich natürlich die Frage, wie groß denn die Menge E^* ist. Die Antwort mag ein wenig überraschen, wenngleich sie einfach zu beweisen ist.

Theorem 3.20 *$f \in E^*$ genau dann wenn f eine positive, numerische Funktion ist.*

Beweis. Eine Richtung ist klar. Für die andere sei

$$A_{in} = \begin{cases} \{f \geq \frac{i}{2^n}\} \cap \{f < \frac{i+1}{2^n}\} & i = 0, ..., n2^n - 1 \\ \{f \geq n\} & i = n2^n \end{cases}$$

und

$$u_n = \sum_{i=0}^{n2^n} \frac{i}{2^n} 1_{A_{in}}.$$

Dann ist $u_n \in E$ und $(u_n)_n$ ist wachsend mit $\sup u_n = f$. ■

Schließlich breiten wir das Integral auf beliebige numerische Funktionen aus. Wir definieren für eine numerische Funktion f

$$f^+ := \max(f, 0) \quad \text{und} \quad f^- := \max(-f, 0).$$

Dann gilt $f = f^+ - f^-$. Aus der Additivitätsforderung für Integrale erhalten wir die folgende Definiton:

Definition 3.21 *Eine numerische Funktion f heißt integrierbar, falls $\int f^+ d\mu$ und $\int f^- d\mu$ reelle Zahlen sind. In diesem Falle sei*

$$\int f d\mu := \int f^+ d\mu - \int f^- d\mu. \tag{3.13}$$

Bemerkung 3.22

1. *(3.13) ist auch dann sinnvoll, wenn entweder $\int f^+ d\mu$ oder $\int f^- d\mu$ unendlich sind (aber nicht beide). In diesem Falle spricht man von Quasi-Integrierbarkeit.*

2. *Falls $\Omega = \mathbb{R}^d$,$\mathcal{A} = B^d$ und $\mu = \lambda^d$ spricht man auch vom Lebesgue–Integral.*

Nachfolgend diskutieren wir einige Eigenschaften des Integrals:

Theorem 3.23 *Für eine numerische Funktion f sind äquivalent.*

1. *f is integrierbar.*

2. *Es gibt eine integrierbare Funktion g, so dass $|f| \leq g$*

3. *$|f|$ ist integrierbar.*

Beweis. Der Beweis ist nicht schwierig und bleibt dem Leser überlassen. ■

Übung 3.24 *Seien f,g integrierbar und $\alpha \in \mathbb{R}$. Man zeige, dass αf,$f + g$,* $\max(f, g)$, $\min(f, g)$ *integrierbar sind und dass*

$$\int \alpha f d\mu = \alpha \int f d\mu \quad \text{und} \quad \int (f + g)\, d\mu = \int f d\mu + \int g d\mu.$$

gilt.

Proposition 3.25 *f, g seien integrierbar. Dann gilt*

$$f \leq g \Longrightarrow \int f d\mu \leq \int g d\mu \tag{3.14}$$

$$\left| \int f d\mu \right| \leq \int |f|\, d\mu \tag{3.15}$$

(3.15) heißt Dreiecksungleichung.

Beweis. $f \leq g$ impliziert $f^+ \leq g^+$ und $f^- \geq g^-$. Daher folgt (3.14) die Monotonie des Integrals auf E^*.
(3.15) ist ein Spezialfall von (3.14), da $f \leq |f|$ und $-f \leq |f|$. ■

Definition 3.26 *Sei $(\Omega, \mathcal{A})$ ein messbarer Raum und μ ein Maß darauf; dann sei*

$$\mathcal{L}^1(\mu) := \{f : \Omega \to \mathbb{R}, f \textit{ is integrierbar}\}$$

der Raum aller integrierbarer Funktionen.

Bemerkung 3.27 *Bemerke, dass die obige Definition von $\mathcal{L}^1(\mu)$ die Messbarkeit von f einschließt.*

Fakt 3.28 *$\mathcal{L}^1(\mu)$ mit der punktweisen Addition und Multiplication mit reellen Zahlen ist ein Vektorraum.*

Beispiel 3.29

1. *Sei $(\Omega, \mathcal{A}, \mu)$ so, dass μ endlich ist, d.h. $\mu(\Omega) < \infty$. Dann*

   $$\{f : f \equiv const\} \subseteq \mathcal{L}^1/\mu.$$

   Aus den obigen Überlegungen folgt dann, dass auch

   $$\{f : \Omega \to \mathbb{R}, f \textit{ ist messbar und beschränkt } \} \subseteq \mathcal{L}^1(\mu).$$

2. *Seien μ, ν zwei Maße auf $(\Omega, \mathcal{A})$. Eine numerische Funktion $f : \Omega \to \bar{\mathbb{R}}$ ist $(\mu + \nu)$–integrierbar, genau dann wenn sie μ–integrierbar und ν -integrierbar ist. Dann gilt*

   $$\int f d(\mu + \nu) = \int f d\mu + \int f d\nu. \tag{3.16}$$

   Dies prüft man leicht nach für Treppenfunktionen und zieht es dann über die Standardargumente hoch auf allgemeine numerische Funktionen. Insbesondere gilt

   $$\mathcal{L}^1(\mu + \nu) = \mathcal{L}^1(\mu) \cap \mathcal{L}^1(\nu).$$

Bislang haben wir nur Integrale über Ω betrachtet. Integrale über beliebige messbare Teilmengen $A \subseteq \Omega$ definiert man wie folgt:

Definition 3.30 *Sei $f \in E^* \cup \mathcal{L}^1(\mu)$. Dann ist für $A \in \mathcal{A}$ das Integral über A definiert als*

$$\int_A f d\mu := \int f 1_A d\mu.$$

Insbesondere $\int f d\mu = \int_\Omega f d\mu$.

Übung 3.31 *Es sei $(\Omega, \mathcal{A})$ ein messbarer Raum und μ ein endliches Maß auf $(\Omega, \mathcal{A})$. Zeigen Sie, dass falls f der uniforme Limes einer Folge $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{L}^1(\mu)$ ist, dies $f \in \mathcal{L}^1(\mu)$ impliziert. Warum ist die Endlichkeit von μ notwendig?*

3.2 Fast überall bestehende Eigenschaften

Ein genauer Blick auf den vorhergehenden Abschnitt zeigt, dass sich am Wert eines Integrals nichts ändert, wenn wir die zu integrierende Funktion auf einer Menge vom Maß 0 abändern. Formal definieren wir

Definition 3.32 *Wir sagen, dass eine Eigenschaft μ–fast überall (μ–f.ü.) auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$ gilt, falls es eine Menge N mit $\mu(N) = 0$, so dass die Eigenschaft auf $\Omega \backslash N$ gilt. Falls μ ein Wahrscheinlichkeitsmaß ist, d.h. falls $\mu(\Omega) = 1$ ist, schreiben wir auch μ–fast sicher (μ–f.s.).*

Beispiel 3.33 *Die Dirichlet–Funktion $1_{\mathbb{Q}}$ is null λ^1–f.ü., wir schreiben auch*

$$1_{\mathbb{Q}} = 0 \qquad \lambda^1 - f.ü.$$

Der folgende Satz ist in modfizierter Form schon aus der Theorie des Riemann–Integrals bekannt.

Theorem 3.34 *Sei $f \in E^*(\Omega, \mathcal{A})$. Dann gilt*

$$\int f d\mu = 0 \Leftrightarrow f = 0 \qquad \mu\text{–f.ü.}$$

Beweis. Da f messbar ist, ist $N := \{f \neq 0\} = \{f > 0\} \in \mathcal{A}$. Wir zeigen

$$\int f d\mu = 0 \Leftrightarrow \mu(N) = 0.$$

Sei $\int f d\mu = 0$. Nun ist $A_n := \left\{ f \geq \frac{1}{n} \right\} \in \mathcal{A}$ für alle $n \in \mathbb{N}$. Des weiteren gilt $A_n \uparrow N$. Offensichtlich gilt $f \geq \frac{1}{n} 1_{A_n}$ und daher

$$0 = \int f d\mu \geq \int \frac{1}{n} 1_{A_n} d\mu = \frac{1}{n} \mu(A_n).$$

Daher gilt $\mu(A_n) = 0$ für alle n, was $\mu(N) = 0$ beweist.
Ist andererseits $\mu(N) = 0$, dann ist $u_n := n 1_N \in E$ für alle $n \in \mathbb{N}$ und $\int u_n d\mu = 0$. Setze $g := \sup u_n$, dann ist definitonsgemäß $g \in E^*, u_n \uparrow g$ und $\int g d\mu = \sup \int u_n d\mu = 0$. Aber es ist $f \leq g$, was $\int f d\mu = 0$ zeigt. ■

Übung 3.35 *Sei f $\mathcal{A}$-messbar und $N \in \mathcal{A}$ mit $\mu(N) = 0$. Zeigen Sie, dass*

$$\int_N f d\mu = 0.$$

Theorem 3.36 *Seien f, g messbare numerische Funktionen, so dass $f = g$ μ–f.ü. auf Ω. Dann gilt*
a) Falls $f \geq 0$ und $g \geq 0$, dann folgt $\int f d\mu = \int g d\mu$
b) Falls f integrierbar ist, dann ist g ebenso integrierbar und es gilt $\int f d\mu = \int g d\mu$.

Beweis. a)Es gilt $f - g = 0$ μ-f.ü. und daher

$$\int (f - g) d\mu = 0 \Longrightarrow \int f d\mu = \int g d\mu.$$

b) Aus $f = g$ μ-f.ü. und a) folgt $\int f^+ d\mu = \int g^+ d\mu$ und $\int f^- d\mu = \int g^- d\mu$ und somit die Behauptung. ■

Korollar 3.37 *f, g seien numerische Funktionen mit $|f| \leq g$ μ-f.ü. Dann ist mit g auch f μ - integrierbar.*

Beweis. Sei $g' = \max(g, |f|)$. Dann gilt $g = g'$ μ-f.ü. Daher ist auch g' μ - integrierbar. Aber $|f| \leq g'$ auf ganz Ω. Die Behauptung folgt. ■
Es erscheint offensichtlich, dass integrierbare Funktionen nicht zu groß werden können. Dies ist im folgenden Satz formalisiert.

Theorem 3.38 *Sei f μ - integrierbar. Dann gilt $|f| < \infty$ μ-f.ü.*

Beweis. Sei $N := \{|f| = \infty\} \in \mathcal{A}$. Dann gilt für alle $\alpha \in \mathbb{R}$ $\alpha 1_N \leq |f|$ und somit $\alpha\mu(N) \leq \int |f| d\mu < \infty$. Daher folgt $\mu(N) = 0$. ■

3.3 Die Räume $\mathcal{L}^p(\mu)$

Oben haben wir schon die Menge der integrierbaren Funktionen $\mathcal{L}^1(\mu)$ kennenglernt und gesehen, dass dies ein Vektorraum ist. Damit es auch ein Körper ist, müsste $\mathcal{L}^1(\mu)$ abgeschlossen sein unter Produktbildung, d.h. mit $f, g \in \mathcal{L}^1(\mu)$ müsste auch $f \cdot g \in \mathcal{L}^1(\mu)$ sein. Dies ist nicht unmöglich, denn das Produkt zweier messbarer Funktionen ist wieder messbar. Das folgende Beispiel zeigt, dass die Inegrierbarkeit unter Produktbildung nicht vererbt wird.

Beispiel 3.39 *Sei $(\Omega, \mathcal{A}) = (\mathbb{N}, \mathcal{P}(N))$, $\mu(\{u\}) = \alpha_n = n^{-p-1}$ und $f(n) = n$. Für $1 < p < \infty$ ist $f(n)$ integrierbar, aber nicht $f^p(n)$, z.B. ist für $p = 2$, f^2 nicht integrierbar.*

Wir werden im folgenden untersuchen, wann $|f|^p$ integrierbar ist. Sei dafür stets $p \geq 1$. Bemerke, dass für eine numerische Funktion $f : \Omega \to \bar{\mathbb{R}}$ auch $|f|^p$ eine numerische Funktion ist, da $\{|f|^p \geq \alpha\}$ entweder Ω oder $\left\{|f| \geq \alpha^{\frac{1}{p}}\right\}$ ist. Daher ist für all f die Größe

$$N_p(f) := \left(\int |f|^p \, d\mu\right)^{\frac{1}{p}} \tag{3.17}$$

definiert mit $0 \leq N_p(f) \leq +\infty$. Offensichtlich gilt

$$N_p(\alpha f) = |\alpha| \, N_p(f) \, .$$

Die beiden folgenden Ungleichungen sind zentral für $N_p(\cdot)$.

Theorem 3.40 (Hölder Ungleichung) *Sei $p > 1$ und q so, dass*

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{3.18}$$

Dann gilt für alle numerischen Funktionen f, g auf Ω

$$N_1(fg) \leq N_p(f) \, N_q(g) \tag{3.19}$$

Beweis. O.B.d.A. $f \geq 0$ und $g \geq 0$. Sei $\sigma := N_p(f)$ und $\tau := N_q(g)$. Wir können annehmen, dass $\sigma > 0$ und $\tau > 0$, sonst ist nämlich $f = 0$ oder $g = 0$ μ–f.ü. und daher auch $f \cdot g = 0$ μ–f.ü. und somit $N_1(fg) = 0$. Andererseits können wir auch annehmen, dass $\sigma < +\infty$ und $\tau < +\infty$.

Nach der Bernoulli Ungleichung gilt

$$(1+\eta)^{\frac{1}{p}} \leq \frac{\eta}{p} + 1 \quad \text{für alle } \eta \in \mathbb{R}^+ \tag{3.20}$$

und mit $\xi := 1 + \eta$

$$\xi^{\frac{1}{p}} \leq \frac{\xi}{p} + \frac{1}{q} \quad \text{für alle } \xi \geq 1. \tag{3.21}$$

Nun ist für zwei Zahlen $x, y \in \mathbb{R}^+$ entweder $x \leq y$ oder $y > x$, daher ist entweder $xy^{-1} \geq 1$ oder $x^{-1}y \geq 1$. Setzen wir $\xi := \max\{xy^{-1}, x^{-1}y\}$, erhalten wir $\xi \geq 1$. Setzt man dies in (3.21) ein (wobei man im zweiten Fall die Bernoulli–Ungleichung mit q statt p anwendet), erhält man

$$x^{\frac{1}{p}} y^{\frac{1}{q}} \leq \frac{1}{p}x + \frac{1}{q}y \quad \text{für alle } x, y \in \mathbb{R}^+. \tag{3.22}$$

Wir wählen $x := (f(\omega)/\sigma)^p$ und $y := (g(\omega)/\tau)^q$ für $\omega \in \Omega$ mit $f(\omega) < \infty$ und $g(\omega) < \infty$. Dann folgt

$$\frac{1}{\sigma\tau} fg \leq \frac{1}{\sigma^p p} f^p + \frac{1}{\tau^q q} g^q. \tag{3.23}$$

(3.23) ist offensichtlich auf $\{f = +\infty\} \cup \{g = +\infty\}$. Integriert man (3.23), erhält man

$$\int fg d\mu \leq \sigma\tau$$

also (3.19). ■

Theorem 3.41 (Minkowski Ungleichung) *Seien f, g numerische Funktionen, so dass $f + g$ auf ganz Ω definiert ist. Dann ist für $1 \leq p < \infty$:*

$$N_p(f+g) \leq N_p(f) + N_p(g). \tag{3.24}$$

Beweis. Da $|f + g| \leq N_p |f| + |g|$, gilt

$$N_p(f+g) \leq N_p(|f| + |g|)$$

und daher genügt es $f \geq 0$ und $g \geq 0$ zu betrachten. Zunächst sei bemerkt, dass für $p = 1$ (3.24) wahr ist mit Gleichheit. Wir können also $1 < p < +\infty$ annehmen. Wähle wieder q so, dass $\frac{1}{p} + \frac{1}{q} = 1$. Schließlich können wir auch noch annehmen, dass $N_p(f)$ und $N_p(g)$ endlich sind. Aber dann folgt aus

$$(f+g)^p \leq (2\max(f,g))^p = 2^p \max(f^p, g^p) \leq 2^p (f^p + g^p),$$

dass $(f+g)^p$ integrierbar ist und somit $N_p(f+g) < \infty$. Aber nun gilt:

$$\int (f+g)^p \, d\mu = \int f (f+g)^{p-1} \, d\mu + \int (f+g)^{p-1} g d\mu. \qquad (3.25)$$

Wendet man die Hölder Ungleichung auf beide Summanden auf der rechten Seite von (3.25) an, erhält man

$$\begin{aligned} \int (f+g)^p \, d\mu &\leq N_p(f) \, Nq\left((f+g)^{p-1}\right) \\ & \quad N_p(g) \, Nq\left((f+g)^{p-1}\right) \\ &= (N_p(f) + N_p(g)) \, Nq\left((f+g)^{p-1}\right) \end{aligned}$$

Wegen $q(p-1) = p$ ergibt sich

$$(N_p(f+g))^p \leq [N_p(f) + N_p(g)] \, (N_p(f+g))^{p-1}$$

Wegen $N_p(f+g) < \infty$ ergibt dies (3.24). ■
Die Funktionen, mit denen wir uns beschäftigt haben, bekommen einen Namen.

Definition 3.42 *$f : \Omega \to \overline{\mathbb{R}}$ heißt p-fach integrierbar, falls f messbar und $|f|^p$ integrierbar ist. Die Menge aller p-fach integrierbaren Funktionen nennen wir $\mathcal{L}^p(\mu)$. Für $p = 2$ sprechen wir auch von quadratischer Integrierbarkeit.*

Die Hölder Ungleichung impliziert

Theorem 3.43 *Das Produkt einer p-fach integrierbaren Funktion und einer q-fach integrierbaren Funktion ist integrierbar, falls $\frac{1}{p} + \frac{1}{q} = 1$.*

Korollar 3.44 *Sei $\mu(\Omega) < \infty$. Dann ist jede p-fach integrierbare Funktion 1-fach integrierbar für alle $1 < p < \infty$.*

Beweis. Da $\mu(\Omega) < \infty$ ist die konstante Funktion 1 q-fach integrierbar. Daher ist $f = f \cdot 1$ integrierbar. ■
Ebenso folgern wir

Theorem 3.45 *Sei $f : \Omega \to \overline{\mathbb{R}}$ p-fach integrierbar, $(1 \leq p < \infty)$ und sei $g : \Omega \to \overline{\mathbb{R}}$ beschränkt durch ein $\alpha \in \mathbb{R}_+$. Dann ist $f \cdot g$ p-fach integrierbar.*

Beweis. Wir wissen $|g| \leq \alpha$. Aber dann folgt, dass $|gf| \leq \alpha |f|$, und $\alpha |f|$ ist p-fach integrierbar. Dies zeigt die Behauptung. ∎
Schließlich betrachten wir auch den Fall $p = \infty$. Definiere

$$\mathcal{L}^{\infty}(\mu) := \{f : \Omega \to \mathbb{R},\ f \text{ ist messbar und } \mu - f.\ddot{u}. \text{ beschränkt}\}$$

Trivialerweise ist $\mathcal{L}^{\infty}(\mu)$ ein Vektorraum über $\mathbb{R}$.

3.4 Konvergenzsätze

Zum Eingang dieses Abschnitts sei bemerkt, dass das oben definierte $N_p(f)$ eine Seminorm auf dem Raum ist $\mathcal{L}^p(\mu)$ ist, d.h.

$$\begin{aligned} N_p &: \mathcal{L}^p(\mu) \to \mathbb{R}^+ \\ N_p(\alpha f) &= |\alpha|\, N_p(f) \qquad (3.26) \\ N_p(f+g) &= N_p(f) + N_p(g) \qquad (3.27) \end{aligned}$$

Letzteres impliziert die Dreiecksungleichung für N_p:

$$d_p(f,g) := N_p(f-g) \qquad f, g \in \mathcal{L}^p(\mu).$$

Der Grund, dass $N_p(\cdot)$ nur eine Seminorm und $d_p(\cdot,\cdot)$ nur eine Pseudometrik ist, ist der, dass $N_p(f) = 0$ nicht impliziert, dass $f \equiv 0$, sondern nur $f = 0$ μ-f.ü. (entsprechend impliziert $d_p(f,g) = 0$ nur, dass $f = g$ μ-f.ü.). Wir betrachten nun Konvergenz bzgl. dieser Seminorm, wobei wir sagen, dass f_n gegen f in $\mathcal{L}^p$ konvergiert (in Symbolen $f_n \to^{\mathcal{L}^p} f$), falls $d_p(f_n, f) \to 0$. Es gibt eine kleine Schwierigkeit mit dieser Definition, denn der Limes ist, wie oben erwähnt, nicht eindeutig. Formal geht man daher zum Quotientenraum $\frac{\mathcal{L}^p(\mu)}{\simeq_\mu}$ über, wobei $f \simeq_\mu g :\Longleftrightarrow f = g \quad \mu - f.\ddot{u}$. In diesem ist der Limes bzgl. $d_p(f_n, f)$ eindeutig, denn $N_p(\cdot)$ist eine Norm auf $\frac{\mathcal{L}^p(\mu)}{\simeq_\mu}$.
In einem ersten Schritt etablieren wir eines der zentralen Lemmata der Integrationstheorie:

Lemma 3.46 (Fatou) *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Dann gilt für alle Folgen $(f_n)_n$ messbarer, positiver Funktionen:*

$$\int \liminf f_n d\mu \leq \liminf \int f_n d\mu. \qquad (3.28)$$

Beweis. Wir wissen schon, dass $f := \liminf_{n\to\infty} f_n$ und $g_n := \inf_{m\geq n} f_n$ in E^* sind. Definitionsgemäß gilt $g_n \uparrow f$ und daher

$$\int f d\mu = \sup_{n\in\mathbb{N}} \int g_n d\mu = \lim_{n\to\infty} \int g_n d\mu. \tag{3.29}$$

Schließlich ist $f_m \geq g_n$ für alle $m \geq n$ und somit

$$\int g_n d\mu \leq \inf_{m\geq n} \int f_m d\mu \tag{3.30}$$

(3.30) zusammen mit (3.29) ergibt (3.28). ∎
Wählt man $f_n = 1_{A_n}, A_n \in \mathcal{A}$, dann ist $\liminf_{n\to\infty} 1_{A_n}$ gegeben durch

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m. \tag{3.31}$$

Daher ist $\liminf A_n$ die Menge aller $\omega \in \Omega$, die in allen bis auf endlich viele der A_m sind. Ähnlich definiert man

$$\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m, \tag{3.32}$$

als die Merge aller $\omega \in \Omega$, die in unendlich vielen der An sind. Offenbar ist

$$(\limsup A_n)^c = \liminf \left(A_n^c\right).$$

Aus dem Fatouschen Lemma leitet man ab:
a) $\mu\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty} \mu\left(A_n\right)$
b) Falls μ endlich ist, gilt

$$\mu\left(\limsup_{n\to\infty} A_n\right) \geq \limsup_{n\to\infty} \mu\left(A_n\right).$$

Aus dem Lemma von Fatou folgt unser erster Konvergenzsatz auf überraschend einfache Art.

Theorem 3.47 (Riesz) *Sei $(f_n)_n$ eine Folge in $\mathcal{L}^p(\mu)$ mit $f_n \to f \in \mathcal{L}^p(\mu)$ μ-f.ü. Dann konvergiert f_n in $\mathcal{L}^p(\mu)$ gegen f, genau dann wenn*

$$\lim_{n\to\infty} \int |f_n|^p d\mu = \int |f|^p d\mu \tag{3.33}$$

Beweis. Zunächst bemerken wir, dass $N_p(f) = N_p(f-g+g) \leq N_p(f-g) + N_p(g)$ und $N_p(-g) = N_p(g)$ implizieren, dass

$$| N_p(f) - N_p(g) | \leq N_p(f \pm g). \tag{3.34}$$

Falls daher $f_n \to^{\mathcal{L}^p(\mu)} f$ gilt, so auch $N_p(f_n - f) \to 0$ und somit folgt (**??**) aus (3.34).
Für die Umkehrung, bemerke man, dass für $\alpha, \beta \in \mathbb{R}^+$ gilt $|\alpha - \beta|^p \leq (\alpha + \beta)^p \leq 2^p(\alpha^p + \beta^p)$. Also definiert

$$g_n := 2^p(|f_n|^p + |f|^p) - |f_n - f|^p\,, \quad n \in \mathbb{N}$$

eine Folge nicht-negativer Funktionen in $\mathcal{L}^1(\mu)$. Weil $f_n \to f \quad \mu - f.ü.$, wissen wir $g_n \to 2^{p+1}|f|^p \; \mu - f.ü.$. Dies ist auch der $\liminf$ der g_n's. Somit ergibt (**??**) zusammen mit dem Fatouschen Lemma

$$\begin{aligned} \int 2^{p+1}|f|^p\, d\mu \qquad &= \qquad \int \liminf_{n\to\infty} g_n d\mu \leq \liminf_{n\to\infty} \int g_n d\mu \\ &=^{(??)} 2^{p+1} \int |f|^p - \limsup \int |f_n - f|^p\, d\mu. \end{aligned}$$

Also

$$\limsup_{n\to\infty} \int |f_n - f|^p\, d\mu \leq 0.$$

■
Um einen wesentlichen Vorteil des Maßintegrals gegenüber dem Riemannintegral zu diskutieren, benötigen wir einen vorbereitenden Schritt:

Lemma 3.48 *Sei $(f_n)_n$ eine Folge in $E^*(\Omega, \mathcal{A})$. Dann gilt*

$$N_p(\sum_{i=1}^{\infty} f_n) \leq \sum_{n=1}^{\infty} N_p(f_n) \qquad (1 \leq p < \infty)\,. \tag{3.35}$$

Beweis. Der Beweis ist eine Folgerung aus der Minkowski-Ungleichung, der dem Leser überlassen bleibt. ■
Der folgende satz geht auf H. Lebesgue zurück.

Theorem 3.49 (Lebesgue, dominierte Konvergenz) *Sei (f_n) eine Folge in $\mathcal{L}^p(\mu), 1 \leq p < +\infty$, die $\mu - f.ü.$ auf Ω konvergiert. Sei $g : \Omega \to \bar{\mathbb{R}}^+$ in $\mathcal{L}^p(\mu)$ mit*

$$|f_n| \leq g, \qquad n \in \mathbb{N}. \tag{3.36}$$

Dann gibt es ein $f : \Omega \to \mathbb{R}$ mit $f_n \to f \quad \mu. - f.ü.$ Jedes solches f ist in $\mathcal{L}^p(\mu)$ und $f_n \to^{\mathcal{L}^p} f$.

Beweis. Zunächst werfen wir die Mengen weg, die unsere Berechnungen stören: Es gibt Nullmengen M_1, M_2, so dass $\lim f_n \in \overline{\mathbb{R}}$ existiert auf M_1^c und so dass $g < \infty$ auf M_2^c (hierzu bemerke, dass g^p integrierbar ist). Definiere

$$f(\omega) = \begin{cases} \lim_{n\to\infty} f_n(\omega) & \omega \in (M_1 \cup M_2)^c \\ 0 & \omega \in M_1 \cup M_2 \end{cases}$$

Dann ist $f : \Omega \to \mathbb{R}$ und f ist $\mathcal{A}$ - messbar. $f_n \to f \; \mu - f.ü.$ Da $|f| \leq g$ $\mu - f.ü.$ $\quad |f|^p \in \mathcal{L}^1(\mu)$. Wir setzen

$$g_n := |f_n - f|^p$$

und wir wollen zeigen, dass $\lim_n \int g_n d\mu = 0$. Definitionsgemäß gilt

$$0 \leq g_n \leq (|f_n| + |f|)^p \leq (|f| + g)^p \, .$$

Daher ist mit $h := (|f| + g)^p$ auch g_n integrierbar. Mit Fatou angewandt auf $(h - g_n)_n$ erhalten wir

$$\begin{aligned} \int h d\mu = \int \liminf (h - g_n) \, d\mu & \leq \liminf \int (h - g_n) \, d\mu \\ &= \int h d\mu - \limsup \int g_n d\mu. \quad (3.37) \end{aligned}$$

Somit gilt $\limsup \int g_n d\mu \leq 0$. Dies zeigt die Behauptung. ■

Die Sätze von der monotonen und majorisierten Konvergenz sind zwei wesentliche Vorteile des Lebesgueintegrals gegnüber dem Riemannintegral, für welches man gleichmäßige Konvergenz benötigt, um Integration und Limesbildung zu vertauschen.
Der letzte Satz zeigte, dass eine Folge, die f_n, die gegen einen Limes f konvergiert, der durch eine p-integrierbare Funktion beschränkt ist, auch in $\mathcal{L}^p$ gegen f konvergiert. Eine natürliche Frage ist, wann so ein f existiert. Die Antwort gibt der folgende Satz, der für $p = 2$ auf F. Riesz und E. Fischer zurückgeht.

Theorem 3.50 *Jede Cauchyfolge $(f_n)_n$ in $\mathcal{L}^p(\mu)$ konvergiert in $\mathcal{L}^p$ gegen einen Limes $f \in \mathcal{L}^p$. Ferner gibt es eine Teilfolge von $(f_n)_n$, die gegen f μ -f.ü. konvergiert.*

Beweis. Der Beweis soll hier nicht gegeben werden. Er findet sich z.B. im Buch von Bauer [**?**]. ■

Interessanterweise impliziert weder $\mathcal{L}^p$–Konvergenz $\mu-f.s.$–Konvergenz, noch umgekehrt. Dies ist der Inhalt der beiden nächsten Beispiele

Beispiel 3.51 *Sei $\Omega = [0,1[$ und $\mu := \lambda^1$ auf $\mathcal{A} := B^1 \cap \Omega$. Für $n \in \mathbb{N}$ wähle $k, h \in \mathbb{N}$ so, dass $n = 2^h + k < 2^{h+1}$. Diese Wahl ist eindeutig. Setze*

$$A_n : [k2^{-h}, (k+1)2^{-h}[\text{ und } f_n := 1_{A_n}.$$

dann gilt

$$\int f_n^p d\mu = \int f_n d\mu = \mu(A_n) = 2^{-h}$$

Da $h \to \infty$ wenn $n \to \infty$, konvergiert f_n gegen null in $\mathcal{L}^p(\lambda^1)$ für alle p. Andererseits ist (f_n) für kein $\omega \in \Omega$ konvergent. In der Tat, für $\omega \in \Omega$ und $h = 0, 1, 2, ...$ gibt es ein eindeutig bestimmtes k mit $\omega \in A_{2^h+k}$. Falls $k < 2^h - 1$, folgt $\omega \notin A_{2^h+k+1}$, falls $h = 2^h - 1$ und $h \geq 1$, gilt $\omega \notin A_{2^{h+1}}$

Beispiel 3.52 *Sei wieder $\Omega = [0,1[, \mathcal{A} = \mathcal{B}^1 \cap \Omega$ und $\mu = \lambda^1 \mid \Omega$. Sei $A_n = [0, \frac{1}{n}[$ und $f_n = n^2 1_{[0,\frac{1}{n}[} = n^2 1_{A_n}$. Dann gilt $f_n \to 0 \quad \lambda^1 - a.e.$, aber für jedes $1 \leq p < \infty$ ist*

$$\int |f_n|^p \, d\mu = n^p \to_{n\to\infty} \infty.$$

Daher konvergiert $N_p(f_n)$ nicht gegen null.

Aus dem oben Hergeleiteten folgt für **endliche** Maße –also auch Wahrscheinlichkeiten:
Sei μ endlich. Dann folgt aus der Hölderschen Ungleichung für $1 \leq p' \leq p < \infty$ und eine numerische Funktion f

$$N_{p'}(f) \leq N_p(f)\mu^{\frac{1}{p'}-\frac{1}{p}}(\Omega)$$

und

$$\mathcal{L}^p(\mu) \subseteq \mathcal{L}^{p'}(\mu).$$

Schließlich impliziert Konvergenz in $\mathcal{L}^p(\mu)$ auch Konvergenz in $\mathcal{L}^{p'}(\mu)$.
Abschließend wollen wir den Zusammenhang zwischen Riemann– und Lebesgue–Integration näher betrachten. Sei dafür zunächst $[a, b]$ ein kompaktes Intervall.

Theorem 3.53 *Sei $f : [a,b] \to \mathbb{R}$ eine Borel–messbare Funktion. Ist f Riemann–integrierbar, so ist f auch Lebesgue–integrierbar und die beiden Integrale stimmen überein.*

Beweis. Sei $\tau : a = \alpha_0 \leq \alpha_1 \leq \alpha_2 \leq .. \leq \alpha_n = b$ eine Partition von $[a,b]$. Für das Riemann–Integral müssen wir

$$L_\tau := \sum_{i=1}^{n} \phi_i(\alpha_i - \alpha_{i-1}) \quad \text{und } \mathcal{U}_\tau := \sum_{i=1}^{n} \Phi_i(\alpha_i - \alpha_{i-1})$$

mit

$$\phi_i = \inf\{f(x), x \in [\alpha_{i-1}, \alpha_i]\} \text{ und}$$
$$\Phi_i := \sup\{f(x), x \in [\alpha_{i-1}, \alpha_i]\}.$$

betrachten. Nun sind für $\mu = \lambda^1$ die Funktionen

$$l_\tau := \sum_{i=1}^{n} \phi_i 1_{A_i} \text{ und } u_\tau = \sum_{i=1}^{n} \Phi_i 1_{A_i},$$

wobei $A_i = [\alpha_{i-1}, \alpha_i])$ ist, μ–integrierbar und

$$L\tau = \int l_\tau d\mu \quad \text{und} \quad \mathcal{U}_\tau = \int u_\tau d\mu.$$

Da f Riemann–integrierbar ist, gibt es eine Folge aufsteigender Partitionen von $[a,b]$, so dass $(L_{\tau_n})_n$ und $(\mathcal{U}_{\tau_n})_n$ den gleichen Limes haben. Entsprechend sind die Folgen $(\mathcal{U}_{\tau_n})$ und (L_{τ_n}) fallend bzw. wachsend und die Funktion

$$q := \lim_{n\to\infty} (u_{\tau_n} - l_{\tau_n})$$

existiert. Wegen $u_{\tau_n} \geq l_{\tau_n}$ folgt mit dem Fatouschen Lemma

$$0 \leq \int q d\mu \leq \lim_{n\to\infty} (\mathcal{U}_{\tau_n} - L_{\tau_n}) = 0$$

Daher ist $q = 0 \quad \lambda - f.s.$ und da $l_{\tau_n} \leq f \leq u_{\tau_n}$ zeigt dies, dass $\lim_{n\to\infty} l_{\tau_n} = f$ $\lambda^1 - f.s.$
Nun ist f Riemann–integrierbar und somit beschränkt, also Lebesgue–integrierbar, weil $\lambda([a,b])$ endlich ist. Daher kann man $(|\ell_{\tau_n}|)$ durch eine λ–integrierbare Funktion dominieren. Aus dem Satz über dominierte Konvergenz folgt

$$\int f d\lambda = \int \lim l_{\tau_n} d\lambda = \lim \int l_{\tau_n} d\lambda = \lim_{n\to\infty} L_{\tau_n} = \int f dx.$$

■

Bemerkung 3.54 *Umgekehrt gibt es natürlich Lebesgue-integrierbare Funktionen, die nicht Riemann–integrierbar sind, z.B. die Dirichletsche Sprungfunktion* $1_{\mathbb{Q}}(x)$

Der Fall unbeschränkter Intervalle ist nun leicht:

Korollar 3.55 *Sei* $f \geq 0, f : \mathbb{R} \to \mathbb{R}$ *eine* B^1 *- messbare Funktion.* f *sei Riemann–integrierbar für alle Intervalle* $[a, b], a < b \in \mathbb{R}$*.* f *ist genau dann Lebesgue–integrierbar, falls der Limes der Riemann–Integrale*

$$\rho := \lim_{n \to \infty} \int_{-n}^{n} f(x)dx$$

existiert. In diesem Falle ist ρ *das Lebesgue–Integral.*

Beweis. Sei ρ_n das Riemann–Integral von f über $A_n = [-n, n]$. Aus dem letzten Satz folgt, dass

$$\rho_n = \int_{A_n} f d\lambda^1 = \int_{\mathbb{R}} 1_{A_n} f d\lambda^1.$$

Schließlich gilt $f 1_{A_n} \uparrow f$ und somit

$$\sup \rho_n = \int f d\lambda.$$

f ist Riemann–integrierbar genau dann wenn if $\sup \rho_n < \infty$ und dann gilt $\rho = \sup \rho_n$. Dies beweist die Behauptung. ∎
Schlussendlich bekommt man für beliebige messbare Funktionen von $\mathbb{R}$ nach $\mathbb{R}$, dass Riemann– und Lebesgue–Integral übereinstimmen, falls $|f|$ Riemann–integrierbar ist auf $\mathbb{R}$. Andererseits impliziert die Existenz des uneigentlichen Riemann–Integrals auf $\mathbb{R}$ nicht Lebesgue–Integrierbarkeit:

Übung 3.56 *Betrachte* $f(x) := \frac{1}{x} \sin(x)$*. Zeigen Sie, dass* f *stetig in* 0 *fortgesetzt werden kann, uneigentlich Riemann–integrierbar, aber nicht Lebesgue–integrierbar ist.*

3.5 Stochastische Konvergenz

Wir werden hier noch einen weiteren Konvergenzbegriff neben $\mu - f.s.$–Konvergenz und Konvergenz in $\mathcal{L}^p(\mu)$ betrachten: stochastische Konvergenz. Diese ist motiviert durch das schwache Gesetz der großen Zahlen. Sei, wie immer, $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.

Definition 3.57 *Eine Folge $(f_n)_n$ messbarer, reeller Funktionen heißt stochastisch konvergent gegen $f : \Omega \to \mathbb{R}$, falls für alle $\varepsilon > 0$ und $A \in \mathcal{A}$ mit $\mu(A) < \infty$ gilt*

$$\lim_{n\to\infty} \mu\left(\{|f_n - f| \geq \varepsilon\} \cap A\right) = 0. \tag{3.38}$$

Ist μ endlich, ist (3.38) äquivalent zu

$$\lim_{n\to\infty} \mu\left(\{|f_n - f| \geq \varepsilon\}\right) = 0 \tag{3.39}$$

für alle $\varepsilon > 0$.

Für nicht–endliche Maße sind (3.38) und (3.39) nicht äquivalent. Als erstes fragen wir uns, ob stochastische Limiten eindeutig sind.

Theorem 3.58 *a)Sei $(f_n)_n$ stochastisch konvergent gegen f und $f^* = f$ $\mu - f.s.$ Dann konvergiert $(f_n)_n$ auch stochastisch gegen f^*.*
b) Wenn μ σ–endlich ist, sind je zwei Limiten einer stochastisch konvergenten Folge $\mu - f.s.$ gleich.

Beweis. a) Ist klar, denn $\{f_n - f\} \cap A$ und $\{f_n - f^*\} \cap A$ unterscheiden sich nur um eine n-unabhängige Nullmenge.
b) Beweisen wir nicht, sondern verweisen auf den Beweis im Buch von Bauer [?]. ■
Die folgende Ungleichung ist zentral in der Wahrscheinlichkeitstheorie und das Verbindungsglied zwischen stochastischer Konvergenz und Konvergenz in $\mathcal{L}^p$.

Lemma 3.59 (Chebyshev–Markov Ungleichung) *Sei $f : \Omega \to \mathbb{R}$ messbar und $g : \mathbb{R} \to \mathbb{R}^+$ wachsend. Dann gilt für alle $\varepsilon > 0$*

$$\mu\left(f \geq \varepsilon\right) \leq \frac{1}{g(\varepsilon)} \int g(f) d\mu. \tag{3.40}$$

Beweis. Sei $A_\varepsilon := \{f \geq \varepsilon\} \in \mathcal{A}$. Dann gilt

$$\int g(f) d\mu \geq \int_{A_\varepsilon} g(f) d\mu \geq \int_{A_\varepsilon} g(\varepsilon) = g(\varepsilon)\mu\left(A_\varepsilon\right).$$

■
Eine unmittelbare Folge ist

Theorem 3.60 *Sei $(f_n)_n$ eine Folge in $\mathcal{L}^p(\mu)$. Falls (f_n) gegen $f \in \mathcal{L}^p(\mu)$ in $\mathcal{L}^p$ konvergiert, dann auch μ - stochastisch.*

Beweis. Aus der Chebyshev–Markov Ungleichung angewandt auf $|f_n - f|$ mit $g(x) = x^p$ folgt

$$\mu\left(\{|f_n - f| \geq \varepsilon\} \cap A\right) \leq \mu\left(\{|f_n - f| \geq \varepsilon\}\right) \leq \varepsilon^{-p} \int |f_n - f|^p \, d\mu$$

Die rechte Seite konvergiert gegen Null nach Voraussetzung. ■
Wir sehen nun, dass stochastische Konvergenz auch schwächer ist als f.s.–Konvergenz.

Theorem 3.61 *Sei (f_n) eine Folge messbarer Funktionen auf Ω, die $\mu - f.s.$ gegen eine messbare, reelle Funktion f auf Ω konvergiert. Dann konvergiert (f_n) auch also μ - stochastisch gegen f.*

Beweis. Es gilt

$$\{|f_n - f| \geq \varepsilon\} \subseteq \left\{\sup_{m \geq n} |f_m - f| \geq \varepsilon\right\}$$

und daher

$$\mu\left(\{|f_n - f| \geq \varepsilon\} \cap A\right) \leq \mu\left(\left\{\sup_{m \geq n} |f_m - f| \geq \varepsilon\right\} \cap A\right)$$

für alle $\varepsilon > 0$ und $A \in \mathcal{A}$ gilt. Die Behauptung folgt daher aus dem folgenden Lemma, denn für A mit $\mu(A) < \infty$ das Maß $\mu\,|_{A \subset \mathcal{A}}$ endliche Masse hat. ■

Lemma 3.62 *μ sei endlich. Die Folge $(f_n)_n$ messbarer, reeller Funktionen konvergiert $\mu - f.s.$ gegen 0, wenn eine der folgenden äquivalenten Bedingungen erfüllt ist.*

$$\lim_{n \to \infty} \mu(\{\sup_{m \geq n} |f_m| \geq \varepsilon\}) = 0 \quad \textit{für alle } \varepsilon > 0 \tag{3.41}$$

$$\lim_{n \to \infty} \mu(\{\sup_{m \geq n} |f_m| > \varepsilon\}) = 0 \quad \textit{für alle } \varepsilon > 0 \tag{3.42}$$

$$\mu(\{\limsup_{n \to \infty} \{|f_n| > \varepsilon\}) = 0 \quad \textit{für alle } \varepsilon > 0. \tag{3.43}$$

Beweis. Wir zeigen nur den Teil, den wir auch anwenden, nämlich die Äquivalenz von (3.41) und der $\mu - f.s.$ Konvergenz.
Für $\varepsilon > 0$ und $n \in \mathbb{N}$ setze

$$A_n^\varepsilon := \left\{ \sup_{m \geq n} |f_m| \geq \varepsilon \right\}.$$

Offensichtlich sind $n \mapsto A_n^\varepsilon$ und $\varepsilon \mapsto A_n^\varepsilon$ fallend und daher ist $k \mapsto A_n^{\frac{1}{k}}$ wachsend. Definieren wir schließlich

$$A := \{\omega : \lim_{n\to\infty} |f_n(\omega)| = 0\} = \{\omega : \lim_{n\to\infty} \sup |f_n(\omega)| = 0\}$$

Dann ist $A \in \mathcal{A}$ da $\limsup f_n$ messbar ist. Offensichtlich gilt

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \left(A_n^{\frac{1}{k}}\right)^c = \bigcap_{\alpha>0} \bigcup_{n=1}^{\infty} (A_n^\alpha)^c$$

und somit

$$A^c = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} A_n^{\frac{1}{k}} = \bigcup_{\alpha>0} \bigcap_{n=1}^{\infty} A_n^\alpha.$$

Daher folgt

$$\bigcap_{n=1}^{\infty} A_n^{\frac{1}{k}} \to A^c \text{ and } A_n^{\frac{1}{k}} \downarrow \bigcap_{m=1}^{\infty} A_m^{\frac{1}{k}}.$$

Also

$$\mu(A^c) = \sup_k \mu\left(\bigcap_{n=1}^{\infty} A_n^{\frac{1}{k}}\right) = \sup_k \inf \mu(A_n^{\frac{1}{k}}) \tag{3.44}$$

da μ als endliches Maß stetig von oben und unten ist. Also konvergiert f_n gegen null $\mu - f.s.$ genau dann wenn die Zahl in (3.44) Null ist. Dies ist aber genau dann der Fall, wenn

$$\inf_n \mu\left(A_n^{\frac{1}{k}}\right) = \lim_{n\to\infty} \mu\left(A_n^{\frac{1}{k}}\right) = 0 \quad \text{für alle } k \in \mathbb{N}.$$

Dies zeigt die Äquivalenz von (3.41) und $\mu - f.s.$ Konvergenz. ■
Die Beispiele im vorigen Abschnitt zeigen nun, dass es in der Tat Folgen gibt, die stochastisch konvergieren, aber nicht f.s. oder in $\mathcal{L}^p$.

4 Produktmaße und der Satz von Fubini

Bislang haben wir nur den Fall eines einzigen Maßraumes $(\Omega, \mathcal{A}, \mu)$ betrachtet. Wir wir in der Folge sehen werden kann dies für den Fall, dass μ Masse eins hat mit einem Zufallsexperiment identifiziert werden. Natürlich ist man in der Wahrscheinlichkeitstheorie nicht nur an einem Zufallsexperiment interessiert, sondern an der vielfachen Hintereinanderausführung eines solchen – darüber hinaus nimmt man häufig an, dass die zugehörigen Experimente unabhängig sind.
Wie kann man das modellieren? Schon in der Stochastikvorlesung lernt man, dass die Unabhängigkeitsannahme mit einer Multiplikationsvorschrift für die zugehörigen Wahrscheinlichkeiten korrespondiert. Auf der Maßebene entspricht dem der Übergang zu einem Produktmaß. In diesem Kapitel lernen wir, wie ein solches Maß geeignet definiert werden kann und wie man bezüglich eines solchen Maßes integriert. Das generische Beispiel ist hierbei das des d–dimensionalen Lebesgue Maßes λ^d. Tatsächlich sind ja die Mengen, auf denen λ^d zunächst definiert ist, die d–dimensionalen Intervalle. Diese lassen sich als die Produkte eindimensionaler Intervalle auffassen, ihr Inhalt als das Produkt der eindimensionalen Inhalte.
Wir werden sehen, dass der Fall eines allgemeinen Produktmaßes sehr ähnlichen Ideen folgt. Man definiert zunächst den Inhalt von "Rechtecken" und erweitert diesen auf die Produkt–σ-Algebra .
Gegeben seien für diese Kapitel stets messbare Räume $(\Omega_i, \mathcal{A}_i), i = 1, ...n$. Wir definieren

$$\Omega := \prod_{i=1}^{n} \Omega_i = \Omega_1 \times ... \times \Omega_n,$$

und die Projektionen

$$p_i : \Omega \to \Omega_i$$

die $(\omega_1, ...\omega_n)$ auf ω_i abbilden. Die Produkt–σ–Algebra ist definiert durch

$$\otimes_{i=1}^{n} \mathcal{A}_i := \mathcal{A}_1 \otimes ... \otimes \mathcal{A}_n := \sigma(p_1, ..., p_n).$$

Sie ist durch die Projektionen p_i erzeugt, d.h. sie ist die kleinste σ–Algebra, so dass alle Projektionen p_i messbar sind. Die folgende Art $\mathcal{A}$ zu erzeugen wird zentral für den Rest des Kapitels sein.

Theorem 4.1 *Sei $\mathbb{E}_i$ eine Erzeuger von $\mathcal{A}_i$ für $i = 1, \ldots, n$, so dass für jedes i eine Folge $(E_{ik})_k$ in $\mathbb{E}_i$ mit $E_{ik} \uparrow \Omega_i$ wenn $k \to \infty$ existiert. Dann wird $\mathcal{A}_1 \otimes .. \otimes \mathcal{A}_n$ erzeugt durch*

$$\{E_1 \times \cdots \times E_n, \qquad E_i \in \mathbb{E}_i\}.$$

Beweis. Sei $\mathcal{A}$ eine beliebige σ–Algebra über Ω. p_i ist $\mathcal{A} - \mathcal{A}_i$–messbar genau dann, wenn gilt $p_i^{-1}(E_i) \in \mathcal{A}$ für alle $E_i \in \mathbb{E}_i$. Wenn dies aber für alle i wahr ist, dann ist auch

$$E_1 \times ... \times E_n = \bigcap_{i=1}^{n} p_i^{-1}(E_i) \in \mathcal{A}.$$

Ist andererseits $E_1 \times ... \times E_n \in \mathcal{A}$ für alle $E_i \in \mathbb{E}_i$, dann gilt auch

$$F_k := E_{1k} \times E_{2k} \times \cdots \times E_{(i-1)k} \times E_i \times E_{(i+1)k} \times ... \times E_{nk \in} \mathcal{A}$$

für alle k. Aber (F_k) konvergiert gegen

$$\Omega_1 \times \ldots \times E_i \times \Omega_{i+1} \times .. \times \Omega_k = p_i^{-1}(E_i),$$

also ist auch $p_i^{-1}(E_i) \in \mathcal{A}$. ■

Beispiel 4.2 *Wählt man $\Omega_i = \mathbb{R}$ für alle i, $\mathcal{A}_i = \mathcal{B}^1$ für alle i und $\mathbb{E}_i = \mathcal{J}^1$, dann ist offenbar*

$$\{E_1 \times .. \times E_n, E_i \in \mathbb{E}_i\} = \mathcal{J}^d.$$

Theorem 4.1 ergibt daher $\mathcal{B}^d = \mathcal{B}^1 \times .. \times \mathcal{B}^1$ (d Mal). Wie wir beim Studium des Lebesgue–Maßes gesehen haben, ist λ^n das eindeutige Maß auf $\mathcal{B}^n$ mit

$$\lambda^n(I_1 \times .. \times I_n) = \lambda^1(I_1)....\lambda^1(I_n)$$

für alle $I_i \in \mathcal{J}^1$.

Dieses Beispiel führt unmittelbar zu den Fragen: Gegeben eine Familie von Maßen μ_i auf $(\Omega_i, \mathcal{A}_i)$. Wann gibt es ein Maß π auf $(\Omega, \mathcal{A})$ so dass

$$\pi(E_1 \times .. \times E_n) = \mu(E_1)...\mu(E_n) \tag{4.1}$$

für alle $E_i \in \mathbb{E}_i$ (wobei angenommen sei, dass die $\mathbb{E}_i$ die $\mathcal{A}_i$ erzeugen) und wann ist dieses π eindeutig?
Die zweite Frage beantworten wir sofort.

Theorem 4.3 *Falls jeder Erzeuger $\mathbb{E}_i$ von $\mathcal{A}_i$ $\bigcap$–stabil ist und eine Folge (E_{ik}) mit $E_{ik} \uparrow \Omega_i$ und $\mu(E_{ik}) < \infty$ enthält, gibt es höchstens ein π wie in* (4.1)

Beweis. Sei

$$\mathbb{E} := \{E_1 \times ... \times E_n, E_i \in \mathbb{E}_i\}.$$

Theorem 4.1 zeigt, dass $\mathbb{E}$ $\mathcal{A}_1 \otimes .. \otimes \mathcal{A}_n$ erzeugt. Da

$$\left(\prod_{i=1}^{n} E_i\right) \cap \left(\prod_{i=1}^{n} F_i\right) = \prod_{i=1}^{n} (E_i \cap F_i)$$

ist mit $\mathbb{E}_i$ auch $\mathbb{E}$ $\cap$–stabil. Darüber hinaus gilt

$$E_k := E_{1k} \times .. \times E_{nk} \uparrow \Omega.$$

Die Behauptung folgt daher aus dem Eindeutigkeitsteil des Satzes von Carathéodory, da

$$\pi(E_k) = \mu_1(E_{1k}) ....\mu_n(E_{nk}) < \infty.$$

■

Nun beantworten wir auch die erste Frage, d.h. wir konstruieren das Produktmaß auf dem Produkt zweier Maßräume $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$. Die Verallgemeinerung auf größere n ist eine Standardinduktion.
Für $Q \subseteq \Omega_1 \times \Omega_2$ und $\omega_i \in \Omega_i, i = 1, 2$, definieren wir zuerst

$$Q_{\omega_1} := \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in Q\}$$

und

$$Q_{\omega_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in Q\},$$

die sogenannten ω_1– bzw. ω_2–Schnitte von Q. Wir erhalten

Lemma 4.4 *Sei $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Dann gilt für beliebige $\omega_1 \in \Omega$ und $\omega_2 \in \Omega_2$, $Q_{\omega_1 \in} \mathcal{A}_2$ und $Q_{\omega_2} \in \mathcal{A}_1$.*

Beweis. Für $Q, Q_1, Q_2, ... \subseteq \Omega_1 \times \Omega_2$ und beliebige $\omega_1 \in \Omega$ gilt $(\Omega \backslash Q)_{\omega_1} = \Omega_2 \backslash Q_{\omega_1}$ und

$$\left(\bigcup_{i=1}^{\infty} Q_i\right)_{\omega_1} = \bigcup_{i=1}^{\infty} (Q_i)_{\omega_1}.$$

Darüberhinaus gilt $\Omega_{\omega_1} = \Omega_2$ und allgemeiner

$$(A_1 \times A_2)_{\omega_1} = \begin{cases} A_2 & \text{falls } \omega_1 \in A_1 \\ \emptyset & \text{andernfalls,} \end{cases}$$

wobei $A_i \subset \Omega_i$ sei. Daher erhalten wir für alle $\omega_1 \in \Omega$, dass

$$\mathcal{A}' := \{Q \subset \Omega : Q_{\omega_1} \in \mathcal{A}_2\}$$

eine σ–Algebra über Ω ist. $\mathcal{A}'$ enthält alle Mengen der Form $A_1 \times A_2, A_i \in \mathcal{A}_i$. Aus Theorem 4.1 wissen wir, dass $\mathcal{A}_1 \otimes \mathcal{A}_2$ die kleinste solche σ-Algebra ist. Dies zeigt das Lemma für Q_{ω_1}. Der Beweis für Q_{w_2} geht analog. ■
Demzufolge können wir Q_{ω_1} mit μ_2 und Q_{ω_2} mit μ_1 messen. Weiter gilt

Lemma 4.5 *Sind μ_1, μ_2 σ–endlich, so sind für alle $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ die Funktionen*

$$\omega_1 \longmapsto \mu_2(Q_{\omega_1}) \text{ und } \omega_2 \longmapsto \mu_1(Q_{\omega_2})$$

(definiert auf Ω_1 bzw. Ω_2) $\mathcal{A}_1$– bzw. $\mathcal{A}_2$–messbar.

Beweis. Sei $\varsigma_Q(\omega_1) = \mu_2(Q_{\omega_1})$. Angenommen, dass $\mu_2(\Omega_2) < \infty$, definieren wir

$$\mathcal{D} := \{D \in \mathcal{A}_1 \otimes \mathcal{A}_2, \varsigma_D \text{ is } \mathcal{A}_1\text{–messbar}\}. \tag{4.2}$$

Dann ist $\mathcal{D}$ ein Dynkin–System, das alle Mengen der Form $A_1 \times A_2, A_i \in \mathcal{A}_i, i = 1, 2$ enthält (Übung). Das System $\mathbb{E}$ aller Mengen $A_1 \times A_2, A_i \in \mathcal{A}_i, i = 1, 2$ ist $\cap$–stabil und erzeugt $\mathcal{A}_1 \otimes \mathcal{A}_2$. Daher gilt $\mathcal{D}(\mathbb{E}) = \mathcal{A}_1 \otimes \mathcal{A}_2$ und da $\mathbb{E} \subseteq \mathcal{D} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$, erhalten wir $\mathcal{D} = \mathcal{A}_1 \otimes \mathcal{A}_2$.
Der Beweis für den Fall, dass μ_2 nur σ–endlich ist, befindet sich im Buch von Bauer [**?**].
Die andere Behauptung folgt analog. ■
Nun erhält man schnell die Existenz von Produktmaßen.

Theorem 4.6 *Seien $(\Omega_i, \mathcal{A}_i, \mu_i), i = 1, 2$ be σ–endliche Maßräume. Dann gibt es ein eindeutig bestimmtes Maß π auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ mit*

$$\pi(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2) \tag{4.3}$$

für alle $A_i \in \mathcal{A}_i, i = 1, 2$. Für jedes $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ gilt

$$\pi(Q) = \int \mu_2(Q_{\omega_1})\mu_1(d\omega_1) = \int \mu_1(Q\omega_2)\,\mu_2(d\omega_2). \tag{4.4}$$

Offensichtlich ist mit μ_1 und μ_2 auch π σ–endlich.

Beweis. Sei wieder $\varsigma_Q(\omega_1) = \mu_2(Q_{\omega_1})$. Definiere $\pi(Q) := \int \varsigma_Q d\mu_1$. Für jede Folge (Q_n) paarweise disjunkter Mengen in $\mathcal{A}_1 \otimes \mathcal{A}_2$ gilt $\varsigma_{\cup Q_n} = \sum \varsigma_{Q_n}$ und somit wegen monotoner Konvergenz

$$\pi\left(\bigcup_{n=1}^{\infty} Q_n\right) = \sum_{n=1}^{\infty} \pi(Q_n).$$

Wegen $\varsigma_\emptyset = 0$ gilt auch $\pi(\emptyset) = 0$ und somit ist π ein Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$. π hat die Eigenschaft (4.3), weil

$$\varsigma_{A_1 \times A_2} = \mu_2(A_2) 1_{A_i}$$

und daher

$$\pi(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2).$$

Genauso definiert man ein Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ vermöge

$$\pi^{'}(Q) = \int \mu_1(Q_{\omega_2}) \mu_2(d\omega_2)$$

Wegen des obigen Eindeutigkeitssatzes angewandt auf $\mathbb{E}_1 = \mathcal{A}_1$ und $\mathbb{E}_2 = \mathcal{A}_2$, gibt es höchstens ein solches Maß, also ist $\pi^{'} = \pi$ und die zweite Gleichung in (4.4) folgt. ■

Definition 4.7 *Das Maß π im vorigen Satz heißt das Produktmaß von μ_1 und μ_2 und wird mit $\mu_1 \otimes \mu_2$ bezeichnet.*

Das bekannteste Beispiel eines Produktmaßes ist, wie schon eingangs erwähnt, das Lebesgue–Maß. Für dieses gilt

$$\lambda^2 = \lambda^1 \otimes \lambda^1 \text{ oder allgemeiner } \lambda^{m+n} = \lambda^m \otimes \lambda^n.$$

Nun betrachten wir die Integration bezüglich eines Produktmaßes. Um die Notation einfach zu halten, schreiben wir

$$\omega_2 \longmapsto f_{\omega_1}(\omega_2) := f(\omega_1, \omega_2)$$

und

$$\omega_1 \longmapsto f_{\omega_2}(\omega_1) := f(\omega_1, \omega_2)$$

für eine Funktion $f : \Omega_1 \times \Omega_2 \to \Omega^{'}$ (für eine Menge $\Omega^{'}$) und $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$.
Für $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ gilt offenbar

$$(1_Q)_{\omega_1} = 1_{Q_{\omega_1}} \text{ und } (1_Q)_{\omega_2} = 1_{Q_{\omega_2}}. \tag{4.5}$$

Aus dem oben gezeigten ergibt sich schnell das folgende:
Ist $\left(\Omega^{'}, \mathcal{A}^{'}\right)$ ein messbarer Raum und

$$f : \Omega_1 \times \Omega_2 \to \Omega^{'}$$

eine messbare Abbildung, so sind auch die Abbildungen f_{ω_1} bzw. f_{ω_2} $\mathcal{A}_2$ - $\mathcal{A}^{'}$– bzw. $\mathcal{A}_1$ - $\mathcal{A}^{'}$–messbar.
Schon aus (4.4) erhält man eine Idee, wie die Integration bzgl. eines Produktmaßes $\mu_1 \otimes \mu_2$ vonstatten gehen sollte. Die folgenden beiden Sätze, die auf Tonelli und Fubini zurückgehen, erweitern (4.4) auf $\mu_1 \otimes \mu_2$–integrierbare Funktionen.

Theorem 4.8 (Tonelli) *Seien* $(\Omega_i, \mathcal{A}_i, \mu_i)\,, i = 1, 2$ *zwei* σ*–endliche Maßräume und sei*

$$f : \Omega_1 \otimes \Omega_2 \to \bar{\mathbb{R}}^+$$

$\mathcal{A}_1 \otimes \mathcal{A}_2$*–messbar. Dann sind*

$$\omega_2 \longmapsto \int f_{\omega_2} d\mu_1 \text{ und } \omega_1 \longmapsto \int f_{\omega_2} d\mu_2$$

$\mathcal{A}_1$*– bzw.* $\mathcal{A}_2$*–messbar. Ferner gilt:*

$$\begin{aligned} \int f d\left(\mu_1 \otimes \mu_2\right) &= \int \left(\int f_{\omega_2} d\mu_1\right) \mu_2(d_{\omega_2}) \\ &= \int \left(\int f_{\omega_1} d\mu_2\right) \mu_1\left(d\omega_1\right). \end{aligned} \tag{4.6}$$

Beweis. Setze $\Omega := \Omega_1 \times \Omega_2, \mathcal{A} := \mathcal{A}_1 \otimes \mathcal{A}_2$ und $\pi := \mu_1 \otimes \mu_2$. Für Treppenfunktionen

$$f = \sum_{i=1}^{n} \alpha_i 1_{Q^i} \quad , \alpha_i \geq 0, Q^i \in \mathcal{A},$$

gilt $f_{\omega_2} = \sum_i \alpha_i 1_{Q^i_{\omega_2}}$ und wegen (4.5) folgt

$$\int f_{\omega_2} d\mu_1 = \sum_{i=1}^{n} \alpha_i \mu_1(Q^i_{\omega_2}).$$

Also ist $\omega_2 \longmapsto \int f_{\omega_2} d\mu$ $\mathcal{A}_2$–messbar. Mit (4.4) erhält man

$$\int \left(\int f_{\omega_2} d\mu_1 \right) \mu_2(d\omega_2) = \sum_{i=1}^{n} \alpha_i \pi(Q^i) = \int f d\pi$$

also die erste Gleichung in (4.6).
Für beliebige, positive $\mathcal{A}$–messbare Funktionen $f \geq 0$ sei (u^n) eine Folge von Elementarfunktionen mit $u^n \uparrow f$. Dann ist nach dem ersten Teil des Beweises $\left(u_{\omega_2}^n\right)$ eine Folge von Elementarfunktionen bzgl. $\mathcal{A}_2$ mit $u_{\omega_2}^n \uparrow f_{\omega_2}$. Somit erhalten wir, dass

$$\omega_2 \mapsto \varphi^n(\omega_2) := \int u_{\omega_2}^n d\mu_1$$

wachsend ist mit Supremum

$$\omega_2 \longmapsto \int f_{\omega_2} d\mu_1. \tag{4.7}$$

Also ist auch die Funktion, die wir soeben in (4.7) definiert haben, messbar und aus dem Satz von der monotonen Konvergenz erhalten wir

$$\int \left(\int f_{\omega_2} d\mu_2 \right) \mu_2(d\omega_2) \quad = \quad \sup_n \int \varphi^n d\mu_2 = \sup_n \int u^n d\pi,$$

wobei wir für die zweilte Gleichheit den ersten Beweisschritt verwandt haben.
Wegen der Wahl von (u^n) erhalten wir

$$\int \varphi^n d\mu_2 \uparrow \int f d\pi$$

also

$$\int (\int f\omega_2 d\mu_1) \mu_2(d\omega_2) = \int f d\pi.$$

Analoge Beweisschritte beenden den Beweis. ■

Theorem 4.9 (Fubini) *Es seien* $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, 2$ *zwei* σ*–endliche Maßräume und sei*

$$f : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$$

messbar und integrierbar bzgl. $\mu_1 \otimes \mu_2$. *Dann ist* f_{ω_1} μ_2*–integrierbar* $\mu_1 - f.s.$ *in* ω_1 *und* $f\omega_2$ *ist* μ_1*–integrierbar* $\mu_2 - f.s.$ *Daher sind die folgenden Funktionen f.s. definiert:*

$$\omega_1 \mapsto \int f\omega_1 d\mu_2 \ \textit{und} \ \omega_2 \mapsto \int f\omega_2 d\mu_1.$$

Beide Funktionen sind integrierbar (bzgl. μ_1 *bzw.* μ_2*) und es gilt (4.6).*

Beweis. Offensichtlich gilt

$$\left|f_{\omega_i}\right| = \left|f\right|_{\omega_i}, \left(f^+_{\omega_i}\right) = \left(f^+\right)_{\omega_i}, \left(f^-_{\omega_i}\right) = \left(f^-\right)_{\omega_i}.$$

Wendet man (4.6) auf $|f|$ und $\mu_1 \otimes \mu_2$ an, so erhält man

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

Daher ist $\omega_1 \mapsto \int \left|f_{\omega_1}\right| d\mu_2$ $\mu_1 - f.s.$ endlich, also auch $\mu_1 - f.s.$ μ_2–integrierbar. Daher ist

$$\omega_1 \mapsto \int f_{\omega_1} d\mu_2 = \int f^+_{\omega_1} d\mu_2 - \int f^-_{\omega_i} d\mu_2$$

$\mu_1 - f.s.$ definiert und $\mathcal{A}_1$–messbar. Wendet man den Satz von Tonelli f^+ und f^- an, so erhält man:

$$\begin{aligned}\int \left(\int f_{\omega_1} d\mu_2 \right) d\mu_1 &= \int \left(\int f^+_{\omega_1} d\mu_2 \right) d\mu_1 - \int \left(\int f^-_{\omega_1} d\mu_2 \right) d\mu_1 \\ &= \int f^+ d\pi - \int f^- d\pi = \int f d\pi.\end{aligned}$$

Vertauscht man die Rollen von ω_1 und ω_2, erhält man die fehlende Richtung.
■

Die Verallgemeinerung auf den den Fall von mehr (aber endlich vieler) Faktoren, bleibt dem Leser überlassen. Der Beweis erfolgt per Induktion.

Übung 4.10 *Seien $\mu_1, ..., \mu_n$ σ–endliche Maße auf $(\Omega_1, \mathcal{A}_1), ..., (\Omega_n, \mathcal{A}_n)$. Dann gibt es ein eindeutig bestimmtes Maß π auf $\mathcal{A}_1 \otimes ... \otimes \mathcal{A}_n$ mit*

$$\pi(A_1 \times .. \times A_n) = \mu_1(A_1)...\mu(A_n)$$

für alle $A_i \in \mathcal{A}_i$. Beweisen Sie dies.
Dieses Maß heißt das Produktmaß der $\mu_1, ..., \mu_n$. Man schreibt dafür auch

$$\otimes_{i=1}^n \mu_i = \mu, \otimes ... \otimes \mu_n.$$

Weiter sei $f : \Omega, \times .. \Omega_n \to \bar{\mathbb{R}}$ eine $\otimes_{i=1}^n \mu_i$–integrierbare Funktion. Dann gilt für jede Permutation $i_1, ..., i_n$ der Indizes $1, ...n$:

$$\int f d\left(\mu_1 \otimes .. \otimes \mu_n\right) = \int \left(.. \left(f\left(\omega_1 ..., \omega_n\right) \mu_{i_1} d\omega_i\right) \mu_{i_n} d\omega_i\right.$$

Beweisen Sie auch dies.

5 Der Satz von Radon und Nikodym

Schon in einem ersten Stochastikkurs lernt man absolut stetige Wahrscheinlichkeiten kennen. Für solche Wahrscheinlichkeiten μ gilt, dass es eine Funktion $h : \mathbb{R} \to \mathbb{R}^+$ (die Dichte) gibt, die λ^1–integrierbar ist (und so, dass $\int_{\mathbb{R}} h(x) d\lambda(x) = 1$ ist), so dass für alle $A \in \mathcal{B}^1$

$$\nu(A) = \int_A h(x) dx.$$

gilt. Der Vorteil solcher Maße liegt auf der Hand: Man kann sie mit dem Lebesgue-Maß vergleichen. Insbesondere sind sie stetig bzgl. des Lebesgue Maßes, d.h. für alle $\varepsilon > 0$ gibt es $\delta > 0$, so dass, falls

$$\lambda(A) \leq \delta \quad \text{auch} \quad \nu(A) \leq \varepsilon \tag{5.1}$$

gilt. Der prominenteste Vertreter absolut stetiger Maße ist die Normalverteilung, deren Dichte durch

$$h(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

gegeben ist. In diesem Abschnitt fragen wir uns, unter welchen Umständen für zwei Maße μ und ν (auf dem gleichen Messraum $(\Omega, \mathcal{A})$) so eine $\mathcal{A}$-messbare Funktion h existiert, so dass für alle $A \in \mathcal{A}$

$$\nu(A) = \int_A h(x) d\mu(x). \tag{5.2}$$

Es stellt sich heraus, dass die Antwort auf diese Frage eng mit der Stetigkeitseigenschaft (5.1) verbunden ist.

Definition 5.1 *Es sei $h : (\Omega, \mathcal{A}) \to \overline{\mathbb{R}}^+$ eine messbare Funktion. Das Maß ν, das wir durch* (5.2) *definiert haben heißt das Maß mit Dichte h bezüglich μ. Es wird auch mit*

$$\nu = h\mu \tag{5.3}$$

abgekürzt

Übung 5.2 *Zeigen Sie, dass $\nu = h\mu$ in der obigen Situation wirklich ein Maß ist.*

Bevor wir die obige Definition genauer studieren, wollen wir die Eigenschaft (5.1) taufen.

Definition 5.3 *Ein Maß ν auf $\mathcal{A}$ heißt stetig bezüglich eines Maßes μ auf $\mathcal{A}$ (man schreibt auch $\nu \ll \mu$), falls für jedes $N \in A$ mit $\mu(N) = 0$ auch gilt $\nu(N) = 0$.*

Nun untersuchen wir Folgerungen aus den beiden obigen Definitionen

Theorem 5.4 *Sei $\nu = f\mu$ mit $f \in E^*$. Dann gilt für alle $\varphi \in E^*$*

$$\int \varphi d\nu = \int \varphi f d\mu. \tag{5.4}$$

Weiter ist $\varphi : \Omega \to \overline{\mathbb{R}}$ ν–integrierbar, genau dann wenn $\varphi \cdot f$ μ–integrierbar ist und in diesem Falle gilt (5.4) .

Beweis. Man überprüft die Behauptung zunächst für Treppenfunktionen

$$\varphi = \sum_{i=1}^{n} \alpha_i 1_{A_i}, A_i \in \mathcal{A}.$$

Für diese gilt

$$\int \varphi d\nu = \sum_{i=1}^{n} \alpha_i \nu(A_i) = \sum_{i=1}^{n} \alpha_i \int 1_{A_i} f d\mu = \int \varphi f d\mu. \tag{5.5}$$

Ein beliebiges $\phi \in E^*$ kann durch eine Folge (u_n) in E approximiert werden: $u_n \uparrow \varphi$.
Dann gilt auch $u_n f \uparrow \varphi f$. Dies zusammen mit (5.5) impliziert (5.4) mit Hilfe des Satzes von der monotonen Konvergenz. Für ein beliebiges integrierbares φ erhält man (5.4), indem man φ wie üblich in φ^+ und φ^- zerlegt. ■

Übung 5.5 *Zeigen Sie, dass, für $\nu = f\mu$ und $\varrho = g\nu$ mit Funktionen $f, g \in E^*$, $\varrho = (gf)\mu = g(f\mu)$ gilt.*

Der nächste Satz untersucht die Eindeutigkeit von Dichten:

Theorem 5.6 *Für $f, g \in E^*$ gilt:*

$$f = g \; \mu\text{-f.ü.} \Longrightarrow f\mu = g\mu \tag{5.6}$$

Falls f oder g μ–integrierbar sind, gilt auch die Umkehrung von (5.6).

Beweis. $f = g\,\mu$ -f.ü. impliziert $f 1_A = g 1_A \mu$ -f.ü. für alle $A \in \mathcal{A}$. Daher gilt

$$\int_A f d\mu = \int_A g d\mu \quad \text{für alle } A \in \mathcal{A}$$

d.h. $f\mu = g\mu$.
Nehmen wir nun an, dass f μ–integrierbar ist und $f\mu = g\mu$. Dann ist auch g μ–integrierbar. Betrachte $N := \{f > g\} \in \mathcal{A}$ und

$$h := 1_N f - 1_N g.$$

Da $1_N f \leq f, 1_N g \leq g$, sind die Funktionen $1_N f$ und $1_N g$ μ–integrierbar und wegen $f\mu = g\mu$ haben sie dasselbe μ–Integral. Daher folgt

$$\int h d\mu = \int_N f d\mu - \int_N g d\mu = 0$$

Dies zeigt, dass $\mu(N) = 0$. Vertauscht man f und g, erhält man auch $\mu(f \neq g) = 0$. ■
Wir studieren nun die Stetigkeitseigenschaft:

Theorem 5.7 *Seien μ, ν zwei Maße auf $(\Omega, \mathcal{A})$, von denen ν endlich sei. ν ist μ–stetig genau dann wenn für jedes $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass*

$$\mu(A) \leq \delta \Longrightarrow \nu(A) \leq \varepsilon \quad \textit{für alle } A \in \mathcal{A}. \tag{5.7}$$

Beweis. Eine Richtung ist einfach: Falls (5.7) gilt, so folgt $\nu(A) \leq \varepsilon$ für alle A mit $\mu(A) = 0$ und alle $\varepsilon > 0$. Daher ist $\nu(A) = 0$ für alle A mit $\mu(A) = 0$, d.h. ν ist μ–stetig.
Sei umgekehrt angenommen, dass (5.7) falsch wäre. Dann gibt es $\varepsilon > 0$ und eine Folge $(A_n)_n$ in $\mathcal{A}$ mit

$$\mu(A_n) \leq 2^{-n} \text{ und } \nu(A_n) > \varepsilon, \quad n \in \mathbb{N}.$$

Definiere

$$A := \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \in \mathcal{A},$$

dann gilt andererseits

$$\mu(A) \leq \mu\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} \mu(A_m) = 2^{-n+1}, \quad n \in \mathbb{N}$$

und somit $\mu(A) = 0$. Andererseits

$$\nu(A) \geq \limsup \nu(A_n) \geq \varepsilon > 0.$$

Für die erste Ungleichung haben wir hierbei die Endlichkeit von ν benutzt. Daher ist ν nicht μ–stetig. ∎

Wir betrachten nun die zentrale Frage dieses Kapitels: Was ist die Beziehung der Definitionen 5.1 und 5.3? Hierzu zitieren wir zunächst einen wichigten Satz aus der Theorie der Hilberträume. Der Prototyp eines solchen Raumes ist der Raum $L^2(\mu)$ mit dem inneren Produkt $< f, g >= \int fgd\mu$.).

Theorem 5.8 (Rieszscher Darstellungssatz) *Sei $\lambda : H \to \mathbb{R}$ ein lineares, stetiges Funktional über einem Hilbertraum H. Dann gibt es ein $\overline{\lambda} \in H$ mit $\lambda(x) =< x, \overline{\lambda} >$ für alle $x \in H$.*

Beweis. Sei $H_\lambda := \{x \in H : \lambda(x) = 0\}$ (falls $\lambda \not\equiv 0$ – dann wäre die Behauptung trivial). H_λ is abgeschlossen, da λ stetig ist. Sei $a \in H \backslash H_\lambda$ und $a_0 \in H_\lambda$ seine Projektion auf H_λ. Offensichtlich ist $0 \neq a - a_0 \in H_\lambda^+$ (das orthogonale Komplement von H_λ). Setze $a_1 := \frac{a-a_o}{\|a-a_0\|} \in H_\lambda^+$. Dann ist

$$\lambda(a_1) = \frac{1}{\|a - a_0\|}\lambda(a) \neq 0$$

und somit für $x \in H$ der Ausruck $x - \frac{\lambda(x)}{\lambda(a_1)}a_1 \in H_\lambda$ wohldefiniert und

$$< x - \frac{\lambda(x)}{\lambda(a_1)}a_1, a_1 >= 0.$$

Löst man dies bezüglich $\lambda(x)$ erhält man

$$\lambda(x) = \lambda(a_1) < x, a_1 > .$$

Definiert man $\overline{\lambda} := \lambda(a_1)\, a_1$, so erhält man $\lambda(x) =< x, \overline{\lambda} >$. ∎

Der folgende Satz gehört zu den Herzstücken in Maß– und Wahrscheinlichkeitstheorie. Er hat auch den Namen Radon–Nikodym–Dichte für die Dichte eines μ–stetigen Maßes ν geprägt. Man schreibt auch $\nu \ll \mu$, falls ν μ–stetig ist und $\frac{d\nu}{d\mu}$ für die Dichte von ν bzgl. μ.

Theorem 5.9 (Radon–Nikodym) *Es seien μ, ν Maße auf $(\Omega, \mathcal{A})$. Falls μ σ–endlich ist, dann sind äquivalent:*
(i) ν hat eine Dichte bzgl. μ
(ii) $\nu \ll \mu$.

Beweis. $(i) \Longrightarrow (ii)$ wurde schon gezeigt.
$(ii) \Longrightarrow (i)$: Wir beginnen mit dem Fall, dass μ, ν beide endlich sind. Setze $\lambda := \mu + \nu$. Da λ endlich ist, gilt $L^2(\lambda) \subseteq L^2(\nu) \subseteq L^1(\nu)$. Für $f \in L^2(\lambda)$ setze

$$\Lambda(f) := \int f d\nu$$

Dann gilt

$$|\Lambda(f)| \leq \nu(\Omega)^{\frac{1}{2}} \left(\int \left| f^2 \right| d\nu \right)^{\frac{1}{2}} \leq \nu(\Omega)^{\frac{1}{2}} \left(\int \left| f^2 \right| d\lambda \right)^{\frac{1}{2}} = \nu(\Omega)^{\frac{1}{2}} ||f||_2.$$

Daher ist $\Lambda : L^2(\lambda) \to \mathbb{R}$ linear, beschränkt und somit eine stetige Funktion. Aus dem Rieszschen Darstellungssatz wissen wir somit, dass eine Funktion $f_0 \in L^2(\lambda)$ existiert mit

$$\Lambda(f) = \int f d\nu = \int f \cdot f_0 d\lambda = < f, f_0 > \tag{5.8}$$

für alle $f \in L^2(\lambda)$. Insbesondere für $f = 1_E, E \in \mathcal{A}$ folgt

$$\nu(E) = \int_E f_0 d\lambda \geq 0.$$

Somit ist $f_0 \geq 0, \nu$ -f.ü. Andererseits:

$$\int_E (1 - f_0) d\lambda = \lambda(E) - \nu(E) = \mu(E) \geq 0,$$

für alle $E \in \mathcal{A}$, also auch $f_0 \leq 1$. Wir wählen ein $0 \leq f_0 \leq 1$ mit (5.8). Definiere $\Omega_1 = \{f_0 = 1\}$,$\Omega_2 := \{0 < f_0 < 1\}$ und $\Omega_3 := \{f_0 = 0\}$. Dann gilt für alle $E \in \mathcal{A}$

$$\int_E (1 - f_0) \, d\nu = \nu(E) - \int_E f_0 d\nu = \int_E f_0 d\lambda - \int_E f_0 d\nu = \int_E f_0 d\mu.$$

Für $E = \Omega_1$ impliziert dies $\mu(\Omega_1) = 0$ und somit, dass $\nu(\Omega_1) = 0$ (da $\nu \ll \mu$). Mit den üblichen Approximationstechniken erhalten wir für alle $f : \Omega_2 \to \mathbb{R}^+$

$$\int_{\Omega_2} f(1 - f_0) \, d\nu = \int_{\Omega_2} f f_0 d\mu.$$

Wendet man das auf $f = \frac{1_E}{(1-f_0)}, E \subseteq \Omega_2, E \in \mathcal{A}$ an, erhält man

$$\nu(E) = \int_E \frac{f_0}{1-f_0} d\mu.$$

Bedenkt man, dass $\nu(\Omega_3) = 0$, kommt man zu

$$\nu(E) = \nu(E \cap \Omega_2) = \int_{E \cap \Omega_2} \frac{f_0}{1-f_0} d\mu.$$

Definiert man

$$\frac{d\nu}{d\mu}(\omega) := f(\nu) := \begin{cases} \frac{f_0}{1-f_0} & \omega \in \Omega_2 \\ 0 \text{ sonst} \end{cases}$$

so zeigt das, dass f eine Dichte für ν bzgl. μ ist.
In einem zweiten Schritt nehmen wir nun an, dass $\mu(\Omega) < \nu(\Omega) = \infty$. Wir geben eine Partition von Ω bestehend aus paarweise disjunkten Mengen $\Omega_0, \Omega_1, ... \in \mathcal{A}$ an, wobei $\Omega = \bigcup_{i=0}^{\infty} \Omega_i$ und
a) Für $A \in \Omega_0 \cap \mathcal{A}$ gilt entweder $\mu(A) = \nu(A) = 0$ oder $\mu(A) > 0$ und $\nu(A) = +\infty$
b) $\nu(\Omega_n) < +\infty$, für alle $n \in \mathbb{N}$.
Hierzu definieren wir

$$\begin{aligned} \mathcal{Q} \; &: = \{Q \in \mathcal{A} : \nu(Q) < +\infty\} \text{ und} \\ \alpha \; &: = \sup_{Q \in \mathcal{Q}} \mu(Q). \end{aligned}$$

Definitionsgemäß gibt es eine Folge (Q_m) in $\mathcal{Q}$ mit $\alpha = \lim \mu(Q_m)$. Diese Folge kann man wachsend wählen. Dannt hat $Q_0 = \bigcup_{n=1}^{\infty} Q_n \in \mathcal{A}$ das μ–Maß $\mu(Q_0) = \alpha$. Unser Kandidat für Ω_0 ist Q_0^c. In der Tat: nehmen wir $A \in Q_0^c \cap \mathcal{A}$ mit $\nu(A) < +\infty$, dann ist $Q_m \cup A \in \mathcal{Q}$ für alle m, somit $\mu(Q_m \cup A) \leq \alpha$ für alle und somit

$$\mu(Q_0 \cup A) = \lim \mu(Q_m \cup A) \leq \alpha.$$

Nun ist $A \cap Q_0 = \emptyset$ und daher $\mu(Q_0 \cup A) = \mu(Q_0) + \mu(A) = \alpha + \mu(A) \leq \alpha$. Dies impliziert $\mu(A) = 0$. Da schließlich $\nu \ll \mu$, gilt auch $\nu(A) = 0$. Dies zeigt, dass a) für $\Omega_0 = Q_0^c$ gilt. b) ist ebenso wahr für $\Omega_m := Q_m \backslash Q_{m-1}$ und

$$\Omega_1 = Q_1.$$

Definieren wir $\mu_n := \mu \mid \Omega_n \cap \mathcal{A}$ und $\nu_n \mid \Omega_n \cap \mathcal{A}$, dann wissen wir $\nu_n \ll \mu_n$, für $n = 0, 1, 2....$ Für $n \geq 1$ sind die Maße ν_n und μ_n endlich. Aus dem vorhergehenden wissen wir, dass es eine messbare Funktion $f_n \geq 0$ auf Ω_n mit $\nu_n = f_n \mu_n$ gibt. Auf Ω_0 können wir $f_0 \equiv +\infty$ setzten, um $\nu_0 = f_0 \mu_0$ zu erhalten. Kleben wir die f_i's zusammen vermoge

$$f(\omega) = f_i(\omega) 1_{\Omega_i}(\omega)$$

sehen wir, dass $\nu = f\mu$.
Im letzten Schritt, nehen wir auch μ als σ–endlich. Dann existiert eine μ–integrierbare Funktion h auf Ω mit $0 < h(\omega) < +\infty$ für alle $\omega \in \Omega$. In der Tat: es gibt eine Folge $(A_n)_n$ in $\mathcal{A}$ mit $A_n \uparrow \Omega$ und $\mu(A_n) < +\infty$. Dann gibt es ein $0 < \eta_n \leq 2^{-n}$ mit $\eta_n \mu(A_n) \leq 2^{-n}$. Daher leistet

$$h = \sum_{n=1}^{\infty} \eta_n 1_{A_n}$$

das Verlangte. Aber das Maß $h\mu$ ist endlich und hat die gleichen Nullmengen wie μ. Daher gilt $\nu \ll h\mu$. Aus dem obigen wissen wir, dass es eine Funktion $f \geq 0$ gibt, so dass $\nu = f(h\mu) = (fh)\mu$. Somit ist (fh) eine Dichte für ν bzgl. μ. Das beweist den Satz. ■

Abschließend sei noch ein Satz aufgeführt, den wir hier nicht beweisen wollen (ein Beweis findet sich etwa im Maßtheoriebuch von Bauer). Um ihn zu formulieren beötigen wir neben dem Begriff der Stetigkeit zweier Maße gewissermaßen als Gegenstück den Begriff der relativen Singularität.

Definition 5.10 *Es seien μ, ν zwei Maße auf einem messbaren Raum $(\Omega, \mathcal{F})$. ν heißt singulär bezüglich μ (in Zeichen $\nu \perp \mu$), wenn es eine μ–Nullmenge N gibt, mit $\mu(N) = 0 = \nu(N^c)$.*

Singularität ist offensichtlich eine symmetrische Beziehung (man sagt auch μ und ν sind wechselseitig singulär). ν und μ sind offenbar genau dann singulär, wenn alles was für ν wesentlich ist, d.h. alle Mengen mit positiver ν–Masse, von μ "nicht gesehen wird".
Der Lebesguesche Zerlegungssatz sagt nun, dass Stetigkeit und Singularität die beiden Extreme sind, zwischen denen sich alles bewegt: Jedes maß lässt sich in einen μ–stetigen und einen μ–singulären Anteil zerlegen.

Theorem 5.11 (Lebesguescher Zerlegungssatz) *Es seien μ, ν zwei σ–endliche Maße auf einem messbaren Raum $(\Omega, \mathcal{F})$. Dann gibt es Maße ν_1 und ν_2 auf $(\Omega, \mathcal{F})$ mit* *i)* $\nu = \nu_1 + \nu_2$.

ii) $\nu_1 \ll \mu$

iii) $\nu_2 \perp \mu$.

6 Grundbegriffe der Wahrscheinlichkeitstheorie

Wie eingangs erwähnt, versteht man seit Kolmogorov unter einer Wahrscheinlichkeit ein Maß auf dem Raum der möglichen Versuchsausgänge mit Masse eins. Somit werden wir im folgenden stets ein Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ betrachten und es einen Wahrscheinlichkeitsraum nennen, wenn das folgende erfüllt ist:

Definition 6.1 *Ein Wahrscheinlichkeitsraum ist ein Tripel* $(\Omega, \mathcal{F}, \mathbb{P})$*, wobei*

- Ω *eine Menge ist*
- $\mathcal{F}$ *eine* σ*-Algebra über* Ω *ist*
- *und* $\mathbb{P}$ *ein Maß auf* $\mathcal{F}$ *ist mit* $\mathbb{P}(\Omega) = 1$

Einen Wahrscheinlichkeitsraum kann man als eine Art Experiment mit zufälligen Ausgang betrachten (beispielsweise eine Messung, die mit einem Messfehler behaftet ist). Natürlich ist es bei einem solchen Experiment (beispielsweise in der Physik) nicht immer interessant alles, was messbar ist, auch zu messen. Man mag beispielsweise an der Wellenlänge eines Lichtsrahls interessiert sein, aber nicht noch gleichzeitig an der Temperatur, unter der das Experiment durchgeführt wurde. Dieses Konzentrieren auf einige relevante Informationen geschieht in der Wahrscheinlichkeitstheorie mit Hilfe sogenannter Zufallsvariablen.

Definition 6.2 *Eine Zufallsvariable* X *ist eine messbare Abbildung*

$$X : \Omega \to \mathbb{R}^d$$

Hierbei versehen wir den $\mathbb{R}^d$ *mit seiner Borelschen* σ*–Algebra* $\mathcal{B}^d$.

Wesentlich bei der Behandlung von Zufallsvariablen ist nun, dass der zugrunde liegende Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ eigentlich keine Rolle spielt. Betrachten wir beispielsweise die folgenden beiden Zufallsvariablen . Es seien

$$\Omega_1 = \{0, 1\}, \mathcal{F}_1 = \mathcal{P}(\Omega_1), \text{ und } \mathbb{P}_1\{0\} = \frac{1}{2}$$

und

$$\Omega_2 = \{1, 2, ..., 6\}, \mathcal{F}_2 = \mathcal{P}(\Omega_2), \text{ und } \mathbb{P}_2\{i\} = \frac{1}{6}, \quad i \in \Omega_2.$$

Betrachte auf diesen W.-Räumen die Zufallsvariablen

$$\begin{aligned} X_1 : \ & \Omega_1 \to \mathbb{R} \\ & \omega \mapsto \omega \end{aligned}$$

und

$$\begin{aligned} X_2 : \quad & \Omega_2 \to \mathbb{R} \\ & \omega \mapsto \begin{cases} 0 & i \text{ ist gerade} \\ 1 & i \text{ ist ungerade.} \end{cases} \end{aligned}$$

Dann ist

$$\mathbb{P}_1(X_1 = 0) = \mathbb{P}_2(X_2 = 0) = \frac{1}{2}.$$

Somit haben X_1 und X_2 dasselbe Verhalten, obschon sie auf vollständig anderen Räumen defineirt sind Was man aus diesem Beispiel lernen kann, ist, das alles was wirklich zählt die "Verteilung" $P \circ X^{-1}$ der Zufallsvariablen ist:

Definition 6.3 *Die Verteilung $\mathbb{P}_X$ einer Zufallsvariable $X : \Omega \to \mathbb{R}^d$ ist das folgende W.–Maß auf $(\mathbb{R}^d, \mathcal{B}^d)$:*

$$\mathbb{P}_X(A) := \mathbb{P}(X \in A) = \mathbb{P}\left(X^{-1}(A)\right), \qquad A \in \mathcal{B}^d.$$

Somit ist die Verteilung einer Zufallsvariable ihr Bildmaß in $\mathbb{R}^d$.

Beispiel 6.4 *Wichtige Beispiele fürZufallsvariablen , von denen wir einige schon in der Stochastikvorlesung kennengelernt haben, sind:*

- *Die* Dirac Verteilung *mit Atom $a \in \mathbb{R}$. EineZufallsvariable X mit der folgenden Verteilung ist Dirac-verteilt zum Parameter a*

$$\mathbb{P}(X = b) = \delta_a(b) = \begin{cases} 1 & b = a \\ 0 & \textit{sonst.} \end{cases}$$

Hierbei ist natürlich $a \in \mathbb{R}$.

- *Die* Bernoulli–Verteilung *zum Parameter $p \in (0,1)$. Eine Zufallsvariable X heißt Bernoulli-verteilt zum Parameter p wenn gilt*

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p.$$

- *Die* Binomial Verteilung *zu den Parametern with n und p. Eine Zufallsvariable X heißt Binomial-verteilt zu den Parametern n und p, falls gilt*

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad 0 \le k \le n.$$

Die Binomial–Verteilung ist die Verteilung der Summe n unabhängiger Bernoulli–verteilter Zufallsvariablen zum Paramter p.

- *Die* Laplace–Verteilung *auf einer endlichen Menge $\Omega' \subset \mathbb{R}$. Eine Zufallsvariable X heißt Laplace-verteilt auf Ω', falls*

$$\mathbb{P}(X = a) = \frac{1}{|\Omega'|}$$

für alle $a \in \Omega'$ und $\mathbb{P}(X = a) = 0$ für alle anderen a

- *Die* Gleichverteilung *oder* Rechteckverteilung *auf einem Intervall $I = [a,b]$: Eine Zufallsvariable X heißt rechteckverteilt auf $I = [a,b]$, falls für alle $x \in \mathbb{R}$ gilt*

$$\mathbb{P}(X \le x) = \frac{1}{b-a} \int_a^x dx.$$

- *Die* Normal–Verteilung *zu den Parametern μ und σ^2. Eine Zufallsvariable X heißt Normal-verteilt zu den Parametern μ und σ^2 ($\mathcal{N}(\mu, \sigma^2)$-verteilt), falls*

$$\mathbb{P}(X \le a) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^a e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Hierbei ist wieder $a \in \mathbb{R}$. Für den Fall $\mu = 0$ und $\sigma^2 = 1$ spricht man auch von der Standardnormalverteilung.

- *Die Poisson Verteilung zum Parameter* $\lambda \in \mathbb{R}^+$. *Eine Zufallsvariable* X *ist Poisson–verteilt zum Parameter* $\lambda \in \mathbb{R}^+$, *falls*

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad k \in \mathbb{N}.$$

- *Die Exponentialverteilung zum Paramter* $\lambda > 0$. *Eine Zufallsvariable* X *heißt exponentialverteilt zum Parameter* λ, *falls* X *fast sicher positiv ist und*

$$\mathbb{P}(X \leq a) = \int_0^a \lambda e^{-\lambda x} dx$$

für $a \in \mathbb{R}, a > 0$.

- *Die multidimensionale Normal Verteilung zu den Parametern* $\mu \in \mathbb{R}^d$ *und* Σ, *d.h. eineZufallsvariable* X *heißt normal–verteilt in Dimension* d *und zu den Parametern* μ *und* Σ ($\mathcal{N}(\mu, \Sigma)$*-verteilt), falls* $\mu \in \mathbb{R}^d$, Σ *eine symmetrische, positiv definite* $d \times d$ *Matrix ist und für* $A = (-\infty, a_1] \times \ldots \times (-\infty, a_d]$ *gilt*

$$\mathbb{P}(X \in A) = \\ \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_d} \exp\left(-\frac{1}{2}\langle \Sigma^{-1}(x - \mu), (x - \mu)\rangle\right) dx_1 \ldots dx_d.$$

7 Erwartungswerte, Momente und die Jensensche Ungleichung

So, wie Zufallsvariablen das Wichtigste eines Zufallsexperiments herausfiltern, geben die folgenden Kennzahlen die wichtigsten Eigenschaften der Verteilung einer Zufallsvariablen wieder.

Definition 7.1 *Der Erwartungswert einer Zufallsvariablen ist als die folgende Zahl (bzw. Vekor, wenn der Bildraum mehrdimensional ist) definiert*

$$\mathbb{E}(X) := \mathbb{E}_P(X) := \int X(\omega) d\mathbb{P}(\omega),$$

falls dieses Integral wohldefiniert ist.

Schon in der Vorlesung "Stochastik" haben wir gesehen, wie man den Erwartungswert einer Funktion einer Zufallsvariablen berechnet. Hier kommt nun der "allgemein Weg" diesen Erwartungswert zu berechnen.

Proposition 7.2 *Sei $X : \Omega \to \mathbb{R}^d$ eine Zufallsvariable und $f : \mathbb{R}^d \to \mathbb{R}$ eine messbare Funktion, so dass f integrierbar ist bzgl. $\mathbb{P}_X$. Dann ist*

$$\mathbb{E}[f \circ X] = \int f d\mathbb{P}_X. \tag{7.1}$$

Beweis. Falls von der Gestalt $f = 1_A, A \in \mathcal{B}^d$ ist, erhalten wir

$$\mathbb{E}[f \circ X] = \mathbb{E}(1_A \circ X) = \mathbb{P}(X \in A) = \mathbb{P}_X(A) = \int 1_A d\mathbb{P}_X.$$

Also gilt (7.1) für Funktionen $f = \sum_{i=1}^n \alpha_i 1_{A_i}, \alpha_i \in \mathbb{R}, A_i \in \mathcal{B}^d$. Durch die üblichen Zerlegungs– und Approximationstechniken erhalten wir (7.1) für allgemeine integrierbare f. ■
Insbesondere ergalten wir mit $f(x) = x$.

Korollar 7.3 *Es gilt für X mit Werten in $\mathbb{R}$*

$$\mathbb{E}[X] = \int x d\mathbb{P}_X.$$

Ist insbesondere $X : \Omega \to \mathbb{N}_0$ eine Zufallsvariable mit Werten in $\mathbb{N}_0$, so erhalten wir die schon bekannte Formel

$$\mathbb{E}X = \sum_{n=0}^{\infty} n\mathbb{P}(X = n)$$

Übung 7.4 *Sei $X : \Omega \to \mathbb{R}$ ein Zufallsvariable ; dann gilt*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X| \geq n) \leq \mathbb{E}(|X|) \leq \sum_{n=0}^{\infty} \mathbb{P}(|X| \geq n).$$

Somit hat X einen Erwartungswert, genau dann, $\sum_{n=0}^{\infty} \mathbb{P}(|X| \geq n)$ konvergiert.

Der Erwartungswert gibt ein erstes Anzeichen, was wir von einer Zufallsvariable "im Mittel erwarten". (Die wird genauer durch die Gesetze der großen Zahlen beschrieben, die wir später kennenlernen werden). Es ist klar, dass der Erwartungswert 0 beispielsweise mehr über eine Zufallsvariable aussagt, die konstant gleich Null ist, als über eine, die die mit gleicher Wahrscheinlichkeit die Werte $\pm 10^6$ annimmt und sonst keine. Im folgenden geben wir ein Maß für die Aussagekraft des Erwartungswertes.

Definition 7.5 *Für die folgenden Definitonen sei stets angenommen, dass die zugehörigen Integrale existieren.*

1. *Für $p \geq 1$ ist das p'te Moment einer Zufallsvariablen definiert als $\mathbb{E}(X^p)$.*
2. *Das zentrierte p'te Moment einer Zufallsvariablen ist definiert als $\mathbb{E}[(X - \mathbb{E}X)^p]$.*
3. *Die* **Varianz** *einer Zufallsvariablen X ist ihr zentriertes, zweites Moment also*
$$\mathbb{V}(X) := \mathbb{E}\left[(X - \mathbb{E}X)^2\right].$$
Ihre Standardabweichung σ ist gegeben durch
$$\sigma := \sqrt{\mathbb{V}(X)}.$$

Proposition 7.6 *Es ist $\mathbb{V}(X) < \infty$ genau dann, wenn $X \in \mathcal{L}^2(P)$. In diesem Falle gilt*
$$\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2 \tag{7.2}$$
und
$$\mathbb{V}(X) \leq \mathbb{E}\left(X^2\right) \tag{7.3}$$
ebenso wie
$$(\mathbb{E}X)^2 \leq \mathbb{E}\left(X^2\right) \tag{7.4}$$

Beweis. Falls $\mathbb{V}(X) < \infty$ ist, so ist $X - \mathbb{E}X \in \mathcal{L}^2(P)$. Nun ist aber $\mathcal{L}^2(P)$ ein Vektorraum und die Konstante $\mathbb{E}X$ ist darin enthalten, also:
$$\mathbb{E}X \in \mathcal{L}^2(P) \Rightarrow X = (X - \mathbb{E}X) + \mathbb{E}X \in \mathcal{L}^2(P).$$
Ist andererseits $X \in \mathcal{L}^2(P)$, so folgt $X \in \mathcal{L}^1(P)$. Daher existiert $\mathbb{E}X$ (und ist eine Konstante). Daher gilt auch
$$X - \mathbb{E}X \in \mathcal{L}^2(P),$$

also ist $\mathbb{V}X$ endlich. Aus der Linearität des Erwartungswertes folgt:

$$\mathbb{V}(X) = E(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - 2(\mathbb{E}X)^2 + (\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 .$$

Dies impliziert (7.3) und (7.4). ∎
(7.2) impliziert sofort, dass X und $X - \mathbb{E}X$ die gleiche Varianz haben.

Es wird sich nun im nächsten Schritt herausstellen, dass (7.4) ein Spezialfall eines allgemeineren Prinzips ist. Um die herzuleiten, müssen wir Eigenschaften konvexer Funktionen ausnutzen. Wir erinnern uns, dass eine Funktion $\varphi : \mathbb{R} \to \mathbb{R}$ konvex ist, wenn für alle $\alpha \in (0, 1)$ und alle $x, y \in \mathbb{R}$ gilt

$$\varphi(\alpha x + (1 - \alpha) y) \leq \alpha\varphi(x) + (1 - \alpha)\varphi(y)$$

In Analysis I lernt man, dass Konvexität aus $\varphi''(x) \geq 0$ für alle x folgt. Andererseits müssen konvexe Funktionen noch nicht einmal differenzierbar sein. Allerdings sind sie "beinahe differenzierbar". In der Tat zeigt man in Analysis I (wenn nicht: Übung!) das folgende
Sei I ein Intervall und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion. Dann existiert die rechtsseitige Ableitung $\varphi_+^{'}(x)$ für alle $x \in \mathring{I}$ (dem Inneren von I) ebenso wie die linksseitige Ableitung $\varphi_-^{'}(x)$. Somit ist φ stetig auf $\mathring{I}$. Desweiteren ist $\varphi_+^{'}$ wachsend auf $\mathring{I}$ und es gilt:

$$\varphi(y) \geq \varphi(x) + \varphi_+^{'}(x)(y - x)$$

für $x \in \mathring{I}, y \in I$.
Wendet man diese Überlegungen an, ergibt sich die eingangs erwähnte Verallgemeinerung von (7.4).

Theorem 7.7 (Jensensche Ungleichung) *Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$ und sei ferner X $\mathbb{P}$–integrierbar und nehme Werte in einem offenen Intervall $I \subseteq \mathbb{R}$ an. Dann gilt $\mathbb{E}X \in I$ und für jede konvexe Funktion*

$$\varphi : I \to \mathbb{R}$$

ist $\varphi \circ X$ eine Zufallsvariable . Falls $\varphi \circ X$ $\mathbb{P}$-integrierbar ist, gilt:

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi \circ X).$$

Beweis. Sei $I = (\alpha, \beta)$, $-\infty \leq \alpha < \beta \leq \infty$. Somit ist $X(\omega) < \beta$ für alle $\omega \in \Omega$. Dann ist auch $\mathbb{E}(X) \leq \beta$. Es gilt sogar $\mathbb{E}(X) < \beta$, denn wäre $\mathbb{E}(X) = \beta$, wäre die strikt positive Zufallsvariable $\beta - X(\omega)$ gleich 0 $\mathbb{P}$-f.s. Dies ist ein Widerspruch. Analog zeigt man $\mathbb{E}X > \alpha$.
Nach den oben angestellten Überlegungen ist φ stetig auf $\mathring{I} = I$ und somit Borel–messbar. Weiter haben wir gesehen, dass

$$\varphi(y) \geq \varphi(x) + \varphi_{+}^{'}(x)(y-x) \tag{7.5}$$

für alle $x, y \in I$ mit Gleichheit nur für den Fall, dass $x = y$. Somit gilt

$$\varphi(y) = \sup_{x \in I} \left[\varphi(x) + \varphi_{+}^{'}(x)(y-x) \right] \tag{7.6}$$

für alle $y \in I$. Setzt man $y = X(\omega)$ in (7.5), ergibt sich

$$\varphi \circ X(\omega) \geq \varphi(X(\omega)) + \varphi_{+}^{'}(x)\left(X(\omega) - x\right).$$

Integration bzgl. $\mathbb{P}$ liefert

$$\mathbb{E}(\varphi \circ X) \geq \varphi(x) + \varphi_{+}^{'}(x)\left(\mathbb{E}(X) - x\right)$$

für alle $x \in I$. Zusammen mit (7.6) ergibt dies

$$\mathbb{E}(\varphi \circ X) \geq \sup_{x \in I} \left[\varphi(x) + \varphi_{+}^{'}(\mathbb{E}X - x) \right] = \varphi(\mathbb{E}(X)).$$

Dies ist die Aussage der Jensenschen Ungleichung. ■

Korollar 7.8 *Sei $X \in \mathcal{L}^p(\mathbb{P})$ für ein $p \geq 1$. Dann existiert $\mathbb{E}|X|$ und es gilt*

$$|\mathbb{E}(X)|^p \leq \mathbb{E}(|X|^p).$$

Übung 7.9 *Es sei I ein offenes Intervall und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion. Zeigen Sie, dass für $x_1, ..., x_n \in I$ und $\lambda_1, ...\lambda_n \in \mathbb{R}^+$ mit $\sum_{i=1}^n \lambda_i = 1$ gilt*

$$\varphi\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i \varphi(x_i).$$

8 Konvergenz von Zufallsvariablen

Schon in dem Teil über Maßtheorie haben wir drei veschiedene Konvergenztypen kennengelernt.

Definition 8.1 *Sei (X_n) eine Folge von Zufallsvariablen .*

1. *X_n ist stochastisch konvergent gegen eine Zufallsvariable X, falls für jedes $\varepsilon > 0$ gilt:*
$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0$$
2. *X_n konvergiert fast sicher gegen eine Zufallsvariable X, falls für jedes $\varepsilon > 0$ gilt:*
$$\mathbb{P}\left(\limsup_{n\to\infty} |X_n - X| \geq \varepsilon\right) = 0$$
3. *X_n konvergiert gegen eine Zufallsvariable X in $\mathcal{L}^p$ oder in p–Norm, falls gilt:*
$$\lim_{n\to\infty} E\left(|X_n - X|^p\right) = 0.$$

Schon im Maßtheorie–Teil wurde gezeigt:

Theorem 8.2 1. *Falls X_n fast sicher gegen X konvergiert oder in $\mathcal{L}^p$, dann konvergiert (X_n) auch stochastisch gegen X.*

   Keine der beiden Umkehrungen gilt.

2. *Fast sichere Konvergenz impliziert nicht die Konvergenz $\mathcal{L}^p$ und umgekehrt.*

Nun führen wir noch einen weiteren Konvergenzbegriff ein:

Definition 8.3 *Sei $(\Omega, \mathcal{F})$ ein messbarer, topologischer Raum, ausgestattet mit seiner Borelschen σ-Algebra $\mathcal{F}$. Das heißt, dass $\mathcal{F}$ durch die Topologie, nennen wir sie τ auf Ω erzeugt wird. Weiter sei $(\mu_n)_n$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{F})$ und μ sei ein weiteres solches Wahrscheinlichkeitsmaß. Wir sagen, dass μ_n schwach gegen μ konvergiert, wenn für jede beschränkte, stetige und reellwertige Funktion $f : \Omega \to \mathbb{R}$ (wir werden für die Menge all solcher Funktion fortan $\mathcal{C}^b(\Omega)$ schreiben) gilt:*

$$\mu_n(f) := \int f d\mu_n \to_{n\to\infty} \int f d\mu =: \mu(f) \tag{8.1}$$

Der folgende Satz klärt die Beziehung zwischen stochastischer Konvergenz (wir werden dies auch "Konvergenz in Wahrcheinlichkeit" nennen) und schwacher Konvergenz der Verteilungen:

Theorem 8.4 *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge reellwertiger Zufallsvariablen über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. (X_n) konvergiere stochastisch gegen eine Zufallsvariable X. Dann konvergiert $\mathbb{P}_{X_n}$ (die Folge der Verteilungen der Zufallsvariablen) schwach gegen $\mathbb{P}_X$, d.h.*

$$\lim_{n\to\infty} \int f d\mathbb{P}_{X_n} = \int f d\mathbb{P}_X$$

oder äquivalent

$$\lim_{n\to\infty} \mathbb{E}(f \circ X_n) = \mathbb{E}(f \circ X)$$

für alle $f \in \mathcal{C}^b(\mathbb{R})$.
Falls X $\mathbb{P}$–f.s. konstant ist, gilt hiervon auch die Umkehrung.

Beweis. Sei $f \in \mathcal{C}^b(\Omega)$. Wir wollen zunächst annehmen, dass f sogar gleichmäßig stetig ist. Dann gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so dass für jede Wahl von $x', x'' \in \mathbb{R}$gilt:

$$\left|x' - x''\right| < \delta \Rightarrow \left|f\left(x'\right) - f\left(x''\right)\right| < \varepsilon.$$

Definiere die Menge $A_n := \{|X_n - X| \geq \delta\}, n \in \mathbb{N}$. Dann folgt

$$\begin{aligned}
\left|\int f d\mathbb{P}_{X_n} - \int f d\mathbb{P}_X\right| &= |\mathbb{E}(f \circ X_n) - \mathbb{E}(f \circ X)| \\
&\leq \mathbb{E}(|f \circ X_n - f \circ X|) \\
&= \mathbb{E}(|f \circ X_n - f \circ X| \circ 1_{A_n}) + \mathbb{E}\left(|f \circ X_n - f \circ X| \circ 1_{A_n^c}\right) \\
&\leq 2\|f\| \mathbb{P}(A_n) + \varepsilon\mathbb{P}(A_n^c) \\
&\leq 2\|f\| \mathbb{P}(A_n) + \varepsilon.
\end{aligned}$$

Hierbei haben wir die folgende Abkürzung für die Supremumsnorm benutzt

$$\|f\| := \sup\{f(x) : x \in \mathbb{R}\}$$

und die folgende Abschätzung ausgenutzt:

$$|f(X_n) - f(X)| \leq |f(X)| + |f(X_n)| \leq 2\|f\|.$$

Da nun aber $X_n \to X$ stochastisch, folgt dass

$$\mathbb{P}(A_n) \to 0$$

wenn $n \to \infty$. Wenn also n groß genug ist, können wir die Wahrscheinlichkeit für A_n beliebig klein machen, z.B. auch $P(A_n) \leq \frac{\varepsilon}{\alpha \|f\|}$. Somit gilt für solche n, dass

$$\left| \int f d\mathbb{P}_{X_n} - \int f d\mathbb{P}_X \right| \leq 2\varepsilon.$$

Nun sei $f \in \mathcal{C}^b(\Omega)$ beliebig. Wir bezeichnen mit $I_n := [-n, n]$. Da $I_n \uparrow \mathbb{R}$, wenn $n \to \infty$, gilt $\mathbb{P}_{X_n}(I_n) \uparrow 1$, wenn $n \to \infty$. Somit gibt es für alle $\varepsilon > 0$ ein $n_0(\varepsilon) =: n_0$ so dass

$$1 - \mathbb{P}_X(I_{n_0}) = \mathbb{P}_X(\mathbb{R} \setminus I_{n_0}) < \varepsilon.$$

Wir wählen nun eine Funktion u_ε wie folgt:

$$u_\varepsilon(x) = \begin{cases} 1 & x \in I_{n_0} \\ 0 & |x| \geq n_0 + 1 \\ -x + n_0 + 1 & n_0 < x < n_0 + 1 \\ x + n_0 + 1 & -n_0 - 1 < x < -n_0 \end{cases}$$

Schließlich setzen wir $\overline{f} := u_\varepsilon f$. Da $\overline{f} \equiv 0$ außerhalb der kompakten Menge $[-n_0 - 1, n_0 + 1]$ gilt, ist die die Funktion $\overline{f}$ gleichmäßig stetig. Daher gilt

$$\lim_{n\to\infty} \int \bar{f} d\mathbb{P}_{X_n} = \int \bar{f} d\mathbb{P}_X$$

ebenso wie

$$\lim_{n\to\infty} \int u_\varepsilon d\mathbb{P}_{X_n} = \int u_\varepsilon d\mathbb{P}_X;$$

also auch ´

$$\lim_{n\to\infty} \int (1 - u_\varepsilon)\, d\mathbb{P}_{X_n} = \int (1 - u_\varepsilon)\, d\mathbb{P}_X.$$

Mit der Dreiecksungleichung erhalten wir

$$\left| \int f d\mathbb{P}_{X_n} - \int f d\mathbb{P}_X \right| \leq \tag{8.2}$$
$$\int \left| f - \bar{f} \right| d\mathbb{P}_{X_n} + \left| \int \bar{f} d\mathbb{P}_{X_n} - \int \bar{f} d\mathbb{P}_X \right| + \int \left| f - \bar{f} \right| d\mathbb{P}_X.$$

Aus der Tatsache, dass $0 \leq 1 - u_\varepsilon \leq 1_{\mathbb{R}\setminus I_{n_0}}$, können wir überdies schließen,

$$\int (1 - u_\varepsilon)\, d\mathbb{P}_X \leq \mathbb{P}_X \left(\mathbb{R} \setminus I_{n_0}\right) < \varepsilon,$$

so dass für alle $n \geq n_1(\varepsilon)$ auch gilt:

$$\int (1 - u_\varepsilon)\, d\mathbb{P}_{X_n} < 2\varepsilon.$$

Dies ergibt

$$\int \left|f - \bar{f}\right| d\mathbb{P}_X = \int |f|\,(1 - u_\varepsilon)\, d\mathbb{P}_X \leq \|f\|\, \varepsilon$$

einerseits und

$$\int \left|f - \bar{f}\right| d\mathbb{P}_{X_n} \leq 2\, \|f\|\, \varepsilon$$

für alle $n \geq n_1(\varepsilon)$ andererseits. Also erhalten wir aus (8.2):

$$\left|\int f d\mathbb{P}_{X_n} - \int f d\mathbb{P}_X\right| \leq 3\, \|f\|\, \varepsilon + \varepsilon.$$

Dies beweist die schwache Konvergenz der Verteilungen.

Für die Umkehrung sei $\eta \in \mathbb{R}$ und $X \equiv \eta$ (genauer sei X identisch gleich $\eta \in \mathbb{R}$ $\mathbb{P}$–fast sicher). Dies bedeutet für die Verteilung von X, dass $\mathbb{P}_X = \delta_\eta$ wobei δ_η das Dirac–Maß in η ist. Für das offene Intervall $I := (\eta - \varepsilon, \eta + \varepsilon)\,, \varepsilon > 0$ können wir eine stetige und beschränkte Funktion $f \in \mathcal{C}^b(\Omega)$ finden, die $f \leq 1_I$ und $f(\eta) = 1$ erfüllt.
Dann folgt

$$\int f d\mathbb{P}_{X_n} \leq \mathbb{P}_{X_n}(I) = P\left(X_n \in I\right) \leq 1.$$

Da wir die schwache Konvergenz von $\mathbb{P}_{X_n}$ gegen $\mathbb{P}_X$ voraussetzen, können wir folgern

$$\int f d\mathbb{P}_{X_n} \to \int f d\mathbb{P}_X = f(\eta) = 1$$

wenn $n \to \infty$. Daraus folgt aber nun

$$P\left(X_n \in I\right) \to 1$$

wenn $n \to \infty$. Nun ist aber

$$\{X_n \in I\} = \{|X_n - \eta| \leq \varepsilon\}$$

und daher

$$\begin{aligned}
\mathbb{P}\left(|X_n - X| \geq \varepsilon\right) &= \mathbb{P}\left(|X_n - \eta| \geq \varepsilon\right) \\
&= 1 - \mathbb{P}\left(|X_n - \eta| < \varepsilon\right) \to 0
\end{aligned}$$

für alle $\varepsilon > 0$. Dies bedeutet, dass X_n stochastisch gegen X konvergiert. ■

Definition 8.5 *Es seien für jedes $n \in \mathbb{N}$ und X Zufallsvariablen über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Wenn $\mathbb{P}_{X_n}$ schwach $\mathbb{P}_X$ konvergiert, so sagen wir auch, dass X_n gegen X* in Verteilung *konvergiert*

Bemerkung 8.6 *Wenn die Zufallsvariablen im vorhergehenden Theorem $\mathbb{R}^d$–wertig sind und $f : \mathbb{R}^d \to \mathbb{R}$ stetig und beschränkt ist, bleibt die Aussage des Theorems wahr. Dies folgt, da Konvergenz in $\mathbb{R}^d$ gleichbedeutend mit der Konvergenz des Supremums der Koordinaten ist.*

Beispiel 8.7 *Für jede Folge $\sigma > 0$ mit $\lim \sigma_n = 0$ gilt für die Normalverteilung mit Erwartungswert 0 und Varianz σ_n^2, dass*

$$\lim_{n\to\infty} \mathcal{N}(0, \sigma_n^2) = \delta_0.$$

Hierbei ist δ_0 das Dirac-Maß mit Masse in $0 \in \mathbb{R}$. In der Tat, ersetzt man $x = \sigma y$, so ergibt sich für eine stetig, beschränkte rellwertige Funkktion f

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(x) e^{-\frac{x^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} f(\sigma y) dy$$

Nun gilt für alle $y \in \mathbb{R}$ und alle $\sigma \in \mathbb{R}^+$

$$\left| \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} f\left(\sigma y\right) \right| \leq \|f\| \, e^{-\frac{y^2}{2}}.$$

Dies ist eine integrierbare Funktion. Aus dem Satz über dominierte Konvergenz folgt

$$\begin{aligned}
&\lim_{n\to\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{x^2}{2\sigma_n^2}} f(x) dx \\
= &\int_{-\infty}^{\infty} \lim_{n\to\infty} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{x^2}{2\sigma_n^2}} f(x) dx \\
= &f(0).
\end{aligned}$$

Wie das folgende Beispiel zeigt, ist die Annahme der Konstanz des Limes für die Umkehrung in Theorem 8.2 notwendig.

Beispiel 8.8 *Betrachte eine Folge von $X_1, X_2, \dots$ von Bernoulli–verteilten Zufallsvariablen zum Parameter $p = 1/2$. Es ist für jede Funktion $f \in \mathcal{C}^b(\mathbb{R})$*

$$\int f d\mathbb{P}_{X_n} = \int f d\mathbb{P}_{X_1}$$

und somit konvergiert X_n in Verteilung gegen X_1. Andererseits ist für $0 < \varepsilon < \frac{1}{2}$ und jedes $n \geq 2$

$$\mathbb{P}(|X_n - X_1| < \varepsilon) = \mathbb{P}(X_n \neq X_1) = \frac{1}{2}.$$

Man mag sich fragen, ob Verteilungskonvergenz von Zufallsvariablen wirklich gleichbedeutend mit der Konvergenz der Verteilungsfunktionen der Zufallsvariablen ist. Letztere ist für eine Zufallsvariable X definiert als

$$F_X(t) := \mathbb{P}_X(X \leq t).$$

Aus den im Maßtheorieteil besprochenen Stetigkeitseigenschaften für Maße folgt sofort, dass F_X wachsend und rechtsseitig stetig ist. Ferner gilt klarerweise

$$\lim_{t \to -\infty} F(t) = 0 \quad \text{und} \quad \lim_{t \to \infty} F(t) = 1.$$

Der folgende Satz, dass diese Vermutung nur für Zufallsvariablen mit stetigen Verteilungsfunktionen wahr ist:

Theorem 8.9 *Es sei (X_n) eine Folge reellwertiger Zufallsvariablen und es sei X eine weitere reellwertige Zufallsvariable . Genau dann konvergiert (X_n) in Verteilung gegen X wenn*

$$F_{X_n}(t) \to F_X(t) \tag{8.3}$$

für $n \to \infty$ und alle Stetigkeitspunkte t von F_X.

Beweis. Zunächst nehmen wir an, dass X_n in Verteilung gegen X konvergiert. Ferner sei t ein Stetigkeitspunkt von F_X. Sei $\varepsilon > 0$. Dann gibt es ein $\delta > 0$ mit

$$|F_X(t \pm \delta) - F_X(t)| < \varepsilon.$$

Weiter gibt es Funktionen $g_{t,\delta} \in \mathcal{C}^b(\mathbb{R})$ mit

$$1_{(-\infty,t-\delta]} \leq f_{t,\delta} \leq 1_{(-\infty,t]} \leq g_{t,\delta} \leq 1_{(-\infty,t+\delta]}.$$

Aus der vorausgesetzten Verteilungskonvergenz folgern wir

$$\limsup_{n\to\infty} F_{X_n}(t) \leq \lim_{n\to\infty} \mathbb{E}g_{t,\delta}(X_n) \leq F_X(t+\delta) < F_X(t) + \varepsilon$$

und

$$F_X(t) - \varepsilon < F_X(t-\delta) \leq \mathbb{E}f_{t,\delta}(X) = \lim_{n\to\infty} \mathbb{E}f_{t,\delta}(X_n) \leq \liminf_{n\to\infty} F_{X_n}(t).$$

Diese beiden Abschätzungen zusammen ergeben (8.3).
Für die Umkehrung sei wieder $\varepsilon > 0$ und $f \in \mathcal{C}^b(\mathbb{R})$ mit $0 \leq f \leq 1$ gegeben. Bemerke zunächst, dass aus der die Menge der Unstetigkeitsstellen von F_X abzählbar ist. In der Tat ist ja die Menge $\{t : |F_X(t) - F_X(t-)| \geq \frac{1}{n}\}$ höchstens n–elementig. Somit ist die Menge der Stetigkeitspunkte von F_X dicht in $\mathbb{R}$. Wählen wir nun zwei solche Stetigkeitspunkte $a < b$, so dass

$$F_X(a) + (1 - F_X(b)) = \mathbb{P}(X \notin (a,b]) < \varepsilon.$$

Aus der Wahl von a und b als stetigkeitspunkten folgt mit (8.3)

$$\limsup_{n\to\infty} \mathbb{E}f 1_{(a,b]^c}(X_n) \leq \mathbb{P}(X_n \notin (a,b]) = \mathbb{P}(X \notin (a,b]) < \varepsilon. \tag{8.4}$$

Nun ist f als stetige Funktion auf $[a,b]$ gleichmäßig stetig. Somit können wir $(a,b]$ in Intervalle $I_k := (a_k, b_k], k = 1, \ldots m$ zerlegen, so dass für

$$g := \sum_{k=1}^{m} (\inf_{x\in I_k} f(x)) 1_{I_k} \quad \text{und } h := \sum_{k=1}^{m} (\sup_{x\in I_k} f(x)) 1_{I_k}$$

gilt

$$h - \varepsilon \leq g \leq f \leq h \leq g + \varepsilon.$$

Dies zusammen mit (8.3) und (8.4) ergibt

$$\begin{aligned}
\limsup \mathbb{E}f(X_n) &\leq 2\varepsilon + \limsup \mathbb{E}g(X_n) 1_{(a,b]}(X_n) \\
&\leq 2\varepsilon + \sum_{k=1}^{m} (\inf_{x\in I_k} f(x)) \lim (F_{X_n}(b_k) - F_{X_n}(a_k)) \\
&= 2\varepsilon + \sum_{k=1}^{m} (\inf_{x\in I_k} f(x)) \lim (F_X(b_k) - F_X(a_k)) \\
&= 2\varepsilon + \mathbb{E}g(X) 1_{(a,b]}(X) \\
&\leq 2\varepsilon + \mathbb{E}f(X)
\end{aligned}$$

und ähnlich

$$\liminf \mathbb{E}f(X_n) \geq \mathbb{E}f(X) - 2\varepsilon.$$

Dies zeigt die Behauptung. ■

Übung 8.10 *Eine Folge von Zufallsvariablen (X_n) über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ erfülle:*

$$\mathbb{P}(|X_n| > \varepsilon) < \varepsilon$$

für alle genügend großen n und für alle $\varepsilon > 0$. Ist dies äquivalent zur stochastischen Konvergenz von X_n gegen 0 ?

Übung 8.11 *Zeigen Sie, für eine Folge von Poisson Verteilungen (π_{α_n}) mit Parametern $\alpha_n > 0$, dass*

$$\lim_{n\to\infty} \pi_{\alpha_n} = \delta_0,$$

falls $\lim_{n\to\infty} \alpha_n = 0$. Gibt es ein Wahrscheinlichkeitsmaß μ auf $\mathcal{B}^1$ mit

$$\lim_{n\to\infty} \pi_{\alpha_n} = \mu \qquad \textit{(schwach)}$$

falls $\lim \alpha_n = \infty$?

9 Unabhängigkeit

Das Konzept der Unabhängigkeit steht im Zentrum der Wahrscheinlichkeit (so sehr, dass böse Zungen behaupten, Wahrscheinlichkeitstheorie sei im wesentlichen Maßtheorie mit einem Maß der Masse eins plus das Konzept der Unabhängigkeit). Wir haben es bereits in der Stochastikvorlesung kennengelernt. Grob gesprochen geht es auf die folgende Idee zurück:
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum . Für zwei Ereignisse $A, B \in \mathcal{F}$ mit $\mathbb{P}(B) > 0$ kann man sich nun fragen, wie sehr sich die Wahrscheinlichkeit für A ändert, wenn man schon weiß, dass B eintreten wird. Offenbar ist der interessante Teil von A dann nur derjenige, der auch in B liegt, also $A \cap B$. Um allerdings wieder eine Wahrscheinlichkeit zu erhalten, müssen wir $\mathbb{P}(\Omega) = 1$ sicherstellen, die geht nur, wenn wir durch die Wahrscheinlichkeit von B devidieren. Man definiert daher die bedingte Wahrscheinlichkeit von A gegeben B wie folgt

Definition 9.1 *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum . Für zwei Ereignisse* $A, B \in \mathcal{F}$ *mit* $\mathbb{P}(B) > 0$ *ist die bedingte Wahrscheinlichkeit von A gegeben B definiert als*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Verallgemeinerungen dieses Begriffs werden uns in einem späteren Kapitel noch beschäftigen.
Nun ist es naheliegend A und B unabhängig zu nennen, wenn das Wissen, dass B eintritt nichts an der Wahrscheinnlichkeit für B ändert, d.h. wenn $\mathbb{P}(A \mid B)$ und $\mathbb{P}(A)$ das gleiche ergeben, d.h. falls

$$\mathbb{P}(A \mid B) = \mathbb{P}(A).$$

Will man dies in A und B symmetrisch gestalten und zudem auf die für die Unabhängigkeit überflüssige Forderung, dass $\mathbb{P}(B) > 0$ sei verzichten, so definiert man A und B als unabhängig, wenn

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Will man dies auf mehr als zwei Ereignisse verallgemeinern, gelangt man zu der folgenden Definition

Definition 9.2 *Eine Familie* $(A_i)_{i \in I}$ *von Ereignissen aus* $\mathcal{F}$ *heißt unabhängig, falls für jede Folge* $i_1, ..., i_n \in I$ *von Indizes und jedes* $n \in \mathbb{N}$ *gilt:*

$$\mathbb{P}(A_{i_1} \cap .. \cap A_{i_n}) = \mathbb{P}(A_{i_1}) ... \mathbb{P}(A_{i_n}) \tag{9.1}$$

Bemerkung 9.3

- *Wie schon in der Stochastikvorlesung betont, definiert man Unabhängigkeit wie oben — und nicht etwa durch*

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i=1}^{n} P(A_i))$$

um die Unabhängigkeit zu einer vererbbaren Eigenschaft zu machen: Sind $(A_i)_{i \in I}$ *unabhhängig und ist* $J \subseteq I$*, so soll auch* $(A_i)_{i \in J}$ *unabhängig sein.*

- *Ein Beispiel, das wir schon in der Stochastikvorlesung kennengelernt haben, zeigt, dass Unabhängigkeit und paarweise Unabhängigkeit (also die Unabhängigkeit von je zwei Mengen) für $|I| \geq 3$ unterschiedliche Begriffe sind. Beispielsweise sei $\Omega = \{1,2,3,4\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ und $\mathbb{P}(\{i\}) = \frac{1}{4}$ für alle $i \in \Omega$. Man überprüft schnell, dass die Mengen $A = \{1,2\}$, $B = \{2,3\}$ und $C = \{1,3\}$ zwar paarweise unabhängig sind aber nicht unabhängig sind.*

Es liegt nun auf der Hand, wie man allgemein die Unabhängigkeit von Mengensystemen erklärt.

Definition 9.4 *Es sei I eine nichtleere Indexmenge. Für jedes $i \in I$ sei $\mathcal{E}_i \subseteq \mathcal{A}$ eine Mengen von Ereignissen. Die $(\mathcal{E}_i)_{i\in I}$ heißen unabhängig, wenn* (9.1) *für jedes $i_1, ..., i_n \in I$, jedes $n \in \mathbb{N}$ und alle $A_{i_\nu} \in \mathcal{E}_{i_\nu}, \nu = 1, ..., n$ gilt.*

Aus der Definition ergibt sich unmittelbar:

- Eine Familie $(\mathcal{E}_i)_{i\in I}$ ist unabhängig genau dann, wenn jede endliche Teilfamilie unabhängig ist.
- Unabhängigkeit von $(\mathcal{E}_i)_{i\in I}$ bleibt erhalten, wenn wir die Familien $(\mathcal{E}_i)$ verkleinern. Genauer: seien $(\mathcal{E}_i)$ unabhängig und für jedes $i \in I$ sei $\mathcal{E}_i' \subseteq \mathcal{E}_i$ eine Teilfamilie der vorgelegten Familie. Dann sind auch die Familien $(\mathcal{E}_i')$ unabhängig.

Beweistechnisch is der folgende Satz von einiger Wichtigkeit:

Theorem 9.5 *Falls die Familien $(\mathcal{E}_i)_{i\in I}$ unabhängig sind über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, dann sind auch die von ihnen erzeugten Dynkinsysteme $(\mathcal{D}(\mathcal{E}_i))_{i\in I}$ unabhängig.*

Beweis. Nach der obigen Bemerkung genügt es, sich beim Beweis auf endliche Indexmengen I zurückzuziehen. Für $i_0 \in I$ sei dann $\mathcal{D}_{i_0}$ definiert als die Menge aller Ereignisse $E \in \mathcal{F}$ mit der folgenden Eigenschaft:
"Ersetzt man in $(\mathcal{E}_i)_{i\in I}$ die Familie $\mathcal{E}_{i_0}$ durch die einelementige Familie E, so ist diese neu entstehende Familie wieder unabhängig."
$\mathcal{D}_{i_0}$ ist nun ein Dynkin–System: Zunächst ist $\Omega \in \mathcal{D}_{i_0}$, denn ist

$$\emptyset \neq \{i_1, \ldots, i_n\} \subseteq I \setminus \{i_0\},$$

so gilt für $A_{i_j} \in \mathcal{E}_{i_j}, j = 1, ...n$

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n} \cap \Omega) = \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n})\mathbb{P}(\Omega).$$

Ist weiter $E \in \mathcal{D}_{i_0}$, so ergibt sich (mit den gleichen A_{i_ν})

$$\begin{aligned} & \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n} \cap E^c) \\ = \; & \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n} \cap \Omega) - \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n} \cap E) \\ = \; & \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n})\mathbb{P}(\Omega) - \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n})\mathbb{P}(E) \\ = \; & \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n})\mathbb{P}(E^c). \end{aligned}$$

Analog zeigt man die dritte zu überprüfende Eigenschaft unter Ausnutzung der σ–Additivität von $\mathbb{P}$.
Nun ist $\mathcal{D}_{i_0}$ gerade so konstruiert, dass die Menge $(\mathcal{E}_i)_{i\in I}$ unabhängig bleibt, wenn man $\mathcal{E}_{i_0}$ durch $\mathcal{D}_{i_0}$ ersetzt. Nun ist $\mathcal{E}_{i_0} \subset \mathcal{D}_{i_0}$ und daher auch $\mathcal{D}(\mathcal{E}_{i_0}) \subset \mathcal{D}_{i_0}$. Da dies für jedes i_0 wahr ist, kann man sukzessive alle $\mathcal{E}_{i_0}$ durch $\mathcal{D}(\mathcal{E}_{i_0})$ ersetzen und ist wegen der Endlichkeit von I nach endlich vielen Schritten beim gewünschten Ergebnis. ■
Ist von den Familien $(\mathcal{E}_i)_{i\in I}$ auch noch die Durchschnittsstabilität vorausgesetzt, lässt sich noch mehr sagen:

Korollar 9.6 *Seien $(\mathcal{E}_i)_{i\in I}$ unabhängige Familien und jede Familie $\mathcal{E}_i \subseteq \mathcal{F}$ sei durchschnittsstabil. Dann sind auch die Familien $(\sigma(\mathcal{E}_i))_{i\in I}$ der von den $\mathcal{E}_i$ erzeugten σ–Algebren unabhängig.*

Beweis. Der Beweis folgt, da unter den Voraussetzungen des Korollars die von den $\mathcal{E}_i$ erzeugten Dynkinsysteme und σ-Algebren übereinstimmen. ■
Als letztes zeigen wir noch, dass Unabhängigkeit auch unter dem Zusammenfassen unabhängiger σ–Algebren erhalten bleibt.

Theorem 9.7 *Die Familien $(\mathcal{E}_i)_{i\in I}$ seien unabhängig und durchschnittsstabil. mit $\mathcal{E}_i \subseteq \mathcal{F}$ für jedes $i \in I$. Weiter sei die Indexmenge I wie folgt zerlegbar:*

$$I = \dot\cup_{j\in J} I_j$$

mit $I_i \cap I_j = \emptyset, i \neq j$. Es sei $\mathcal{A}_j := \sigma\left(\cup_{i\in I_j}\mathcal{E}_i\right)$ die von I_j erzeugte σ-Algebra . Dann sind auch die σ–Algebren $(\mathcal{A}_j)_{j\in J}$ unabhängig.

Beweis. Für $j \in J$ sei $\tilde{\mathcal{E}}_j$ das Mengensystem aller Mengen der Form

$$E_{i_1} \cap ... \cap E_{i_n}$$

mit $\emptyset \neq \{i_1, ..., i_n\} \subseteq I_j$ und beliebigen $E_{i_\nu} \in \mathcal{E}_{i_\nu}$ für $\nu = 1, ..., n$. Dann ist $\tilde{\mathcal{E}}_j$ nach Konstruktion $\cap$ - stabil. Als sofortige Folgerung aus der Unabhängigkeit der $(\mathcal{E}_i)_{i\in I}$ erhält man auch die Unabhängigkeit der $\left(\tilde{\mathcal{E}}_j\right)_{j\in J}$. Schließlich ist $\mathcal{A}_j = \sigma\left(\tilde{\mathcal{E}}_j\right)$. Daher folgt die Behauptung aus dem vorangegangenen Korollar. ■

Im nächsten Schritte begegnet uns erstmals ein Stück Wahrscheinlichkeitstheorie im eigentlichen Sinne: Natürlich kann auch ein Wahrscheinlichkeitstheoretiker den Ausgang eines Zufallsexperiments nicht vorhersagen (außer in sehr einfach gelagerten Fällen). Er kann aber, unter günstigen Umständen, das Verhalten eines solchen Experiments (im Mittel) für sehr lange Zeiten antizipieren.
Das unten folgende 0–1–Gesetz von Kolomogorov ist ein Beispiel hierfür. Ist eine unendliche Folge unabhängiger σ–Algebren vorgelegt und hängt ein Ereignis von unendliche vielen dieser σ–Algebren ab, so hat das Ereignis schon entweder Wahrscheinlichkeit 0 oder 1.
Hierzu benötigen wir zunächst eine Definition.

Definition 9.8 *Sei $(\mathcal{A}_n)_n$ eine Folge von Teil–σ–Algebren von $\mathcal{F}$ und sei*

$$\mathcal{T}_n := \sigma\left(\bigcup_{m=n}^{\infty} \mathcal{A}_m\right)$$

die σ–Algebra , die durch die $\mathcal{A}_n, \mathcal{A}_{n+1}, ...$erzeugt wird. Dann heißt

$$\mathcal{T}_\infty := \bigcap_{n=1}^{\infty} \mathcal{T}_n$$

die σ-Algebra der terminalen Ereignisse.

Bemerkung 9.9 *$\mathcal{T}_\infty$ ist als Durchschnitt von σ–Algebren natürlich wieder eine σ-Algebra .*

Nun das angekündigte Resultat

Theorem 9.10 (Kolmogorovs Null-Eins-Gesetz) *Es sei $(\mathcal{A}_n)$ eine unabhängige Folge von σ–Algebren mit $\mathcal{A}_n \subseteq \mathcal{F}$. Dann gilt für jedes terminale Ereignis $A \in \mathcal{T}_\infty$*

$$\mathbb{P}(A) = 0 \text{ oder } \mathbb{P}(A) = 1.$$

Beweis. Es sei $A \in \mathcal{T}_\infty$ und es sei $\mathcal{D}$ das System aller Mengen $D \in \mathcal{A}$ die unabhängig A sind. Ziel ist es zu zeigen, dass A in $\mathcal{D}$ ist:
Man überlegt sich leicht, dass $\mathcal{D}$ ein Dynkin–System ist (das geht wie im Beweis von Theorem 9.5 oder ist sonst eine Übung). Nach Theorem 9.7 ist die σ-Algebra $\mathcal{T}_{n+1}$ für jedes $n \in \mathbb{N}$ unabhängig von der σ-Algebra

$$\overline{\mathcal{A}}_n := \sigma\left(\mathcal{A}_1 \cup \ldots \cup \mathcal{A}_n\right).$$

Da für jedes $n \in \mathbb{N}$ die Menge $A \in \mathcal{T}_{n+1}$ ist, folgt dass $\overline{\mathcal{A}}_n \subseteq \mathcal{D}$ für jedes $n \in \mathbb{N}$ gilt. Daraus ergibt sich

$$\overline{\mathcal{A}} := \bigcup_{n=1}^{\infty} \overline{\mathcal{A}}_n \subseteq \mathcal{D}$$

Offensichtlich ist die Folge von σ–Algebren $\left(\overline{\mathcal{A}}_n\right)$ wachsend. Für $E, F \in \overline{\mathcal{A}}$ gibt es somit ein n, so dass sowohl $E \in \overline{\mathcal{A}}_n$ als auch $F \in \overline{\mathcal{A}}_n$ und daher auch $E \cap F \in \overline{\mathcal{A}}_n$ gilt. Daher ist auch $E \cap F \in \overline{\mathcal{A}}$ und dies bedeutet, dass $\overline{\mathcal{A}}$ durchschnittsstabil ist.
Es folgt, dass

$$\sigma\left(\overline{\mathcal{A}}\right) = \mathcal{D}\left(\overline{\mathcal{A}}\right) \subseteq \mathcal{D},$$

(wobei $\mathcal{D}\left(\overline{\mathcal{A}}\right)$ das von $\overline{\mathcal{A}}$ erzeugte Dynkin–System bezeichnet), da $\overline{\mathcal{A}} \subseteq \mathcal{D}$.
Darüber hinaus ist $\mathcal{D} \supseteq \sigma(\overline{\mathcal{A}}) \supseteq \mathcal{T}_n \supseteq \mathcal{T}_\infty$ und insbesondere $\mathcal{T}_\infty \subseteq \sigma\left(\overline{\mathcal{A}}\right)$.
Somit folgt

$$\mathcal{T}_\infty \subseteq \sigma\left(\overline{\mathcal{A}}\right) \subseteq \mathcal{D}.$$

Dies impliziert $A \in \mathcal{D}$.
Daher ist nun A unabhängig von A. Dies kann nur sein, wenn

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A) \cdot \mathbb{P}(A) = (\mathbb{P}(A))^2.$$

gilt. Daher ist $\mathbb{P}(A) \in \{0, 1\}$ wie behauptet. ∎

Als unmittelbare Folgerung erhalten wir

Theorem 9.11 (Borels Null-Eins-Gesetz) *Für jede unabhängige Folge* $(A_n)_n$ *von Ereignissen* $A_n \in \mathcal{F}$ *gilt*

$$\mathbb{P}(A_n \textit{ für unendlich viele } n) = 0$$

oder

$$\mathbb{P}(A_n \textit{ für unendlich viele } n) = 1,$$

d.h.

$$\mathbb{P}\left(\limsup A_n\right) \in \{0, 1\}.$$

Beweis. Wir definieren die σ–Algebren $\mathcal{A}_n = \sigma(A_n)$. D.h.

$$\mathcal{A}_n = \{\emptyset, \Omega, A_n, A_n^c\}.$$

Es folgt, dass $(\mathcal{A}_n)_n$ unabhängig ist. Mit der obigen Bezeichnung für $\mathcal{T}_n$ und $\mathcal{T}_\infty$ ist also

$$Q_n := \bigcup_{m=n}^{\infty} A_m \in \mathcal{T}_n.$$

Da die Folge von σ–Algebren $(\mathcal{T}_n)_n$ fallend ist, gilt auch $Q_m \in \mathcal{T}_n$ für alle $m \geq n, n \in \mathbb{N}$. Da aber auch die Mengenfolge $(Q_n)_n$ fallend ist, erhalten wir

$$\limsup_{n \to \infty} A_n = \bigcap_{k=1}^{\infty} Q_k = \bigcap_{k=j}^{\infty} Q_k \in \mathcal{T}_j$$

für alle $j \in \mathbb{N}$. Somit ist $\limsup A_n \in \mathcal{T}_\infty$. Die Behauptung folgt nun aus dem Kolmogorovschen Null-Eins-Gesetz. ∎

Übung 9.12 *Für jeden W-Raum $(\Omega, \mathcal{F}, \mathbb{P})$ sind für $A \in \mathcal{F}$ stets die Paare $\{A, \emptyset\}$, und $\{A, \Omega\}$ Paare unabhängiger Ereignisse. Sind dies die einzigen solchen Paare, nennen wir die σ-Algebra $\mathcal{F}$ unabhängigkeitsfrei. Man zeige, dass der folgende Raum unabhängikeitsfrei ist.*

$$\Omega = \mathbb{N}, \mathcal{A} = \mathcal{P}(\Omega), \quad \textit{und} \quad \mathbb{P}(\{k\}) = 2^{-k!}$$

für jedes $k \geq 2, \mathbb{P}(\{1\}) = 1 - \sum_{k=2}^{\infty} \mathbb{P}(k)$.

Noch wichtiger als das Konzept der Unabhängigkeit von σ–Algebren ist das Konzept der Unabhängigkeit von Zufallsvariablen, das als nächstes vorgestellt werden soll. Man führt es auf das Konzept der Unabhängigkeit von σ–Algebren zurück. Wir werden wieder stets über einem W-Raum $(\Omega, \mathcal{F}, \mathbb{P})$ arbeiten. Wir definieren

Definition 9.13 *Eine Familie von Zufallsvariablen $(X_i)_i$ heißt unabhängig, wenn die von ihnen erzeugten σ–Algebren $(\sigma(X_i))_i$ unabhängig sind.*

Für endliche Familien von Zufallsvariablen gibt es ein anderes Kriterium für ihre Unabhängigkeit (was schon allein deshalb wichtig ist, weil man offenbar für die Uanbhängigkeit einer Familie von Zufallsvariablen nur die Unabhängigkeit der endlichen Teilfamilien überprüfen muss). Dieses Kriterium ist im wesentlichen schon aus der Stochastikvorlesung bekannt.

Theorem 9.14 *Sei $X_1, ..., X_n$ eine Folge von n Zufallsvariablen mit Werten in messbaren Räumen $(\Omega_i, \mathcal{A}_i)$, $i = 1, \ldots n$, mit durchschnittsstabilen Erzeugern $\mathcal{E}_i$ of $\mathcal{A}_i$. Die Zufallsvariablen $X_1, ..., X_n$ sind genau dann unabhängig, wenn die folgende Gleichheit gilt*

$$\mathbb{P}\left(X_1 \in E_1, ..., X_n \in E_n\right) = \prod_{i=1}^{n} \mathbb{P}\left(X_i \in E_i\right)$$

für alle $E_i \in \mathcal{E}_i, i = 1...n$.

Beweis. Wir setzen

$$\mathcal{G}_i := \left\{X_i^{-1}\left(E_i\right), E_i \in \mathcal{E}_i\right\}.$$

Eine kleine Übung in der Maßtheorie zeigt, dass die $\mathcal{G}_i$ die σ–Algebren $\sigma\left(X_i\right)$ erzeugen. $\mathcal{G}_i$ ist durchschnittsstabil und es ist $\Omega \in \mathcal{G}_i$. Nach Korollar 9.6 müssen wir zeigen, dass die Unabhängigkeit der $\left(\mathcal{G}_i\right)_{i=1...n}$ gleichbedeutend ist mit

$$\mathbb{P}\left(G_1 \cap ... \cap G_n\right) = \mathbb{P}\left(G_1\right) ... \mathbb{P}\left(G_1\right)$$

für alle Wahlen von $G_i \in \mathcal{G}_i$.
Aber dies ist offensichtlich, denn wir können stets $G_i = \Omega_i$ für geeignete Indizes i wählen. ■

Korollar 9.15 *Eine endliche Menge von Zufallsvariablen $X_1, ..., X_{n+1}$ mit Werten in messbaren Räumen $(\Omega_i, \mathcal{F}_i)$ ist genau dann unabhängig, wenn $X_1, ..., X_n$ unabhängig sind und X_{n+1} unabhängig ist von $\sigma\left(X_1, ..., X_n\right)$.*

Beweis. Da die disjunkte Zusammenfassung unabhängiger Familien wieder unabhängige Familien liefert ist die genannte Bedingung notwendig.
Dass sie auch hinreichend ist, sieht man wie folgt ein: Die σ-Algebra $\sigma(X_1, \ldots, X_n)$ wird vom System aller Mengen

$$Q := \{X_1 \in Q_1, \ldots, X_n \in Q_n\}$$

mit $Q_i \in \mathcal{F}_i$ und $i = 1, \ldots, n$ erzeugt. Dabei gilt

$$\mathbb{P}(Q) = \prod_{i=1}^{n} \mathbb{P}(X_i \in Q_i)$$

und für $Q_{n+1} \in \mathcal{F}_{n+1}$

$$\mathbb{P}(Q \cap \{X_{n+1} \in Q_{n+1}\}) = \mathbb{P}(Q)\mathbb{P}(\{X_{n+1} \in Q_{n+1}\}).$$

Zusammen ergibt sich

$$\mathbb{P}\left(\bigcap_{i=1}^{n+1}\{X_i \in Q_i\}\right) = \prod_{i=1}^{n+1}\mathbb{P}(X_i \in Q_i)$$

für alle $Q_i \in \mathcal{F}_i$, was nach obigem Satz gleichbedeutend mit der Unabhängigkeit der X_i ist. ■

Beispiel 9.16 *Es seien X und Y Zufallsvariablen definiert auf einem gemeinsamen W-Raum $(\Omega, \mathcal{F}, \mathbb{P})$. X nehme Werte in einem messbaren Raum $\Omega', \mathcal{F}')$ an und sei dort Dirac–verteilt in einem Punkt $x \in \Omega'$, d.h. $\mathbb{P}(X \in A') = 1$ ($A' \in \mathcal{F}'$), falls $x \in A'$ und ansonsten ist $\mathbb{P}(X \in A') = 0$. Y habe einen beliebigen messbaren Bildraum $(\Omega'', \mathcal{F}'')$ und dort eine beliebige Verteilung. Dann ist X von Y unabhängig.*
In der Tat: Zu zeigen ist

$$\mathbb{P}(X \in A', Y \in A'') = \mathbb{P}(X \in A')\mathbb{P}(Y \in A'')$$

für beliebige $A' \in \mathcal{F}'$ und $A'' \in \mathcal{F}''$. Ist nun $x \notin A'$ so ist diese Gleichheit offensichtlich wahr, beide Seiten sind null. Ist $x \in A'$, so erhalten wir

$$\begin{aligned}
\mathbb{P}(X \in A', Y \in A'') &= \mathbb{P}(X \in A') + \mathbb{P}(Y \in A'') - \mathbb{P}(\{X \in A'\} \cup \{Y \in A''\}) \\
&= 1 + \mathbb{P}(Y \in A'') - 1 \\
&= \mathbb{P}(X \in A')\mathbb{P}(Y \in A'').
\end{aligned}$$

Der folgende Satz zeigt, dass eine messbare Deformation (Transformation) einer Familie unabhängiger Zufallsvariablen wieder eine Familie unabhängiger Zufallsvariablen ergibt.

Theorem 9.17 *Es sei $(X_i)_{i\in I}$ eine Familie unabhängiger Zufallsvariablen mit Werten in messbaren Räumen $(\Omega_i, \mathcal{A}_i)$ und es seien*

$$f_i : (\Omega_i, \mathcal{A}_i) \to \left(\Omega_i^{'}, \mathcal{A}_i^{'}\right)$$

messbare Abbildungen. Dann ist auch die Familie der Zufallsvariablen $(f_i(X_i))_{i\in I}$ unabhängig.

Beweis. Zunächst sei bemerkt, dass aus der Messbarkeit der f_i folgt, dass auch die $(f_i(X_i))_{i\in I}$ messbar und somit Zufallsvariablen sind.
Es sei $i_1, ..., i_n \in I$. Dann gilt für jede Wahl messbarer Mengen $A_{i_1} \in \mathcal{A}_{i_1}, \ldots, A_{i_n} \in \mathcal{A}_{i_n}$ die folgende Identität

$$\begin{aligned}
& \mathbb{P}\left(f_{i_1}(X_{i_1}) \in A'_{i_1}, ..., f_{i_n}(X_{i_n}) \in A'_{i_n}\right) \\
= & \mathbb{P}\left(X_{i_1} \in f_{i_1}^{-1}\left(A'_{i_1}\right), ...X_{i_n} \in f_{i_n}^{-1}\left(A'_{i_n}\right)\right) \\
= & \prod_{\nu=1}^{n} \mathbb{P}\left(X_{i_\nu} \in f_{i_\nu}^{-1}\left(A'_{i_\nu}\right)\right) \\
= & \prod_{\nu=1}^{n} \mathbb{P}\left(f_{i_\nu}(X_{i_\nu}) \in A_{i_\nu}\right)
\end{aligned}$$

aufgrund der Unabhängigkeit der $(X_i)_{i\in I}$. ■

Schon aus Theorem 9.14 lässt sich ein Zusammenhang von Unabhängigkeit mit Produktmaßen erahnen. Dies soll im folgenden präzisiert werden. Zu diesem Zwecke seien $X_1, ..., X_n$ Zufallsvariablen mit

$$X_i : (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{A}_i)$$

Wir definieren die neue Zufallsvariable Y vermöge

$$Y := X_1 \otimes ... \otimes X_n : \Omega \to \Omega_1 \times ... \times \Omega_n$$

Dann erfüllt die Verteilung von Y, die wir mit $\mathbb{P}_Y$ bezeichnen (natürlich) die Gleichheit $\mathbb{P}_Y = \mathbb{P}_{X_1\otimes ...\otimes X_n}$. Insbesondere ist $\mathbb{P}_Y$ ein Wahrscheinlichkeitsmaß auf $\otimes_{i=1}^{n}\mathcal{A}_i$. Der oben erahnte Zusammenhang von Unabhängigkeit und Produktmaßen lässt sich nun wie folgt formalisieren:

Theorem 9.18 *DieZufallsvariablen $X_1, ..., X_n$ sind dann und nur dann unabhängig, wenn ihre Verteilung das Produktmaß der Einzelverteilungen ist, d.h. wenn gilt*

$$\mathbb{P}_{X_1\otimes ...\otimes X_n} = \mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n}$$

Beweis. Es seien $A_i \in \mathcal{A}_i, i = 1, \ldots, n$. Dann erhalten wir mit T

$$Y = X_1 \otimes ... \otimes X_n$$

die Gleichheit

$$\mathbb{P}_Y\left(\prod_{i=1}^{n} A_i\right) = \mathbb{P}\left(Y \in \prod_{i=1}^{n} A_i\right) = \mathbb{P}\left(X_1 \in A_1, ..., X_n \in A_n\right)$$

sowie

$$\mathbb{P}_{X_i}\left(A_i\right) = \mathbb{P}\left(X_i \in A_i\right) \qquad i = 1, \ldots, n.$$

Nun ist $\mathbb{P}_Y$ genau dann das Produktmaß der $\mathbb{P}_{X_i}$, wenn für beliebige Wahlen der $A_i \in \mathcal{A}_i$ gilt

$$\mathbb{P}_Y\left(\prod_{i=1}^{n} A_i\right) = \mathbb{P}_Y\left(A_1 \times ... \times A_n\right) = \mathbb{P}_{X_1}\left(A_1\right) ... \mathbb{P}_{X_n}\left(A_n\right).$$

Nach dem oben berechneten ist dies gleichbedeutend mit

$$\mathbb{P}\left(X_1 \in A_1, ..., X_n \in A_n\right) = \prod_{i=1}^{n} \mathbb{P}\left(X_i \in A_i\right).$$

Aber nach Theorem 9.14 ist diese Forderung äquivalent mit der Unabhängigkeit der X_i. ■
Der Nutzen dieses Satzes liegt darin, dass er uns erlaubt, unabhängige Familien von Zufallsvariablen mit vorgegebener Verteilung zu konstruieren. Für endliche Familien ist dies unmittelbar klar, denn man weiß bereits aus der Maßtheorie, dass endliche Produktmaße existieren. Die Existenz unendlicher Produke von Wahrscheinlichkeitsmaßen ist in der Tat noch zu zeigen. Dies soll in einem der nächsten Kapitel geschehen.

10 Produkte und Summen unabhängiger Zufallsvariablen

Zeil dieses Abschnitts ist es einige Eigenschaften unabhängigerZufallsvariablen genauer zu studieren. Der erste Satz zeigt uns, dass sich die Faktorisierungseigenschaft unabhängier Zufallsvariablen auf deren Erwartungswerte überträgt.

Theorem 10.1 *Es seien $X_1, ..., X_n$ unabhängige, reellwertige Zufallsvariablen . Der Erwartungswert von $\prod_{i=1}^n X_i$ existiert genau dann, wenn alle Erwartungswerte $\mathbb{E}(X_i)$ existieren. In diesem Falle gilt:*

$$\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i) \tag{10.1}$$

Beweis. Da die $X_1, ..., X_n$ unabhängige Zufallsvariablen sind, wissen wir, dass $Q := \otimes_{i=1}^n \mathbb{P}_{X_i}$ ihre gemeinsame Verteilung ist. Aus dem Satz von Fubini erhalten wir daher

$$\begin{aligned}
\mathbb{E}\left(\left|\prod_{i=1}^n X_i\right|\right) &= \int |X_1 \cdots X_n| \, Q(dx) \\
&= \int \dots \int |x_1| \cdots |x_n| \, d\mathbb{P}_{X_1}(dx_1) \dots d\mathbb{P}_{X_n}(x_n) \qquad (10.2) \\
&= \int |x_1| \, d\mathbb{P}_{X_1}(x_1) \dots \int |x_n| \, d\mathbb{P}_{X_n}(x_n)
\end{aligned}$$

Dies zeigt, dass die Integrierbarkeit der einzelnen X_i die Integrierbarkeit von $\prod_{i=1}^n X_i$ zur Folge hat und umgekehrt. In diesem Falle dürfen wir auch die Rechnung (10.2) ohne die Betragsstriche wiederholen. Dies beweist die Behauptung. ■

Bemerkung 10.2 *Das obige Theorem führt notwendig zu der Frage, ob für zwei Zufallsvariablen mit existenten Erwartungswerten* (10.1) *vielleicht gleichbedeutend mit deren Unabhängigkeit ist, d.h. ob zwei integrierbare Zufallsvariablen X, Y genau dann unabhängig sind, wenn*

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\,\mathbb{E}(Y)$$

gilt. Das folgende Beipsiel zeigt, dass dies nicht wahr ist: Sei $\Omega = \{1, 2, 3\}$ und $\mathcal{F} = \mathcal{P}(\Omega)$, sowie $\mathbb{P}$ die Gleichverteilung (die Laplacesche Wahrscheinlichkeit) auf Ω. Seien die Zufallsvariablen X und Y wie folgt definiert: $X(1) = 1, X(2) = 0, X(3) = -1$ und $Y(1) = Y(3) = 0$ und $Y(2) = 1$. Dann ist $XY \equiv 0$ da entweder X oder Y null sind. Da auch $\mathbb{E}X = 0$ ist, folgt

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\,\mathbb{E}(Y).$$

Andererseits ist

$$\mathbb{P}(X = 1, Y = 1) = 0 \text{ aber } \mathbb{P}(X = 1) = \mathbb{P}(Y = 1) = \frac{1}{3},$$

so dass X und Y nicht unabhängig sein können.

Dennoch spielt die soeben diskutierte Gleichheit der Erwartungswerte eine wichtige Rolle in der Wahrscheinlichkeitstheorie, da sie einen ersten Hinweis auf Unabhängigkeit liefert. Sie bekommt auch einen eigenen Namen.

Definition 10.3 *Für zwei integrierbare Zufallsvariablen X, Y mit integrierbarem Produkt definieren wir die Kovarianz von X and Y als*

$$\begin{aligned} cov\,(X,Y) &= \mathbb{E}\left[(X-\mathbb{E}X)\,(Y-\mathbb{E}Y)\right] \\ &= \mathbb{E}\,(XY) - \mathbb{E}X\mathbb{E}Y \end{aligned}$$

X und Y heißen unkorreliert*, falls $cov\,(X,Y)=0$.*

Offensichtlich gilt

Bemerkung 10.4 *Sind X und Y unabhängige Zufallsvariablen , so gilt $cov(X,Y) = 0$.*

Für quadratintegrierbare Zufallsvariablen lässt nun die Varianz der Summe wie folgt berechnen

Theorem 10.5 *Es seien $X_1, ..., X_n$ quadratintegrierbare Zufallsvariablen . Dann gilt*

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i) + \sum_{i\neq j} cov\,(X_i, X_j) \tag{10.3}$$

Sind insbesondere die $X_1, ..., X_n$ paarweise unkorreliert, so erhält man

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}\,(X_i)\,. \tag{10.4}$$

Beweis. Es gilt

$$\begin{aligned} \mathbb{V}\left(\sum_{i=1}^{n} X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^{n} (X_i - \mathbb{E}X_i)\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^{n} (X_i - \mathbb{E}X_i)^2 + \sum_{i\neq j} (X_i - \mathbb{E}X_i)\,(X_j - \mathbb{E}X_j)\right)\right] \\ &= \sum_{i=1}^{n} \mathbb{V}\,(X_i) + \sum_{i\neq j} cov\,(X_i, X_j)\,. \end{aligned}$$

Dies beweist (10.3). Für (10.4) genügt es zu bemerken, dass für unkorrelierte Zufallsvariablen X und Y definitionsgemäß $cov(X,Y) = 0$ gilt. ■
Schlußendlich wollen wir auch die Verteilung der Summe einer Folge unabhängiger Zufallsvariablen berechnen. Es ist naheliegend, dass dabei die folgende Definition eine wichtige Rolle einnimmt.

Definition 10.6 *Es sei A_n die folgende Abbildung vom $\mathbb{R}^{nd}$ in den $\mathbb{R}^d$:*

$$\begin{aligned} A_n : \mathbb{R}^{nd} &\to \mathbb{R}^d \\ (x_1, \ldots, x_n) &\mapsto x_1 + \cdots + x_n \end{aligned} \tag{10.5}$$

(Bemerke, dass A_n stetig und somit $\mathcal{B}^{nd} - -\mathcal{B}^d$–messbar ist.) Für Maße $\mu_1, \ldots \mu_n$ auf $(\mathbb{R}^d, \mathcal{B}^d)$ heißt das Bildmaß von $\mu_1 \otimes \ldots \otimes \mu_n$ unter A_n das Faltungsprodukt oder die Faltung der Maße $\mu_1, \ldots \mu_n$. Man schreibt auch

$$\mu_1 * \ldots * \mu_n := A_n(\mu_1 \otimes \ldots \otimes \mu_n).$$

Durch Kombination der Ergebnisse über Bild– und Produktmaße erhält man die wichtigsten Eigenschaften der Faltung: $\mu_1 * \ldots * \mu_n$ ist wieder ein Maß auf $(\mathbb{R}^d, \mathcal{B}^d)$ mit

$$\mu_1 * \ldots * \mu_n(\mathbb{R}^d) = \mu_1 \otimes \ldots \otimes \mu_n(\mathbb{R}^{nd}) = \|\mu_1\| \cdots \|\mu_n\|.$$

Viele Sätze über Faltungsprodukte lassen sich induktiv über die Anzahl der zu faltenden Maße herleiten, denn es gilt:

$$\mu_1 * \ldots * \mu_{n+1} = (mu_1 * \ldots * \mu_n) * \mu_{n+1}.$$

Insbesondere gilt nach dem Transformationssatz und dem Satz von Fubini für je zwei Maße μ und ν auf $(\mathbb{R}^d, \mathcal{B}^d)$ und eine nichtnegative, $\mathcal{B}^d$–messbare Funktion f:

$$\begin{aligned} \int f d(\mu * \nu) &= \int f \circ A_2 d(\mu \otimes \nu) \\ &= \int f(x+y)\mu(dx)\nu(dy) = \int f(x+y)\nu(dy)\mu(dx). \end{aligned} \tag{10.6}$$

Wählt man speziell für f eine Indikatorfunktion auf einer messbaren Menge, so sieht man dass die Faltung assoziativ ist (die Kommutativität hatten wir gerade zuvor gesehen).

Ferner überprüft man leicht, dass für Maße μ, ν, μ_1, ν_1 auf $(\mathbb{R}^d, \mathcal{B}^d)$ und ein Skalar $\alpha \geq 0$ gilt:

$$\mu * (\nu_1 + \nu_2) = \mu * \nu_1 + \mu * \nu_2 \tag{10.7}$$
$$\mu * (\alpha\nu) = (\alpha\mu) * \nu = \alpha(\mu * \nu). \tag{10.8}$$

Die Faltung ist also auch distributiv und verträglich mit positiven Skalaren. Den Zusammenhang zur Verteilung von Summen unabhängiger Zufallsvariablen klärt das folgende Theorem.

Theorem 10.7 *Es seien $X_1, ..., X_n$ unabhängige und $\mathbb{R}^d$–wertige Zufallsvariablen . Dann ist die Verteilung der Summe $S_n := X_1 + ... + X_n$ gegeben durch das Faltungsprodukt der Verteilungen der X_i. Somit gilt*

$$\mathbb{P}_{S_n} := \mathbb{P}_{X_1} * \mathbb{P}_{X_2} ... * \mathbb{P}_{X_n}$$

Beweis. Wieder sei

$$Y := X_1 \otimes ... \otimes X_n : \Omega \to \left(\mathbb{R}^d\right)^n$$

und

$$A_n : \mathbb{R}^d \times \ldots \times \mathbb{R}^d \to \mathbb{R}^d$$

Dann ist $S_n = A_n \circ Y$ und somit ist S_n eine Zufallsvariable . Nun ist $\mathbb{P}_{S_n}$ das Bildmaß von $\mathbb{P}$ unter $A_n \circ Y$ (was wir wie üblich mit $(A_n \circ Y)(\mathbb{P})$ bezeichnen). Somit gilt

$$\mathbb{P}_{S_n} = (A_n \circ Y)(\mathbb{P}) = A_n(\mathbb{P}_Y).$$

Nun ist $\mathbb{P}_Y = \otimes \mathbb{P}_{X_i}$. Damit gilt nach Definition des Faltungsprodukts

$$\mathbb{P}_{X_i} * ... * \mathbb{P}_{X_n} = A_n(\mathbb{P}_Y) = \mathbb{P}_{S_n}$$

■

Einige Beispiele für Faltungsprodukte haben wir schon in der Stochastikvorlesung kennengelernt.

Beispiel 10.8 *1. In der Stochastikvorlesung haben wir schon gesehen, dass die Summe einer Binomial verteilten Zufallsvariablen X mit den Parametern n und p (also einer $B(n,p)$–verteilten Zufallsvariablen) und einer davon unabhängigen Binomial–verteilten Zufallsvariablen Y, die $B(m,p)$–verteilt ist, eine $B(n+m,p)$–Verteilung besitzt. Also:*

$$B(n.p) * B(m,p) = B(n+m,p).$$

2. *Ebenso haben wir in der Stochastikvorlesung schon gesehen, dass die Faltung einer $\mathcal{P}(\lambda)$–Verteilung (also einer Poisson Verteilung zum Parameter λ) mit einer $\mathcal{P}(\mu)$–Verteilung eine $\mathcal{P}(\mu+\lambda)$–Verteilung ergibt. In Zeichen:*
$$\mathcal{P}(\lambda) * \mathcal{P}(\mu) = \mathcal{P}(\lambda+\mu)$$

3. *Schlußendlich hatten wir uns in der Stochastikvorlesung auch mit der Summe unabhängiger normal–verteilter Zufallsvariablen beschäftigt. Übersetzt man die Resulatet von dort in die Sprache der Faltungsprodukte, so haben wir gesehen, dass die folgende Identität gilt:*
$$\mathcal{N}\left(\mu,\sigma^2\right) * \mathcal{N}\left(\nu,\tau^2\right) = \mathcal{N}\left(\mu+\nu,\sigma^2+\tau^2\right).$$

11 Unendliche Produkte von Wahrscheinlichkeitsräumen

Schon in der Stoachastikvorlesung hatten wir das schwache Gesetz der großen Zahlen kennengelernt, das besagte, dass für eine Folge $X_1, \ldots X_n$ (damals noch diskreter) unabhängiger, identisch verteilter Zufallsvariablen mit bestehendem Erwartungswert $\mathbb{E}X_1$ und endlichen Varianzen die Wahrscheinlichkeit $\mathbb{P}(|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}X_1| \geq \varepsilon)$ für jedes positive $\varepsilon > 0$ gegen 0 konvergiert, wenn man n nach unendlich streben lässt. Schon damals hatten wir uns gefragt, ob nicht etwa auch
$$\mathbb{P}(\lim \frac{1}{n}\sum_{i=1}^n X_i \text{ exisitiert und ist gleich } \mathbb{E}X_1) = 1$$
gilt. Damals hatten wir gesehen (oder waren überredet worden), dass dies nicht nur eventuell schwer zu beweisen ist, sondern, dass die Problematik früher beginnt: Wir können noch nicht einmal die Existenz einer solchen unendlichen Folge von Zufallsvariablen sicherstellen.
In der Tat beginnen viele Sätze in der Wahrscheinlichkeitstheorie mit dem Satz "Es sei $X_1, \ldots, X_n, \ldots$ eine Folge unabhängiger, identisch verteilter Zufallsvariablen ..." (wobei wir im folgendenden "unabhängiger, identisch verteilt" mit dem englischen "i.i.d." abkürzen werden).
Aber: Woher wissen wir, dass es eine solche Folge wirklich gibt?
Dies soll in diesem Abschnitt geklärt werden. Wir werden somit den Rahmen für die wesentlichen Sätze der Wahrscheinlichkeitstheorie setzen.

In derselben Weise wie die Existenz endlicher Folgen von i.i.d. Zufallsvariablen mit der Existenz endlicher Produktmaße zusammenhängt, müssen wir uns für die eingangs gestellte Frage mit der Existenz unendlicher Produkte von Wahrscheinlichkeitsräumen befassen.
Wir werden also im folgenden stets eine Folge messbarer Räume (Ω_n, A_n, μ_n) vorgeben, von denen wir darüber hinaus annehmen, dass die μ_n Wahrscheinlichkeitsmaße sind, also, dass

$$\mu_n(\Omega_n) = 1 \text{ für alle } n.$$

gilt.
Wir konstruieren daraus folgendermaßen den Produktraum $(\Omega, \mathcal{A})$: Wir möchten, dass jedes $\omega \in \Omega$ sich als Folge $(\omega_n)_n$ auffassen lässt, wobei die $\omega_n \in \Omega_n$ gewählt sind. Wir setzen daher

$$\Omega := \prod_{n=1}^{\infty} \Omega_n.$$

Darüber hinaus sollte ein Wahrscheinlichkeitsmaß auf Ω dadurch festgelegt sein, dass wir wissen, was die Wahrscheinlichkeit der ersten n Koordinaten ist (und dies natürlich für jedes $n \in \mathbb{N}$). Es liegt daher nahe zu fordern, dass für $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2, ..., A_n \in \mathcal{A}_n, n \in \mathbb{N}$ die Mengen

$$A := A_1 \times ... \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times ... \tag{11.1}$$

in der σ-Algebra $\mathcal{A}$ liegen.
Da die μ_i in der Anwendung die Verteilung unabhängiger Zufallsvariablen X_i darstellen sollen, sollte das zu definierende Maß μ auf $(\Omega, \mathcal{A})$ einer Menge A, wie wir sie in (11.1) definiert haben die Masse

$$\mu(A) = \mu_1(A)...\mu_n(A_n).$$

geben.
Wir werden das Problem der Existenz eines solchen μ in größerer Allgemeinheit lösen.
Wir nehmen an, dass I eine beliebige Indexmenge ist (also nicht notwendig abzählbar) und dass $(\Omega_i, \mathcal{A}_i, \mu_i)_{i\in I}$ Massräume sind mit Wahrscheinlichkeitsmaßen μ_i: $\mu_i(\Omega_i) = 1$. Für $\emptyset \neq K \subseteq I$ definieren wir das Produkt der zugehörigen Ω_i als

$$\Omega_K := \prod_{i\in K} \Omega_i. \tag{11.2}$$

Insbesondere schreiben wir

$$\Omega := \Omega_I.$$

Mit p_J^K definieren wir für $J \subseteq K$ die kanonischen Projektionen von Ω_K in Ω_J, d.h.

$$p_J^K((\omega_k)_{k \in K}) = (\omega_j)_{j \in J}.$$

Ist $J = \{i\}$ eine einelementige Menge, so schreiben wir auch p_i^K statt $p_{\{i\}}^K$ und p_i statt p_i^I. Offenbar sind p_J^K mit einander verträglich, d.h. es gilt:

$$p_J^L = p_J^K \circ p_K^L \quad \text{für } J \subseteq K \subseteq L \tag{11.3}$$

sowie

$$p_J := p_J^I = p_J^K \circ p_K \quad \text{für } J \subseteq K. \tag{11.4}$$

Schließlich bezeichnen wir mit

$$\mathcal{H}(I) := \{J \subseteq I, J \neq \emptyset, |J| \text{ is finite}\}.$$

Für $J \in \mathcal{H}(I)$ wissen wir nun wie wir die σ–Algebren und Maße über Ω_J definieren müssen. Wir setzen nämlich

$$\mathcal{A}_J := \otimes_{i \in J} \mathcal{A}_i \text{ and } \mu_J := \otimes_{i \in J} \mu_i$$

wie im Satz Fubini aus der Maßtheorie.

In Analogie zum Fall endlicher Indexmengen definieren wir nun

Definition 11.1 *Die Produkt σ-Algebra $\otimes_{i \in I} \mathcal{A}_i$ der σ–Algebren $(\mathcal{A}_i)_{i \in I}$ ist definiert als die kleinste σ-Algebra $\mathcal{A}$ über Ω, so dass alle Projektionen*

$$p_i : \Omega \to \Omega_i$$

$\mathcal{A} - \mathcal{A}_i$–messbar sind. Also ist

$$\otimes_{i \in I} \mathcal{A}_i := \sigma(p_i, i \in I). \tag{11.5}$$

Lemma 11.2 *Es gilt*

$$\otimes_{i \in I} \mathcal{A}_i := \sigma(p_J, J \in \mathcal{H}(I)). \tag{11.6}$$

Beweis. Für jedes $J \in \mathcal{H}$ ist die Projektionsabbildung p_J auch $\mathcal{A}$–$\mathcal{A}_J$–messbar, da $\mathcal{A}_J = \sigma(p_i^J, i \in J)$ gilt und $p_i = p_i^J \circ p_J$ für jedes $i \in J$. Somit folgt die behauptete Gleichheit. ■

Nach der obigen Motivation suchen wir also ein Maß μ auf $(\Omega, \mathcal{A})$, das einer Menge A wie in (11.1) die Masse (Wahrscheinlichkeit) $\mu_1(A_1)\ldots\mu_n(A_n)$ zuweist. Mit anderen Worten μ sollte so, sein ,dass

$$\mu\left(p_J^{-1}\left(\prod_{i\in J} A_i\right)\right) = \mu_J\left(\prod_{i\in J} A_i\right).$$

gilt.
Die Frage, ob ein solches Maß existiert, wird im folgenden Satz geklärt:

Theorem 11.3 *Auf $\mathcal{A} := \otimes_{i\in I}\mathcal{A}_i$ gibt es eindeutig bestimmtes Maß μ mit*

$$p_J(\mu) = \mu_J \tag{11.7}$$

für alle $J \in \mathcal{H}(I)$. Darüber hinaus ist μ ein Wahrscheinlichkeitsmaß, d.h. es gilt $\mu(\Omega) = 1$.

Beweis. Zunächst könne wir annehmen, dass I eine unendliche Menge ist, d.h. dass $|I| = \infty$ gilt, da das Resultat ansonsten aus dem Satz von Fubini folgt.
Wir beginnen den Beweis mit einigen vorbereitenden Überlegungen

Für je zwei Mengen $J, K \in \mathcal{H}$ mit $J \subseteq K$ ist die Abbildung

$$p_J^K : \Omega_K \to \Omega_J$$

$\mathcal{A}_K$ - $\mathcal{A}_J$ - messbar. In der Tat erzeugen ja die Mengen $\prod_{i\in J} A_i$ mit $A_i \in \mathcal{A}_i$ und $i \in J$ ja $\mathcal{A}_J$ und es gilt

$$(p_J^K)^{-1}(\prod_{i\in J} A_i) = \prod_{i\in K} A_i'$$

mit $A_i' = A_i$ für $i \in J$ und $A_i' = \Omega_i$ für $i \in K \setminus J$.
Da wir es ferner mit Wahrscheinlichkeitsmaßen zu tun haben, gilt

$$\prod_{i\in K} \mu_i(A_i') = \prod_{i\in J} \mu_i(A_i)$$

d.h.
$$p_J^K(\mu_K) = \mu_J \text{ für } J \subseteq K \text{ und } J, K \in \mathcal{H}(I).$$

Führen wir also die σ-Algebra der J–Zylindermengen

$$\mathcal{Z}_J := p_J^{-1}(\mathcal{A}_J) \qquad (J \in \mathcal{H}(I)) \tag{11.8}$$

ein, so besagt die soeben festgestellte Messbarkeit von p_J^K, dass

$$\left(p_J^K\right)^{-1}(\mathcal{A}_J) \subset \mathcal{A}_K$$

und somit
$$\mathcal{Z}_J \subseteq \mathcal{Z}_K \qquad (J \subseteq K, J, K \in \mathcal{H}(I)) \tag{11.9}$$
gilt.
Schließlich definieren wir das System aller Zylindermengen als

$$\mathcal{Z} := \bigcup_{J \in \mathcal{H}(I)} \mathcal{Z}_J.$$

Wir bemerken, dass wegen (11.9) für $\mathcal{Z}_1, \mathcal{Z}_2 \in \mathcal{Z}$ gilt, dass $\mathcal{Z}_1, \mathcal{Z}_2 \in \mathcal{Z}_J$ gilt, wenn man $\mathcal{J} \in \mathcal{H}(I)$ geeignet wählt. Somit sehen wir, dass $\mathcal{Z}$ eine Algebra ist (aber i.a keine σ-Algebra). Aufgrund der Definition (11.5) und der Identität (11.6) erhalten wir nun
$$\mathcal{A} = \sigma(\mathcal{Z}).$$

Nun kommen wir zum Hauptteil des Beweises. Dieser ist in vier Schritte unterteilt.

1. Wenn das gestellte Problem überhaupt lösbar ist, so muss jeder Menge $Z = p_J^{-1}(A)$. die Masse $\mu(Z) = \mu_J(A)$ zugeordnet werden. Dies muss für alle $J \in \mathcal{H}(I)$ und alle $A \in \mathcal{A}_J$ gelten.

   Zunächst zeigen wir, dass dies überhaupt wohldefiniert ist, d.h. die Masse von Z darf nur von Z selbst abhängen, nicht aber von der speziellen wahl von A oder J. Sei also

   $$Z = p_J^{-1}(A) = p_K^{-1}(B)$$

   für $J, K \in \mathcal{H}(I)$, $A \in \mathcal{A}_J, B \in \mathcal{A}_K$.

Falls nun $J \subseteq K$ ist, erhalten wir aus den obigen Überlegungen

$$p_J^{-1}(A) = p_K^{-1}\left(\left(p_J^K\right)^{-1}(A)\right)$$

und somit

$$p_K^{-1}(B) = p_K^{-1}(B') \quad \text{mit} \quad B' := \left(p_J^K\right)^{-1}(A)\,.$$

Da p_K eine Projektion ist, gilt $p_K(\Omega) = \Omega_K$. Somit ergibt sich

$$B = B' = \left(p_J^K\right)^{-1}(A)\,.$$

und somit mit den einleitenden Überlegungen

$$\mu_K(B) = \mu_J(A)\,.$$

Sind nun J und K beliebig, so definieren wir $L := J \cup K$. Da nun gilt, dass $J, K \subseteq L$, impliziert (11.9) die Existenz eines $C \in \mathcal{A}_L$ mit $p_L^{-1}(C) = p_J^{-1}(A) = p_K^{-1}(B)$.

Somit folgt aus dem soeben Gezeigten, dass

$$\mu_L(C) = \mu_J(A) \quad \text{und} \quad \mu_L(C) = \mu_K(B)$$

gilt. Also ist

$$\mu_J(A) = \mu_K(B).$$

Somit ist die Funktion

$$\mu_0\left(p_J^{-1}(A)\right) := \mu_J(A), \tag{11.10}$$

für $J \in \mathcal{H}(I)$ und $A \in \mathcal{A}_J$ wohldefiniert auf $\mathcal{Z}$.

2. Wir zeigen nun, dass das soeben in (11.10) definierte μ_0 ein Volumen (ein Inhalt) auf $\mathcal{Z}$ ist.

   Trivialerweise gilt dass, $\mu_0 \geq 0$ und $\mu_0(\emptyset) = 0$.

   Darüber hinaus gilt aufgrund der vorherigen Überlegungen, dass für $Y, Z \in \mathcal{Z}$ mit $Y \cap Z = \emptyset$ eine Menge $J \in \mathcal{H}(I)$ und $A, B \in \mathcal{A}_J$ existieren, so dass

$$Y = p_J^{-1}(A) \text{ und } Z = p_J^{-1}(B)$$

gilt. Ist zudem $Y \cap Z = \emptyset$, so folgt auch $A \cap B = \emptyset$. Und wegen

$$Y \cup Z = p_J^{-1}(A \cup B)$$

erhalten wir

$$\mu_0(Y \cup Z) = \mu_J(A \cup B) = \mu_J(A) + \mu_J(B) = \mu_0(Y) + \mu_0(Z)$$

und dies ist die endliche Additivität von μ_0.

Es bleibt zu zeigen, dass μ_0 auch σ-additiv ist. Dann kann μ_0 nach dem Satz von Carathéodory auf genau eine Art zu einem Maß μ auf $\sigma(\mathcal{Z}) = \mathcal{A}$ fortgesetzt werden. Wenn so ein μ existiert, ist es auch ein Wahrscheinlichkeitsmaß, weil

$$\Omega = p_J^{-1}(\Omega_J) \text{ für alle } J \in \mathcal{H}$$

gilt und daher

$$\mu(\Omega) = \mu_0(\Omega) = \mu_J(\Omega_J) = 1.$$

Um die gewünschte σ–Additivität von μ_0 nachzuweisen, zeigen wir zunächst:

3. Es sei $Z \in \mathcal{Z}$ und $J \in \mathcal{H}$. Dann ist für alle $\omega_J \in \Omega_J$ die Menge

$$Z^{\omega_J} := \{\omega \in \Omega : (\omega_J, p_{I\setminus J}(\omega)) \in Z\}$$

eine Zylindermenge. Sie besteht aus allen $\omega \in \Omega$ mit der folgenden Eigenschaft: Wenn wir die Koordinaten ω_i mit $i \in J$ durch die entsprechenden Koordinaten von ω_J ersetzen, erhalten wir einen Punkt aus Z. Darüber hinaus gilt

$$\mu_0(Z) = \int \mu_0(Z^{\omega_J})\mu_J(d\omega_J), \tag{11.11}$$

was die folgende Überlegung zeigt: Für $Z \in \mathcal{Z}$ gibt es $K \in \mathcal{H}$ und $A \in \mathcal{A}_K$, so dass $Z = p_K^{-1}(A)$ und dies bedeutet $\mu_0(Z) = \mu_K(A)$. Da I unendlich ist, können wir annehmen, dass $J \subset K$ und $J \neq K$ gilt. Für den ω_J–Schnitt von A in Ω_K, den wir A_{ω_J} nennen wollen, also für die Menge aller $\omega' \in \Omega_{K\setminus J}$ mit $(\omega_J, \omega') \in A$ gilt

$$Z^{\omega_J} = p_{K\setminus J}^{-1}(A_{\omega_J}).$$

Nach dem Satz von Fubini ist $A_{\omega_J} \in \mathcal{A}_{K\setminus J}$ und somit sind die $Z^{\omega_J} = p_{K\setminus J}^{-1}(A_{\omega_J})$ $(K \setminus J)$-Zylindermengen. Da nun $\mu_K = \mu_J \otimes \mu_{K\setminus J}$ gilt, impliziert der Satz von Fubini

$$\mu_0(Z) = \mu_K(A) = \int \mu_{K\setminus J}(A^{\omega_J})d\mu_J(\omega_J). \tag{11.12}$$

Aber dies ist gleichbedeutend mit (11.11), da

$$\mu_0(Z^{\omega_J}) = \mu_{K\setminus J}(A_{\omega_J})$$

(wegen $Z^{\omega_J} = p_{K\setminus J}^{-1}(A_{\omega_J})$).

4. Schließlich zeigen wir noch, dass μ_0 stetig ist in $\emptyset$ und somit σ-additiv. Hierzu sei (Z_n) eine fallende Familie von Zylindermengen in $\mathcal{Z}$ mit

$$\alpha := \inf_n \mu_0(Z_n) > 0.$$

Wir werden zeigen, dass dann auch

$$\bigcap_{n=1}^{\infty} Z_n \neq \emptyset \tag{11.13}$$

gilt.

Nun ist jedes Z_n von der Form

$$Z_n = p_{J_n}^{-1}(A_n) \qquad \text{mit } J_n \in \mathcal{H} \text{ und } A_n \in \mathcal{A}_{J_n}.$$

Aufgrund von (11.9) können wir annehmen, dass $J_1 \subseteq J_2 \subseteq J_3 \ldots$ gilt. Wir wollen und das im vorigen Schritt bewiesene Resultat mit $J = J_1$ und $Z = Z_n$ verwenden. Da die Abbildung

$$\omega_{J_1} \longmapsto \mu_0\left(Z_n^{\omega_{J_1}}\right)$$

$\mathcal{A}_1$–messbar ist, erhalten wir, dass

$$Q_n := \left\{\omega_{J_1} \in \Omega_{J_1} : \mu_0\left(Z_n^{\omega_{J_1}}\right) \geq \frac{\alpha}{2}\right\} \in \mathcal{A}_{J_1}.$$

gilt. Da alle μ_J's Masse eins haben, erhalten wir aus (11.11):

$$\alpha \leq \mu_0\left(Z_n\right) \leq \mu_{J_1}\left(Q_n\right) + \frac{\alpha}{2},$$

und somit, dass

$$\mu_{J_1}(Q_n) \geq \frac{\alpha}{2} > 0$$

für alle $n \in \mathbb{N}$ gilt. Zusammen mit (Z_n) ist auch (Q_n) fallend. μ_{J_1} ist zudem als endliches Maß stetig in $\emptyset$. Dies impliziert, dass $\bigcap_{n=1}^{\infty} Q_n \neq \emptyset$ gilt. Somit gibt es ein $\omega_{J_1} \in \bigcap_{n=1}^{\infty} Q_n$ mit

$$\mu_0\left(Z_n^{\omega_{J_1}}\right) \geq \frac{\alpha}{2} > 0.$$

Ist nun $J_2 \neq J_1$, so ergibt eine erneute Anwendung von 3. die Existenz eines $\omega_{J_2 \setminus J_1}$ mit

$$\mu_0(Z_n^{\omega_{J_1}})^{\omega_{J_2 \setminus J_1}} \geq 2^{-2}\alpha$$

für alle $n \in \mathbb{N}$. Setzen wir $\omega_{J_2} := (\omega_{J_1}, \omega_{J_2 \setminus J_1})$, so gilt

$$\omega_{J_2} \in \Omega_{J_2}.$$

Es ist

$$(Z_n^{\omega_{J_1}})^{\omega_{J_2 \setminus J_1}} = Z_n^{\omega_{J_2}}$$

und daher

$$\mu_0(Z_n^{\omega_{J_2}}) \geq 2^{-2}\alpha > 0 \text{ sowie } p_{J_1}^{J_2}(\omega_{J_2}) = \omega_{J_1}$$

für alle $n \in \mathbb{N}$.

Im Falle $J_1 = J_2$ wählen wir natürlich gleich $\omega_{J_1} = \omega_{J_2}$.

Induktiv erhält man durch wiederholte Anwendung von Schritt 3. für jedes $k \in \mathbb{N}$ die Existenz eines $\omega_{J_k} \in \Omega_{J_k}$ mit

$$\mu_0\left(Z_n^{\omega_{J_k}}\right) \geq 2^{-k}\alpha > 0 \text{ und } p_{J_k}^{J_{k+1}}\left(\omega_{J_{k+1}}\right) = \omega_{J_k}. \tag{11.14}$$

Aufgrund dieser letzten Eigenschaft, existiert ein $\omega_0 \in \Omega$ mit

$$p_{J_k}(\omega_0) = \omega_{J_k}.$$

Wegen (11.14) gilt $Z_n^{\omega_{J_n}} \neq \emptyset$. Somit gibt es ein $\tilde{\omega}_n \in \Omega$ mit

$$\left(\omega_{J_n}, p_{I \setminus J_n}(\tilde{\omega}_n)\right) \in Z_n.$$

Aber dann ist auch

$$\left(\omega_{J_n}, p_{I\setminus J_n}(\omega_0)\right) = \omega_0 \in Z_n.$$

Somit ist $\omega_0 \in \bigcap_{n=1}^{\infty} \mathcal{Z}_n$. Dies beweist (11.13).

Daher ist μ_0 σ–additiv und hat somit eine eindeutig definierte Fortsetzung nach $\mathcal{A}$ (gemäß dem Satz von Caratheodory). Die Fortsetzung μ von μ_0 hat Masse eins (d.h. $\mu(\Omega) = 1$), da für $J \in \mathcal{H}(I)$ gilt $\Omega = p_J^{-1}(\Omega_y)$ und somit

$$\mu(\Omega) = \mu_0(\Omega) = \mu_J(\Omega_J) = 1.$$

Dies beweist den Satz.

■

12 Null–Eins–Gesetze

Schon in Kapitel 9 haben wir den Prototyp eines Null–Eins–Gesetzes angetroffen, das Borelsche 0-1-Gesetz:
Für eine Folge von unabhängigen Ereignissen $(A_n)_n$ gilt

$$\mathbb{P}(\limsup A_n) \in \{0, 1\}.$$

Dies ist an sich naürlich eine interessante Tatsache, noch interessanter aber ist es, welche der beiden Möglichkeit zutrifft: Ist die in Frage stehende Wahrscheinlichkeit null oder ist sie eins? Dies kann in manchen Fällen dadurch gelöst werden, dass man die andere Möglichkeit durch Abschätzungen ausschließt. Noch häufiger aber wird das folgende Lemma gebrauch, das zu den wesentlichen Hilfsmitteln der Wahrscheinlichkeitstheorie zählt.

Lemma 12.1 (Borel–Cantelli–Lemma) *Es sei (A_n) eine Folge von Ereignissen definiert über einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Dann gilt die Implikation*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \Rightarrow \mathbb{P}(\limsup A_n) = 0 \tag{12.1}$$

Sind die Ereignisse (A_n) zumindest paarweise unabhängig, dann gilt auch die Umkehrung:

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \Rightarrow \mathbb{P}(\limsup A_n) = 1. \tag{12.2}$$

Bemerkung 12.2
- *Am häufigsten wird das Borel–Cantelli–Lemma in der Form* (12.1) *benutzt. Bemerkenswerterweise benötigen wir hier keinerlei Kenntnisse über die Abhängigkeitsstruktur der A_n.*
- *Die zweite Abschätzung* (12.2) *war zunächst nur unter der Voraussetzung der vollständigen Unabhängigkeit bekannt. Die hier zitierte Form ist natürlich stärker (d.h. gilt unter einer schwächeren Voraussetzung).*

Proof of Lemma 12.1. (12.1) ist einfach. Wir setzen

$$A := \limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i.$$

Dies impliziert

$$A \subseteq \bigcup_{i=n}^{\infty} A_i \quad \text{für alle } n \in \mathbb{N}.$$

und somit

$$\mathbb{P}(A) \leq \mathbb{P}\left(\bigcup_{i=n}^{\infty} A_i\right) \leq \sum_{i=n}^{\infty} \mathbb{P}(A_i) \tag{12.3}$$

Da $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$ gegen einen endlichen Wert konvergiert, konvergiert $\sum_{i=n}^{\infty} \mathbb{P}(A_i)$ gegen null, wenn $n \to \infty$. Daraus folgt $\mathbb{P}(A) = 0$ und somit (12.1).

Für (12.2) setzen wir $A := \limsup A_n$ und weiter

$$I_n := 1_{A_n}, \quad S_n := \sum_{j=1}^{n} I_j$$

und schließlich

$$S := \sum_{j=1}^{\infty} I_j.$$

Da die A_n als paarweise unabhängig vorausgesetzt sind, sind sie auch unkorreliert. Also folgt

$$\begin{aligned} \mathbb{V}(S_n) &= \sum_{j=1}^{n} \mathbb{V}(I_j) = \sum_{j=1}^{n} \left[\mathbb{E}\left(I_j^2\right) - \mathbb{E}(I_j)^2\right] \\ &= \mathbb{E}(S_n) - \sum_{j=1}^{n} \mathbb{E}(I_j)^2 \leq \mathbb{E}S_n, \end{aligned}$$

wobei wir in der letzten Gleichung Gebrauch davon gemacht haben, dass die I_j Indikatoren sind und daher $I_j^2 = I_j$ gilt. Laut Voraussetzung ist $\sum_{n=1}^{\infty} \mathbb{E}(I_n) = +\infty$. Da nun $S_n \uparrow S$ ist dies äquivalent mit

$$\lim_{n\to\infty} \mathbb{E}(S_n) = \mathbb{E}(S) = \infty \tag{12.4}$$

Andererseits ist $\omega \in A$ dann und nur dann, wenn $\omega \in A_n$ für unendlich viele n. Dies wiederum ist dann und nur dann der Fall, wenn $S(\omega) = +\infty$. Die Behauptung ist somit

$$\mathbb{P}(S = +\infty) = 1.$$

Dies sehen wir wie folgt ein. Nach der Chebyschevschen Ungleichung erhalten wir

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \le \eta) \ge 1 - \frac{\mathbb{V}(S_n)}{\eta^2}$$

für alle $\eta > 0$. Wegen (12.4) dürfen wir annehmen, dass $\mathbb{E}S_n > 0$ gilt und

$$\eta = \frac{1}{2}\mathbb{E}S_n$$

wählen. Also folgt

$$\mathbb{P}\left(S_n \ge \frac{1}{2}\mathbb{E}(S_n)\right) \ge \mathbb{P}\left(|S_n - \mathbb{E}S_n| \le \frac{1}{2}\mathbb{E}S_n\right) \ge 1 - 4\frac{\mathbb{V}(S_n)}{\mathbb{E}(S_n)^2}$$

Aber es ist

$$\mathbb{V}(S_n) \le E(S_n)$$

und $E(S_n) \to \infty$. Daher gilt

$$\lim \frac{\mathbb{V}(S_n)}{E(S_n)^2} = 0.$$

Somit haben wir für alle $\varepsilon > 0$ und alle n groß genug

$$\mathbb{P}\left(S_n \ge \frac{1}{2}\mathbb{E}S_n\right) \ge 1 - \varepsilon.$$

Nun aber ist $S \ge S_n$ und somit auch

$$\mathbb{P}\left(S \ge \frac{1}{2}ES_n\right) \ge 1 - \varepsilon$$

für alle $\varepsilon > 0$. Aber dies bedeutet

$$\mathbb{P}(S = +\infty) = 1,$$

was wir zeigen wollten. ■

Beispiel 12.3 *Es sei (X_n) eine Folge reellwertigerZufallsvariablen , die der Bedingung*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > \varepsilon) < +\infty \tag{12.5}$$

für alle $\varepsilon > 0$ genügt. Dann konvergiert

$$X_n \to 0 \qquad \mathbb{P} - f.s.$$

In der Tat, besagt das Borel–Cantelli–Lemma, dass (12.5)

$$\mathbb{P}(|X_n| > \varepsilon \textit{ unendlich oft in } n) = 0$$

impliziert. Dies aber ist ganz genau die Definition der, $\mathbb{P} - f.s.$–Konvergenz von X_n gegen 0.

Übung 12.4 *Ist* (12.5) *äquivalent zur $\mathbb{P}$–f.s. Konvergenz von X_n gegen 0?*

Hier ist nun die Übersetzung des Kolomogorovschen 0-1-Gesetzes auf Zufallsvariablen:

Theorem 12.5 *(Kolmogorovsches* 0–1*–Gesetz) Sei (X_n) eine Folge unabhängiger Zufallsvariablen mit Werten in beliebeigen messbaren Räumen. Dann gilt für jedes terminale Ereignis A, d.h. für jedes A mit*

$$A \in \bigcap_{n=1}^{\infty} \sigma(X_m, m \geq n),$$

dass $\mathbb{P}(A) \in \{0, 1\}$.

Beweis. Der Beweis folgt unmittelbar aus dem Kolmogorovsches 0–1–Gesetz für σ–AlgebrenX ■

Korollar 12.6 *Es sei* $(X_n)_{n\in\mathbb{N}}$ *eine Folge unabhängiger, reelvertiger Zufallsvariablen . Wir setzen*

$$\mathcal{T}_\infty := \bigcap_{m=1}^{\infty} \sigma\left(X_i, i \geq m\right)$$

als die Tail σ*–Algebra. Falls* T *eine reellwertigeZufallsvariable ist, die ferner als messbar bezüglich* $\mathcal{T}_\infty$ *angenommen wird, dass ist* T $\mathbb{P}$*–f.s. konstant, d.h. es gibt ein* $\alpha \in \mathbb{R}$ *so dass*

$$\mathbb{P}\left(T = \alpha\right) = 1.$$

Solche Zufallsvariablen $T : \Omega \to \mathbb{R}$*, die* $\mathcal{T}_\infty$*–messbar sind, heißen auch Tailfunktionen.*

Beweis. Für jedes $\gamma \in \bar{\mathbb{R}}$ gilt

$$\{T \leq \gamma\} \in \mathcal{T}_\infty.$$

Dies impliziert, dass

$$\mathbb{P}\left(T = \gamma\right) \in \{0, 1\}$$

gilt. Wählen wir $\gamma = +\infty$, so erhalten wir

$$\mathbb{P}\left(T \leq \gamma\right) = \mathbb{P}\left(\Omega\right) = 1.$$

Es sei $\alpha := \inf_{\gamma\in\bar{\mathbb{R}}} \{\mathbb{P}\left(T \leq \gamma\right) = 1\}$. Wir bezeichnen mit $\mathcal{C}$ die folgende Menge

$$\mathcal{C} := \{\gamma \in \mathbb{R} : \mathbb{P}\left(T \leq \gamma\right) = 1\}.$$

Dann gilt für einen geeignet gewählte fallende Folge $(\gamma_n) \in \mathcal{C}$, dass $\gamma_n \downarrow \alpha$ und da

$$\{T \leq \gamma_n\} \downarrow \{T \leq \alpha\}$$

gilt, folgt $\alpha \in \mathcal{C}$. Somit ist $\alpha = \min\{\gamma \in \mathcal{C}\}$. Daraus folgt

$$\mathbb{P}\left(T < \alpha\right) = 0,$$

was wiederum

$$\mathbb{P}\left(T = \alpha\right) = 1$$

impliziert ■

Beispiel 12.7 *Eine Münze wird unendlich oft geworfen. Dann tritt jede endliche Folge*

$$(\omega_1, ..., \omega_k), \omega_i \in \{K, Z\}, k \in \mathbb{N}$$

unendlich oft in den Beobachtungen $(X_i)_{i\in\mathbb{N}}$*auf. In der Tat: Definieren wir für* $n \in \mathbb{N}_0$

$$A_n := \{(X_{kn+1}, X_{k(n+1)}) = (\omega_1, ..., \omega_k)$$

so ist für jedes $n \in \mathbb{N}_0$

$$\mathbb{P}(A_n) \equiv \varrho \in (0, 1).$$

Somit ist

$$\sum_{n=0}^{\infty} \mathbb{P}(A_n) = \infty$$

und die Behauptung folgt aus dem Borel–cantelli–Lemma.

Übung 12.8 *Beweisen Sie den zweiten Teil des Borel–Cantelli–Lemas* (12.2) *unter der stärkeren Voraussetzung, dass die Ereignisse* $(A_n)_n$ *unabhängig sind. Folgen Sie dabei den folgenden Schritten:*

1. *Für jede Folge* (α_n) *reeller Zahlen mit* $0 \leq \alpha_n \leq 1$ *gilt*

$$\prod_{i=1}^{n} (1 - \alpha_i) \leq \exp\left(-\sum_{i=1}^{n} \alpha_i\right)$$

   Dies impliziert

$$\sum_{n=1}^{\infty} \alpha_n = \infty \Rightarrow \lim_{n\to\infty} \prod_{i=1}^{n} (1 - \alpha_i) = 0$$

2. *Für* $A := \limsup A_n$ *gilt*

$$\begin{aligned} 1 - \mathbb{P}(A) &= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) \\ &= \lim_{n\to\infty} \prod_{m=n}^{\infty} (1 - \mathbb{P}(A_m)). \end{aligned}$$

3. *Da* $\sum \mathbb{P}(A_n)$ *divergiert, gilt wegen 1.*

$$\lim_{N\to\infty} \prod_{m=n}^{\infty} (1 - \mathbb{P}(A_m)) = 0$$

und somit $\mathbb{P}(A) = 1$ *wegen 2.*

Wir gehen nun noch einmal auf das Kolmogorovsche 0–1–Gesetz zurück. Sind die zugrundeliegenden Zufallsvariablen nicht nur unabhängig, sondern auch identisch verteilt, so lässt sich ein stärkeres 0–1–Gesetz beweisen, das auf Hewitt und Savage zurückgeht. Es nutzt Symmetrieeigenschaften der zugrunde liegenden Folge von i.i.d. Zufallsvariablen aus.
Für die Formulierung benötigen wir eine Definition.

Definition 12.9 *Es sei* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ *eine Folge reellwertiger Zufallsvariablen über einem W-Raum* $(\Omega, \mathcal{F}, \mathbb{P})$.

- *Eine endliche Permutation* π *der natürlichen Zahlen ist ein* $\pi \in S(\mathbb{N})$ *mit* $\pi(n) = n$ *für alle* $n \geq n_0$ *und* n_0 *genügend groß.*

- *Für eine endliche Permutation* π *sei* $\mathbf{X}_\pi = (X_{\pi(n)})_{n\in\mathbb{N}}$.

- *Eine messbare, numerische Funktion* g *auf* $\overline{\mathbb{R}} \to \overline{\mathbb{R}}$ *(wobei beide Mengen mit ihren Borelschen* σ*–Algebren ausgestattet seien) heißt bezüglich* $\mathbf{X}$ *symmetrisch, falls für alle endlichen Permutationen* π *gilt*

$$g(\mathbf{X}_\pi) = g(\mathbf{X}).$$

- $A \in \overline{\mathcal{B}}^{\mathbb{N}}$ *heißt bezüglich* $\mathbf{X}$ *symmetrisch, falls der Indikator* 1_A *bezüglich* $\mathbf{X}$ *symmetrisch ist.*

- *Die* σ*-Algebra aller bezüglich* X *symmetrischen Mengen bezeichnen wir mit* $\mathcal{S}$.

- *Ist* $\mathbf{I} = (I_n)$ *die Identität auf* $\overline{\mathbb{R}}^{\mathbb{N}}$, *so ist jede bezüglich* $\mathbf{I}$ *symmetrische Menge* A *auch symmetrische bezüglich eines jeden anderen Zufallsvektors* $\mathbf{X}$. *In der Tat gilt für alle endlichen Permutationen* π

$$\{\mathbf{X}_\pi \in A\} = \{\mathbf{I}_\pi \circ \mathbf{X} \in A\} = \{\mathbf{X} \in \mathbf{I}_\pi^{-1} A\} = \{\mathbf{X} \in \mathbf{I}_{\pi^{-1}} A\} = \{\mathbf{X} \in A\}.$$

- *Die bezüglich* $\mathbf{I}$ *symmmetrsichen Mengen bezeichnen wir auch als universell symmetrisch; die* σ*-Algebra aller universell symmetrischen Mengen bezeichnen wir mit* $\mathcal{S}_u$.

Lemma 12.10 $\mathcal{S}$ *und* $\mathcal{S}_u$ *sind* σ*–Algebren, deren Urbilder* $\mathfrak{S} := \mathbf{X}^{-1}(\mathcal{S})$ *und* $\mathfrak{S}_u := \mathbf{X}^{-1}(\mathcal{S}_u)$ *unter* $\mathbf{X}$ *die* σ*-Algebra der terminalen Ereignisse* $\mathcal{T}_\infty$ *unter* $\mathbf{X}$ *enthalten, also*

$$\mathfrak{S} \supseteq \mathfrak{S}_u \supseteq \mathcal{T}_\infty.$$

Beweis. Dass $\mathcal{S}$ und $\mathcal{S}_u$ σ–Algebren sind, ist eine Übung.
Sei nun $A \in \mathcal{T}_\infty = \bigcap_{n=1}^\infty \sigma(X_n, X_{n+1}, \ldots)$. Somit existieren messbare Mengen $B \in \overline{\mathcal{B}}^{\mathbb{N}}$, für $n \geq 1$, so dass $A = \{(X_{n+k})_{k\geq 0} \in B_n\} = \{\mathbf{X} \in \mathbb{R}^{n-1} \times B_n\}$ für alle $n \geq 1$. D.h.

$$A = \bigcap_{k\geq n} \{\mathbf{X} \in \mathbb{R}^{k-1} \times B_k\} = \{\mathbf{X} \in \bigcap_{k\geq n} \mathbb{R}^{k-1} \times B_k\}$$

für jedes n. Nimmt man den Limes $n \to \infty$ ergibt sich

$$A = \liminf_{n\to\infty} \{\mathbf{X} \in \mathbb{R}^{n-1} \times B_n\} = \{\mathbf{X} \in \liminf_{n\to\infty} \mathbb{R}^{n-1} \times B_n\}.$$

Nun ist aber $\{\liminf_{n\to\infty} \mathbb{R}^{n-1} \times B_n\}$ eine universell symmetrische Menge, also $A \in \mathfrak{S}_u$. ■

Hier nun das angekündigte 0–1–Gesetz von Hewitt und Savage.

Theorem 12.11 (0–1–Gesetz von Hewitt und Savage) *Gegeben sei eine i.i.d. Folge* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$. *Dann gilt für jedes* $A \in \mathfrak{S}$, *dass*

$$\mathbb{P}(\mathbf{X}) \in \{0, 1\}.$$

Beweis. Wie beim Beweis des Kolmogorovschen 0–1–Gesetzes ist die Idee zu zeigen, dass für ein symmetrisches Ereignis A $\mathbb{P}(A) = \mathbb{P}(A)^2$ gilt.
Sei also $A \in \mathfrak{S}$. Da die σ-Algebra $\overline{\mathcal{B}}^{\mathbb{N}}$ von den Zylindermengen $\sigma(I_1, \ldots, I_n)$ erzeugt wird, kann man nach dem Approximationssatz für endliche Maße eine Folge $A_n \in \sigma(X_1, \ldots, X_n)$ finden mit

$$\mathbb{P}^{\mathbf{X}}(A \Delta A_n) \to 0 \text{ wenn } n \to \infty.$$

Hierbei bezeichnet $A \Delta B = (A \setminus B) \cup (B \setminus A)$ die symmetrische Differenz zwischen A und B. Insbesondere gilt also $\mathbb{P}^{\mathbf{X}}(A_n) \to \mathbb{P}(A)$. Die A_n sind von der Form

$$A_n = \{\omega : (\omega_1, \ldots, \omega_n) \in B_n\} \qquad \text{mit } B \in \mathcal{B}^n.$$

Wir betrachten nun die endliche Permutation π mit

$$\pi(j) = \begin{cases} j + n & \text{für } 1 \leq j \leq n \\ j - n & \text{für } n + 1 \leq j \leq 2n \\ j & \text{für } j \geq 2n + 1 \end{cases}$$

Bemerke, dass π^2 die Identität ergibt und dass mit $\mathbf{X}$ auch $\mathbf{X}_\pi$ i.i.d. Koordinaten hat. Aus letzterem folgt

$$\mathbb{P}^{\mathbf{X}}(A_n \Delta A) = \mathbb{P}^{\mathbf{X}_\pi}(A_n \Delta A).$$

Nun ist $\{\mathbf{X} \in A\} = \{\mathbf{X}_\pi \in A\}$, da A symmetrisch ist und

$$\{\mathbf{X}_\pi \in A_n\} = \{(X_{n+1}, \ldots, X_{2n}) \in B_n\} =: \{\mathbf{X} \in A'_n\},$$

wobei wir mit A'_n die Menge

$$A'_n := \{\omega : (\omega_{n+1}, \ldots, \omega_{2n}) \in B_n\}.$$

Damit ist

$$\{\mathbf{X}_\pi \in A_n \Delta A)\} = \{\mathbf{X} \in A'_n \Delta A)\}.$$

Also wegen $\mathbb{P}^{\mathbf{X}} = \mathbb{P}^{\mathbf{X}_\pi}$

$$\mathbb{P}^{\mathbf{X}}(A_n \Delta A) = \mathbb{P}^{\mathbf{X}}(A'_n \Delta A).$$

Hieraus folgt

$$\mathbb{P}^{\mathbf{X}}(A \Delta (A_n \cap A'_n)) \leq \mathbb{P}^{\mathbf{X}}(A_n \Delta A) + \mathbb{P}^{\mathbf{X}}(A'_n \Delta A) = 2\mathbb{P}^{\mathbf{X}}(A_n \Delta A) \to 0.$$

(Die rechte Seite geht nach der Kosntruktion von A_n und A'_n gegen 0.) Somit folgt

$$\mathbb{P}^{\mathbf{X}}(A \Delta (A_n \cap A'_n)) \to 0.$$

Dies bedeutet insbesondere auch

$$\mathbb{P}^{\mathbf{X}}(A) - \mathbb{P}^{\mathbf{X}}(A'_n \cap A) \to 0.$$

Andererseits gilt

$$\mathbb{P}^{\mathbf{X}}(A_n' \cap A_n) \to \mathbb{P}^{\mathbf{X}}(A_n)\mathbb{P}^{\mathbf{X}}(A_n')$$

und

$$\mathbb{P}^{\mathbf{X}}(A_n) = \mathbb{P}^{\mathbf{X}}(A_n') \to \mathbb{P}^{\mathbf{X}}(A).$$

Somit folgt $\mathbb{P}^{\mathbf{X}}(A) = \mathbb{P}^{\mathbf{X}}(A)^2$, was die Behauptung zeigt. ∎

Korollar 12.12 *Gegeben sei eine i.i.d. Folge* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$*. Dann ist jede bezüglich* $\mathbf{X}$ *symmetrische, reellwertige Funktion* g *auf* $\overline{\mathbb{R}}^{\mathbb{N}}$ $\mathbb{P}^{\mathbf{X}}$*–f.s. konstant und somit auch* $g \circ \mathbf{X}$ $\mathbb{P}^{\mathbf{X}}$*–f.s. konstant.*

Beweis. Für jedes $t \in \overline{\mathbb{R}}$ ist die Menge $\{g \le t\}$ symmetrisch bezüglich $\mathbf{X}$, denn für jede endliche Permutation π gilt

$$\{\mathbf{X} \in \{g \le t\}\} = \{g(\mathbf{X}) \le t\} = \{g(\mathbf{X}_\pi) \le t\} = \{\mathbf{X}_\pi \in \{g \le t\}\}.$$

Also ist $\mathbb{P}^{\mathbf{X}}(g \le t)$ für jedes t entweder 0 oder 1. Der Rest des Beweises folgt ebenso wie im Beweis von Korollar 12.6. ∎

Die folgenden beiden Sätze beleuchten besonders die Bedeutung des 0–1–Gesetzes von Hewitt und Savage.

Theorem 12.13 *Gegeben sei eine i.i.d. Folge* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ *mit Partialsummen* $S_n = \sum_{i=1}^n X_i$*,* $n \ge 1$*. Dann sind*

$$\limsup_{n\to\infty} S_n, \liminf_{n\to\infty} S_n \text{ und } \lim_{n\to\infty} S_n \text{ (falls existent)}$$

messbar bezüglich $\mathfrak{S}$ *und fast sicher konstant.*

Beweis. Offensichtlich sind die Funktionen

$$(x_n)_{n\ge 1} \mapsto \limsup_{n\to\infty} \sum_{k=1}^n x_k$$

und

$$(x_n)_{n\ge 1} \mapsto \liminf_{n\to\infty} \sum_{k=1}^n x_k$$

symmetrisch bezüglich $\mathbf{X}$. Die Behauptung folgt daher aus dem vorhergehenden Korollar. ∎

Zu beachten ist, dass die o.g. Funktionen nicht messbar sind bezüglich der terminalen σ-Algebra der Folge $\mathbf{X}$, man den Satz also nicht dem Kolmogorovschen 0–1–Gesetz ableiten kann.
Als Konsequenz ergibt sich eine vollständige Klassifikation des Partialsummenverhaltens für i.i.d. Zufallsgrößen.

Theorem 12.14 *Gegeben sei eine i.i.d. Folge* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ *mit Partialsummen* $S_n = \sum_{i=1}^n X_i$, $n \geq 1$. *Dann trifft stest eine der folgenden Alternativen zu:*

1. $S_n = 0$, $\mathbb{P}$*-f.s für alle* $n \geq 1$.
2. $\lim S_n = \infty$ $\mathbb{P}$*-f.s.*
3. $\lim S_n = -\infty$ $\mathbb{P}$*-f.s.*
4. $\limsup S_n = \infty$ $\mathbb{P}$*-f.s. und* $\liminf S_n = -\infty$ $\mathbb{P}$*-f.s.*

Beweis. Wenn die X_i $\mathbb{P}$-f.s. verschwindne liegt Alternative 1. vor. Ansonsten argumentieren wir wie folgt: Aus Satz 12.13 wissen wir schon, dass es ein $c \in \overline{\mathbb{R}}$ mit $\limsup S_n = c$ $\mathbb{P}$-f.s. gibt. Da $(X_{n+1})_{n\geq 1}$ dieselbe Verteilung hat wie (X_n), gilt auch $\limsup S_n - X_1 = c$ $\mathbb{P}$-f.s. Also

$$c = \limsup S_n = X_1 + \limsup(S_n - X_1) = X_1 + c \quad \mathbb{P} - f.s.$$

Da X_1 nicht fast sicher verschwindet, folgt $c \in \{-\infty, \infty\}$. Analog zeigt man, dass $\liminf S_n \in \{-\infty, \infty\}$ $\mathbb{P}$-f.s. Das beweist die Behauptung. ■
Natürlich stellt sich die Frage, wann welche der Alternativen gilt (wobei leicht einzusehen ist, dass 1. dann und nur dan gilt wenn X_1 fast sicher verschwindet). Unter der Voraussetzung, dass $\mathbb{E}X_1$ existiert, lässt sich dies beantworten. Der Beweis benutzt das im folgenden Kapitel zu zeigende Gesetz der großen Zahlen. Da wir den Satz jedoch (leider) soweiso nicht vollständig zeigen können, führen wir den Satz schon jetzt an.

Theorem 12.15 *Gegeben sei eine i.i.d. Folge* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ *mit Partialsummen* $S_n = \sum_{i=1}^n X_i$, $n \geq 1$. $\mu = \mathbb{E}X_1$ *existiere. Dann gilt*

1. $S_n = 0$, $\mathbb{P}$*-f.s für alle* $n \geq 1 \Leftrightarrow X_1 \equiv 0$ $\mathbb{P}$*-f.s.*
2. $\lim S_n = \infty$ $\mathbb{P}$*-f.s.* $\Leftrightarrow \mu > 0$.
3. $\lim S_n = -\infty$ $\mathbb{P}$*-f.s.*$\Leftrightarrow \mu < 0$.

4. $\limsup S_n = \infty$ $\mathbb{P}$*-f.s. und* $\liminf S_n = -\infty$ $\mathbb{P}$*-f.s.*$\Leftrightarrow \mu = 0$ *und* $\mathbb{P}(X_1 = 0) < 1$.

Beweis. 1. is klar
2. und 3. folgt aus dem noch zu zeigenden starken Gesetz der großen Zahlen
4. Ist das sogenannte Chung–Fuchs–Theorem (siehe z.B. [?], Satz 2.2.7.), das hier nicht bewiesen werden kann.
■

13 Gesetze der großen Zahlen

Das zentrale Ziel der Wahrscheinlichkeitstheorie kann beschrieben werden als das asymptotische Verhalten von Folgen Zufallsvariablen zu analysieren. Diese Analyse kann auf mannigfache Weise geschehen: eine erste Form haben wir schon im vorhergehenden Kapitel gesehen, eine andere in der Vorlesung über Stochastik.
Das von dort bekannte Gesetz der großen Zahlen werden wir in einem ersten Schritt etwas verallgemeinern.

Theorem 13.1 *[Khintchines schwaches Gesetz der großen Zahlen] Es sei* $(X_n)_{n\in\mathbb{N}}$ *eine Folge integrierbarer, reellwertiger Zufallsvariablen , die paarweise unkorreliert sind. Angenommen, dass*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}(X_i) = 0$$

gilt was natürlich die quadratische Integrierbarkeit der X_i *impliziert). Dann gilt das **schwache Gesetz der großen Zahlen** , d.h.*

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\mathbb{E}\sum_{i=1}^{n} X_i \right| > \varepsilon \right) = 0$$

für alle $\varepsilon > 0$.

Beweis. Gemäß der Tschebyscheffschen Ungleichung gilt für $\varepsilon > 0$:

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right| > \varepsilon\right) &\leq \frac{1}{\varepsilon^2}\mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right) \\
&= \frac{1}{\varepsilon^2}\frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right) \\
&= \frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i - \mathbb{E}X_i) \\
&= \frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i).
\end{aligned}$$

Hierbei haben wir die Annahme, dass die vorgelegten Zufallsvariablen paarweise unkorreliert sind ausgenutzt. Nach Voraussetzung konvergiert der rechtsstehende Ausdruck gegen 0. ■

Bemerkung 13.2 *Wie wir in diesem Kapitel lernen werden, kann die Voraussetzung der Quadratintegrierbarkeit für Folgen von i.i.d. Zufallsvariablen fallengelassen werden.*

Theorem 13.1 ist eine befriedigende Aussage. In eine gewissen Sinne gilt, was wir intuitiv schon zu wissen meinen: die normierten Partialsummen geeigneter Zufallsvariablen konvergieren gegen ihren Erwartungswert. Allerdings würde man Konvergenz eigentlich als

$$\mathbb{P}(\lim \frac{1}{n}S_n \text{ existiert und ist gleich } \mathbb{E}(\frac{1}{n}S_n)) = 1$$

formulieren, also als fast sichere Konvergenz. Dass dies für Folgen von i.i.d. Zufallsvariablen in der Tat wahr ist, wird in der Folge gezeigt werden. Ein solcher Satz heißt dann ein **starkes Gesetz der großen Zahlen**. Es wurde erstmals (unter stärkeren als den folgenden Bedingungen) 1930 von Kolomogorov bewiesen. ir folgen einem jüngeren und sehr eleganten Beweis von Etemadi aus dem Jahr 1981.

Theorem 13.3 (Starkes Gesetz der großen Zahlen – (Etemadi 1981)) *Für jede Folge reellwertiger, paarweise unabhängiger identisch verteilter Zufallsvariablen mit existierendem Erwartungswert gilt das starkes Gesetz der* *großen Zahlen, d.h.*

$$\mathbb{P}\left(\limsup\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X_1\right| > \varepsilon\right) = 0$$

für jedes $\varepsilon > 0$.

Als unmittelbare Folgerung leiten wir das starke Gesetz der großen Zahlen von Kolomogorov ab:

Korollar 13.4 (Starkes Gesetz d. großen Zahlen (Kolomogorov 1930)) *Für jede Folge reellwertiger, i.i.d. Zufallsvariablen mit existierendem Erwartungswert gilt das starkes Gesetz der großen Zahlen.*

Bevor wir uns an den Beweis von Theorem 13.3 machen, ein paar einleitende Bemerkungen, die die Struktur des folgenden ein wenig erhellen sollen.

1. Bezeichne mit $S_n = \sum_{i=1}^n X_i$. Dann behauptet Theorem 13.3, dass $\frac{S_n}{n} \to \eta =: \mathbb{E}X_1$ $\mathbb{P}$–f.s.

2. Mit X_n genügen auch X_n^+ und X_n^- (wobei wir mit $X_n^+ = \max(X_n, 0)$ und $X_n^- = (-X_n)^+$) den Voraussetzungen von Theorem 13.3. Da $X_n = X_n^+ - X_n^-$ gilt, genügt es daher das Starke Gesetz der großen Zahlen Theorem 13.3 für positive Zufallsvariablen zu beweisen. Wir nehmen daher in der Folge $X_n > 0$ an.

3. Alle Beweise des Starken Gesetzes der großen Zahlen, benutzen eine Abschneidetechnik. Wir schneiden, daher X_n dort ab, wo es zu groß ist. Dies geschieht im wesentlichen deshalb, da die einzige vernünftige Kontrolle, die wir über das abweichende Verhalten von $\frac{1}{n}\sum_{i=1}^n X_i$ erhalten können, die Tschebyscheffsche Ungleichung ist. Da wir aber noch nicht einmal die Existenz eines zweiten Moments voraussetzen, müssen wir es uns künstlich verschaffen. Wor führen daher

   $$Y_n := X_n 1_{\{|X_n|<n\}} = X_n 1_{\{X_n<n\}}$$

   ein.

   Natürlich gilt, wenn μ die Verteilung von X_n und μ_n die Verteilung von X_n, dass dann $\mu_n \neq \mu$. In der Tat ist ja $\mu_n = f_n(\mu)$, wobei wir

   $$f_n(x) := \begin{cases} x & \text{falls } 0 \leq x < n \\ 0 & \text{sonst} \end{cases}$$

definieren. Die so definierten Y_n haben natürlich endliche Varianz:

$$\mathbb{E}\left(Y_n^2\right) = \mathbb{E}\left(f_n^2 \circ X_n\right) = \int f_n^2 d\mu = \int_0^n x^2 d\mu < \infty.$$

4. Natürlich müssen wir jede Information, die wir über die Folge der Y_n erhalten wieder zurückübersetzen in Aussagen über die we X_n. Zu diesem Zwecke wenden wir das Borel–Cantelli Lemma an und zeigen, dass

$$\sum_{n=1}^{\infty} \mathbb{P}\left(X_n \neq Y_n\right) < \infty$$

gilt. Dies impliziert, dass $X_n \neq Y_n$ höchstens endlich oft geschehen kann mit Wahrscheinlichkeit eins. Insbensondere mitteln sich die endlich vielen Male bei der Mittelwertbildung aus, d.h., wenn wir zeigen können, dass $\frac{1}{n}\sum_{i=1}^{n} Y_i \to \eta \quad \mathbb{P}-f.s.$, so haben wir auch gezeigt, dass $\frac{1}{n}\sum_{i=1}^{n} X_i \to \eta \quad \mathbb{P}-f.s.$ gilt.

5. Zum Zwecke des Beweises sei noch das folgende bemerkt: Sei $\alpha > 1$ und für $n \in \mathbb{N}$ sei

$$k_n := [\alpha^n]$$

die größte ganze Zahl, die $\leq \alpha^n$ ist. Somit ist $k_n \in \mathbb{N}$ und

$$k_n \leq \alpha^n < k_n + 1.$$

Da

$$\lim_{n\to\infty} \frac{\alpha^n - 1}{\alpha^n} = 1$$

gilt, gibt es ein $c_\alpha, 0 < c_\alpha < 1$ so dass

$$k_n > \alpha^n - 1 \geq c_\alpha \alpha^n \quad \text{für alle } n \in \mathbb{N}. \tag{13.1}$$

Wir wenden uns nun dem Beweis des starken Gesetzes der großen Zahlen zu.

Beweis von Theorem 13.3.
Der Beweis des starken Gesetzes der großen Zahlen erfolgt in mehreren Schritten:
Schritt 1: O.B.d.A werden wir (wie oben erwähnt) annehmen, dass $X_n > 0$. Wir definieren die trunkierten (abgeschnittenen, gekappten) Versionen der

X_n vermöge $Y_n := 1_{\{X_n<n\}} X_n$. Dann erben die Y_n von den X_n die paarweise Unabhängikeit. Weiter sind sie quadratisch integrierbar. Wir setzen

$$S_n^1 := \sum_{i=1}^{n} (Y_i - \mathbb{E}Y_i) \, .$$

Sei nun $\varepsilon > 0$. Mit Hilfe der Chebyshevschen Ungleichung und der Unabhängigkeit der Y_n erhalten wir für deren empirische Mittwelwerte

$$\mathbb{P}\left(\left|\frac{1}{n}S_n^1\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} V\left(\frac{1}{n}S_n^1\right) = \frac{1}{\varepsilon^2}\frac{1}{n^2} V\left(S_n^1\right) = \frac{1}{n^2}\frac{1}{\varepsilon^2}\sum_{i=1}^{n} V\left(Y_i\right) .$$

Unter Beachtung von $V(Y_i) = \mathbb{E}(Y_i^2) - (\mathbb{E}(Y_i))^2 \leq \mathbb{E}(Y_i^2)$ gelangen wir daher zu

$$\mathbb{P}\left(\left|\frac{1}{n}S_n^1\right| > \varepsilon\right) \leq \frac{1}{n^2}\frac{1}{\varepsilon^2}\sum_{i=1}^{n} E\left(Y_i^2\right) .$$

Sei nun $\alpha > 1$ der Parameter aus Anmerkung 5. oben. Für $k_n = [\alpha^n]$ ergibt dann die obige Rechnung

$$\mathbb{P}\left(\left|\frac{1}{k_n}S_{k_n}^1\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2 k_n^2}\sum_{i=1}^{k_n} \mathbb{E}\left(Y_i^2\right)$$

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

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{1}{k_n}S_{k_n}^1\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\sum_{i=1}^{k_n} \frac{1}{k_n^2}\mathbb{E}\left(Y_i^2\right) .$$

Nun ändern wir auf der rechten Seite die Summationsreihenfiolge wie folgt: Man kann die Doppelsumme rechts als eine Summe über sämtliche Zeilensummen einer $\mathbb{N}\times\mathbb{N}$-Matrix auffassen, wobei in der n-ten Zeile nur die ersten k_n ersten Glieder verschieden sind von Null (nämlich $\frac{1}{k_n^2}\mathbb{E}(Y_1{}^2), \ldots, \frac{1}{k_n^2}\mathbb{E}(Y_{k_n}^2)$). Summieren wir stattdessen über die Spaltensummen (dies ist erlaubt und liefert dasselbe Ergbenis, weil die Matrix nur nicht–negative Einträge besitzt), erhalten wir

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{1}{k_n}S_{k_n}^1\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sum_{j=1}^{\infty} t_j \mathbb{E}\left(Y_j^2\right)$$

wobei

$$t_j := \sum_{n=n_j}^{\infty} \frac{1}{k_n^2}$$

ist und n_j das kleinste n mit $k_n \geq j$ ist. Aus (13.1) erhalten wir

$$t_j \leq \frac{1}{c_\alpha^2} \sum_{n=n_j}^{\infty} \frac{1}{\alpha_n^{2n}} = \frac{1}{c_\alpha^2} \alpha^{-2n_j} \frac{1}{1-\frac{1}{\alpha^2}} = d_\alpha \alpha^{-2n_j}$$

wobei

$$d_\alpha := \frac{1}{c_\alpha^2} \frac{1}{(1-\alpha^{-2})} > 0$$

gesetzt ist. Hieraus folgt

$$t_j \leq d_\alpha j^{-2}$$

da ja $\alpha^{n_j} \geq k_{n_j} \geq j$ gilt. Berechnet man die Erwartungswerte der Quadrate der Y_j wie in der Vorbemerkung, erhält man

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{1}{k_n} S_{k_n}^{1}\right| > \varepsilon\right) \leq \frac{d_\alpha}{\varepsilon^2} \sum_{j=1}^{\infty} \frac{1}{j^2} \sum_{k=1}^{j} \int_{k-1}^{k} x^2 d\mu.$$

Vertauschen wir wieder die Summationsreihenfolge wie oben, gelangen wir zu

$$\sum_{j=1}^{\infty} \frac{1}{j^2} \sum_{k=1}^{j} \int_{k-1}^{k} x^2 d\mu = \sum_{k=1}^{\infty} \left(\sum_{j=k}^{\infty} \frac{1}{j^2}\right) \int_{k-1}^{k} x^2 d\mu.$$

Da nun

$$\begin{aligned}
\sum_{j=k}^{\infty} \frac{1}{j^2} &< \frac{1}{k^2} + \frac{1}{k(k+1)} + \frac{1}{(k+1)(k+2)} + \ldots \\
&= \frac{1}{k^2} + \left(\frac{1}{k} - \frac{1}{k+1}\right) + \left(\frac{1}{k+1} - \frac{1}{k+2}\right) + \ldots = \frac{1}{k^2} + \frac{1}{k} \leq \frac{2}{k},
\end{aligned}$$

gilt, bekommen wir schließlich die folgende Abschätzung

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{1}{k_n} S_{k_n}^{1}\right| > \varepsilon\right) \leq \frac{2d_\alpha}{\varepsilon^2} \sum_{k=1}^{\infty} \int_{k-1}^{k} \frac{x}{k} x d\mu \leq \frac{2d_\alpha}{\varepsilon^2} \sum_{k=1}^{\infty} \int_{k-1}^{k} x d\mu = \frac{2d_\alpha}{\varepsilon^2} E(X_1) < \infty.$$

Wir können somit das Borel–Cantelli Lemma einsetzen und erhalten

$$\mathbb{P}\left(\left|\frac{1}{k_n}S^{1}_{k_n}\right| > \varepsilon \text{ unendlich oft in } n\right) = 0.$$

Dies aber ist äquivalent zur fast sicheren Konvergenz von

$$\lim_{n\to\infty}\frac{1}{k_n}S^{1}_{k_n} = 0 \qquad \mathbb{P}-f.s. \tag{13.2}$$

Schritt 2: Nun ,da wir die Konvergenz von $\frac{1}{k_n}S^{1}_{k_n}$ festgestellt haben, vermuten wir natürlich, dass dies auch die Konvergenz von $\frac{1}{k_n}\sum_{i=1}^{k_n}Y_i$ impliziert. Wir werden in diesem Schritt lernen, dass $\frac{1}{k_n}\sum_{i=1}^{k_n}Y_i$ nur gegen $\mathbb{E}(X_1)$ konvergieren kann.
Definitionsgemäß gilt für Y_n

$$\mathbb{E}(Y_n) = \int x d\mu_n(x) = \int_0^n x d\mu(x).$$

Wenden wir den Satz von der monotonen Konvergenz hierauf an, sehen wir dass

$$\mathbb{E}(X_1) = \lim_{n\to\infty}\mathbb{E}(Y_n)$$

gilt. Nach einer Übungsaufgabe aus der Analysis (siehe Übung 13.6 unten) folgt hieraus

$$\mathbb{E}(X_1) = \lim_{n\to\infty}\frac{1}{n}(\mathbb{E}Y_1 + ... + \mathbb{E}Y_n). \tag{13.3}$$

Setzen wir nun die Definition der Summen S^{1}_{n} ein, ergibt das

$$\frac{1}{k_n}S^{1}_{k_n} = \frac{1}{k_n}\sum_{i=1}^{k_n}Y_i - \frac{1}{k_n}\sum_{i=1}^{k_n}E(Y_i),$$

(13.2) und (13.3) zusammen ergeben

$$\lim_{n\to\infty}\frac{1}{k_n}\sum_{i=1}^{k_n}Y_i = \lim_{n\to\infty}\frac{1}{k_n}\sum_{i=1}^{k_n}\mathbb{E}Y_i = \mathbb{E}X_1,$$

was wir in diesem Schritt zeigen wollten.

Schritt 3: Nun wollen wir die Trunkierung (also die Kappung) der X_n loswerden. Betrachten wir die folgende Reihe:

$$\sum_{n=1}^{\infty} \mathbb{P}\left(X_n \neq Y_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(X_n \geq n\right) = \sum_{n=1}^{\infty} \mu\left([n, \infty)\right) = \sum_{n=1}^{\infty} \sum_{i=n}^{\infty} \mu\left([i, i+1)\right).$$

Vertauschen wir (wieder einmal) die Summationsreihenfolge, ergibt sich aus dieser Rechnung

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}\left(X_n \neq Y_n\right) &= \sum_{i=1}^{\infty} i\mu([i, i+1)] = \sum_{i=1}^{\infty} \int_i^{i+1} id\mu \\ &\leq \sum_{i=1}^{\infty} \int_i^{i+1} xd\mu = \int_1^{\infty} xd\mu \leq \int_0^{\infty} xd\mu = \mathbb{E}\left(X_1\right) < \infty. \end{aligned}$$

Mit Hilfe des Borel–Cantelli Lemmas impliziert dies

$$\mathbb{P}\left(X_n \neq Y_n \text{ unendlich oft in } n\right) = 0.$$

Also gibt es ein (zufälliges) n_0, so dass mit Wahrscheinlichkeit eins $X_n = Y_n$ für alle $n \geq n_0$ gilt. Aber diese endlichen vielen n, für die X_n und Y_n verschieden sind, fallen weg, wenn man Mittwelwerte bildet und den Limes $n \to \infty$ nimmt. Also gilt

$$\lim_{n \to \infty} \frac{1}{k_n} S_{k_n} = \mathbb{E}X_1 \qquad \mathbb{P}-f.s.$$

Schritt 4: Schließlich wollen wir noch zeigen, dass das Gesetz der großen Zahlen nicht nur für Teilfolgen gilt, sondern auch beim Übergang zur ganzen Folge erhalten bleibt.

Für festes $\alpha > 1$ sind die Folgen $\left(k_n\right)_n$ natürlich auch fest und divergieren gegen $+\infty$. Somit gibt es für jedes $m \in \mathbb{N}$ ein $n \in \mathbb{N}$, so dass

$$k_n < m \leq k_{n+1}.$$

Nach Voraussetzung waren die X_i nicht–negativ; hier aus folgt

$$S_{k_n} \leq S_m \leq S_{k_{n+1}}.$$

Also gilt
$$\frac{S_{k_n}}{k_n} \cdot \frac{k_n}{m} \leq \frac{S_m}{m} \leq \frac{S_{k_{n+1}}}{k_{n+1}} \cdot \frac{k_{n+1}}{m}.$$
Aus der Definition der k_n folgt nun für alle $n \in \mathbb{N}$
$$k_n \leq \alpha^n < k_n + 1 \leq m \leq k_{n+1} \leq \alpha^{n+1}.$$
Dies ergibt
$$\frac{k_{n+1}}{m} < \frac{\alpha^{n+1}}{\alpha^n} = \alpha$$
und
$$\frac{k_n}{m} > \frac{\alpha^n - 1}{\alpha^{n+1}}.$$
Nun gilt für alle $n \geq n_1 = n_1(\alpha)$, dass $\alpha^n - 1 \geq \alpha^{n-1}$ ist. Ist daher m genügend groß, etwa $m \geq k_{n_1}$ (und somit auch $n \geq n_1$), erhalten wir
$$\frac{k_n}{m} > \frac{\alpha^n - 1}{\alpha^{n+1}} > \frac{\alpha^{n-1}}{\alpha^{n+1}} = \frac{1}{\alpha^2}$$
Nun gibt es nach dem 3. Beweisschritt für jedes $\alpha > 1$ eine Menge Ω_α mit voller Wahrscheinlichkeit $\mathbb{P}(\Omega_\alpha) = 1$, auf der gilt
$$\lim \frac{1}{k_n} S_{k_n}(\omega) = \mathbb{E}X_1 \text{ für alle } \omega \in \Omega_\alpha.$$
Ohne Beschränkung der Allgemeinheit können wir nun annehmen, dass die X_i nicht $\mathbb{P}$-fast sicher identisch gleich null sind, denn sonst ist die Aussage des Starken Gesetzes der großen Zahlen natürlich trivialerweise erfüllt. Dies aber bedeutet, dass wir o.B.d.A annehmen können, dass $\mathbb{E}X_1 > 0$ gilt. Da nun $\alpha > 1$ gewählt war, gilt dann
$$\frac{1}{\alpha}\mathbb{E}X_1 < \frac{1}{k_n} S_{k_n}(\omega) < \alpha \mathbb{E}X_1$$
für alle $\omega \in \Omega_\alpha$ und alle genügend großen m. Für solche m und ω bedeutet dies nichts anderes als
$$\left(\alpha^{-3} - 1\right) \mathbb{E}X_1 < \frac{1}{m} S_m(\omega) - \mathbb{E}X_1 < \left(\alpha^2 - 1\right) \mathbb{E}X_1.$$
Definieren wir nun
$$\Omega_1 := \bigcap_{n=1}^{\infty} \Omega_{1+\frac{1}{n}},$$

so ist natürlich $\mathbb{P}(\Omega_1) = 1$ (das Komplement liegt in einer abzählbare Vereinigung von Nullmengen) und

$$\lim_{m\to\infty} \frac{1}{m} S_m(\omega) = \mathbb{E}X_1$$

für alle $\omega \in \Omega_1$. Dies beweist das Starke Gesetz der großen Zahlen. ■

Bemerkung 13.5 *Theorem 13.3 besagt insbesondere, dass für i.i.d. Folgen reelwertiger Zufallsvariablen (X_n) mit existierendem Erartungswert das Starke Gesetz der großen Zahlen gilt. Da $\mathbb{P}$–fast sichere Konvergenz stochastische Konvergenz impliziert, gilt Theorem 13.1 – das schwache Gesetz der großen Zahlen – auch für solche Folgen. Die dort erhobenen Forderung der endlichen zweiten Momente ist somit nicht notwendig.*

Übung 13.6 *Für jedes $n \in \mathbb{N}$ sei $(a_{n_m})_m$ eine Folge reeller Zahlen, so dass $\lim_{m\to\infty} a_{n_m} = a$ für alle $n \in \mathbb{N}$. Man zeige, dass dies auch impliziert, dass ihr Cesaro–Mittwelwert existiert und*

$$\lim_{n\to\infty} \frac{1}{n}(a_{n_1} + a_{n_2} + ... + a_{n_n}) = a.$$

erfüllt.

Übung 13.7 *Beweisen Sie das Starke Gestz der großen Zahlen für Folgen von i.i.d. Zufallsvariablen $(X_n)_n$ mit endlichem vierten Moment, .d..h für Zufallsvariablen mit $\mathbb{E}(X_1^4) < \infty$.*
Benutzen Sie dabie ***nicht*** *die Aussage von Theorem 13.3.*

Man kann sich natürlich fragen, ob die Forderung eines endlichen ersten Moments in Theorem 13.3 notwendiog ist. Dass man darauf nicht verzichten kann, lehrt der folgende Satz

Theorem 13.8 *Sei $(X_n)_{n\geq 1}$ eine Folge reellwertiger, identisch verteilter und paarweise unabhängiger Zufallsvariablen . Konvergiert die Folge*

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)_{n\geq 1}$$

für $n \to \infty$ $\mathbb{P}$–fast sicher gegen eine Zufallsvariable Y, so ist jedes X_n integrierbar und Y $\mathbb{P}$–fast sicher konstant, nämlich

$$Y \equiv \mathbb{E}X_1 \qquad \mathbb{P} - f.s.$$

Beweis. Wir setzen

$$S_n^* := \frac{1}{n}(X_1 + \ldots + X_n).$$

Nach Voraussetzung konvergiert S_n^* fast sicher gegen Y und somit die Folge

$$\frac{1}{n}X_n = S_n^* - \frac{n-1}{n}S_{n-1}^*$$

fast sicher gegen 0. Daher kann nur für endlich viele n $X_n \geq n$ gelten. Definiert man also die Ereignisse $C_n := \{\omega : |X_n(\omega)| \geq n\}$, so gilt

$$\mathbb{P}(\limsup_{n\to\infty} C_n) = 0.$$

Für $m \neq n$ sind C_m und C_n nach Voraussetzung unabhängig. Damit ist (12.2) des Borel–Cantelli Lemmas anwendbar (genauer die Kontraposition), die besagt, dass dann $\sum \mathbb{P}(C_n) < \infty$ gilt. Nun sind die X_n identisch verteilt, es gilt also auch

$$\mathbb{P}(C_n) := \mathbb{P}(\{|X_k| \geq n\})$$

und somit konvergiert auch die Reihe

$$\sum_{n=1}^{\infty} \mathbb{P}(\{|X_k| \geq n\}) < \infty.$$

Nun ist aber für jedes $k \in \mathbb{N}$

$$\sum_{n=1}^{\infty} \mathbb{P}(\{|X_k| \geq n\}) \leq \mathbb{E}(|X_k|) \leq 1 + \sum_{n=1}^{\infty} \mathbb{P}(\{|X_k| \geq n\})$$

also existiert $\mathbb{E}X_k$ für jedes $k \in \mathbb{N}$. Dann aber folgt mit dem Starken Gesetz der großen Zahlen 13.3, dass

$$\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}X_1 \qquad \mathbb{P} - f.s.$$

gilt. Dies war die Behauptung. ∎

Bemerkung 13.9 *Eine natürliche Frage, die man natürlicherweise bei allen Grenzwertsätzen, so auch bei Theorem 13.3 stellt, ist die nach der Konvergenzgeschwindigkeit. Unter geeigneten Vorausstezungen ist diese in den Gesetzen der großen Zahlen exponentiell. Was diese Voraussetzungen sind und was genau die exponentielle Konvergenz bedeutet, klärt ein sogenanntes Prinzip der großen Abweichungen. Es soll in einem gesonderten Kapotel untersucht werden.*

Am Ende des Kapitels geben wir noch eine bekannte Anwendung des Starken Gesetzes der großen Zahlen, das von eigenständigem Interesse finden. Wir begeben uns hierfür auf das Gebiet der Zahlentheorie.

Zu diesem Zwecke sei $(\Omega, \mathcal{F}, \mathbb{P})$ gegeben durch $\Omega = [0,1), \mathcal{F} = \mathcal{B}^1 \left|\Omega\right.$ und $\mathbb{P} = \lambda^1 \left|\Omega\right.$ sei das auf das Intervall $[0,1]$ eingeschränkte Lebesguemaß.

Für jede Zahl $\omega \in \Omega$ betrachten wir ihre g–adische Entwicklung

$$\omega = \sum_{n=1}^{\infty} \xi_n g^{-n} \tag{13.4}$$

Hierfür ist $g \geq 2$ eine natürliche Zahl und die Ziffern ξ_n erfüllen $\xi_n \in \{0, ..., g-1\}$. Diese Darstellung ist eindeutig, wenn wir fordern dass nicht alle bis auf endlich viele Ziffern ξ_n gleich $g-1$ sein dürfen (dies ist aber für das Folgende beinahe gleichgültig, denn nur für eine Lebesgue–Nullmenge von Zahlen aus dem Einheitsintervall gibt es mehr als eine g–adische Darstellung). Für jedes $\varepsilon \in \{0, ..., g-1\}$ sei $S_n^{\varepsilon,g}(\omega)$ die Anzahl aller Indizes $i \in \{1, ..., n\}$ mit $\xi_i(\omega) = \varepsilon$ in der g–adischen Darstellung der Zahl ω (13.4). Eine Zahl $\omega \in [0,1)$ heiße g–**normal**, falls

$$\lim_{n\to\infty} \frac{1}{n} S_n^{\varepsilon,g}(\omega) = \frac{1}{g}$$

für alle $\varepsilon = 0, ..., g-1$ gilt. Somit ist eine Zahl ω genau dann g-**normal**, wenn auf lange Sicht all ihre Ziffern mit derselben relativen Häufigkeit in ihrer g–adischen Darstellung auftreten (dies ist dan automatisch die Gleichverteilung auf den Ziffern). Eine Zahl $\omega \in [0,1)$ heißt **absolut normal**, falls ω g–normal ist, für alle $g \in \mathbb{N}, g \geq 2$.
Wählt man nun eine Zahl $\omega \in [0,1)$ gemäß dem Lebesguemaß, dann sind die Ziffern $\xi_i(\omega)$ i.i.d. Zufallsvariablen mit der uniformen Verteilung auf der Menge der zulässigen Zeichen $\{0, \ldots, g-1\}$. Wir machen uns das am Beispiel $g = 2$ klar. Dann ist nämlich die Menge aller Zahlen mit einer 0 als n'ter Ziffer, also mit $\xi_n(\omega) = 0$, gerade die Menge

$$A_n := [0, \frac{1}{2^n}) \cup [\frac{2}{2^n}, \frac{3}{2^n}) \cup \ldots \ldots \cup [\frac{2^n-2}{2^n}, \frac{2^n-1}{2^n}).$$

Offenbar gilt

$$\mathbb{P}(A_n) = \frac{1}{2}$$

für alle $n \in \mathbb{N}$. Wie man sich leicht klar macht ist dann

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n}) = \frac{1}{2}\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_{n-1}})$$

also

$$\begin{aligned} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n}) &= \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_{n-1}})\mathbb{P}(A_{i_n}) = \ldots \\ &= \mathbb{P}(A_{i_1}) \cdot \ldots \cdot \mathbb{P}(A_{i_n}) \end{aligned}$$

für endlich viele paarweise verschieden Indizes $i_1, \ldots, i_n$. Da dies auch für die Komplemente der A_{i_k} gilt, folgt die Unabhängigkeit. Der Fall allgemeiner g geht analog.
Somit sind die folgendermaßen definierten Zufallsvariablen

$$X_n(\omega) = \begin{cases} 1 & \text{falls } \xi_n(\omega) = \varepsilon \\ 0 & \text{sonst} \end{cases}$$

i.i.d. für jedes $g \geq 2$ i.i.d. Zufallsvariablen . Ferner gilt

$$S_n^{\varepsilon,g}(\omega) = \sum_{i=1}^{n} X_i^{\varepsilon,g}(\omega).$$

Gemäß dem Starken Gesetz der großen Zahlen (Theorem 13.3) folgt

$$\frac{1}{n} S_n^{\varepsilon,g}(\omega) \to \mathbb{E}\,(X_1^{\varepsilon,g}) = \frac{1}{g} \qquad \lambda - f.s.$$

für alle $\varepsilon \in \{0, ..., g-1\}$ und alle $g \geq 2$. Somit ist also λ–fast jede Zahl ω g–normal. Mit anderen Worten gibt es eine Nullmenge N_g mit $\lambda\,(N_g) = 0$, so dass ω g–normal ist für alle $\omega \in N_g^c$.
Nun ist aber auch

$$N := \bigcup_{g=\alpha}^{\infty} N_g$$

als Vereinigung abzählbar vieler Nullmengen Lebesgue Nullmenge. Außerhalb von N sind alle Zahlen absolut normal. Mit anderen Worten:

Theorem 13.10 (E. Borel) *λ–fast jede Zahl $\omega \in [0, 1)$ is absolut normal.*

Das wirklich Überraschende an diesem Resultat ist, dass kaum normale Zahlen bekannt sind. Champernowne hat 1933 gezeigt, dass die Zahl

$$\omega = 0,1234567891011121214...$$

in einem weiteren Sinne 10–normal ist, nämlich, dass nicht nur alle Ziffern ,sondern dass auch alle Ziffernblöcke endlicher Länge asymptotisch eine relative Häufigkeit 10^{-L} haben, wobei L die Länge des Ziffernblocks ist. Ob "natürliche Kandidaten" wie $\sqrt{2}, \log 2, e$ oder π (bzw. die auf (0,1) eingeschränkten Versionen) normal bezüglich irgendeiner Basis sind, ist nicht bekannt. Absolut normale Zahlen sind nur ganz wenige bekannt.

14 Große Abweichungen

Schon im vorigen Kapitel hatten wir die Frage nach der Konvergenzgeschwindigkeit in den Gesetzen der großen Zahlen gestellt. Diese soll in diesem Kapitel unter geeigneten Voraussetzungen beantwortet werden. Dies führt zu dem sogenannten satz von Cramér, den dieser 1938 bewies. Es ist der erste (mathematische) Fall eines Prinzips der großen Abweichungen (physikalisch kann das Boltzmannsche Gesetz $S = k \log W$ als ein Prinzip der großen Abweichungen ansehen).
Wir werden von nun an annehmen, dass die vorgelegten Zufallsvariablen $X_1, X_2, \ldots$ reellwertige, i.i.d. Zufallsvariablen sind mit einer endlichen momenterzeugenden Funktion

$$\varphi(t) := \mathbb{E}e^{tX_1} < \infty \qquad \text{für alle } t \in \mathbb{R} \tag{14.1}$$

(dass diese wichtig ist, wenn man exponentielle Konvergenz im Gesetz der großen Zahlen zeigen will, liegt auf der Hand, wenn man sich vor Augen führt, dass für so eine Konvergenzgeschwindigkeit in der Herleitung des Schwachen Gesetzes der großen Zahlen am besten die gewöhnliche (quadratische) Chebyshev–Ungleichung durch eine exponentielle ersetzt wird; Bei dieser taucht dann automatisch die momenterzeugende Funktion (14.1) auf). Wir werden uns in diesem Kapitel vor allem damit beschäftigen, den Zusammenhang zwischen $\varphi(a)$ und

$$-I(a) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{S_n}{n} \geq na)$$

herzustellen. Hierbei schreiben wir – wie immer – $S_n := \sum_{i=1}^n X_i$. In einem ersten Schritt werden wir zeigen, dass dieser Limes überhaupt existiert. Dazu definieren wir

$$\pi_n := \mathbb{P}(S_n \geq na) \tag{14.2}$$

und bemerken dass

$$\pi_{m+n} \geq \mathbb{P}(S_m \geq a, S_{n+m} - S_m \geq na) = \mathbb{P}(S_m \geq a)\mathbb{P}(S_{n+m} - S_m \geq na) = \pi_m \pi_n$$

aufgrund der Unabhängigkeit und identischen Verteilung der X_i. Definieren wir weiter

$$\gamma_n := \frac{1}{n} \log \mathbb{P}(S_n \geq na), \tag{14.3}$$

so folgt

Lemma 14.1 *Es gilt*

$$\gamma_{m+n} \geq \gamma_m + \gamma_n \tag{14.4}$$

uns daraus folgt, dass

$$\frac{\gamma_n}{n} \to \sup_{m \geq 1} \frac{\gamma_m}{m} \qquad \text{as } n \to \infty. \tag{14.5}$$

Beweis. Offensichtlich gilt $\limsup \frac{\gamma_n}{n} \leq \sup_m \frac{\gamma_m}{m}$. Es genügt also zeigen, dass für jedes m gilt

$$\liminf \frac{\gamma_n}{n} \geq \frac{\gamma_m}{m}.$$

Schreiben wir $n = km + l$ mit $0 \leq l < m$ und benutzen (wiederholt) die (14.4), erhalten wir

$$\gamma_n \geq k\gamma_m + \gamma_l.$$

Division durch n ergibt

$$\frac{\gamma_n}{n} \geq \left(\frac{km}{km+l}\right) \frac{\gamma_m}{m} + \frac{\gamma_l}{n}.$$

Schickt man n gegen ∞ und erinnert sich, dass $0 \leq l < m$ war, erhält man das Resultat. ■

Dieses Lemma impliziert schon, dass

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\frac{S_n}{n} \geq na) = -I(a)$$

existiert und (natürlich) nicht–positiv ist. Aus der Formel, die wir für den Limes der $\frac{\gamma_n}{n}$ gewonnen haben, folgt

$$\mathbb{P}(S_n \geq na) \leq e^{-nI(a)}. \tag{14.6}$$

Übung 14.2 *Man zeige, dass die folgenden Aussagen äquivalent sind:*

1. $I(a) = \infty$.
2. $\mathbb{P}(X_1 \geq a) = 0$.
3. $\mathbb{P}(S_n \geq na) = 0$ *für alle* n.

Übung 14.3 *Man zeige, dass*

$$I\left(\frac{a+b}{2}\right) \leq \frac{1}{2}(I(a) + I(b))$$

gilt, I also konvex ist.

Wir werden nun die oben angekündigte Abschätzung mit einer exponentiellen Chebyshev–Ungleichung durchführen. In der tat gilt ja für jedes $t > 0$:

$$\mathbb{P}(S_n \geq na) \leq e^{-nta}\varphi(t)^n$$

oder mit $\psi(t) := \log \phi(t)$

$$\mathbb{P}(S_n \geq na) \leq e^{-n(ta-\psi(t)}.$$

Die Abschätzung ist natürlich nur dann gut, wenn die rechte Seite wenigstens kleiner ist als eins, der Exponent also negativ.

Lemma 14.4 *Wenn $a > \mathbb{E}X_1$ gilt und t klein genug ist, gilt $at - \psi(t) < 0$.*

Bemerkung 14.5 *Die Existenz des Erwartungswertes $\mathbb{E}X_1$ folgt aus der Annahme* (14.1).

Proof of Lemma 14.4. Bemerke, dass $\psi(0) = \log \varphi(0) = 0$. Können wir also zeigen, dass $\psi'(t) = \frac{\varphi'(t)}{\varphi(t)}$ gegen $\mu := \mathbb{E}X_1$ konvergiert, wenn t gegen 0 geht (denn dann ist $at - \psi(t) \sim (a-\mu)t$ negativ, wenn t klein genug ist). Zunächst zeigen wir, dass die Ableitungen existieren. Sei $F(X) := \mathbb{P}(X_1 \leq x)$. Da gilt

$$\left|e^{hx} - 1\right| = \left|\int_0^{hx} e^y dy\right| \leq |hx|(e^{hx} + 1),$$

bekommt man mit Hilfe des Satzes von der dominierten Konvergenz

$$\varphi'(t) = \int xe^{tx} dF(x) \qquad \text{für } t \in (0, t_0).$$

Nimmt man den Limes $t \to 0$ und wendet für die $x < 0$ den Satz von der monotonen Konvergenz und für die $x > 0$ den Satz von der dominierten Konvergenz, sieht man, dass $\varphi'(t) \to \mu$ für $t \to 0$. Da nun andererseits $\phi(t) \to 1$, wenn $t \to 0$, haben wir somit gezeigt, dass $\psi'(t) \to \mu$ gilt, wenn $t \to 0$, was nach der Eingangsbemerkung die Behauptung beweist. ■

Nun, da wir eine Schranke für $\mathbb{P}(S_n \geq na)$ gefunden haben, liegt es nahe, diese Schranke zu optimieren, also das Minimum von $-ta + \psi(t)$ zu finden. Dazu bilden wir

$$\frac{d}{dt}[ta - \psi(t)] = a - \frac{\phi'(t)}{\phi(t)}$$

und somit sollte das Minimum (wenn alles gut geht) bei $a = \frac{\phi'(t)}{\phi(t)}$ angenommen werden. Um sicherzustellen, dass wirklich alles gut geht, definieren wir

$$F_t(x) = \frac{1}{\varphi(t)} \int_{-\infty}^{x} e^{ty} dF(y).$$

Man beachte, dass $F_t(x)$ eine Verteilungsfunktion ist. Ihr Mittwelwert berechnet sich als

$$\int x DF_t(x) =) = \frac{1}{\varphi(t)} \int_{-\infty}^{\infty} x e^{tx} dF(x) = \frac{\varphi'(t)}{\varphi(t)}.$$

Differentiert man noch einmal, erhält man

$$\frac{d}{dt}\frac{\varphi'(t)}{\varphi(t)} = \frac{\varphi''(t)}{\varphi(t)} \left(\frac{\varphi'(t)}{\varphi(t)}\right)^2 = \int x^2 dF_t(x) - \left(\int x dF_t(x)\right)^2 \geq 0.$$

Diese Ungleichung ist sogar strikt, wenn wir annehmen

$$\text{F ist nicht die Dirac-Verteilung in } \mu. \tag{14.7}$$

Unter (14.7) ist $\frac{\varphi'(t)}{\varphi(t)}$ strikt wachsend. Da $\frac{\varphi'(0)}{\varphi(0)} = \mu$, zeigt dies, dass für $a > \mu$ höchstens ein t_a existiert, das $a = \frac{\varphi'(t_a)}{\varphi(t_a)}$ löst.

Ein solches t_a ist für uns der wesentliche Punkt, um die korrekte Rate für γ_n zu bestimmen. In der Tat gilt:

Theorem 14.6 *Es sei (X_i) eine Folge von i.i.d. Zufallsvariablen , die (14.1) und (14.7) erfüllt. Sei $S_n := \sum_{i=1}^n X_i$. Dann gilt für alle $a > \mathbb{E}X_1$ die folgende Gleichheit*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -I(a), \tag{14.8}$$

wobei
$$I(a) := \sup_{t\in\mathbb{R}} [ta - \psi(t)] \tag{14.9}$$
gilt.

Beweis. O.B.d.A. nehmen wir an, dass $a = 0$ und $\mathbb{E}X_1 < 0$ gilt (substituiert man nämlich $X_1 \to X_1 + a$ so ist auch $\varphi(t) \to e^{at}\varphi(t)$ und somit mit $I(\cdot)$ definiert in (14.9) verschiebt sich auch $I(a) \to I(0)$). Wir schreiben in der Folge
$$g := \inf_{t\in\mathbb{R}} \varphi(t)$$
und bemerken, dass
$$I(0) = -\log g \quad \text{ mit } I(0) = \infty \text{ falls } g = 0$$
gilt.
Nun hatten wir oben schon gesehen, dass mit Hilfe der exponentiellen Tschebyschev–Ungleichung folgt:
$$\mathbb{P}(S_n \geq na) \leq e^{-n(ta-\psi(t)} \tag{14.10}$$
für alle positiven t und somit
$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) \leq -\sup_{t\in\mathbb{R}^+} [ta - \psi(t)]\,. \tag{14.11}$$

Um das Supremum über die ganze reelle Achse auszudehnen, erinnern wir uns daran, dass nach den Vorüberlegungen φ eine strikt konvexe Funktion war. Es ist offenbar $\varphi'(0) = \mathbb{E}X_1 < 0$ (nach Annahme). Wir unterscheiden drei Fälle je nachdem, wo $\mathbb{P}$ seine Masse hat.

- $\mathbb{P}(X_1 < 0) = 1$.

  Dann ist $\phi'(t) = \int xe^{tx}dF(x) < 0$ (wobei F die zu $\mathbb{P}$ gehörige Verteilungsfunktion $F(x) := \mathbb{P}(X_1 \leq x)$ ist) für alle $t \in \mathbb{R}$. Somit ist ϕ strikt fallend. Es ist somit
  $$\lim_{t\to\infty} \varphi(t) = g = \mathbb{P}(X_1 = 0) = 0.$$
  Da auch
  $$\mathbb{P}(S_n \geq 0) = 0$$
  gilt, haben wir in diesem Fall schon (14.8).

- $\mathbb{P}(X_1 \leq 0) = 1$ und $1 \neq \mathbb{P}(X_1 = 0) > 0$.

  Wie oben zeigt man, dass ϕ strikt fallend ist und

$$\lim_{t\to\infty} \varphi(t) = g = \mathbb{P}(X_1 = 0) > 0.$$

  Da in diesem Falle

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}(X_1 = \ldots = X_n = 0) = g^n$$

  gilt, folgt auch hier (14.8)

- $\mathbb{P}(X_1 < 0) > 1$ und $\mathbb{P}(X_1 > 0) > 0$.

  Dann gilt offenbar $\lim_{t\to\pm\infty} \varphi(t) = \infty$ und da φ wie oben bemerkt strikt konvex ist, gibt es ein eindeutiges τ so, dass φ in τ minimal wird. Für diese τ gilt natürlich $\varphi'(tau) = 0$ und $\tau > 0$, denn die Ableitung von φ ist in 0 negativ. Somit gehört τ zu den in (14.10) zulässigen t und es gilt daher

$$\mathbb{P}(S_n \geq 0) \leq \mathbb{E}e^{\tau S_n} = (\varphi(\tau))^n = g^n,$$

  also

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \leq \log g.$$

  Um zu zeigen, dass $\log g$ auch eine untere Schranke ist, verwenden wir eine Technik, die als Tilten oder exponentielle Maßtransformation bekannt ist. Die Idee hierbei ist es, die zugrunde liegende Verteilung der X_i so zu verschieben, dass der Erwartungswert 0 (also unser a) ist. Dann wissen wir aus den Gesetzen der großen Zahlen, dass sich S_n so wie na verhalten wird. Wir kassieren aber einen "Strafterm" dafür, dass wir die Verteilung geändert haben.

  Genauer führen wir eine neue Folge (Y_i) von i.i.d. Zufallsvariablen ein, die die Verteilung

$$G(x) = \frac{1}{g} \int_{-\infty}^{x} e^{\tau y} dF(y)$$

  besitzen. G heißt auch die Cramér–Transformierte von F. Bemerke, dass

$$g = \varphi(\tau) = \int_{-\infty}^{\infty} e^{\tau y} dF(y).$$

  Wir benötigen nun die folgenden drei Lemmata

Lemma 14.7 *Es gilt* $\mathbb{E}Y = 0$ *und* $\mathbb{V}Y \in (0,\infty)$.

Beweis. Wir bezeichnent mit $\hat{\varphi}(t) = \mathbb{E}e^{tY}$. Dann erhalten wir für alle $t \in \mathbb{R}$

$$\hat{\varphi}(t) = \int_{\mathbb{R}} e^{tx} dG(x) = \frac{1}{g}\int_{\mathbb{R}} e^{tx}e^{\tau x} dG(x) = \frac{1}{g}\varphi(t+\tau) < \infty.$$

Dies impliziert, dass mit φ auch $\hat{\varphi}$ eine C^∞–Funktion ist. Damit ergibt sich

$$\begin{aligned}
\mathbb{E}Y &= \hat{\varphi}'(0) = \frac{1}{g}\varphi'(\tau) = 0 \text{ und} \\
\mathbb{V}Y &= \hat{\varphi}''(0) = \frac{1}{g}\varphi''(\tau) \in (0,\infty).
\end{aligned}$$

■

Lemma 14.8 *Es sei* $T_n = \sum_{i=1}^n Y_i$. *Dann gilt*

$$\mathbb{P}(S_n \geq 0) = g^n \mathbb{E}(e^{\tau T_n} 1_{\{T_n \geq 0\}}).$$

Beweis. Beachtet man, dass

$$\begin{aligned}
\mathbb{P}(S_n \geq 0) &= \int_{\sum_{i=1}^n x_i \geq 0} dF(x_1)\dots dF(x_n) \\
&= \int_{\sum_{i=1}^n x_i \geq 0} [ge^{-\tau x_1} dF(x_1)]\dots[ge^{-\tau x_n} dF(x_n)]
\end{aligned}$$

so folgt die Behauptung. ■

Lemma 14.9 *Es gilt*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}(e^{\tau T_n} 1_{\{T_n \geq 0\}}) \geq 0.$$

Beweis. Aufgrund von Lemma 14.7 kann man den Zentralen Grenzwertsatz, den wir im folgenden Kapitel beweisen werden, auf T_n anwenden. Wählen wir nun eine Zahl $C > 0$ so, dass

$$\frac{1}{\sqrt{2\pi}} \int_0^C e^{-x^2} 2dx > \frac{1}{4}$$

gilt, erhalten wir die folgende Schranke

$$\mathbb{E}(e^{\tau T_n}1_{\{T_n\geq 0\}}) \geq e^{-\tau C \mathbb{V}X_1\sqrt{n}}\mathbb{P}\left(\frac{T_n}{\mathbb{V}X_1\sqrt{n}} \in [0,C)\right).$$

Da die Wahrscheinlichkeit rechts für n gegen unendlich gegen eine Zahl $\geq \frac{1}{4}$ konvergiert, folgt die Behauptung. ■

Der Beweis des Theorems folgt nun, da aus Lemma 14.8 zusammen mit Lemma 14.9 folgt, dass

$$\liminf_{n\to\infty}\frac{1}{n}\mathbb{P}(S_n \geq 0) = \log g + \liminf_{n\to\infty}\frac{1}{n}\mathbb{E}(e^{\tau T_n}1_{\{T_n\geq 0\}}) \geq \log g.$$

Dies ist die Aussage des Theorems.

■

Bemerkung 14.10 *Das obige Theorem nennt man auch ein* **Prinzip der großen Abweichungen**. *Genauer sagt man die Folge (S_n) genügt einem Prinzip der großen Abweichungen mit Geschwindeikeit n und Rate (oder Ratenfunktion) I.*

Beispiel 14.11 *Ist X_1 normalverteilt zu den Parametern $\mu = 0$ und $\sigma^2 = 1$, so ist*

$$\varphi(t) = \frac{1}{\sqrt{2\pi\sigma^2}}\int e^{tx}e^{-\frac{1}{2}x^2}dx = \frac{1}{\sqrt{2\pi\sigma^2}}\int e^{-\frac{1}{2}(x+t)^2}dx\, e^{\frac{1}{2}t^2} = e^{\frac{1}{2}t^2}.$$

also

$$I(a) = \sup[ta - t^2/2] = a^2/2.$$

Übung 14.12 *Berechnen Sie die Ratenfunktion, für X_i die $\mathcal{N}(\mu, \sigma^2)$–verteilt sind.*

Übung 14.13
1. *Berechnen Sie die Ratenfunktion, für X_i die Poisson-verteilt sind zum Parameter $\lambda > 0$.*
2. *Berechnen Sie die Ratenfunktion, für X_i die Bernoulli–verteilt sind zum Parameter $p = \mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = 0)$.*

Die Ratenfunktion hat die folgenden Eigenschaften

Lemma 14.14 *Unter den Bedingungen von Theorem 14.6 gilt*

1. *I ist von unten–halbstetig und konves auf $\mathbb{R}$.*
2. *I hat kompakte Niveaumengen $N_L := \{z \in \mathbb{R} : I(z) \leq L\}$ für alle $L \geq 0$.*
3. *I is stetig und strikt konvex auf dem Inneren von $D_I := \{z \in \mathbb{R} : I(z) < \infty\}$*
4. *$I(z) \geq 0$ mit $I(z) = 0$ genau dann, wenn $z = \mathbb{E}X_1$.*

Bemerkung 14.15 *Die untere Halbstetigkeit von I ist äquivalent dazu, dass die Nievaumengen abgeschlossen sind.*
Die Konvexität von I impliziert, dass D_I ein Intervall ist.

Übung 14.16 *Beweisen Sie Lemma 14.14.*

Bemerkung 14.17 *Die Aussagen von Theorem 14.6 bleibt natürlich wahr, wenn wir die Wahrscheinlichkeiten $\mathbb{P}(S_n \leq an)$ für $a < \mathbb{E}X_1$ abschätzen. Dies sieht man leicht durch Übergang von X_1 zu $-X_1$.*

Zum Schluss bemerken wir, dass ein Prinzip der großen Abweichungen natürlich wieder das Gesetz der großen Zahlen zur Folge hat:

Korollar 14.18 *Unter den Bedingungen aus Theorem 14.6 gilt das Starke Gesetz der großen Zahlen für die Folge der (S_n)*

Beweis. Ohne Einschränkung können wir annehmenm, dass $\mathbb{E}X_1 = 0$ ist. Man bemerke, dass für jedes $\delta > 0$

$$\mathbb{P}(|S_n| \geq \delta) = \mathbb{P}(S_n \geq \delta) + \mathbb{P}(S_n \leq -\delta)$$

gilt und dass aus Theorem 14.6 und der obigen Bemerkung über die Wahrscheinlichkeit einer unteren Abweichung folgt, dass für genügend große n

$$\mathbb{P}(S_n \geq \delta) \leq e^{-\frac{1}{2}nI(\delta)}$$

und

$$\mathbb{P}(S_n \leq -\delta) \leq e^{-\frac{1}{2}nI(\delta)}$$

gilt. Somit ist

$$\mathbb{P}(|S_n| \geq \delta) \leq 2e^{-\frac{1}{2}nI(\delta)}$$

und daher

$$\sum_{n=1}^{\infty} \mathbb{P}(|S_n| \geq \delta) \leq 2 \sum_{n=1}^{\infty} e^{-\frac{1}{2}nI(\delta)} < \infty$$

endlich. Aus dem Borel–Cantelli Lemma folgt daher, dass $|S_n|$ für jedes $\delta > 0$ mit Wahrscheinlichkeit eins nur für endlich viele n größer ist als δ. Dies aber heißt, dass S_n fast sicher gegen 0 – also seinen Erwartungswert – konvergiert. ■

15 Der zentrale Grenzwertsatz

In den vorhergehenden Abschnitten haben wir eines der zentralen Gesetze der Wahrscheinlichkeitstheorie kennengelernt: das Gesetz der großen Zahlen: Wenn für eine Folge von i.i.d. Zufallsvariablen (X_n) der Erwartungswert $\mathbb{E}X_1$ existiert, dann konvergiert der empirische Mittelwert $\frac{1}{n}\sum_{i=1}^{n} X_i$ gegen eben diesen Erwartungswert $\mathbb{E}X_1$ (fast sicher oder stochastisch). Dies ist ein sehr wichtiges udn auch ein sehr befriedigendes Resultat. Betrachtet man andererseits die Histogramme der Verteilung der *absoluten* Häufigkeiten einer Folge von i.id. Zufallsvariablen mit nichtdegenerierter Verteilung so lässt sich viel mehr Struktur erkennen. Nicht nur ein Punkt – $\mathbb{E}X_1$ – ist wichtig, sondern die Punkte in "nicht zu großem Abstand" (überlicherweise etwas in der Größenordnung $\sqrt{n}$ bei n Zufallsvariablen) bekommen auch einen gehörigen Teil der Masse ab.
Die Analyse dieser "Feinstruktur" soll uns in diesem Kapitel beschäftigen. Der *zentrale Grenzwertsatz*, der die entsprechenden Aussagen mathematisch fomuliert wird uns einige Zeit begleiten. Zum einen aufgrund seiner zentralen Rolle für die gesamte Wahrscheinlichkeitstheorie, zum anderen, weil wir zwei wichtige (und grundverschiedene Beweistechniken kennenlernen wollen: den Steinschen Ansatz und die Methode der Fouriertransformierten, die historisch am Anfang der "echten" Wahrscheinlichkeitstheorie stand und eine Verbindung von komplexer Analysis und Wahrscheinlichkeitstheorie darstellt).
Der Name (zentraler Grenzwertsatz) des folgenden Satzes geht auf den ungarischen Mathematiker Polya zurück, der erste der von uns präsentierten Beweise auf Charles Stein.

Für die Formulierung bemerken wir noch einmal (wie oben), dass – wollen wir einen echten Grenzwertsatz formulieren, bei dem auf der rechten Seite eine nicht–degenerierte Limesverteilung (mit endlicher Varianz, die verschieden ist von Null) entsteht, die Skala $\frac{1}{n}$ für die Summe $\sum_{i=1}^n X_i$ zu groß ist: die Varianz der Summe geht auf dieser Skala gegen 0. In der Tat überlegt man sich schnell, dass die richtige Skala $\frac{1}{\sqrt{n}}$ sein muss, weil dann die Summe stest von endlicher Varianz ist. Hier ist das Resultat, wenn man diese Skala wählt:

Theorem 15.1 (Zentraler Grenzwertsatz – CLT) *Es sei $X_1, \ldots, X_n, \ldots$ eine Folge von i.i.d. Zufallsvariablen mit endlicher Varianz, d.h. mit $\mathbb{E}X_1^2 < \infty$. Dann gilt der zentrale Grenzwertsatz für sie, d.h.*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n (X_i - \mathbb{E}X_1)}{\sqrt{n\mathbb{V}X_1}} \le a\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx. \tag{15.1}$$

gilt für alle $a \in \mathbb{R}$.

Bevor wir uns an den Beweis dieses Satzes machen, bemerken wir, dass der Satz auch unter den folgenden schwächeren Bedingungen gilt (die insbsondere auf die identische Veretilung der X_i verzichten).

In der Tat gilt der Zentrale Grenzwertsatz auch unter den folgenden schwächeren Bedingugnen. Es sei X_i eine Folge unabhängiger Zufallsvariablen . Für $j = 1, \ldots n$ definiere

$$X_{nj} := \frac{1}{s_n}(X_j - \eta_j)$$

wobei

$$\eta_j = \mathbb{E}X_j$$

der Erwartungswert von X_j ist und

$$s_n := \sqrt{\sum_{i=1}^n \mathbb{V}X_i}$$

die Varianz der Summe.
Wir sagen, dass die Folge der (X_n) der **Lindeberg–Bedingung** genügt, falls

$$L_n(\varepsilon) \to 0 \text{ wenn } n \to \infty$$

für alle $\varepsilon > 0$ gilt. Hierbei ist

$$L_n(\varepsilon) = \frac{1}{s_n^2} \sum_{j=1}^{n} \mathbb{E}\left[(X_j - \eta_j)^2 ; |X_j - \eta_j| \geq \varepsilon s_n\right].$$

Intuitiv gesprochen fordert die Lindeberg–Bedingung, dass keine der Zufallsvariablen die gesamte Summe dominiert. Dies wird in der zweiten Bemerkung unten präziser gemacht.
Die verallgemeinerte Form des zentralen Grenzwertsatzes besagt nun, eine Folge von Zufallsvariablen X_n, die unabhängig ist und der **Lindeberg–Bedingung** genügt, auch dem zentralen Grenzwertsatz genügt, d.h. es gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n (X_i - \eta_i)}{\sqrt{\sum_{i=1}^n \mathbb{V}X_i}} \leq a\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx.$$

Der Beweis folgt im wesentlichen den gleichen Ideen, wie der Beweis, den wir unten angeben werden. Wir sparen uns die zusätzliche technische Arbeit, da sie keine weitere Einsicht liefert. Für einen Beweis entlang der Ideen von Stein verweisen wir auf die Skript von Alsmeyer (siehe [**?**]).
Zunächst wollen wir die Lindeberg Bedingung noch etwas näher beleuchten.

Bemerkung 15.2

- *Sind die Zufallsvariablen (X_n) i.i.d., so gilt die Lindeberg–Bedingung. In der Tat: Sei $\mu := \mathbb{P}_{X_n}$, $\eta := \mathbb{E}X_n$ und $\sigma := \mathbb{V}X_n$. Dann ist $s_n^2 = n\sigma^2$ und somit divergiert s_n nach unendlich, wenn $n \to \infty$. Also folgt*

  $$L_n(\varepsilon) = \frac{1}{\sigma^2} \int_{\{|x-\eta| \geq \varepsilon s_n\}} (x - \eta)^2 \mu(dx) \to 0$$

  für alle $\varepsilon > 0$, wenn n gegen ∞ strebt.

- *Genügt die Folge derZufallsvariablen (X_n) der sogenannten Lyapunov-Bedingung, d.h. existiert ein $\delta > 0$, für das gilt*

  $$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{j=1}^{n} \mathbb{E}(|X_j - \mathbb{E}X_j|)^{2+\delta}) = 0$$

  so genügt sie auch der Lindeberg–Bedingung. In der Tat folgt ja für jedes $\varepsilon > 0$ aus der Ungleichung

  $$|x - \mathbb{E}X_1| \geq \varepsilon s_n$$

die Ungleichung

$$|x - \mathbb{E}X_1|^{2+\delta} \geq |x - \mathbb{E}X_1|^2(\varepsilon s_n)^\delta.$$

Hieraus folgt

$$\begin{aligned} L_n(\varepsilon) &\leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{j=1}^n \int_{\{|x-\mathbb{E}X_j| \geq \varepsilon s_n\}} |x - \mathbb{E}X_j|^{2+\delta} \mathbb{P}_{X_j}(dx) \\ &\leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{j=1}^n \mathbb{E}(|X_j - \mathbb{E}X_j|)^{2+\delta} \end{aligned}$$

und aus der Lyapunov–Bedingung folgt just die Konvergenz dieses Ausdrucks gegen 0.

Ist die Folge der (X_n) *fast sicher gleichmäßig beschränkt und gilt zudem* $s_n \to \infty$ *mit* $n \to \infty$*, so gilt die Lindeberg Bedingung. Gemäß Voraussetzung gibt es nämlich ein* $\alpha > 0$ *mit* $|X_n| \leq \alpha/2$ *und somit auch mit*

$$|\mathbb{E}X_n| \leq \mathbb{E}(|X_n|) \leq \frac{\alpha}{2}$$

für alle n fast sicher. Also ist auch

$$|X_n - \mathbb{E}X_n| \leq \alpha$$

fast sicher für alle n. Dann aber folgt auch für $\delta > 0$ *die Lyapunov Bedingung*

$$\frac{1}{s_n^{2+\delta}} \sum_{j=1}^n \mathbb{E}(|X_j - \mathbb{E}X_j|^{2+\delta}) \leq \frac{\alpha^\delta}{s_n^{2+\delta}} \sum_{j=1}^n \mathbb{E}(|X_j - \mathbb{E}X_j|^2) = \left(\frac{\alpha}{s_n}\right)^\delta$$

und somit die Lindeberg–Bedingung.

Umgekehrt folgt aus der Lindeberg–Bedingung die sogenannte Feller–Bedingung.

Lemma 15.3 *Genügt eine Folge* (X_n) *vonZufallsvariablen der Lindeberg–Bedingung so genügt sie auch der Feller Bedingung*

$$\lim_{n\to\infty} \left(\max_{1\leq j\leq n} \frac{\sqrt{\mathbb{V}X_j}}{s_n} \right) = 0.$$

Beweis. *Für jedes $1 \leq j \leq n$ gilt*

$$\begin{aligned}\sigma_j^2 &:= \mathbb{V}X_j = \int (x - \mathbb{E}X_j)^2 d\mathbb{P}_{X_j}(x) \leq \varepsilon^2 s_n^2 + \int_{\{|x-\mathbb{E}X_j| \geq \varepsilon s_n\}} (x - \mathbb{E}X_j)^2 d\mathbb{P}_{X_j}(x) \\ &\leq \varepsilon^2 s_n^2 + \sum_{k=1}^{n} \int_{\{|x-\mathbb{E}X_k| \geq \varepsilon s_n\}} (x - \mathbb{E}X_k)^2 d\mathbb{P}_{X_k}(x).\end{aligned}$$

Somit folgt

$$\left(\max_{1\leq j\leq n} \frac{\sqrt{\mathbb{V}X_j}}{s_n} \right) = \max_{1\leq j\leq n} \left(\frac{\sqrt{\mathbb{V}X_j}}{s_n} \right)^2 \leq \varepsilon^2 + L_n(\varepsilon)$$

für beliebiges $\varepsilon > 0$. Gilt somit für die (X_n) die Lindeberg–Bedingung, so folgt auch die Feller–Bedingung. ■
Die Feller–Bedingung illustriert besonders deutlich, dass keine der Zufallsvariablen einen zu großen Einfluss auf die Summe haben darf.

Um zu vestehen, was wir im Beweis des Zentralen Grenzwertsatzes veranstalten, müssen wir zunächst den Begriff der Verteilungskonvergenz, die dort behauptet wird, diskutieren. Hierzu sei für eine messbare Menge $A \in \mathcal{B}^1$ die Menge

$$\partial A := \overline{A} \setminus A^\circ$$

der Rand von A. Wir nennen eine messbare Menge A randlos bezüglich einer reellwertigen Zufallsvariablen X mit Verteilung $\mathbb{P}_X$, falls

$$\mathbb{P}_X(\partial C) = \mathbb{P}(X \in \partial C) = 0$$

gilt.

Theorem 15.4 (Portmonteau–Theorem) *Seien $X, X_1, X_2, \ldots$ reellwertige Zufallsvariablen . Dann sind äquivalent:*

1. *$X_n \to X$ in Verteilung, d.h $\int f d\mathbb{P}_{X_n} \to \int f d\mathbb{P}_X$ für alle $f \in \mathcal{C}^b(\mathbb{R})$.*
2. *$\liminf \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ für alle offenen Mengen G.*
3. *$\limsup \mathbb{P}(X_n \in A) \leq \mathbb{P}(X_n \in A)$ für alle abgeschlossenen Mengen A.*
4. *$\lim \mathbb{P}(X \in C) = Q(X \in C)$ für alle bezüglich X randlosen Mengen C.*

5. $\lim_{n\to\infty} F_n(x) = F(x)$ *für alle* x*, in denen* F *stetig ist. Hierbei bezeichnet* F *die Verteilungsfunktion von* X*,* F_n *ist die Verteilungsfunktion von* X_n*.*

6. $\int f d\mathbb{P}_{X_n} \to \int f d\mathbb{P}_X$ *für alle* $f \in \mathcal{C}^b(\mathbb{R})$*, die gleichmäßig stetig sind.*

Beweis. Wir verfolgen ein etwas verwirrendes Beweisschema.
1. $\Rightarrow$ 6. ist offensichtlich.

6. $\Rightarrow$ 3.: Es sei A eine abgeschlossene Menge. Für $m \in \mathbb{N}$ definiere

$$A_m := \{x \in \mathbb{R} : d(x, A) \leq \frac{1}{m}\},$$

wobei $d(x, A) := \inf\{|x - y| : y \in A\}$ der Abstand von x zu A ist. Dann ist G_m fallend und

$$\lim_{m\to\infty} A_m = \{x \in \mathbb{R} : d(x, A) = 0\} = A,$$

da A abgeschlossen ist. Aufgrund der Stetigkeitseigenschaft endlicher Maße folgt somit

$$\lim_{m\to\infty} \mathbb{P}(X \in A_m) = \mathbb{P}(X \in A).$$

Also existiert zu beliebigem $\varepsilon > 0$ eine Nummer m_0 mit

$$P(X \in A_{m_0}) \leq \mathbb{P}(X \in A) + \varepsilon.$$

Umgekehrt können wir eine stetige und beschränkte Funktion f finden (sogar eine gleichmäßig stetige) mit $f_{|A} = 1$ und $f_{|A^c_{m_0}} = 0$ sowie $0 \leq f \leq 1$. In der Tat verknüpft man die gleichmäßig stetigen Funktionen $d(x, A)$ (diese ist wegen der Dreiecks–Ungleichung natürlich gleichmäßig stetig) und

$$g(t) := \begin{cases} 1 & t \leq 0 \\ 1 - t & 0 < t < 1 \\ 0 & t \geq 1 \end{cases}$$

zu

$$f(x) = g(m_0 \cdot d(x, A)),$$

so ist f gleichmäßig stetig und es gilt $f_{|A} = 1$ und $f_{|A^c_{m_0}} = 0$ sowie $0 \leq f \leq 1$. Für dieses f gilt nun $1_A \leq f \leq 1_{A_{m_0}}$ und nach Voraussetzung 6

$$\lim_{n\to\infty} \int f d\mathbb{P}_{X_n} = \int f d\mathbb{P}_X.$$

Insgesamt folgt

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in A) \leq \lim_{n\to\infty} \int f d\mathbb{P}_{X_n} = \int f d\mathbb{P}_X \leq \mathbb{P}(X \in A_{m_0}) \leq \mathbb{P}(X \in A) + \varepsilon$$

also 3.

3. $\Rightarrow$ 1.: Es sei $f \in \mathcal{C}^b(\mathbb{R})$. Durch Übergang zu

$$\tilde{f}(x) := \frac{f(x) + \|f\|}{2\|f\| + 1}$$

(wobei $\|f\|$ die Supremumsnorm von f ist) können wir annehmen, dass $0 \leq f < 1$ gilt.
Wir zeigen nun, dass

$$\limsup_{n\to\infty} \int f d\mathbb{P}_{X_n} \leq \int f d\mathbb{P}_X \tag{15.2}$$

gilt.
Hierzu sei für $\mu \in \mathbb{N}$

$$G_i := \{x \in \mathbb{R} : f(x) \geq i/m\}, \quad 0 \leq i \leq m$$

mit $G_0 = \mathbb{R}$ und $G_m = \emptyset$. Bemerke, dass die G_i abgeschlossene Mengen sind, da f stetig ist. Nun gilt für beliebige Wahrscheinlichkeitsverteilungen Q über $(\mathbb{R}, \mathcal{B})$

$$\sum_{i=1}^{m} \frac{i-1}{m} Q(\{x : \frac{i-1}{m} \leq f(x) < \frac{i}{m}\}) \leq \int f dQ \leq \sum_{i=1}^{m} \frac{i}{m} Q(\{x : \frac{i-1}{m} \leq\leq f(x) < \frac{i}{m}\})$$

sowie

$$\begin{aligned}
Q(\{x : \frac{i-1}{m} \leq f(x) < \frac{i}{m}\}) &= Q(G_{i-1}) - Q(G_i) \\
\sum_{i=1}^{m} \frac{i-1}{m}(Q(G_{i-1}) - Q(G_i)) &= \frac{1}{m} \sum_{i=1}^{m} Q(G_i) \quad \text{und} \\
\sum_{i=1}^{m} \frac{i}{m}(Q(G_{i-1}) - Q(G_i)) &= \frac{1}{m} + \frac{1}{m} \sum_{i=1}^{m} Q(G_i).
\end{aligned}$$

Wendet man dies auf $Q = \mathbb{P}_X$ bzw. $Q = \mathbb{P}_{X_n}$ an, erhält man

$$\begin{aligned}
\limsup_{n\to\infty} \int f d\mathbb{P}_{X_n} &\leq \frac{1}{m} + \frac{1}{m} \limsup_{n\to\infty} \sum_{i=1}^{m} \mathbb{P}(X_n \in G_i) \\
&\leq \frac{1}{m} + \frac{1}{m} \sum_{i=1}^{m} \mathbb{P}(X \in G_i) \\
&\leq \frac{1}{m} + \int f d\mathbb{P}_X .
\end{aligned}$$

Da dies für jedes m gilt, folgt (15.2).
Wendet man dies auf $-f$ an, so erhält man

$$\liminf_{n\to\infty} \int f d\mathbb{P}_{X_n} = -\limsup_{n\to\infty} \int (-f) d\mathbb{P}_{X_n} \geq -\int (-f) d\mathbb{P}_X = \int f d\mathbb{P}_X .$$

Dies ergibt zusammen 1.

2. $\Leftrightarrow$ 3. ist offensichtlich, wegen der Dualität von offenen und abgeschlossenen Mengen.

3. $\Rightarrow$ 4.: Es sei C mit $\mathbb{P}(X \in \partial C) = 0$. Wendet man 2. bzw. 3. auf die offene (bzw. abgeschlossene) Menge C° (bzw. $\overline{C}$) an, so ergibt sich:

$$\begin{aligned}
\mathbb{P}(X \in \overline{C}) &\geq \limsup \mathbb{P}(X_n \in \overline{C}) \geq \limsup \mathbb{P}(X_n \in C) \\
&\geq \liminf \mathbb{P}(X_n \in C) \geq \liminf \mathbb{P}(X_n \in C^\circ) \geq \mathbb{P}(X \in C^\circ),
\end{aligned}$$

d.h.

$$\lim_{n\to\infty} \mathbb{P}(X_n \in C) = \mathbb{P}(X \in C).$$

4. $\Rightarrow$ 5.: Es sei x ein Stetigkeitspunkt von F. Nun ist F als Verteilungsfunktion immer rechtsseitig stetig (dies hatten wir schon in der Stochastikvorlesung bemerkt). Also ist F in x stetig, genau dann, wenn es in x linksseitig stetig ist, d.h.

$$\mathbb{P}(X \leq x) = \mathbb{P}(X < x)$$

gilt. Also $\mathbb{P}(X \in \partial(-\infty, x]) = 0$. Also gilt

$$\begin{aligned}
F(x) &= \mathbb{P}(X \in (-\infty, x]) \\
&= \lim_{n\to\infty} \mathbb{P}(X_n \in (-\infty, x]) = \lim_{n\to\infty} F_n(x),
\end{aligned}$$

wobei wir beim zweiten Gleichheitszeichen gerade die Voraussetzung 4. verwenden.

5. $\Rightarrow$ 2.: Es sei $G \subseteq \mathbb{R}$ offen. Dann gibt es eine Folge (J_m) von Intervallen

$$J_m = (c_m, c_m + g_m] \quad c_m, g_m \in \mathbb{R}, g_m > 0,$$

mit $G = \bigcup_m J_m$. Nun liegt die Menge

$$\{x \in \mathbb{R} : \mathbb{P}(X = x) = 0\}$$

dicht in $\mathbb{R}$. Bezeichnen wir mit

$$\mathcal{U} := \{(a,b] \subset \mathbb{R} : a \leq b, \mathbb{P}(X = a) = \mathbb{P}(X = b) = 0\},$$

so können wir also wiederum jedes Intervall J_m als eine abzählbare Vereinigung von Elementen aus $\mathcal{U}$ darstellen. Somit erhalten wir insgesamt die Existenz einer Folge $(I_m)_m$ mit $I_m \in \mathcal{U}$ für alle m,

$$I_m = (a_m, a_m + h_m], \quad a_m, h_m \in \mathbb{R}, h_m > 0$$

so dass

$$G = \bigcup_m I_m$$

gilt. Da die I_m als Randpunkte Stetigkeitspunkte von F besitzen, gilt für jedes m

$$\begin{aligned}\mathbb{P}(X \in I_m) &= F(a_m + h_m) - F(a_m) \\ &= \lim_{n\to\infty} F_n(a_m + h_m) - F_n(a_m) = \mathbb{P}(X_n \in I_m).\end{aligned}$$

Da $\mathcal{U}$ durchschnittsstabil ist, erhält man aus dieser Rechnung zusammen mit der Einschluss– Ausschlussformel für jedes $M \in \mathbb{N}$

$$\begin{aligned}\mathbb{P}(X \in \bigcup_{m=1}^{M} I_m) &= \sum_{j=1}^{M} (-1)^{j+1} \sum_{1 \leq i_1 < \ldots < i_j \leq M} \mathbb{P}(X \in I_{i_1} \cap \ldots \cap I_{i_j}) \\ &= \lim_{n\to\infty} \sum_{j=1}^{M} (-1)^{j+1} \sum_{1 \leq i_1 < \ldots < i_j \leq M} \mathbb{P}(X_n \in I_{i_1} \cap \ldots \cap I_{i_j}) \\ &= \lim_{n\to\infty} P(X_n \in \bigcup_{m=1}^{M} I_m).\end{aligned}$$

Wählt man nun M so groß, dass

$$\mathbb{P}(X \in G) \leq \mathbb{P}(X \in \bigcup_{m=1}^{M} I_m) + \varepsilon > 0$$

gilt (wobei $\varepsilon > 0$ beliebig, aber fest gewählt ist), dann folgt die Behauptung aus

$$\mathbb{P}(X \in G) - \varepsilon \leq \mathbb{P}(X \in \bigcup_{m=1}^{M} I_m) = \lim_{n\to\infty} \mathbb{P}(X_n \in \bigcup_{m=1}^{M} I_m) \leq \liminf_{n\to\infty} \mathbb{P}(X_n \in G)$$

(wobei die letzte Ungleichung gilt, da $G \supset \bigcup_{m=1}^{M} I_m$ gilt).

Nach einer Kontrolle, ob wir alle zu beweisenden Richtungen auch bewiesen haben, zeigt dies die Behauptung ■

Wir werden uns nun an den eigentlichen Beweis des zentralen Grenzwertsatzes machen. Wir präsentieren zunächst einen Beweis, dessen Grundidee auf Charles Stein zurückgeht. Wir vergewissern uns zunächst dessen, was zu tziegen ist.

<u>Fakt 1:</u> Es genügt Satz 15.1 für den Fall von i.i.d. zentrierten Zufallsvariablen (X_i), d.h. solchen mit $\mathbb{E}X_1 = 0$, zu beweisen. Andernfalls subtrahiert man einfach $\mathbb{E}X_1$ von X_i.

<u>Fakt 2:</u> Es genügt Satz 15.1 für den Fall von i.i.d. Zufallsvariablen mit Varianz 1 zu beweisen. Andernfalls ersetzt man X_i einfach durch $X_i/\sqrt{\mathbb{V}X_1}$.

<u>Fakt 3:</u> Setzen wir

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

so behauptet Theorem 15.1 nach dem obigen Portmanteau–Theorem die Verteilungskonvergenz von S_n gegen eine standard–normalverteilt Zufallsvariable . Daher müssen wir zeigen (auch nach dem obigen Portmanteau–Theorem), dass

$$\mathbb{E}\left[f\left(\frac{S_n}{\sqrt{n}}\right)\right] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = \mathbb{E}[f(Y)] \tag{15.3}$$

für $n \to \infty$ für alle $f : \mathbb{R} \to \mathbb{R}$, die gleichmäßig stetig und beschränkt sind. Hie ist Y eine standard–normalverteilte, d.h. eine $\mathcal{N}(0,1)$–verteilte Zufallsvariable .

Wir bereiten den Beweis in zwei Lemmata vor.

Lemma 15.5 *Sei $f : \mathbb{R} \to \mathbb{R}$ beschränkt und gleichmäßig stetig. Setze*

$$\mathcal{N}(f) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y) e^{-y^2/2} dy$$

und

$$g(x) := e^{x^2/2} \int_{-\infty}^{x} (f(y) - \mathcal{N}(f))\, e^{-y^2/2} dy.$$

Dann genügt g der Differentialgleichung

$$g'(x) - xg(x) = f(x) - \mathcal{N}(f). \tag{15.4}$$

Beweis. Leitet man g ab, so ergibt sich:

$$\begin{aligned} g'(x) &= xe^{x^2/2} \int_{-\infty}^{x} (f(y) - \mathcal{N}(f))\, e^{-y^2/2} dy + e^{x^2/2} (f(x) - \mathcal{N}(f))\, e^{-x^2/2} \\ &= xg(x) + f(x) - \mathcal{N}(f). \end{aligned}$$

■

Die Rolle des obigen Lemmas wird deutlich, wenn man in (15.4) eine Zufallsvariable einsetzt und den Erwartungswert nimmt.

$$\mathbb{E}[g'(X) - Xg(X)] = \mathbb{E}[f(X) - \mathcal{N}(f)].$$

Ist $X \sim \mathcal{N}(0, 1)$ standard–normalverteilt, ist die rechte Seite dieser Gleichung null, also auch die linke Seite. Anstatt zu zeigen, dass

$$\mathbb{E}[f(S_n) - \mathcal{N}(f)]$$

gegen 0 konvergiert, kann man selbiges auch für

$$\mathbb{E}[g'(S_n) - S_n g(S_n)]$$

zeigen.

Im nächsten Schritt diskutieren wir die Eigenschaften der oben definierten Funktion g.

Lemma 15.6 *Sei $f : \mathbb{R} \to \mathbb{R}$ beschränkt und gleichmäßig stetig und g die Lösung der Differentialgleichung*

$$g'(x) - xg(x) = f(x) - \mathcal{N}(f). \tag{15.5}$$

Dann sind $g(x), xg(x)$ und $g'(x)$ beschränkt und gleichmäßig stetig.

Beweis. Offenbar ist nach der vorigen Lemma g sogar differenzierbar, also auch stetig. Dann ist aber auch $xg(x)$ stetig. Schließlich ist

$$g'(x) = xg(x) + f(x) - \mathcal{N}(f)$$

als die Summe stetiger Funktionen auch persönlich stetig.
Für den Nachweis der Beschränktheit bemerken wir zunächst, dass stetige Funktionen auf kompakten Mengen natürlcih beschrankt sind. Somit ist die Beschränktheit von g, xg und g' nur für $x \to \pm\infty$ nachzuweisen.
Hierfür sei zunächst bemerkt, dass

$$\begin{aligned} g(x) &= e^{x^2/2} \int_{-\infty}^{x} (f(y) - \mathcal{N}(f))\, e^{-y^2/2} dy \\ &= -e^{x^2/2} \int_{x}^{\infty} (f(y) - \mathcal{N}(f))\, e^{-y^2/2} dy \end{aligned}$$

(da das Integral über ganz $\mathbb{R}$ ja null ergeben muss).
Für $x \leq 0$ erhalten wir

$$g(x) \leq \sup_{y \leq 0} |f(y) - \mathcal{N}(f)|\, e^{x^2/2} \int_{-\infty}^{x} e^{-y^2/2} dy$$

während für $x \geq 0$ gilt

$$g(x) \leq \sup_{y \geq 0} |f(y) - \mathcal{N}(f)|\, e^{x^2/2} \int_{x}^{\infty} e^{-y^2/2} dy.$$

Nun ist für $x \leq 0$

$$e^{x^2/2} \int_{-\infty}^{x} e^{-y^2/2} dy \leq e^{x^2/2} \int_{-\infty}^{x} \frac{-y}{|x|} e^{-y^2/2} dy = \frac{1}{|x|}$$

und ähnlich für $x \geq 0$

$$e^{x^2/2} \int_{x}^{\infty} e^{-y^2/2} dy \leq e^{x^2/2} \int_{x}^{\infty} \frac{y}{x} e^{-y^2/2} dy \leq \frac{1}{|x|}.$$

Somit sehen wir, dass $x \geq 1$ gilt

$$|g(x)| \leq |xg(x)| \leq \sup_{y \geq 0} |f(y) - \mathcal{N}(f)|$$

und ebenso für $x \leq -1$

$$|g(x)| \leq |xg(x)| \leq \sup_{y\leq 0} |f(y) - \mathcal{N}(f)|\,.$$

Also sind g und xg beschränkt. Dann ist aber auch g' beschränkt, da

$$g'(x) = xg(x) + f(x) - \mathcal{N}(f).$$

■

Nun machen wir uns an den Beweis des Zentralen Grenzwertsatzes

Beweis von Theorem 15.1. Wir nehmen o.B.d.A. an, dass $\mathbb{E}X_i = 0$ und $\mathbb{V}X_i = 1$ für alle i gilt.
Wir schreiben

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

und erinnern, dass wir für alle beschränkten und gleichmäßig stetigen Funktionenen f

$$\mathbb{E}\left[g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{S_n}{\sqrt{n}} g\left(\frac{S_n}{\sqrt{n}}\right)\right] \to 0$$

wenn $n \to \infty$ zeigen willen. Hierbei ist g wie oben definiert.
Eine Anwendung des Hauptsatzes der Integral– und Differentialrechnung ergibt

$$\begin{aligned}\int_0^1 &\frac{X_j}{\sqrt{n}}\left[g'\left(\frac{S_n}{\sqrt{n}} - (1-s)\frac{X_j}{\sqrt{n}}\right) - g'\left(\frac{S_n - X_j}{\sqrt{n}}\right)\right] ds \\ &= g\left(\frac{S_n}{\sqrt{n}}\right) - g\left(\frac{S_n - X_j}{\sqrt{n}}\right) - \frac{X_j}{\sqrt{n}} g'\left(\frac{S_n - X_j}{\sqrt{n}}\right).\end{aligned}$$

Wenden wir dies an, so erhalten wir

$$\begin{aligned}&\mathbb{E}\left[g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{S_n}{\sqrt{n}} g\left(\frac{S_n}{\sqrt{n}}\right)\right] \\ = &\sum_{j=1}^n \mathbb{E}\left[\frac{1}{n} g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{X_j}{\sqrt{n}} g\left(\frac{S_n}{\sqrt{n}}\right)\right] \\ = &\sum_{j=1}^n \mathbb{E}\left[\frac{1}{n} g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{X_j}{\sqrt{n}} g\left(\frac{S_n - X_j}{\sqrt{n}}\right) - \frac{X_j^2}{n} g'\left(\frac{S_n - X_j}{\sqrt{n}}\right)\right. \\ &\qquad \left. - \frac{X_j^2}{n}\int_0^1 g'\left(\frac{S_n}{\sqrt{n}} - (1-s)\frac{X_j}{\sqrt{n}}\right) - g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) ds\right].\end{aligned}$$

Wendet man die Linearität des Erwartungswertes aus, dass $\mathbb{E}X_i = 0$ für alle i gilt, dass X_i und $S_n - X_i$ für alle i unabhängig sind und schließlich dass $\mathbb{E}X_i^2 = 1$ für alle i gilt, ergibt sich

$$\begin{aligned}
&\mathbb{E}\left[g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{S_n}{\sqrt{n}} g\left(\frac{S_n}{\sqrt{n}}\right)\right] \\
= &\sum_{j=1}^{n} \mathbb{E}\left[\frac{1}{n} g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{1}{n} g'\left(\frac{S_n - X_j}{\sqrt{n}}\right)\right. \\
&\qquad \left. - \frac{X_j^2}{n} \int_0^1 g'\left(\frac{S_n}{\sqrt{n}} - (1-s)\frac{X_j}{\sqrt{n}}\right) - g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) ds\right].
\end{aligned}$$

Wir haben somit alles auf die Funktion g' zurückgespielt.
Wir setzen

$$\begin{aligned}
&\sum_{j=1}^{n} \mathbb{E}\left[\frac{1}{n} g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{1}{n} g'\left(\frac{S_n - X_j}{\sqrt{n}}\right)\right. \\
&\qquad \left. - \frac{X_j^2}{n} \int_0^1 g'\left(\frac{S_n}{\sqrt{n}} - (1-s)\frac{X_j}{\sqrt{n}}\right) - g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) ds\right] \\
=: &\sum_{j=1}^{n} \Gamma_j.
\end{aligned}$$

Die Idee ist nun, dass die stetige Funktion g' auf jeder kompakten Menge gleichmäßig stetig ist. Ist nun $\frac{X_j}{\sqrt{n}}$ klein, dann sind auch die Differenzen

$$g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{1}{n} g'\left(\frac{S_n - X_j}{\sqrt{n}}\right)$$

und

$$g'\left(\frac{S_n}{\sqrt{n}} - (1-s)\frac{X_j}{\sqrt{n}}\right) - g'\left(\frac{S_n - X_j}{\sqrt{n}}\right)$$

klein, wenn $\frac{S_n}{\sqrt{n}}$ in die gewählte kompakte Menge fällt. Andererseits ist die Wahrscheinlichkeit, dass $\frac{S_n}{\sqrt{n}}$ außerhalb einer großen kompakten Menge realisiert wird klein – ebenso wie die Wahrscheinlichkeit, dass $\frac{X_j}{\sqrt{n}}$ sehr groß wird. Dies zusammen mit der Beschränktheit von g und g' ergibt im wesentlichen den Beweis (wobei mit der Reihenfolge der Argumente ein wenig aufpassen muss).

Für $K > 0, \delta > 0$ schreiben wir

$$\Gamma_j^1 := \Gamma_j 1_{|\frac{X_j}{\sqrt{n}}| \le \delta} 1_{|\frac{S_n}{\sqrt{n}}| \le K}$$
$$\Gamma_j^2 := \Gamma_j 1_{|\frac{X_j}{\sqrt{n}}| \le \delta} 1_{|\frac{S_n}{\sqrt{n}}| > K}$$
$$\Gamma_j^3 := \Gamma_j 1_{|\frac{X_j}{\sqrt{n}}| > \delta}.$$

Daher gilt

$$\begin{aligned}\sum_{j=1}^n \Gamma_j &= \sum_{j=1}^n \Gamma_j^1 + \Gamma_j^2 + \Gamma_j^3 \\ &= \sum_{j=1}^n \Gamma_j^1 + \sum_{j=1}^n \Gamma_j^2 + \sum_{j=1}^n \Gamma_j^3.\end{aligned}$$

Um diese Terme abzuschätzen bemerken wir zunächst, dass g' auf jedem Intervall $[-K, K]$ gleichmäßig steti ist, d.h. zu jedem $\varepsilon > 0$ existiert ein $\delta > 0$, so dass für alle $x, y \in [-K, K]$ gilt:

$$|x - y| < \delta \Rightarrow |g'(x) - g'(y)| < \varepsilon.$$

Da wir das so gefundene δ natürlich beliebig kleiner (aber niemals größer) machen dürfen, können wir annhemen, dass $\delta < K/2$ gilt.
Wir bestimmen zunächst die Größe von K. Dies geschieht durch Abschätzung des zweiten Terms:
Nach der Tschebyshevschen Ungleichung gilt

$$\mathbb{P}(|\frac{S_n - X_1}{\sqrt{n}}| > K - \delta) \le \frac{V\left(\frac{S_n - X_1}{\sqrt{n}}\right)}{(K - \delta)^2} \le \frac{4}{K^2}.$$

Somit können wir für gegebens $\varepsilon > 0$ ein K finden, so dass

$$\mathbb{P}(|\frac{S_n - X_1}{\sqrt{n}}| > K - \delta) \le \varepsilon.$$

K sei so gewählt. Für dieses K gilt dann auch

$$\mathbb{P}(|\frac{S_n}{\sqrt{n}}| > K) \le \frac{V\left(\frac{S_n}{\sqrt{n}}\right)}{K^2} \le \frac{1}{K^2} \le \varepsilon.$$

Da g' durch $||g'|| := \sup_{x\in\mathbb{R}} |g'(x)|$ beschränkt ist, dürfen wir wie folgt anschätzen:

$$\begin{aligned}
\sum_{j=1}^{n} |\Gamma_j^2| &= \Bigg|\mathbb{E}\Bigg[\frac{1}{n}g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{1}{n}g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) \\
&\quad -\frac{X_j^2}{n}\int_0^1 g'\left(\frac{S_n}{\sqrt{n}} - (1-s)\frac{X_j}{\sqrt{n}}\right) - g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) ds 1_{|\frac{X_j}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n}{\sqrt{n}}|>K}\Bigg]\Bigg| \\
&\le \sum_{j=1}^{n} \mathbb{E}\left[\frac{1}{n}2||g'||1_{|\frac{X_j}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n}{\sqrt{n}}|>K}\right] + \sum_{j=1}^{n} \frac{2||g'||}{n}\mathbb{E}\left[X_j^2 1_{|\frac{X_j}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n}{\sqrt{n}}|>K}\right].
\end{aligned}$$

Beachte nun, dass auf der Menge, die wir für Γ^2 betrachten $|\frac{S_n}{\sqrt{n}}| > K$ und $|X_i/\sqrt{n}| \le \delta$ für alle $1 \le i \le n$ gilt, also auch

$$\left|\frac{S_n - X_i}{\sqrt{n}}\right| + \delta \ge \left|\frac{S_n - X_i}{\sqrt{n}}\right| + \left|\frac{X_i}{\sqrt{n}}\right| \ge \left|\frac{S_n}{\sqrt{n}}\right| > K$$

und somit

$$\left|\frac{S_n - X_i}{\sqrt{n}}\right| > K - \delta.$$

Also ist

$$\begin{aligned}
&\sum_{j=1}^{n} \mathbb{E}\left[\frac{1}{n}2||g'||1_{|\frac{X_j}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n}{\sqrt{n}}|>K}\right] + \sum_{j=1}^{n} \frac{2||g'||}{n}\mathbb{E}\left[X_j^2 1_{|\frac{X_j}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n}{\sqrt{n}}|>K}\right] \\
\le\ & 2||g'||\mathbb{E}\left[1_{|\frac{X_j}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n}{\sqrt{n}}|>K}\right] + 2||g'||\mathbb{E}\left[X_1^2 1_{|\frac{X_1}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n - X_1}{\sqrt{n}}|>K-\delta}\right] \\
\le\ & 2||g'||\mathbb{E}\left[1_{|\frac{S_n}{\sqrt{n}}|>K}\right] + 2||g'||\mathbb{E}\left[X_1^2 1_{|\frac{X_1}{\sqrt{n}}|\le\delta}\right]\mathbb{E}\left[1_{|\frac{S_n - X_1}{\sqrt{n}}|>K-\delta}\right] \\
\le\ & 2||g'||\mathbb{P}\left[\left|\frac{S_n}{\sqrt{n}}\right| > K\right] + 2||g'||\mathbb{P}\left[\left|\frac{S_n - X_1}{\sqrt{n}}\right| > K - \delta\right] \\
\le\ & 4||g'||\varepsilon
\end{aligned}$$

Für die Γ_j^1-Terme erinnern wir uns an das eingang Erwähnte, dass nämlich für festes $K > 0$ (deas wir wie im ersten schritt fest wählen) die stetige Funktion g' auf $[-K, K]$ gleichmäßig stetig ist. Also gibt es zu gegebenem $\varepsilon > 0$ ein $\delta > 0$ so dass

$$|x - y| < \delta \Rightarrow |g'(x) - g'(y)| < \varepsilon.$$

Mit diesem δ rechnen wir

$$\begin{aligned}
\sum_{j=1}^n |\Gamma_j^1| &= \sum_{j=1}^n \Bigg| \mathbb{E}\Bigg[\frac{1}{n}g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{1}{n}g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) \\
&\quad -\frac{X_j^2}{n}\int_0^1 g'\left(\frac{S_n}{\sqrt{n}} - (1-s)\frac{X_j}{\sqrt{n}}\right) - g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) ds 1_{|\frac{X_j}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n}{\sqrt{n}}|\le K}\Bigg]\Bigg| \\
&\le \sum_{j=1}^n \Bigg| \mathbb{E}\Bigg[\frac{1}{n}g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{1}{n}g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) 1_{|\frac{X_j}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n}{\sqrt{n}}|\le K}\Bigg]\Bigg| \\
&\quad + \Bigg|\Bigg[\frac{X_j^2}{n}\int_0^1 g'\left(\frac{S_n}{\sqrt{n}} - (1-s)\frac{X_j}{\sqrt{n}}\right) - g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) ds 1_{|\frac{X_j}{\sqrt{n}}|\le\delta} 1_{|\frac{S_n}{\sqrt{n}}|\le K}\Bigg]\Bigg| \\
&\le \sum_{j=1}^n \left(\frac{1}{n}\varepsilon + \frac{\mathbb{E}X_j^2}{n}\varepsilon\right) \\
&= n\left(\frac{2\varepsilon}{n}\right) = 2\varepsilon.
\end{aligned}$$

Schließlich wenden wir uns den Γ_j^3–Termen zu:
Da $\mathbb{E}X_1^2 = 1 < \infty$ ist, gibt es zu gegebenem $\varepsilon > 0$ ein n_0, so dass für alle $n \ge n_0$ und δ wie im letzten Schritt

$$\mathbb{E}X_1^2 1_{X_1 > \sqrt{n}\delta} < \varepsilon.$$

gilt. Weiter ist für $n \ge n_1$ auch

$$\mathbb{P}\left(\left|\frac{X_j}{\sqrt{n}}\right| > \delta\right) \le \varepsilon$$

Hieraus folgt für $n \ge \max\{n_0, n_1\}$

$$\begin{aligned}
\sum_{j=1}^n |\Gamma_j^3| &= \Bigg| \mathbb{E}\Bigg[\frac{1}{n}g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{1}{n}g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) \\
&\quad -\frac{X_j^2}{n}\int_0^1 g'\left(\frac{S_n}{\sqrt{n}} - (1-s)\frac{X_j}{\sqrt{n}}\right) - g'\left(\frac{S_n - X_j}{\sqrt{n}}\right) ds 1_{|\frac{X_j}{\sqrt{n}}|>\delta}\Bigg]\Bigg| \\
&\le \sum_{j=1}^n \left(\frac{2}{n}\mathbb{E}1_{|\frac{X_j}{\sqrt{n}}|>\delta}||g'|| + \frac{2}{n}||g'||\mathbb{E}\left[X_j^2 1_{|\frac{X_j}{\sqrt{n}}|>\delta}\right]\right) \\
&\le 2||g'||\mathbb{P}\left(\left|\frac{X_j}{\sqrt{n}}\right| > \delta\right) + 2||g'||\mathbb{E}\left[X_j^2 1_{|\frac{X_j}{\sqrt{n}}|>\delta}\right] \\
&\le 4\varepsilon||g'||
\end{aligned}$$

Somit erhalten wir für gegebenes $\varepsilon > 0$ mit der obigen Wahl von $\delta > 0$ und $K > 0$:

$$\begin{aligned}
& \mathbb{E}\left[g'\left(\frac{S_n}{\sqrt{n}}\right) - \frac{S_n}{\sqrt{n}} g\left(\frac{S_n}{\sqrt{n}}\right)\right] \\
= & \sum_{j=1}^n \Gamma_j^1 + \sum_{j=1}^n \Gamma_j^2 + \sum_{j=1}^n \Gamma_j^3 \\
\leq & 2\varepsilon + 8\varepsilon ||g'||.
\end{aligned}$$

Da $\varepsilon > 0$ beweist dies das Theorem. ■

Wir werden im folgenden eine zweite Methode (die historisch erste) darstellen, den Zentralen Grenzwertsatz zu beweisen. Das wichtigste Hiflsmittel hierür ist die sogenannte Fouriertransformierte eines Maßes bzw. einer Zufallsvariablen . Sie wird wie folgt definiert

Definition 15.7 *Es sei μ ein endliches Maß auf $\mathbb{R}$. Dann heißt die durch*

$$\hat{\mu}(x) := \int e^{ixy} \mu(dy) \qquad x \in \mathbb{R}$$

gegebene (komplexe) Funktion die Fouriertransformierte von μ.
Ist X eine reellewertige Zufallsvariable mit Verteilung $\mathbb{P}_X$, so heißt $\hat{\mathbb{P}}_X$ die charakteristische Funktion von X. Sie wird auch mit φ_X bezeichnet. Es gilt nach dem Transformationssatz

$$\varphi_X(x) = \mathbb{E} e^{ixX}.$$

Zunächst einige Beispiele.

Beispiel 15.8 *1. Wen $a \in \mathbb{R}$ ist und $\mu(a) = \delta_a$ das Dirac–Maß in a, dann gilt*

$$\hat{\delta}_a(x) = e^{ixa}, \qquad x \in \mathbb{R}.$$

2. Für jede diskrete Verteilung $\mu = \sum_{n=1}^\infty \alpha_n \delta_{a_n}$ mit $\alpha_n \geq 0$, $a_n \in \mathbb{R}$ und $\sum \alpha_n = 1$ gilt also

$$\hat{\mu}(x) = \sum_{n=1}^\infty \alpha_n e^{ixa_n}.$$

Insbesondere ist für die Binomialverteilung zu den Parametern $n \in \mathbb{N}$ und $p \in (0,1)$

$$\mu = B(n,p) = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k$$

die Fouriertransformierte durch

$$\hat{\mu}(x) = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{ixk} = ((1-p) + pe^{ix})^n \qquad x \in \mathbb{R}$$

gegeben.

Für die Poisson–Verteilung

$$\mu = \pi_\lambda = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k$$

folgt analog

$$\hat{\mu}(x) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{ixk} = e^{\lambda(e^{ix}-1)} \qquad x \in \mathbb{R}.$$

3. *Als nächstes berechnen wir die Fouriertransformierte der Standard–Normalverteilung. Sei*

$$\mu(dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx.$$

Dann ist die Fouriertransformierte $\hat{\mu}$ gegeben durch

$$\hat{\mu}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixy} e^{-\frac{1}{2}y^2} dy.$$

Dieses Integral berechnen wir im folgenden. Sei hierzu $\Psi := \hat{\mu}$ und $\Phi = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Dann löst Φ die Differentialgleichung

$$y' + xy = 0,$$

wie man durch nachrechenen verifziert. Selbiges gilt auch für Ψ. In der Tat folgt aus

$$\Psi(x) = \int_{-\infty}^{\infty} e^{ixy} \Phi(y) dy$$

durch Differnzieren unter dem Integral

$$\Psi'(x) = i\int_{-\infty}^{\infty} y e^{ixy}\Phi(y)dy$$

und mittels partieller Integration

$$\begin{aligned} x\Psi(x) &= \int_{-\infty}^{\infty} x\, e^{ixy}\Phi(y)dy = [-ie^{ixy}\Phi(y)]_{-\infty}^{\infty} + i \inf_{-\infty}^{\infty} e^{ixy}\Phi'(y)dy \\ &= i \inf_{-\infty}^{\infty} e^{ixy}\Phi'(y)dy, \end{aligned}$$

da Φ im Unendlichen verrschwindet. Da nun

$$\Phi'(y) + y\Phi(y) = 0$$

gilt, folgt $x\Psi(x) = -\Psi'(x)$ für alle $x \in \mathbb{R}$. Da nun $\sqrt{2\pi}\Phi$ und Ψ die vorgelegte Differentialgleichung lösen und beide für $x = 0$ den Wert 1 annehmen, folgt aus dem klassischen Existenz– und Eindeutigkeitssatz für lineare Differentialgleichungen, dass $\Psi = \sqrt{2\pi}\Phi$ gilt, also

$$\hat{\mu} = e^{-\frac{1}{2}x^2}$$

gilt.

Ähnlich zeigt man, dass für die allgemeine Normalverteilung

$$\mu(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2} dx.$$

gilt

$$\hat{\mu}(x) = e^{iax - \frac{1}{2}\sigma^2 x^2}.$$

Wir werden im folgenden noch einige Eigenschaften der Fouriertransformierten sammeln.

Theorem 15.9 *Für die Fouriertransformierte $\hat{\mu}$ eines jeden endlichen Maßes μ auf $\mathbb{R}$ gilt*

1. *$\hat{\mu}$ ist gleichmäßig stetig.*
2. *$|\hat{\mu}(x)| \leq \|\mu\| = \hat{\mu}(0)$ für alle $x \in \mathbb{R}$. Hierbei bezeichnet*

$$\|\mu\| := \mu(\mathbb{R}).$$

3. *$\hat{\mu}$ ist positiv semi–definit, d.h. für endlich viele Punkte $x_1, \ldots, x_n \in \mathbb{R}$ und $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ gilt*

$$\sum_{s,t=1}^{n} \lambda_s \overline{\lambda}_t \hat{\mu}(x_s - x_t) \geq 0.$$

Beweis. 1.) μ ist als endliches Maß auf $\mathbb{R}$ regulär, d.h. es gibt zu jedem $\varepsilon > 0$ eine kompakte Menge $K \subset \mathbb{R}$ mit $\mu(K^c) < \varepsilon$. Somit ist

$$\alpha := \sup\{|y| : y \in K\} < \infty.$$

Damit gilt für $y \in K$, $x_1, x_2 \in \mathbb{R}$ mit $|x_1 - x_2| \leq \delta := \varepsilon/\alpha$

$$|e^{ix_1y} - e^{ix_2y}| \leq |y||x_1 - x_2| \leq \alpha|x_1 - x_2| \leq \varepsilon.$$

Dies impliziert schon die Stetigkeit von $\hat{\mu}$, denn

$$\begin{aligned}
|\hat{\mu}(x_1) - \hat{\mu}(x_2)| &\leq \int |e^{ix_1y} - e^{ix_2y}|\mu(dy) \\
&= \int_K |e^{ix_1y} - e^{ix_2y}|\mu(dy) + \int_{K^c} |e^{ix_1y} - e^{ix_2y}|\mu(dy) \\
&\leq \varepsilon\mu(K) + 2\mu(K^c) \\
&\leq \varepsilon(\|\mu\| + 2)
\end{aligned}$$

2.) Dies folgt unmittelbar aus der Definition der Fouriertransformierten.
3.) Auf der linken Seite der zu zeigenden Ungelichung steht ein μ–Integral mit Integrand

$$f(y) := \sum_{s,t=1}^{n} \lambda_s \overline{\lambda}_t e^{i(x_s - x_t)y} = \sum_{s=1}^{n} \lambda_s e^{ix_sy} \overline{\sum_{t=1}^{n} \lambda_t e^{ix_ty}} = \left| \sum_{s=1}^{n} \lambda_s e^{ix_sy} \right|^2 \geq 0.$$

Somit ist auch $\int f d\mu \geq 0$. ∎

Bemerkung 15.10 *Ein Satz von Bochner besagt, dass umgekehrt auch jede stetig, positiv semidefinite Funktion Fouriertransformierte eines endlichen positiven Maßes auf $\mathbb{R}$ ist. Dies soll hier nicht gezeigt werden.*

Die folgenden Aussagen zeigen, wie sich Operationen mit Maßen in den zugehörigen Fouriertransfomierten widerspiegeln.

Theorem 15.11 *Es seien μ und ν Maße auf $(\mathbb{R}, \mathcal{B})$ und $\alpha \in \mathbb{R}$. Dann gilt:*

i) $\widehat{\mu + \nu} = \hat{\mu} + \hat{\nu}$.

ii) $\widehat{\alpha\mu} = \alpha\hat{\mu}$.

iii) $\widehat{\mu * \nu} = \hat{\mu} \cdot \hat{\nu}$.

Beweis. i) und ii) sind klar.
Für iii) erinnere man sich daran, dass $\mu * \nu$ die Verteilung der Summe zweier unabhängiger nach μ bzw. ν verteilter Zufallsvariablen ist. Also

$$\begin{aligned} \widehat{\mu * \nu} &= \int e^{itx} d(\mu * \nu)(x) = \int \int e^{it(x+y)} d\mu(x) d\nu(y) \\ &= \int e^{itx} d\mu(x) \int e^{ity} d\nu(y) = \hat{\mu} \cdot \hat{\nu} \end{aligned}$$

■

Korollar 15.12 *Es seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen und $S_n = \sum_{j=1}^n$. Dann gilt*

$$\varphi_{S_n} = \prod_{j=1}^n \varphi_{X_j}.$$

Ein wesentlicher Grund für die Nützlichkeit der Fouriertransfomierten ist der folgende Satzt, der besagt, dass die Fouriertransformierte charakteristisch ist für das zugrunde liegende Maß (daher wird sie auch ofr charakteristische Funktion genannt).

Theorem 15.13 (Eindeutigkeitssatz für Fouriertransfomierte) *Es seien μ und ν endliche Maße auf $(\mathbb{R}, \mathcal{B})$ mit $\hat{\mu} = \hat{\nu}$. Dann gilt $\mu = \nu$.*

Korollar 15.14 *Es seien X und Y reelle Zufallsvariablen mit Verteilungen $\mathbb{P}^X$ und $\mathbb{P}^Y$. Gilt dann $\varphi_X = \varphi_Y$, so gilt auch die Identität $\mathbb{P}^X = \mathbb{P}^Y$.*

Zum Beweis von Satz 15.13 benötigen wir vorbereitend das folgende (technische) Lemma.

Lemma 15.15 (Fouriersches Integraltheorem) *Es sei $f : \mathbb{R} \to \mathbb{R}$ Lebesgue–integrierbar und stetig. Definiere*

$$\hat{f}(t) = \int f(x) e^{itx} d\lambda(x).$$

Ist $\hat{f}$ Lebesgue–integrierbar und stetig. Definiere

$$\hat{f}(t) = \int f(x) e^{itx} d\lambda(x).$$

Ist $\hat{f}$ Lebesgue–integrierbar, so gilt für jedes $x \in \mathbb{R}$:

$$f(x) = \frac{1}{2\pi} \int \hat{f}(x) e^{-itx} d\lambda(t).$$

Beweis. Für feste $s, x \in \mathbb{R}$ und alle $t \in \mathbb{R}$ gilt

$$|\hat{f}(x) e^{-itx} e^{-\frac{s^2t^2}{2}}| = |\hat{f}(x)||e^{-\frac{s^2t^2}{2}}| \leq |\hat{f}(x)|.$$

Da $\hat{f}$ Lebesgue–integrierbar ist, definiert

$$\Phi_x(s) = \frac{1}{2\pi} \int \hat{f}(x) e^{-itx} e^{-\frac{s^2t^2}{2}} d\lambda(t)$$

für jedes $x \in \mathbb{R}$ eine Abbildung von den reellen Zahlen in die komplexe Ebene. Wendet man den Satz über dominierte Konvergenz an, so erhält man

$$\lim_{s \to 0} \Phi_x(s) = \frac{1}{2\pi} \int \hat{f}(x) e^{-itx} d\lambda(t).$$

Ferner erhalten wir aus dem Satz von Fubini

$$\begin{aligned} \Phi_x(s) &= \frac{1}{2\pi} \int \left(\int f(y) e^{ity} d\lambda(y) \right) e^{-itx} e^{-s^2t^2/2} d\lambda(t) \\ &= \frac{1}{2\pi} \int f(y) \int e^{it(y-x)} e^{-s^2t^2/2} d\lambda(t) d\lambda(y) \end{aligned}$$

Aus dem obigen Beispiel wissen wir, dass $e^{-\frac{x^2}{2s^2}}$ die Fouriertransformierte der $\mathcal{N}(0, 1/s^2)$–Verteilung, die bekanntlich die Dichte

$$g_{0,1/s^2}(t) = \frac{|s|}{\sqrt{2\pi}} e^{-s^2t^2/2}$$

besitzt. Daher ergibt sich für $x \in \mathbb{R}$ und $s \neq 0$

$$\begin{aligned}
\Phi_x(s) &= \frac{1}{2\pi}\frac{\sqrt{2\pi}}{|s|}\int f(y)\int e^{it(y-x)}d\mathcal{N}(0,1/s^2)(t)d\lambda(y) \\
&= \frac{1}{\sqrt{2\pi}|s|}\int f(y)\hat{\mathcal{N}}(0,1/s^2)(y-x)d\lambda(y) \\
&= \frac{1}{\sqrt{2\pi}|s|}\int f(y)e^{-\frac{(y-x)^2}{2s^2}}d\lambda(y)
\end{aligned}$$

Sei nun $\varepsilon > 0$ vorgelegt. Aus der vorausgesetzten Stetigkeit von f folgt, dass es zu $x \in \mathbb{R}$ ein $\delta_x > 0$ gibt, so dass

$$|f(y) - f(x)| < \varepsilon \qquad \text{falls } y \in A_x := (x - \delta_x, x + \delta_x).$$

Da

$$\frac{1}{\sqrt{2\pi}|s|}\int e^{-\frac{(y-x)^2}{2s^2}}d\lambda(y) = 1$$

folgt für $x \in \mathbb{R}$ und $s \neq 0$

$$\begin{aligned}
|\Phi_x(s) - f(x)| &= \frac{1}{\sqrt{2\pi}|s|}\left|\int (f(y)-f(x))e^{-\frac{(y-x)^2}{2s^2}}d\lambda(y)\right| \\
&\leq \frac{1}{\sqrt{2\pi}|s|}\left[\int_{A_x}|(f(y)-f(x))|\,e^{-\frac{(y-x)^2}{2s^2}}d\lambda(y)\right. \\
&\qquad \left.+\int_{A_x^c}|f(y)|\,e^{-\frac{(y-x)^2}{2s^2}}d\lambda(y) + \int_{A_x^c}|f(x)|\,e^{-\frac{(y-x)^2}{2s^2}}d\lambda(y)\right] \\
&\leq \frac{1}{\sqrt{2\pi}|s|}\left[\int_{A_x}|\varepsilon|\,e^{-\frac{(y-x)^2}{2s^2}}d\lambda(y)\right. \\
&\qquad \left.+\int_{A_x^c}|f(y)|\,e^{-\frac{\delta_x^2}{2s^2}}d\lambda(y) + |f(x)|\int_{A_x^c}e^{-\frac{(y-x)^2}{2s^2}}d\lambda(y)\right] \\
&\leq \frac{1}{\sqrt{2\pi}|s|}\int_{A_x}|\varepsilon|\,e^{-\frac{(y-x)^2}{2s^2}}d\lambda(y) \\
&\qquad +\frac{1}{\sqrt{2\pi}|s|}\int |f(y)|\,e^{-\frac{\delta_x^2}{2s^2}}d\lambda(y) + |f(x)|\,\frac{s^2}{\delta_x^2} \\
&= \varepsilon + \frac{1}{\sqrt{2\pi}|s|}\int |f(y)|\,e^{-\frac{\delta_x^2}{2s^2}}d\lambda(y) + |f(x)|\,\frac{s^2}{\delta_x^2},
\end{aligned}$$

wobei wir in der letzten Ungleichung die Tschbyschev–Ungleichung verwandt haben. Da für jedes $x \in \mathbb{R}$

$$\lim_{s\to 0} \frac{e^{-\delta_x^2/2s^2}}{|s|} = 0$$

gilt ebenso wie

$$\lim_{s\to 0} \frac{|f(x)|}{\delta_x^2} s^2 = 0$$

wird die rechte Seite so klein wie wir wollen. Also gilt

$$\lim_{s\to 0} \Phi_x(s) = f(x).$$

Wegen

$$\lim_{s\to 0} \Phi_x(s) = \frac{1}{2\pi} \int \hat{f}(x) e^{-itx} d\lambda(t).$$

zeigt das die Behauptung. ■
Als zweites Lemma benötigen wir

Lemma 15.16 *Es seien $a < b \in \mathbb{R}$ und $\varepsilon > 0$. Definiere den ε–Hut $g_\varepsilon : \mathbb{R} \to \mathbb{R}$ vermöge*

$$g_\varepsilon(x) = \frac{x-a+\varepsilon}{\varepsilon} 1_{(a-\varepsilon,a)}(x) + 1_{[a,b]}(x) + \frac{b-x+\varepsilon}{\varepsilon} 1_{(b,b+\varepsilon)}(x).$$

Dann ist die Funktion

$$\hat{g}_\varepsilon(t) = \int g_\varepsilon(x) e^{itx} d\lambda(x)$$

Lebesgue–integrierbar über der komplexen Ebene $\mathbb{C}$.

Beweis. Sei $c := \frac{a+b}{2}$ und $d := \frac{b-a}{2}$. Für $t \neq 0$ gilt dann

$$\begin{aligned}
\hat{g}_\varepsilon(t) &= \int g_\varepsilon(x) e^{itx} d\lambda(x) = \int_{a-\varepsilon}^{b+\varepsilon} g_\varepsilon(x) e^{itx} d\lambda(x) \\
&= \int_{-d-\varepsilon}^{d+\varepsilon} g_\varepsilon(y+c) e^{ity} d\lambda(y) e^{itc}
\end{aligned}$$

wobei wir $y = x - c$ substituiert haben. Wegen

$$g_\varepsilon(y+c) = g_\varepsilon(-y+c), \qquad e^{ity} = \cos(ty) + i \sin(ty)$$

und den Symmetrien von cos bzw. sin ergibt sich

$$\begin{aligned}\hat{g}_\varepsilon(t) &= 2e^{itc}\int_0^{d+\varepsilon} g_\varepsilon(y+c)\cos(ty)dy \\ &= 2e^{itc}\left[\int_0^d \cos(ty)dy + \int_d^{d+\varepsilon}\frac{d+\varepsilon-y}{\varepsilon}\cos(ty)dy\right]\end{aligned}$$

Wegen $\frac{d}{dy}(\frac{1}{t}\sin ty) = \cos ty$ und

$$\frac{d}{dy}\left(\frac{d+\varepsilon-y}{\varepsilon}\sin ty - \frac{1}{\varepsilon t^2}\cos ty\right) = \frac{d+\varepsilon-y}{\varepsilon}\cos ty$$

folgt

$$\begin{aligned}\hat{g}_\varepsilon(t) &= 2e^{itc}\left[(\frac{1}{t}\sin ty)\Big|_0^d + \left(\frac{d+\varepsilon-y}{\varepsilon}\sin ty - \frac{1}{\varepsilon t^2}\cos ty\right)\Big|_d^{d+\varepsilon}\right] \\ &= 2e^{itc}\left[\frac{1}{t}\sin td - \frac{1}{\varepsilon t^2}\cos t(d+\varepsilon) - \frac{1}{t}\sin td + \frac{1}{\varepsilon t^2}\cos td\right] \\ &= 2e^{itc}\frac{1}{\varepsilon t^2}(\cos td - \cos t(d+\varepsilon)).\end{aligned}$$

Somit ist

$$|\hat{g}_\varepsilon(t)| = \begin{cases} 4/\varepsilon t^2 & \text{für alle } t \neq 0 \\ 2(\varepsilon+d) & \text{für } |t| \leq 1/(d+\varepsilon) \end{cases}$$

also ist $\hat{g}_\varepsilon(t)$ Lebesgue–integrierbar, wie behauptet. ■

Nun können wir den Beweis des Eindeutigkeitssatzes erbringen.

Beweis von Satz 15.13 . Seien $a < b \in \mathbb{R}$. Es seien ferner g_ε und $\hat{g}_\varepsilon$ definiert wie im vorherigen Lemma. Aus der Stetigkeit von g_ε und den vorhergehenden Lemmata folgt:

$$g_\varepsilon(x) = \frac{1}{2\pi}\int \hat{g}_\varepsilon(t)e^{-itx}d\lambda(t).$$

Damit folgt für zwei endliche Maße μ und ν

$$\begin{aligned}\int g_\varepsilon(x)d\mu(x) &= \int\frac{1}{2\pi}\int e^{-itx}d\lambda(t)d\mu(x)\hat{g}_\varepsilon(t)e^{-itx}d\lambda(t) \\ &= \int\frac{1}{2\pi}\int e^{-itx}d\mu(x)hatg_\varepsilon(t)e^{-itx}d\lambda(t) \\ &= \int\frac{1}{2\pi}\int \hat{\mu}(-t)\hat{g}_\varepsilon d\lambda(t)\end{aligned}$$

und

$$\int g_\varepsilon(x)d\nu(x) = \int \frac{1}{2\pi}\int \hat{\nu}(-t)\hat{g}_\varepsilon d\lambda(t).$$

Aus $\hat{\mu} = \hat{\nu}$ folgt also mit dem Satz von der majorisierten Konvergenz für das Intervall $A = [a, b]$:

$$\begin{aligned}
\mu(A) &= \lim_{\varepsilon\to 0}\int g_\varepsilon(x)d\mu(x) = \lim_{\varepsilon\to 0}\int \frac{1}{2\pi}\int \hat{\mu}(-t)\hat{g}_\varepsilon d\lambda(t) \\
&= \lim_{\varepsilon\to 0}\int \frac{1}{2\pi}\int \hat{\nu}(-t)\hat{g}_\varepsilon d\lambda(t) = \lim_{\varepsilon\to 0}\int g_\varepsilon(x)d\nu(x) = \nu(A).
\end{aligned}$$

Da $a, b \in \mathbb{R}$ beliebig waren, folgt der Beweis. ■
Zum Beweis des zentralen Grenzwertsatzes mit Mitteln der Forurieranalysis benötigen wir noch einen zweiten Baustein. Diesen wollen wir folgendermaßen vorbereiten: Wendet man das Protmanteau–Theorem auf die Funktionen $\sin(tx)$ und $\cos(tx)$ an, so erhält man

Korollar 15.17

- *Es seien $\mathbb{P}_N, \mathbb{P}$ Wahrscheinlichkeitsverteilungen über $(\mathbb{R}, \mathcal{B})$, so dass $(\mathbb{P}_N)$ in Verteilung gegen $\mathbb{P}$ konvergiert. Dann gilt auch für die Fouriertransformierten*

$$\hat{\mathbb{P}}_N(t) \to \hat{\mathbb{P}} \qquad \textit{für alle} t \in \mathbb{R}.$$

- *Es seien X_n, X Zufallsvariablen , so dass X_n in Verteilung gegen X konvergiere, dann gilt auch*

$$\varphi_{X_n}(t) \to \varphi_X(t) \qquad \textit{für alle} t \in \mathbb{R}.$$

Wir fragen uns nun, ob wir vielleicht auch umgekehrt von der Konvergenz der Fouriertransfomierten auf die Konvergenz der Verteilungen schließen können. Vorbereitend benötigen wir hierzu

Theorem 15.18 (Satz von Helly) *Es sei (F_n) eine Folge von Verteilungsfunktionen. Dann existiert eine Teilfolge $(F_{n_j})_j$ und eine Funktion $F : \mathbb{R} \to \mathbb{R}$ mit:*

- *F ist rechtsseitig stetig monoton und nimmt Werte in $[0, 1]$ an.*
- *$F_{n_j}(x) \to F(x)$ in allen Stetigkeitspunkten x von F.*

Zum Beweis beötigen wir das folgende Lemma, das nicht bewiesen werden soll.

Lemma 15.19 *Es sei* $\varrho_0(r,s) := \frac{|r-s|}{1+|r-s|}$ *(*$r,s \in \mathbb{R}$*) und für* $x,y \in \mathbb{R}^{\mathbb{N}}$ *sei*

$$\varrho(x,y) := \sum_{j\geq 1} \varrho(x_j,y_j)/2^j.$$

Dann besitzt $A \subseteq \mathbb{R}^{\mathbb{N}}$ *genau dann einen kompakten Abschluss bzgl.* ϱ*, wenn die Mengen* $\{x_j : x \in A\}$ *für alle* j *beschränkt sind.*

Beweis des Satzes von Helly, Theorem 15.18. Es sei $r_1, r_2, \ldots$ eine Abzählung von $\mathbb{Q}$. Da die F_n als Verteilungsfunktionen allesamt beschr änkt sind, existiert nach dem vorigen Lemma eine Folge (n_j) und $y \in \mathbb{R}^{\mathbb{N}}$, so dass

$$\lim_{j\to\infty}(F_{n_j}(r_1), F_{n_j}(r_2), \ldots) = (y_1, y_2, \ldots)$$

in der ϱ–Norm. Definiere

$$F_0(r_i) := y_i,$$

dann ist

$$\lim_{j\to\infty}(F_{n_j}(r) = F_0(r)$$

für alle $r \in \mathbb{Q}$. Nun sind die F_n Verteilungsfunktionen. Wir behaupten, dass daher

$$F(x) := \inf\{F_0(r) : r \in \mathbb{Q}, r > x\}$$

die gewünschten Eigenschaften hat:
1.) Zu $x \in \mathbb{R}$ und $\varepsilon > 0$ gibt es ein $s \in \mathbb{Q}$ mit $F(x) \leq F_0(s) < F(x) + \varepsilon$. Für alle y mit $x \leq y < s$ gilt dann

$$F(y) := \inf\{F_0(r) : r \in \mathbb{Q} r > y\} \geq F(y)$$

und $F(y) \leq F_0(s)$. Also

$$F(x) \leq F(y) < F(x) + \varepsilon.$$

Somit ist F rechtsseitig stetig.
2.) F ist schwach monoton steigend, da F_0 die Eigenschaft besitzt.
3.) Es gilt $0 \leq F(x) \leq 1$, da die F_n die Eigenschaft besitzen.

Es sei nun x ein Stetigkeitspunkt von F und $\varepsilon > 0$. Dann gibt es es $r, s \in Q$ mit $r < x < s$ und

$$F(x) - \varepsilon < F_0(s_1) \leq F_0(s_2) < F(x) + \varepsilon.$$

Dabei gilt für alle j

$$F_{n_j}(s_1) \leq F_{n_j}(x) \leq F_{n_j}(s_2).$$

Also folgt

$$\begin{aligned} F(x) - \varepsilon &< \lim_{j\to\infty} F_{n_j}(s_1) \leq \lim_{j\to\infty} F_{n_j}(x) \\ &\leq \lim_{j\to\infty} F_{n_j}(s_2) < F(x) + \varepsilon. \end{aligned}$$

Da $\varepsilon > 0$ beliebig war, folgt die Behauptung. ■
Zuletzt beötigen wir auch noch dass folgende

Lemma 15.20 *Die im Beweis von Satz 15.18 konstruierte Funktion F hat eine Menge von Steigkeitspunkten $C(F), diedichtin \mathbb{R}$ liegen.*

Beweis. Definiere

$$D := \{y \in \mathbb{R} : F \text{ ist nicht (rechtsseitig) stetig in } y\}.$$

Dann ist D abzählbar, da es höchstens n Sprünge der Höhe $1/n$ geben kann. Daraus aber folgt schon die Behauptung. ■
Nun haben wir alles bereit gestellt, um den folgenden Satz zu beweisen.

Theorem 15.21 (Stetigkeitssatz) *Es sei $\mathbb{P}_n$ eine Folge von Wahrscheinlichkeitsverteilungen über $(\mathbb{R}, \mathcal{B})$. Es gilt: $\mathbb{P}_n$ konvergiert genau dann in Verteilung gegen eine Wahrscheinlichkeitsverteilung $\mathbb{P}$, wenn die Folge der Fouriertransformierten $\hat{\mathbb{P}}_n$ gegen eine Funktion φ konvergiert, die in 0 stetig ist. φ ist dann die Fouriertransformierte von $\mathbb{P}$.*

Beweis. Dass aus der Verteilungskonvergenz die Konvergenz der Fouriertransformierten folgt, hatten wir ja schon aus dem Portmanteautehorem abgeleitet. Die Fouriertransfomierte ist nach den einleitenden Bemerkungen zu Fouriertransfomierten stetig, also auch stetig in 0.
Konvergiere nun umgekehrt die Folge der $\hat{\mathbb{P}}_n$ für alle $t \in \mathbb{R}$ gegen eine Funktion φ, die in 0 stetig ist. Sei F_n die Folge der zugehörigen Verteilungsfunktionen.

Wir behaupten: Zu gegebenem $\varepsilon > 0$ existiert eine kompakte Menge $K \subset \mathbb{R}$ mit

$$\mathbb{P}_n(K) \geq 1 - \varepsilon$$

für alle $n \in \mathbb{N}$ (dies nennt man die Straffheit der Folge $(\mathbb{P}_n)$. Dies leiten wir folgendermaßen her. Da $\varphi(0) = \lim \hat{\mathbb{P}}_n(0) = 1$ und φ stetig ist in 0, existiert ein $\delta > 0$ mit

$$\frac{1}{\delta} \int_{-\delta}^{\delta} |1 - \varphi(t)| d\lambda(t) < \frac{\varepsilon}{2}.$$

Aus dem Satz über dominierte Konvergenz folgt die Existenz eines n_0, so dass

$$\frac{1}{\delta} \int_{-\delta}^{\delta} |1 - \hat{\mathbb{P}}_n(t)| d\lambda(t) < \varepsilon.$$

für alle $n \geq n_0$. Mit Hilfe des Satzes von Fubini ergibt sich also

$$\begin{aligned}
\frac{1}{\delta} \int_{-\delta}^{\delta} |1 - \hat{\mathbb{P}}_n(t)| d\lambda(t) &= \int_{\mathbb{R}} \frac{1}{\delta} \int_{-\delta}^{\delta} |1 - e^{itx}| d\lambda(t) d\mathbb{P}_n(x) \\
&= \int_{\mathbb{R}} \left(2 - \frac{2}{x} \frac{\sin \delta x}{\delta}\right) d\mathbb{P}_n(x) \\
&\geq 2 \int_{\{x : |x| > \frac{2}{\delta}\}} \left(1 - \frac{1}{|x\delta|}\right) d\mathbb{P}_n(x) \\
&\geq \mathbb{P}_n\left(\{x : |x| > \frac{2}{\delta}\}\right).
\end{aligned}$$

Damit ergibt sich für $K' := [-\frac{2}{\delta}, \frac{2}{\delta}]$:

$$\mathbb{P}_n(K') = 1 - \mathbb{P}_n\left(\{x : |x| > \frac{2}{\delta}\}\right) > 1 - \varepsilon$$

für alle $n \geq n_0$. Durch eventuelle Vergrößerung kann man K' nun so zu einem abgeschlossenen Intervall K erweitern, dass auch

$$\mathbb{P}_n(K) > 1 - \varepsilon$$

für alle $n \in \mathbb{N}$ erfüllt. Damit ist die Zwischenbehauptung bewiesen.
Wir setzen den Beweis nun folgendermaßen fort: Nach dem Satz von Helly existiert eine Teilfolge $(F_{n_j})_j$ der Folge $(F_n)_n$ und eine rechtseitig stetige, monotone Funktion $F_0 : \mathbb{R} \to [0, 1]$ mit

$$\lim_{j \to \infty} F_{n_j}(x) = F_0(x) \qquad \text{für alle } x \text{ in denen } F_0 \text{ stetig ist.}$$

Um zu zeigen, dass in der Tat F eine Verteilungsfunktion ist, muss noch gezeigt werden, dass

$$\lim_{x\to-\infty} F_0(x) = 0 \qquad \text{und} \qquad \lim_{x\to\infty} F_0(x) = 1$$

gilt. Dazu sei $\varepsilon > 0$ beliebig. Nach der soeben bewiesenen Zwischenbehauptung gibt es eine kompakte Menge $K \subset \mathbb{R}$ mit $\mathbb{P}_n(K) > 1-\varepsilon$ für alle $n \in \mathbb{N}$. Nach der vorhergehenden Lemma gibt es einen Stetigkeitspunkt z von F_0 mit $F \subset \{x \in \mathbb{R} : x \leq z\}$ und somit mit

$$F_{n_j}(z) \geq 1-\varepsilon \qquad \text{für alle } j \in \mathbb{N}$$

sowie $F_0(z) = \lim_{j\to\infty} F_{n_j}(z) \geq 1-\varepsilon$. Aus der Monotonie folgt

$$F_0(x) \geq 1-\varepsilon \qquad \text{für alle } x \geq z,$$

also

$$\lim_{x\to\infty} F_0(x) = 1.$$

Die zweite Eigenschaft zeigt man analog.
Nun gehört zu F_0 eine Wahrscheinlichkeitsverteilung $\mathbb{P}_0$. Für diese gilt dann, dass

$$\mathbb{P}_{n_j} \to \mathbb{P}_0 \qquad \text{in Verteilung.}$$

Nach dem ersten Teil des Satzes gilt dabei

$$\varphi_0(t) = \lim_{n\to\infty} \hat{\mathbb{P}}_n(t) = \lim_{j\to\infty} \hat{\mathbb{P}}_{n_j}(t) = \hat{\mathbb{P}}_0(t)$$

f ür alle $t \in \mathbb{R}$. Insbesondere ist also φ_0 die Fouriertransfomierte von $\mathbb{P}_0$. Dann folgt aber aus dem Eindeutigkeitssatz für Fouriertransfomierte, dass auch jede andere konvergente Teilfolge von (F_n) für alle Stetigkeitspunkte von F_0 gegen F_0 konvergiert. Somit folgt wegen der Beschränktheit der F_n:

$$\lim_{n\to\infty} F_n(x) = F_0(x)$$

für alle Stetigkeitspunkte x von F_0 und somit

$$\mathbb{P}_n \to \mathbb{P}_0 \qquad \text{in Verteilung,}$$

was zu zeigen war. ∎

Wir können nun einen zweiten Beweis des Zentralen Grenzwertsatzes angeben.

Zweite Beweis des Zentralen Grenzwertsatzes 15.1. Wir schreiben kurz $\mu := \mathbb{E}X_1$ und $\sigma^2 = \mathbb{V}X_1$. Die Beweisstrategie ist es zu zeigen, dass die Fouriertransformierte von

$$S_n := \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n\sigma^2}}$$

gegen die Fouriertransformierte der Standardnormalverteilung konvergiert. Sei φ_{S_n} diese Fouriertransformierte (wir werden ganz generell in diesem Beweis die Fouriertransformierte einer Zufallsvariablen Y mit $\varphi_Y(\cdot)$ bezeichnen). Dann gilt durch eine einfache Transformation und unter Ausnutzung der Unabhängigkeit der X_i:

$$\begin{aligned}
\varphi_{S_n}(t) &= \varphi_{\sum_{i=1}^n (X_i-\mu)}(t/\sqrt{\sigma^2}) \\
&= \prod_{i=1}^n \varphi_{(X_i-\mu)}(t/\sqrt{\sigma^2})
\end{aligned}$$

Taylorentwicklung und Ausnutzung der identischen Verteilung und Zentrierung der $X_i - \mu$ ergibt:

$$\begin{aligned}
\varphi_{S_n}(t) &= \left(1 + i\frac{1}{\sqrt{n\sigma^2}}\mathbb{E}(X_i - \mu) - \frac{1}{2}\frac{t^2}{n\sigma^2}\mathbb{E}((X_i-\mu)^2) + r(t/\sqrt{n\sigma^2})\right)^n \\
&= \left(1 - \frac{1}{2}\frac{t^2}{n\sigma^2}\sigma^2) + r(t/\sqrt{n\sigma^2})\right)^n
\end{aligned}$$

mit einem Restterm $r(t/\sqrt{n\sigma^2})$, der der Identität

$$\lim_{t/\sqrt{n\sigma^2}\to 0} \frac{nr(t/\sqrt{n\sigma^2})}{t^2} = 0$$

genügt. Nun ist

$$\lim_{n\to\infty} \left(1 - \frac{1}{2}\frac{t^2}{n\sigma^2}\sigma^2) + r(t/\sqrt{n\sigma^2})\right)^n = e^{-\frac{1}{2}t^2}$$

also

$$\lim_{n\to\infty} \varphi_{S_n}(t) = e^{-\frac{1}{2}t^2}.$$

Dies aber ist die Fouriertransformierte der $\mathcal{N}(0,1)$–Verteilung. Der Stetigkeitssatz liefert die Behauptung ■

Wir beenden dieses Kapitel mit einer kurzen Diskussion der Approximationsgeschwindigkeit im Zentralen Grenzwertsatz. In der Tat. will man diesen anwenden, beispielsweise für Folgen von i.i.d. Zufallsvariablen , so ist nicht nur die Frage, dass die richtig gewichtete Summe in Verteilung gegen die Normalverteilung konvergiert von Interesse, sondern auch die Frage, wie schnell diese Konvergenz vonstatten geht. Dies ist der Inhalt des Satzes von Berry und Esseen:

Theorem 15.22 *(Berry/Essen): Es sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von i.i.d. Zufallsvariablen mit $\mathbb{E}X_1^3 < \infty$.Dann gilt für jede $\mathcal{N}(0,1)$–verteilte Zufallsvariable Z:*

$$\sup_{a\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\sum_{i=1}^n X_i - \mathbb{E}X_1}{\sqrt{n\mathbb{V}X_1}} \leq a\right) - \mathbb{P}\left(Z \leq a\right)\right| \leq \frac{C}{\sqrt{n}}\mathbb{E}(|X_1 - \mathbb{E}X_1|^3).$$

Es ist relativ eionfach zu beweisen, dass der numerische Wert der obigen Konstante C kleiner ist als 6 und man kann ebenso zeigen, dass er größer ist als 0.4. Kürzlich konnte Chistyakov zeigen, dass der Wert für symmetische Verteilungen $C = \frac{1}{\sqrt{2\pi}}$ ist und damit eine alte Vermutung von Kolomogorov beweisen.