

VO Integration und Stochastik  
Universität Wien, WS 2022

R. Zweimüller

Dieses Werk ist lizenziert unter einer (CC-BY)  
Creative Commons Namensnennung 4.0 International Lizenz.

April 4, 2023



# Contents

---

<b>1</b>	<b>Das Lebesgue-Maß</b>	<b>5</b>
1.1	Vorschau auf die Maßtheorie . . . . .	5
1.2	Das Maßproblem . . . . .	8
1.3	Das Lebesgue-Borel Prämaß . . . . .	10
1.4	Das Lebesguesche äußere Maß . . . . .	15
1.5	Lebesgue-messbare Mengen . . . . .	19
1.6	Algebren und $\sigma$ -Algebren . . . . .	20
1.7	Das Lebesgue-Maß . . . . .	23
<b>2</b>	<b>Ein wenig abstrakte Maßtheorie</b>	<b>27</b>
2.1	Halbringe, Prämaße und Maße . . . . .	27
2.2	Äußere Maße und Fortsetzungssatz . . . . .	30
2.3	Anwendung: Lebesgue-Stieltjes Maße auf $\mathbb{R}$ . . . . .	35
2.4	Erzeugte $\sigma$ -Algebren . . . . .	39
2.5	Messbare Abbildungen . . . . .	42
<b>3</b>	<b>Das Lebesgue-Integral</b>	<b>47</b>
3.1	Die Grundidee - Einfache Funktionen . . . . .	47
3.2	Definition des Integrals . . . . .	49
3.3	Riemann-Integral und Unendliche Reihen . . . . .	54
3.4	Konvergenzsätze und Transformationsformel . . . . .	57
3.5	Mehr Beispiele und Anwendungen . . . . .	61
3.6	Räume integrierbarer Funktionen . . . . .	67
3.7	Produkträume und Satz von Fubini . . . . .	75
3.8	Die Transformationsformel in $\mathbb{R}^d$ . . . . .	82
<b>4</b>	<b>Ein Blick in die Stochastik</b>	<b>91</b>
4.1	Grundbegriffe - Der formale Rahmen . . . . .	91
4.2	Bedingte Wahrscheinlichkeiten und Unabhängigkeit . . . . .	94
4.3	Zufallsvariable, Erwartungswert und Momente . . . . .	98
4.4	Das Gesetz der Großen Zahl . . . . .	108
4.5	Der Zentrale Grenzwertsatz . . . . .	111



# 1. Das Lebesgue-Maß

## 1.1 Vorschau auf die Maßtheorie

Ausgangspunkt der Maßtheorie ist die ganz naheliegende und wichtige Frage, wie man die Größe von Mengen in sinnvoller und konsistenter Weise messen kann. Ist  $X$  eine "Grundmenge",  $\mathcal{P}(X)$  ihre Potenzmenge, so können wir einer beliebigen Teilmenge  $A \in \mathcal{P}(X)$  von  $X$  etwa die Anzahl  $\#A$  ihrer Elemente zuordnen, um eine *Größenfunktion*  $\mu : \mathcal{P}(X) \rightarrow [0, \infty]$  zu erhalten,  $\mu(A) := \#A$ . Evident gilt  $\mu(\emptyset) = 0$  und  $\mu$  ist *paarweise additiv*, dh sind  $A, B \in \mathcal{P}(X)$  disjunkt, so folgt  $\mu(A \cup B) = \mu(A) + \mu(B)$ , und generell *endlich additiv*, dh sind  $n \in \mathbb{N}$  und  $A_1, \dots, A_n$  *paarweise disjunkt* (dh  $A_i \cap A_j = \emptyset$  falls  $i \neq j$ ), dann gilt  $\mu(A_1 \cup \dots \cup A_n) = \mu(A_1) + \dots + \mu(A_n)$ . Ganz allgemein werden wir von einer Funktion welche die Größe von Mengen angibt erwarten, dass sie diese einfachen Grundeigenschaften besitzt. Tatsächlich gilt entsprechendes hier sogar für die Vereinigung unendlich vieler paarweise disjunkter Mengen. Die Funktion  $\mu$  ist *abzählbar additiv* ( $\sigma$ -additiv) indem für jede *disjunkte Folge* (Abkürzung für Folge paarweise disjunkter Mengen)  $A_1, A_2, \dots$  stets  $\mu(\bigcup_{n \geq 1} A_n) = \sum_{n \geq 1} \mu(A_n)$  gilt.

Wir können noch einen weiteren rein mengentheoretischen Größenbegriff ansetzen, diesmal einen der es erlaubt, gewisse unendliche Mengen voneinander zu unterscheiden: definieren wir  $\mu : \mathcal{P}(X) \rightarrow [0, \infty]$  indem wir  $\mu(A) := 0$  setzen falls  $A$  *abzählbar* (dh endlich oder abzählbar unendlich) ist, und  $\mu(A) := \infty$  sonst, so bekommen wir erneut eine endlich additive Größenfunktion. Ist dieses  $\mu$  auch  $\sigma$ -additiv? (Siehe PS.)

**Bemerkung 1.1 (Rechnen mit  $\infty$ )** Wir sollten uns kurz klarmachen, wie mit dem Wert  $\infty$  (den wir ja explizit zulassen wollen um über unendlich große Mengen sprechen zu können), umzugehen ist. Was die Addition betrifft, sei wie üblich  $a + \infty = \infty + a = \infty = \infty + \infty$  für jedes  $a \in \mathbb{R}$ , und natürlich  $\sum_{n \geq 1} a_n = \infty$  falls alle  $a_n \in [0, \infty]$  und wenigstens ein  $a_n = \infty$ . Vor  $\infty - \infty$  werden wir uns hüten. Für die Multiplikation treffen wir neben den naheliegenden Regeln  $a \cdot \infty = \infty$  und  $(-a) \cdot \infty = -\infty$  für  $a \in (0, \infty)$  aber in der Maßtheorie die Vereinbarung, dass  $0 \cdot \infty$  (und  $\infty \cdot 0$ ), wann immer wir diesen Ausdruck zulassen, als 0 zu interpretieren ist:

$$0 \cdot \infty = \infty \cdot 0 := 0.$$

Das ist einfach praktisch, für Summen konstanter Terme  $a_n = a$  gilt dann  $\sum a_n = \#M \cdot a$

$$\sum_{n \in M} a = \#M \cdot a \quad \text{für } a \in [0, \infty] \text{ und } M \subseteq \mathbb{N}$$

auch wenn  $\#M$  oder  $a$  zu  $\{0, \infty\}$  gehört. Prägen Sie sich diese Konvention bitte gut ein.

Ein ganz anderer grundlegender Größenbegriff, speziell für Teilmengen von  $X := \mathbb{R}$ , ist jener der elementargeometrischen Länge. Für ein Intervall  $I := (a, b] \subseteq \mathbb{R}$  ( $a \leq b$ ) ist jedenfalls klar, dass wir unter seiner *Länge* die Zahl  $\lambda(I) := b - a$  verstehen wollen, wir erhalten also eine Größenfunktion  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  die zunächst nur auf der Familie  $\mathcal{I}_{\mathbb{R}}$  solcher Intervalle definiert ist. (Ergänzend vereinbaren wir, dass  $(a, b] := \emptyset$  falls  $a > b$ .) Wahrscheinlich glauben Sie mir gerne, dass auch  $\lambda$  endlich additiv ist. Diese Funktion scheint auch  $\sigma$ -additiv zu sein, jedenfalls klappt es im folgenden

**Beispiel 1.1** Das Intervall  $A := (0, 1]$  können wir als Vereinigung der paarweise disjunkten Teilmengen  $A_n := (\frac{1}{n+1}, \frac{1}{n}]$ ,  $n \geq 1$  ansehen. Dabei finden wir wirklich

$$\sum_{n \geq 1} \lambda(A_n) = \lim_{N \rightarrow \infty} \sum_{n=1}^N \lambda(A_n) = \lim_{N \rightarrow \infty} \sum_{n=1}^N \left( \frac{1}{n} - \frac{1}{n+1} \right) = \lim_{N \rightarrow \infty} \left( 1 - \frac{1}{N+1} \right) = 1 = \lambda(A) \quad (\text{Teleskopsumme}).$$

Allerdings sieht längst nicht jede abzählbare Zerlegung eines Intervalles so einfach aus! Im allgemeinen ist es nicht möglich die Unterteilungspunkte der Lage nach so anzuordnen, dass man eine Teleskopsumme erhält. Für einen Beweis müssen wir uns schon ein wenig mehr bemühen, aber das klären wir demnächst. Ausserdem wollen wir nicht nur Intervalle messen. Die Frage, wie man am besten mit komplizierteren Mengen umgeht, ist das zentrale Thema des ersten Kapitels.

Allgemeiner können wir für halboffene  $d$ -dimensionale Quader in  $\mathbb{R}^d$ , dh für Cartesische Produkte der Form  $\prod_{i=1}^d (a_i, b_i]$  vom elementaren  $d$ -dimensionale Volumen  $\lambda^d(\prod_{i=1}^d (a_i, b_i]) := \prod_{i=1}^d (b_i - a_i)$  sprechen. Wir bekommen damit eine Größenfunktion  $\lambda^d : \mathcal{I}_{\mathbb{R}}^d \rightarrow [0, \infty]$  auf der Familie  $\mathcal{I}_{\mathbb{R}}^d$  solcher Mengen. Viele der unten im Fall von  $\lambda = \lambda^1$  besprochenen Ideen lassen sich mit moderatem Aufwand auf  $\lambda^d$  übertragen.

Aufgrund ihres geometrischen Ursprunges besitzen die Mengenfunktionen  $\lambda^d$  zusätzlich die Eigenschaft der *Translationsinvarianz*, dh verschiebt man eine Menge  $A \in \mathcal{I}_{\mathbb{R}}^d$  mittels eines Vektors  $t \in \mathbb{R}^d$ , so verändert dies deren Größe nicht, es gilt also  $\lambda^d(A + t) = \lambda^d(A)$  für  $t \in \mathbb{R}^d$ , wobei  $A + t := \{a + t : a \in A\}$  die um  $t$  verschobene Menge bezeichnet.

Die Beispiele  $\lambda^d$  illustrieren noch einen wichtigen Punkt: Häufig ist eine Größenfunktion zunächst nicht auf  $\mathcal{P}(X)$ , sondern nur auf einer kleineren Familie  $\mathcal{E} \subseteq \mathcal{P}(X)$  von Teilmengen definiert. daher sollten wir die oben genannten Eigenschaften etwas präziser formulieren. Es geht dabei immer um Vereinigungen paarweise disjunkter Mengen. Der Einfachheit halber sprechen wir dann kurz von *disjunkten Vereinigungen*, und es ist gelegentlich nützlich, diese besondere Situation durch ein eigenes Symbol auszudrücken und anstelle von  $A \cup B$ ,  $\bigcup_{n=1}^N A_n$  und  $\bigcup_{n \geq 1} A_n$  manchmal  $A \uplus B$ ,  $\biguplus_{n=1}^N A_n$  bzw  $\biguplus_{n \geq 1} A_n$  zu schreiben. Durch diese Notation drückt man also zusätzlich aus, dass es zwischen den betreffenden Mengen keine Überlappungen gibt. Alternativ findet man auch die Schreibweise  $A \cup B$  (disjunkt),  $\bigcup_{n=1}^N A_n$  (disjunkt) bzw  $\bigcup_{n \geq 1} A_n$  (disjunkt). (Aber natürlich ist es weiterhin korrekt und zulässig, in solchen Situationen einfach  $A \cup B$  etc zu schreiben.)

**Definition 1.1** Für  $\mathcal{E} \subseteq \mathcal{P}(X)$  nennt man eine Mengenfunktion  $\mu : \mathcal{E} \rightarrow [0, \infty]$

1. paarweise additiv (auf  $\mathcal{E}$ ), falls

$$\mu(A \uplus B) = \mu(A) + \mu(B) \quad \text{für disjunkte } A, B \in \mathcal{E} \text{ mit } A \cup B \in \mathcal{E}, \quad (1.1)$$

2. endlich additiv (auf  $\mathcal{E}$ ), falls

$$\mu \left( \biguplus_{n=1}^N A_n \right) = \sum_{n=1}^N \mu(A_n) \quad \text{für } N \in \mathbb{N} \text{ und paarweise disjunkte } A_1, \dots, A_N \in \mathcal{E} \text{ mit } \bigcup_{n=1}^N A_n \in \mathcal{E}, \quad (1.2)$$

3.  $\sigma$ -additiv (auf  $\mathcal{E}$ ), falls

$$\mu \left( \biguplus_{n \geq 1} A_n \right) = \sum_{n \geq 1} \mu(A_n) \quad \text{für paarweise disjunkte } A_1, A_2, \dots \in \mathcal{E} \text{ mit } \bigcup_{n \geq 1} A_n \in \mathcal{E}. \quad (1.3)$$

Die  $\sigma$ -Additivität erweist sich als die "richtige" Bedingung für einen nützlichen Größenbegriff, es handelt sich um eine Stetigkeitseigenschaft, die ein reibungsloses Zusammenspiel mit Grenzübergängen, also mit der Analysis, sicherstellt. Unter einem *Maß* verstehen wir später eine Größenfunktion, welche unter anderem auch

diese entscheidende Eigenschaft besitzt.

Dadurch haben wir es allerdings sehr häufig mit unendlichen Reihen anstelle endlicher Summen zu tun! Wie Sie wissen, ist da Vorsicht geboten, weil zB der (Grenz-)wert einer konvergenten Reihe i.a. von der Anordnung der Summanden abhängt. Würde solches in Zusammenhang mit (1.3) passieren, wäre dies natürlich haarsträubend, hängt doch die linke Seite nicht von der Reihenfolge der Mengen  $A_n$  ab. Wir überzeugen uns daher lieber gleich davon, dass hier keine Gefahr droht. Dies liegt daran, dass wir uns auf nichtnegative Summanden konzentrieren, und um solche kümmert sich der folgende Satz.

**Satz 1.1 (Nichtnegative Reihen und Doppelreihen)** Für beliebige nichtnegative Zahlen  $a_n, b_{k,l} \in [0, \infty]$  mit  $k, l, n \in \mathbb{N}$  gelten folgende Aussagen:

a) Der (Grenz-)wert der Reihe  $\sum_{n \geq 0} a_n \in [0, \infty]$  ist immer definiert.

b) Es gilt

$$\sum_{n \geq 0} a_n = \sup \left\{ \sum_{n \in M} a_n : M \subseteq \mathbb{N} \text{ endlich} \right\} \in [0, \infty]. \quad (1.4)$$

c) Für jede Umordnung  $(a_{n_k})_{k \geq 0}$  von  $(a_n)_{n \geq 0}$  ist

$$\sum_{k \geq 0} a_{n_k} = \sum_{n \geq 0} a_n.$$

d) Definiert man die Summe der Doppelfolge  $(b_{k,l})_{k \geq 0, l \geq 0}$  durch

$$\sum_{k \geq 0, l \geq 0} b_{k,l} := \sup \left\{ \sum_{(k,l) \in M} b_{k,l} : M \subseteq \mathbb{N} \times \mathbb{N} \text{ endlich} \right\} \in [0, \infty],$$

dann gilt  $\sum_{k \geq 0, l \geq 0} b_{k,l} = \sum_{n \geq 0} a_n$  für jede Anordnung  $(a_n)$  der  $b_{k,l}$ , und

$$\sum_{k \geq 0, l \geq 0} b_{k,l} = \sum_{k \geq 0} \sum_{l \geq 0} b_{k,l} = \sum_{l \geq 0} \sum_{k \geq 0} b_{k,l}. \quad (1.5)$$

**Beweis.** a) Per Definition ist  $a := \sum_{n \geq 0} a_n = \lim_{N \rightarrow \infty} s_N$  mit  $s_N := \sum_{n=0}^N a_n$ . Die Folge der Partialsummen  $s_N$  ist aber monoton wachsend, und daher entweder konvergent in  $[0, \infty)$ , oder sie divergiert gegen  $\infty$ .

b) Wir schreiben  $s$  für das Supremum rechts. Um zu sehen, dass  $s \geq a$  gilt, beobachten wir einfach, dass  $s \geq s_N$  für jedes  $N \geq 1$ . Es ist ja  $s_N = \sum_{n \in \{0, \dots, N\}} a_n$  eine der Summen in der Definition von  $s$ . Umgekehrt gilt auch  $s \leq a$ . Ist nämlich  $\sum_{n \in M} a_n$  irgendeine der Summen aus der Definition von  $s$ , dann wählen wir  $N := \max M$ , sodass  $M \subseteq \{0, \dots, N\}$  und daher  $\sum_{n \in M} a_n \leq s_N \leq a$  gilt.

c) Die rechte Seite in (1.4) ändert sich nicht, wenn die  $a_n$  umgeordnet werden.

d) Die Aussage über Anordnungen folgt sofort aus b) und c). Interessanter (und oft wichtig) ist die Aussage (1.5) über Doppelreihen. Weil sich (per Definition)  $\sum_{k \geq 0, l \geq 0} b_{k,l}$  nicht ändert wenn wir  $(b_{l,k})_{k \geq 0, l \geq 0}$  anstelle von  $(b_{k,l})_{k \geq 0, l \geq 0}$  nehmen, brauchen wir nur die erste Identität zu prüfen. Sie ist ein wenig heikel weil da unterschiedliche Grenzübergänge hintereinander ausgeführt werden. Ausführlicher geschrieben ist ja

$$\sum_{k \geq 0} \sum_{l \geq 0} b_{k,l} = \lim_{K \rightarrow \infty} \sum_{k=0}^K \left( \lim_{L \rightarrow \infty} \sum_{l=0}^L b_{k,l} \right).$$

d<sub>1</sub>) Wir zeigen zuerst:  $\sum_{k \geq 0, l \geq 0} b_{k,l} \leq \sum_{k \geq 0} \sum_{l \geq 0} b_{k,l}$ . Weil links das genannte Supremum steht, brauchen wir nur zu überprüfen, dass für jede endliche Menge  $M \subseteq \mathbb{N} \times \mathbb{N}$  stets  $\sum_{(k,l) \in M} b_{k,l} \leq \sum_{k \geq 0} \sum_{l \geq 0} b_{k,l}$  gilt. Dazu seien einfach  $K := \max\{k : (k, l) \in M\}$  und  $L := \max\{l : (k, l) \in M\}$ , dann haben wir  $M \subseteq \{0, \dots, K\} \times \{0, \dots, L\}$  und somit tatsächlich

$$\sum_{(k,l) \in M} b_{k,l} \leq \sum_{k=0}^K \sum_{l=0}^L b_{k,l} \leq \sum_{k \geq 0} \sum_{l \geq 0} b_{k,l}.$$

d<sub>2</sub>) Umgekehrt gilt auch  $\sum_{k \geq 0, l \geq 0} b_{k,l} \geq \sum_{k \geq 0} \sum_{l \geq 0} b_{k,l}$ . Das ist offensichtlich im Fall  $\sum_{k \geq 0, l \geq 0} b_{k,l} = \infty$ , und wir nehmen daher an, dass  $\sum_{k \geq 0, l \geq 0} b_{k,l} < \infty$  gilt. Dann muss auch  $\sum_{k \geq 0} \sum_{l \geq 0} b_{k,l} < \infty$  sein: Andernfalls gibt es zu jedem  $N \geq 1$  ein  $K$  mit

$$\sum_{k=0}^K \sum_{l \geq 0} b_{k,l} > N + 1$$

und dann ein  $L$  sodass

$$\sum_{l=0}^L b_{k,l} > \sum_{l \geq 0} b_{k,l} - \frac{1}{K+1} \quad \text{für } k \in \{0, \dots, K\}$$

gilt, insgesamt also  $\sum_{k=0}^K \sum_{l=0}^L b_{k,l} > N$ . Weil  $N$  beliebig war, müsste also das Supremum der Summen über endliche Indexmengen auch  $\infty$  sein, im Widerspruch zur Annahme. Somit ist  $\sum_{k \geq 0} \sum_{l \geq 0} b_{k,l} < \infty$  wie behauptet. Sei nun  $\varepsilon > 0$ . Wir können  $K$  so wählen, dass

$$\sum_{k=0}^K \sum_{l \geq 0} b_{k,l} \geq \sum_{k \geq 0} \sum_{l \geq 0} b_{k,l} - \frac{\varepsilon}{2}.$$

Weiters gibt es ein  $L$  derart, dass

$$\sum_{l=0}^L b_{k,l} \geq \sum_{l \geq 0} b_{k,l} - \frac{\varepsilon}{2} \frac{1}{K+1} \quad \text{für } k \in \{0, \dots, K\}.$$

Dann folgt

$$\sum_{k \geq 0, l \geq 0} b_{k,l} \geq \sum_{k=0}^K \sum_{l=0}^L b_{k,l} \geq \sum_{k \geq 0} \sum_{l \geq 0} b_{k,l} - \varepsilon,$$

und daraus, weil  $\varepsilon > 0$  beliebig war, unsere Behauptung. ■

## 1.2 Das Maßproblem

In diesem ersten Kapitel konzentrieren wir uns hauptsächlich auf den Fall der Längenfunktion  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$ . Weil  $\mathcal{I}_{\mathbb{R}} = \{(a, b] : a \leq b \text{ in } \mathbb{R}\}$  doch recht bescheiden ist, stellt sich schnell die Frage, welchen Mengen  $A \subseteq \mathbb{R}$  wir sonst noch eine Länge zuordnen können ohne dabei die natürlichen Eigenschaften der ursprünglichen Größenfunktion zu verlieren. Sobald wir uns interessanter Mathematik zuwenden, haben wir es schnell mit wesentlich komplizierteren und wilderen Mengen zu tun. Wie schaffen wir es also, die Länge einer beliebig schlimmen Teilmenge von  $\mathbb{R}$  zu definieren?

Leider gar nicht! Es gibt einfach viel zu viele wilde Mengen. Es ist nicht möglich, die oben definierte Längenfunktion  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  zu einer Funktion



$\lambda : \mathcal{P}(\mathbb{R}) \rightarrow [0, \infty]$  fortzusetzen<sup>1</sup>, die  $\sigma$ -additiv und translationsinvariant auf ganz  $\mathcal{P}(\mathbb{R})$  ist. Das *Maßproblem* ist unlösbar!

Dies liegt daran, dass es unheimlich komplizierte Teilmengen von  $\mathbb{R}$  gibt. Erinnern Sie sich bitte daran, dass man beim Umgang mit unendlichen Mengen auch ein wenig aufpassen muss, und sich wiederum in einem geeigneten axiomatischen Rahmen bewegen sollte um widersprüchlich Konzepte (wie zB die Menge aller Mengen) zu vermeiden. Wir verwenden im Weiteren immer ZFC, das übliche Zermelo-Fraenkelsche Axiomensystem inklusive des Auswahlaxioms (axiom of choice, AC). Letzteres spielt oft eine entscheidende Rolle wenn es um unendliche Mengen geht. Dabei klingt es eigentlich recht harmlos:

**Auswahlaxiom (AC).** Zu jeder Familie  $\mathcal{F}$  nichtleerer paarweise disjunkter Mengen  $F$  gibt es eine Menge  $V$ , welche aus jedem  $F \in \mathcal{F}$  genau ein Element enthält (und sonst nichts).

Mit seiner Hilfe beweisen wir nun den grundlegenden

**Satz 1.2 (Satz von Vitali)** *Es gibt keine Funktion  $\mu : \mathcal{P}(\mathbb{R}) \rightarrow [0, \infty]$  welche translationsinvariant und  $\sigma$ -additiv ist, und zudem  $\mu((0, 1]) = 1$  erfüllt.*

**Beweis. (i)** Es sei  $Q$  die abzählbar unendliche Menge  $\mathbb{Q} \cap (-1, 1]$ . Wir verschaffen uns eine Menge  $V \subseteq (0, 1]$  mit folgenden seltsamen Eigenschaften: die verschobenen Kopien  $V + t = \{v + t : v \in V\}$ ,  $t \in Q$ , sind paarweise disjunkt, und es gilt

$$(0, 1] \subseteq \biguplus_{t \in Q} (V + t) \subseteq (-1, 2]. \quad (1.6)$$

Nehmen wir nun an,  $\mu$  wäre eine Funktion mit den genannten Eigenschaften, so finden wir, dass  $\mu(A) \leq \mu(A) + \mu(B \setminus A) = \mu(B)$  für  $A \subseteq B$ , und somit

$$\mu((0, 1]) \leq \mu\left(\biguplus_{t \in Q} (V + t)\right) \leq \mu((-1, 2]) = 3\mu((0, 1]).$$

Wegen  $\sigma$ -Additivität und Translationsinvarianz ist ja  $\mu((-1, 2]) = \mu((-1, 0] \uplus (0, 1] \uplus (1, 2]) = \mu((-1, 0]) + \mu((0, 1]) + \mu((1, 2]) = 3\mu((0, 1])$ . Weiters gilt nun  $\mu(\biguplus_{t \in Q} V + t) = \sum_{t \in Q} \mu(V + t) = \sum_{t \in Q} \mu(V) = \infty \mu(V)$ . Daher ist  $\mu(\biguplus_{t \in Q} V + t) \in \{0, \infty\}$ , und somit auch  $\mu((0, 1]) \in \{0, \infty\}$ , ein Widerspruch.

**(ii)** Um  $V$  zu erhalten, gehen wir folgendermaßen vor (Details im PS).

**a)** Durch  $x \sim y$  wenn  $x - y \in \mathbb{Q}$  wird auf  $\mathbb{R}$  eine Äquivalenzrelation definiert. Es bezeichne  $\mathbb{R}/\sim$  die Menge der Äquivalenzklassen, dann gilt  $\mathbb{R} = \biguplus_{C \in \mathbb{R}/\sim} C$  (paarweise disjunkt). Die Äquivalenzklassen sind genau die Mengen der Form  $C_x := x + \mathbb{Q} = \{x + q : q \in \mathbb{Q}\}$  mit  $x \in \mathbb{R}$ .

**b)** Jedes  $C \in \mathbb{R}/\sim$  enthält Punkte aus  $(0, 1]$ , dh  $C \cap (0, 1] \neq \emptyset$ . Als Konsequenz des Auswahlaxioms gibt es eine *Vitali-Menge*, dh eine Menge  $V \subseteq (0, 1]$  welche aus jeder Äquivalenzklasse  $C$  genau einen Punkt  $x_C$  enthält.

**c)** Für  $p, q \in \mathbb{Q}$  mit  $p \neq q$  gilt  $(V + p) \cap (V + q) = \emptyset$ , wobei  $V + q := \{v + q : v \in V\}$  die um  $q$  verschobene Menge  $V$  ist.

**d)** Schreiben wir kurz  $Q := \mathbb{Q} \cap (-1, 1]$ , dann gilt  $(0, 1] \subseteq \biguplus_{t \in Q} (V + t) \subseteq (-1, 2]$ . ■

Das ist nun ein wenig unangenehm. Es kommt aber noch schlimmer! Um zu illustrieren, wie schlecht die Sache tatsächlich steht, insbesondere in höheren Dimensionen, zitieren wir ein oft als *Banach-Tarski-Paradoxon* bezeichnetes Ergebnis:

<sup>1</sup>Ist  $f$  eine Funktion auf einer Menge  $M$ , so versteht man unter einer *Fortsetzung* von  $f$  nach  $N \supseteq M$  bekanntlich eine Funktion  $g$  auf der größeren Menge  $N$ , deren *Einschränkung*  $g|_M$  auf  $M$  mit  $f$  übereinstimmt. Man bezeichnet eine Fortsetzung gerne wieder mit demselben Symbol, was insbesondere dann gefahrlos möglich ist, wenn diese Fortsetzung durch gewisse Forderungen oder explizite Vereinbarungen *eindeutig bestimmt* ist.

**Satz 1.3 (Satz von Banach & Tarski)** *Es seien  $d \in \mathbb{N}$ , und  $A, B \subseteq \mathbb{R}^d$  nichtleere beschränkte offene Mengen. Dann gibt es eine Zerlegung  $A = \bigsqcup_{n \geq 1} C_n$  von  $A$  in abzählbar viele paarweise disjunkte Mengen  $C_n \subseteq \mathbb{R}^d$ ,  $n \geq 1$ , und dazu Bewegungen  $T_n : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $n \geq 1$ , sodass die Bilder  $T_n C_n$  eine Zerlegung  $B = \bigsqcup_{n \geq 1} T_n C_n$  von  $B$  bilden. Ist  $d \geq 3$ , so reichen sogar endlich viele Mengen  $C_1, \dots, C_m$ .*

Insbesondere läßt sich also zB der offene Einheitswürfel des  $\mathbb{R}^3$  in endlich viele Stücke zerschneiden aus welchen man, geeignet verdreht, verschoben und evtl gespiegelt, den offenen Würfel mit Seitenlänge 2022 zusammensetzen kann! Das "können" ist hier natürlich nicht ganz wörtlich zu nehmen. Konstruktiv kommt man an diese schrecklichen Teilmengen nicht heran. Sie sollten diese Aussage aber unbedingt als Warnung im Hinterkopf behalten, und als hervorragende Motivation, im weiteren nur dem sicheren Boden formaler Argumentation zu vertrauen. (Auch wenn Ihnen letztere manchmal übertrieben pedantisch erscheinen mag - sie ist es eben nicht. Selbst der geschulten Intuition dürfen wir hier nicht leichtfertig vertrauen.) Zudem müssen wir offenbar immer genau darauf achten, mit welcher Art von Mengen wir es zu tun haben, und bei jeder Mengenfunktion entsprechend den Definitionsbereich mitdenken.

Trotz dieser Schwierigkeiten ist die Frage viel zu wichtig und zu interessant, um sie enttäuscht ganz wegzuschieben. Der Maßtheorie gelingt es (unter anderem), ein Längenmaß  $\lambda$  (das zurecht berühmte *Lebesgue-Maß*) für eine sehr große Familie  $\mathcal{L}_{\mathbb{R}} \subseteq \mathcal{P}(\mathbb{R})$  zu konstruieren (jene der *Lebesgue-messbaren Mengen*), die alle Mengen enthält, denen man im mathematischen Alltag begegnet. (Und entsprechendes gilt für  $\lambda^d$ .) Diese Funktion ist eines der wichtigsten Objekte der Mathematik, die Maßtheorie ein tragender Pfeiler der modernen Analysis, und der formale Rahmen der mathematischen Stochastik.

## 1.3 Das Lebesgue-Borel Prämaß

Die Ausführungen des vorigen Abschnittes zeigen, dass wir im Umgang mit Mengenfunktionen sehr vorsichtig sein sollten, und insbesondere immer darauf achten müssen mit welcher Art von Menge wir es gerade zu tun haben. Sehen wir uns nun die elementare Längenfunktion genauer an. Wir überprüfen sorgfältig (also durch formale Argumente), dass diese tatsächlich die oben erwähnten Eigenschaften besitzt - wenigstens dann, wenn wir uns auf ganz einfache Mengen, nämlich die Intervalle aus der Familie  $\mathcal{I}_{\mathbb{R}} = \{(a, b] : a, b \in \mathbb{R}, a \leq b\} \subseteq \mathcal{P}(\mathbb{R})$  beschränken. In diesem simplen Rahmen können wir im Detail ausprobieren, wie man dabei korrekt vorgehen kann. (Wenn Ihnen dabei manches ein wenig übertrieben erscheint, so denken Sie bitte an den vorigen Abschnitt zurück. Ausserdem werden wir die folgenden Argumente später nochmals in allgemeinerer Form benötigen. Da ist es vorteilhaft, diese in einfacher Form schon mal kennenzulernen.)

Wir betrachten also  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  mit  $\lambda((a, b]) := b - a$ . Damit dürfen wir  $\lambda$  vorerst auch wirklich nur auf Intervalle dieses speziellen Typs anwenden, und nicht etwa auf die Mengen  $(3, 4)$  oder  $(0, 1] \cup (4, 5]$ . Um den Überblick zu bewahren, sehen wir uns zunächst an, wie sich Elemente von  $\mathcal{I}_{\mathbb{R}}$  unter einfachen Mengenoperationen verhalten. Mit anderen Worten: wir werfen einen Blick auf die Struktur von  $\mathcal{I}_{\mathbb{R}}$ . (Eine Skizze wird Sie schnell von der Richtigkeit folgender Aussagen überzeugen, es schadet aber nicht, wenn Sie das einmal auch genauer überlegen, zB um sicherzugehen, dass Sie in Ihren Skizzen auch wirklich alle Fälle berücksichtigt haben.)

**Bemerkung 1.2** *Die Familie  $\mathcal{I}_{\mathbb{R}}$  hat folgende Eigenschaften:*

*a)  $\emptyset \in \mathcal{I}_{\mathbb{R}}$ .*

- b) Für  $A, B \in \mathcal{I}_{\mathbb{R}}$  ist auch  $A \cap B \in \mathcal{I}_{\mathbb{R}}$ .  
c) Für  $A, B \in \mathcal{I}_{\mathbb{R}}$  ist  $A \cup B$  nicht unbedingt in  $\mathcal{I}_{\mathbb{R}}$ . (Aber solche Mengen sind trotzdem sehr einfach, und wir werden sie gerne verwenden, allerdings ohne über die vorerst nicht definierte Größe  $\lambda(A \cup B)$  zu sprechen.)  
d) Für  $A, B \in \mathcal{I}_{\mathbb{R}}$  ist  $A \setminus B$  nicht unbedingt in  $\mathcal{I}_{\mathbb{R}}$ . Wir können diese Differenz aber als endliche disjunkte Vereinigung von Mengen aus  $\mathcal{I}_{\mathbb{R}}$  darstellen, dh zu  $A, B \in \mathcal{I}_{\mathbb{R}}$  gibt es immer ein  $m \in \mathbb{N}$  und paarweise disjunkte Mengen  $C_1, \dots, C_m \in \mathcal{I}_{\mathbb{R}}$  mit  $A \setminus B = \bigcup_{j=1}^m C_j$ . Wir werden dies künftig kurz so ausdrücken:

$$\text{für } A, B \in \mathcal{I}_{\mathbb{R}}: \quad A \setminus B = \biguplus_{j=1}^m C_j \quad \text{mit } m \in \mathbb{N} \text{ und } C_1, \dots, C_m \in \mathcal{I}_{\mathbb{R}}. \quad (1.7)$$

- e) Eine stärkere Form von (1.7) werden wir demnächst verwenden: Man kann aus  $A \in \mathcal{I}_{\mathbb{R}}$  eine beliebige endliche Zahl von Mengen aus  $\mathcal{I}_{\mathbb{R}}$  herauschneiden, und bekommt wieder eine Menge obigen Typs (Induktionsbeweis, siehe PS),

$$\text{für } A, B_1, \dots, B_n \in \mathcal{I}_{\mathbb{R}}: \quad A \setminus (B_1 \cup \dots \cup B_n) = \biguplus_{j=1}^m C_j \quad \text{mit } m \in \mathbb{N} \text{ und } C_1, \dots, C_m \in \mathcal{I}_{\mathbb{R}}.$$

Solche strukturellen Eigenschaften von Mengenfamilien werden noch eine große Rolle spielen, wir kommen bald darauf zurück. Fürs Erste konzentrieren wir uns aber auf Intervalle. Die Familie  $\mathcal{I}_{\mathbb{R}}$  ist schon sehr eingeschränkt, warum wollen wir nicht gleich beliebige Intervalle behandeln? Nun, je kleiner die Familie der untersuchten Mengen, desto weniger müssen wir überprüfen, wenn wir Grundeigenschaften von  $\lambda$  nachweisen wollen. ZB ersparen wir es uns, auf Überlappungen an den Rändern eingehen zu müssen. Zu Ihrer Beruhigung halten wir aber gleich fest, dass man beliebige Intervalle aus Mengen des Typs  $(a, b]$  gewinnen kann, sobald man abzählbar viele Mengenoperationen verwendet.

**Beispiel 1.2** Wir sehen etwa, dass

$$[a, b] = \bigcap_{n \geq 1} (a - \frac{1}{n}, b] \quad \text{und} \quad (a, \infty) = \bigcup_{n \geq 1} (a, a + n].$$

Zur Übung sollten Sie auch andere Typen von Intervallen so ausdrücken. (PS)

Tatsächlich kommen wir mittels abzählbarer Mengenoperationen noch weit über die simple Welt der Intervalle hinaus. So können wir beispielsweise jede offene Menge in analoger Weise gewinnen:

**Lemma 1.1 (Offene Mengen als abzählbare Vereinigungen aus  $\mathcal{I}_{\mathbb{R}}$ )** Ist  $G \subseteq \mathbb{R}$  offen, dann gibt es eine Folge  $(A_n)_{n \geq 1}$  in  $\mathcal{I}_{\mathbb{R}}$  mit  $G = \bigcup_{n \geq 1} A_n$ .

**Beweis.** Da  $G$  offen ist gibt es zu jedem Punkt  $x \in G$  ein  $\varepsilon_x > 0$  derart, dass  $(x - \varepsilon_x, x + \varepsilon_x) \subseteq G$ . Weil  $\mathbb{Q}$  in  $\mathbb{R}$  dicht liegt, gibt es dann auch rationale Zahlen  $a_x, b_x$  mit  $x - \varepsilon_x < a_x < x < b_x < x + \varepsilon_x$ . Für diese gilt dann

$$G = \bigcup_{x \in G} (a_x, b_x].$$

Damit ist  $G$  als Vereinigung halboffener Intervalle angegeben. Während die Indexmenge  $G$  in dieser Darstellung überabzählbar ist (wenn nicht gerade  $G = \emptyset$ ), handelt es sich doch um eine abzählbare Vereinigung: Indem wir lauter rationale Endpunkte gewählt haben, ist sichergestellt, dass es nur abzählbar viele unterschiedliche Mengen gibt, die auf der rechten Seite auftreten.

Falls Ihnen das letzte Argument noch unheimlich erscheint, können Sie es ein wenig formaler so fassen: Die Familie  $\mathcal{R} := \{(a, b] : a \leq b \text{ in } \mathbb{Q}\}$  der Intervalle mit rationalen Endpunkten ist abzählbar. Erst recht ist die Teilfamilie  $\mathcal{R}(G) := \{A \in \mathcal{R} : A \subseteq G\}$  abzählbar. Nach obiger Überlegung gilt aber  $G = \bigcup_{A \in \mathcal{R}(G)} A$ . ■

Die Familie  $\mathcal{I}_{\mathbb{R}}$  verträgt sich überdies gut mit der Geometrie von  $\mathbb{R}$ , jedenfalls solange wir auf Spiegelungen verzichten. Entsprechendes gilt für die Längenfunktion, was auch kaum verwundert.

**Bemerkung 1.3** Die Familie  $\mathcal{I}_{\mathbb{R}}$  ist translationsinvariant in dem Sinne, dass für  $A \in \mathcal{I}_{\mathbb{R}}$  und  $t \in \mathbb{R}$  stets  $A + t \in \mathcal{I}_{\mathbb{R}}$  gilt. Schreiben wir kurz  $\mathcal{I}_{\mathbb{R}} + t := \{A + t : A \in \mathcal{I}_{\mathbb{R}}\}$ , so gilt  $\mathcal{I}_{\mathbb{R}} + t = \mathcal{I}_{\mathbb{R}}$ . Ebenso ist die elementare Längenfunktion  $\lambda$  translationsinvariant auf  $\mathcal{I}_{\mathbb{R}}$ , dh für  $A \in \mathcal{I}_{\mathbb{R}}$  und  $t \in \mathbb{R}$  gilt immer  $\lambda(A + t) = \lambda(A)$ .

Jetzt aber zur zentralen Aussage dieses Abschnittes. Dass  $\lambda$  additiv ist scheint so selbverständlich, dass Sie vielleicht nicht gleich nach einem Beweis fragen würden. Die erhoffte  $\sigma$ -Additivität ist aber eine Aussage über unendliche Reihen, und da wollen wir schon aufpassen. Um zu sehen, dass die Sache nicht trivial ist, machen Sie sich bitte klar, dass man etwa das Intervall  $(0, 1]$  durchaus in komplizierter Weise in unendlich viele Teilintervalle zerschneiden kann, die sich nicht von links nach rechts anordnen lassen. Es ist das Verdienst von E. Borel, hier Klarheit geschaffen zu haben. Anstatt die Teilintervalle der Reihe nach zu einer unendlichen Teleskopsumme zu ordnen (wie wir es in dem einfachen Beispiel 1.1 getan haben), führt sein Argument den unendlichen Fall auf den endlichen zurück - und verwendet dabei den bei dieser Gelegenheit entdeckten Überdeckungssatz von Heine-Borel.

**Satz 1.4 ( $\sigma$ -Additivität der Längenfunktion)** Die elementare Längenfunktion  $\lambda$  ist  $\sigma$ -additiv auf  $\mathcal{I}_{\mathbb{R}}$ , dh für jede Folge  $(A_n)_{n \geq 1}$  paarweise disjunkter Mengen  $A_n \in \mathcal{I}_{\mathbb{R}}$  für welche auch  $\bigcup_{n \geq 1} A_n$  in  $\mathcal{I}_{\mathbb{R}}$  liegt, gilt

$$\lambda\left(\biguplus_{n \geq 1} A_n\right) = \sum_{n \geq 1} \lambda(A_n).$$

**Beweis.** Zu zeigen ist: Für jede disjunkte Folge  $((a_n, b_n])_{n \geq 1}$  mit  $\bigcup_{n \geq 1} (a_n, b_n] = (a, b] \in \mathcal{I}_{\mathbb{R}}$  gilt

$$\sum_{n \geq 1} (b_n - a_n) = b - a.$$

(i) Wir zeigen zuerst: Für jede disjunkte Folge  $((a_n, b_n])_{n \geq 1}$  in  $\mathcal{I}_{\mathbb{R}}$  mit  $\bigcup_{n \geq 1} (a_n, b_n] \subseteq (a, b] \in \mathcal{I}_{\mathbb{R}}$  gilt

$$\sum_{n \geq 1} (b_n - a_n) \leq b - a.$$

Betrachte endlich viele dieser Intervalle,  $(a_1, b_1], \dots, (a_N, b_N]$ . Da disjunkt, können wir sie so ordnen, dass

$$a \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq \dots \leq a_N \leq b_N \leq b.$$

Dann ist, wenn wir die allen Paaren aufeinanderfolgender Punkte entsprechenden (nicht-negativen) Differenzen anschreiben,

$$\begin{aligned} \sum_{n=1}^N (b_n - a_n) &\leq (a_1 - a) + (b_1 - a_1) + (a_2 - b_1) + \dots + (b_N - a_N) + (b - b_N) \\ &= b - a, \end{aligned}$$

wie behauptet. Weil das für jedes  $N \geq 1$  gilt, folgt die entsprechende Abschätzung auch für den Grenzwert der unendlichen Reihe:

$$\sum_{n \geq 1} (b_n - a_n) = \lim_{N \rightarrow \infty} \sum_{n=1}^N (b_n - a_n) \leq b - a.$$

(ii) Nun zeigen wir: Für jede endliche Überdeckung  $\bigcup_{n=1}^N (a_n, b_n] \supseteq (a, b]$  gilt

$$b - a \leq \sum_{n=1}^N (b_n - a_n).$$

Für  $N = 1$  ist die Behauptung trivial. Um sie induktiv für beliebiges  $N$  nachzuweisen, nehmen wir an sie gelte für  $N - 1$ . Wir ordnen die Intervalle so, dass  $b \in (a_N, b_N]$ .

Falls nun  $(a, b] \subseteq (a_N, b_N]$ , dann gilt sogar  $b - a \leq b_N - a_N$ . Andernfalls wird das Teilintervall  $(a, b] \setminus (a_N, b_N] = (a, a_N]$  von den  $N - 1$  verbleibenden Intervallen überdeckt,  $(a, a_N] \subseteq \bigcup_{n=1}^{N-1} (a_n, b_n]$ , und unsere Induktionsvoraussetzung garantiert, dass  $\sum_{n=1}^{N-1} (b_n - a_n) \geq a_N - a$ . Insgesamt folgt also  $b - a = (b - a_N) + (a_N - a) \leq (b_N - a_N) + \sum_{n=1}^{N-1} (b_n - a_n)$ , wie gefordert.

(iii) Schließlich sehen wir uns den (eigentlich interessanten) Fall unendlicher Überdeckungen an:  $\bigcup_{n \geq 1} (a_n, b_n] \supseteq (a, b]$ . Für beliebiges  $\varepsilon > 0$  betrachten wir die Punkte  $\alpha := a + \varepsilon/2$  und  $\beta_n := b_n + 2^{-(n+1)}\varepsilon$ ,  $n \geq 1$ . Dann ist

$$[\alpha, b] \subseteq \bigcup_{n \geq 1} (a_n, \beta_n),$$

das kompakte Intervall auf der linken Seite wird also von den offenen Mengen rechts überdeckt. Nach dem Satz von Heine und Borel enthält diese offene Überdeckung eine endliche Teilüberdeckung, insbesondere gibt es ein  $N \geq 1$  sodass

$$(\alpha, b] \subseteq \bigcup_{n=1}^N (a_n, \beta_n].$$

Damit befinden wir uns im vorher betrachteten Fall endlicher Überdeckungen, und wissen daher, dass

$$\begin{aligned} b - a &\leq b - \alpha + \frac{\varepsilon}{2} \leq \sum_{n=1}^N (\beta_n - a_n) + \frac{\varepsilon}{2} \\ &= \sum_{n=1}^N (b_n - a_n) + \varepsilon \sum_{n=1}^N 2^{-(n+1)} + \frac{\varepsilon}{2} \leq \sum_{n \geq 1} (b_n - a_n) + \varepsilon, \end{aligned}$$

und da  $\varepsilon > 0$  beliebig war, folgt

$$b - a \leq \sum_{n \geq 1} (b_n - a_n),$$

wie gewünscht. ■

Infolge dieses Satzes wird  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  auch als *Lebesgue-Borelsches Prämaß* bezeichnet. Der Begriff *Prämaß* drückt aus, dass diese Mengenfunktion die entscheidende Eigenschaft der  $\sigma$ -Additivität besitzt, man spricht aber nicht von einem Maß, weil der Definitionsbereich noch nicht die richtige Struktur hat.

Aus dem Satz (und der banalen Eigenschaft  $\lambda(\emptyset) = 0$ ) ergibt sich nun eine Liste weiterer natürlicher Eigenschaften der Längenfunktion. (Im letzten Punkt verwenden wir folgende Notation: Sind  $A_n \in \mathcal{P}(X)$ ,  $n \geq 1$ , so bedeutet  $A_n \nearrow$ , dass  $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$  und  $A_n \nearrow A$  heißt, dass zusätzlich  $A = \bigcup_{n \geq 1} A_n$ . Analog für  $A_n \searrow$  und  $A_n \searrow A (= \bigcap_{n \geq 1} A_n)$ .)

**Proposition 1.1 (Weitere Eigenschaften des Lebesgue-Borelschen Prämaßes)**  
Die Längenfunktion  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  hat folgende Eigenschaften:

a)  $\lambda$  ist endlich additiv auf  $\mathcal{I}_{\mathbb{R}}$ : für paarweise disjunkte  $A_1, \dots, A_N \in \mathcal{I}_{\mathbb{R}}$  mit  $\bigcup_{n=1}^N A_n \in \mathcal{I}_{\mathbb{R}}$  gilt

$$\lambda\left(\biguplus_{n=1}^N A_n\right) = \sum_{n=1}^N \lambda(A_n).$$

b)  $\lambda$  ist monoton (isoton) auf  $\mathcal{I}_{\mathbb{R}}$ : sind  $A, B \in \mathcal{I}_{\mathbb{R}}$ ,  $B \subseteq A$ , so gilt  $\lambda(B) \leq \lambda(A)$ .

c)  $\lambda$  ist endlich subadditiv auf  $\mathcal{I}_{\mathbb{R}}$ : für  $A_1, \dots, A_N \in \mathcal{I}_{\mathbb{R}}$  mit  $\bigcup_{n=1}^N A_n \in \mathcal{I}_{\mathbb{R}}$  gilt

$$\lambda\left(\bigcup_{n=1}^N A_n\right) \leq \sum_{n=1}^N \lambda(A_n).$$

d)  $\lambda$  ist  $\sigma$ -subadditiv auf  $\mathcal{I}_{\mathbb{R}}$ :

$$A_n \in \mathcal{I}_{\mathbb{R}}, n \geq 1, \text{ und } \bigcup_{n \geq 1} A_n \in \mathcal{I}_{\mathbb{R}} \implies \lambda \left( \bigcup_{n \geq 1} A_n \right) \leq \sum_{n \geq 1} \lambda(A_n).$$

e)  $\lambda$  ist stetig von unten auf  $\mathcal{I}_{\mathbb{R}}$ :

$$A_n \in \mathcal{I}_{\mathbb{R}}, n \geq 1, \text{ und } A_n \nearrow A \in \mathcal{I}_{\mathbb{R}} \implies \lim_{n \rightarrow \infty} \lambda(A_n) = \lambda(A).$$

**Beweis.** a) Definiere  $A_n := \emptyset \in \mathcal{I}_{\mathbb{R}}$  für  $n > N$  und verwende die  $\sigma$ -Additivität:

$$\lambda \left( \biguplus_{n=1}^N A_n \right) = \lambda \left( \biguplus_{n \geq 1} A_n \right) = \sum_{n \geq 1} \lambda(A_n) = \sum_{n=1}^N \lambda(A_n).$$

b) Aus  $A, B \in \mathcal{I}_{\mathbb{R}}, B \subseteq A$  folgt  $A = B \uplus (A \setminus B) = B \uplus \biguplus_{k=1}^{m_n} C_k$  mit  $B, C_k \in \mathcal{I}_{\mathbb{R}}$  disjunkt. Daher ist, nach a),  $\lambda(A) = \lambda(B) + \sum_{k=1}^{m_n} \lambda(C_k) \geq \lambda(B)$ . Beachten Sie:  $\lambda(A \setminus B)$  ist hier ia nicht definiert! Wir müssen peinlich genau darauf achten, auf welche Mengen  $\lambda$  angewandt werden kann!

c) Folgt aus d) mit  $A_n := \emptyset$  für  $n > N$ .

d) Nach Bemerkung 1.2 e) ist  $A'_n := A_n \setminus (A_1 \cup \dots \cup A_{n-1}) = \biguplus_{k=1}^{m_n} C_{n,k}$  mit disjunkten  $C_{n,k} \in \mathcal{I}_{\mathbb{R}}$ . Daher gilt ( $\sigma$ -Additivität)

$$\lambda(A) = \sum_{n \geq 1} \sum_{k=1}^{m_n} \lambda(C_{n,k}) \leq \sum_{n \geq 1} \lambda(A_n),$$

wobei der zweite Schritt wieder mit Bemerkung 1.2 e) zu rechtfertigen ist, wonach ja für jedes  $n \geq 1$ ,  $A_n \setminus A'_n = A_n \setminus (\biguplus_{k=1}^{m_n} C_{n,k}) = \biguplus_{j=1}^{l_n} D_{n,j}$  mit  $D_{n,j} \in \mathcal{I}_{\mathbb{R}}$ , sodass

$$\lambda(A_n) = \lambda \left( \biguplus_{k=1}^{m_n} C_{n,k} \uplus \biguplus_{j=1}^{l_n} D_{n,j} \right) = \sum_{k=1}^{m_n} \lambda(C_{n,k}) + \sum_{j=1}^{l_n} \lambda(D_{n,j}) \geq \sum_{k=1}^{m_n} \lambda(C_{n,k}).$$

e) Siehe PS. ■

Zu beachten war in diesem Beweis insbesondere, dass wir nicht voreilig mit  $\lambda(A \setminus B)$  rechnen dürfen. Mithilfe der Beobachtungen d) und e) aus Bemerkung 1.2 konnten wir das Argument aber retten. Interessant (und später wichtig) ist auch folgender Aspekt: Der Beweis der Proposition verwendet eigentlich nirgends explizit, dass die Mengen in  $\mathcal{I}_{\mathbb{R}}$  Intervalle sind. Vielmehr beruht er nur auf der  $\sigma$ -Additivität aus dem vorigen Satz, und eben den strukturellen Eigenschaften von  $\mathcal{I}_{\mathbb{R}}$  aus Bemerkung 1.2. Darauf kommen wir noch zurück. Einprägen sollten Sie sich auch die Idee aus Schritt d), wo wir uns zu einer beliebigen Folge  $(A_n)_{n \geq 1}$  von Mengen eine paarweise disjunkte Folge  $(A'_n)_{n \geq 1}$  mit  $\bigcup_{n=1}^N A_n = \biguplus_{n=1}^N A'_n$  für alle  $N \in \mathbb{N} \cup \{\infty\}$  verschafft haben. Das wird noch öfters nützlich sein.

Am Ende dieses Abschnittes überlegen wir uns noch, dass die Längenfunktion bereits durch ein paar wenige Eigenschaften eindeutig charakterisiert werden kann, also ein besonders natürliches Objekt ist.

**Proposition 1.2 (Charakterisierung der Längenfunktion)** Ist  $\nu : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  eine Funktion mit folgenden Eigenschaften,

- a)  $\nu((0, 1]) = 1$ ,
- b)  $\nu$  ist endlich additiv auf  $\mathcal{I}_{\mathbb{R}}$ , und
- c)  $\nu$  ist translationsinvariant auf  $\mathcal{I}_{\mathbb{R}}$ ,

dann ist  $\nu = \lambda$ .

**Beweis.** Siehe PS. ■

## 1.4 Das Lebesguesche äußere Maß

Von  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  ausgehend könnten wir nun  $\lambda(A)$  für ein paar weitere einfache Typen von Mengen (zB  $(0, 1] \cup (4, 5]$ ) explizit definieren. Das ersparen wir uns, weil wir damit ohnehin nicht recht weit kommen, und weil all das im Rahmen der folgenden Überlegungen automatisch miterledigt wird.

Wir wenden uns der Frage zu, wie man richtig komplizierten Mengen  $A \subseteq \mathbb{R}$  noch in vernünftiger Weise eine Maßzahl zuordnen könnten. Es wird jetzt also ernst! Die Grundidee ist dem Darboux'schen Zugang zum Riemann-Integral ähnlich: Wie dort mithilfe von Obersummen, versuchen wir eine sichere obere Schranke für die Größe von  $A$  dadurch zu bekommen, dass wir  $A$  mittels einfacher Mengen überdecken (dort endlich viele Rechtecke), über deren (Gesamt-)größe wir sprechen können. Die Lebesguesche Variante dieser Idee unterscheidet sich von Darboux' Methode aber in zwei wesentlichen Punkten:

Erstens lassen wir anstelle endlich vieler überdeckender Mengen gleich abzählbar unendlich viele zu. Das ist viel flexibler und führt schon in einfachen Fällen zu komplett unterschiedlichen Antworten (siehe Bsp unten). Dank der Vorarbeiten des letzten Abschnittes sind wir dafür gut gerüstet. Zweitens verzichten wir darauf, die Menge  $A$  auch von innen, dh durch einfache Teilmengen (analog zur Untersumme bei Darboux), einzugrenzen. Unsere Grundbausteine, die Intervalle aus  $\mathcal{I}_{\mathbb{R}}$  eignen sich einfach nicht gut dazu. In nächsten Abschnitt besprechen wir dann, warum die Approximation von aussen genügt.

**Definition 1.2** Eine abzählbare  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung (kurz  $\mathcal{I}_{\mathbb{R}}$ -Ü) von  $A \subseteq \mathbb{R}$ , ist einfach eine Folge  $(E_k)_{k \geq 1}$  von Mengen aus  $\mathcal{I}_{\mathbb{R}}$  deren Vereinigung  $A$  enthält:  $A \subseteq \bigcup_{k \geq 1} E_k$  (wobei ruhig einige  $E_k$  leer sein dürfen).

Zu jeder Menge  $A \subseteq \mathbb{R}$  gibt es eine  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung  $(E_k)_{k \geq 1}$ , zB jene mit  $E_k := (-k, k]$ . Haben wir nun so eine Überdeckung gefunden, dann muss die Länge von  $A$ , sofern sie überhaupt konsistent definiert werden kann, durch die *Längensumme*  $\sum_{k \geq 1} \lambda(E_k)$  der Überdeckung beschränkt sein. Sparsamere Überdeckungen liefern genauere Approximationen von aussen, und damit bessere obere Schranken. Das Optimum an Information über  $A$  das wir auf diese Weise gewinnen können, ist das Infimum dieser Schranken.

**Definition 1.3** Das Lebesguesche äußere Maß auf  $\mathbb{R}$  ist die Mengenfunktion  $\lambda^* : \mathcal{P}(\mathbb{R}) \rightarrow [0, \infty]$  mit

$$\lambda^*(A) := \inf \left\{ \sum_{k \geq 1} \lambda(E_k) : \begin{array}{l} (E_k)_{k \geq 1} \text{ ist eine abzählbare} \\ \mathcal{I}_{\mathbb{R}}\text{-Überdeckung von } A \end{array} \right\}, \quad A \subseteq \mathbb{R}.$$

Man beachte, dass  $\lambda^*(A)$  wirklich für jede Teilmenge von  $\mathbb{R}$  definiert ist.

**Bemerkung 1.4** Wir sehen leicht, dass  $\lambda^*$  auf  $\mathcal{P}(\mathbb{R})$  translationsinvariant ist, dh für  $A \subseteq \mathbb{R}$  und  $t \in \mathbb{R}$  gilt immer  $\lambda^*(A + t) = \lambda^*(A)$ . (Das gilt einfach weil schon  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  diese Eigenschaft hat, und die  $\mathcal{I}_{\mathbb{R}}$ -Überdeckungen von  $A + t$  genau die Folgen  $(E_k + t)_{k \geq 1}$  sind, wobei  $(E_k)_{k \geq 1}$  eine  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung von  $A$ .)

Wenn wir  $\lambda^*$  als guten Größenbegriff auffassen wollen, sollte  $\lambda^*$  wenigstens auf  $\mathcal{I}_{\mathbb{R}}$  sinnvolle Werte liefern. Zum Glück funktioniert das:

**Proposition 1.3** ( $\lambda^*$  setzt  $\lambda$  fort) Für  $A \in \mathcal{I}_{\mathbb{R}}$  ist  $\lambda^*(A) = \lambda(A)$ .

**Beweis.** Sei  $A \in \mathcal{I}_{\mathbb{R}}$  fest. Da  $(A, \emptyset, \emptyset, \dots)$  eine  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung von  $A$  ist, gilt jedenfalls  $\lambda^*(A) \leq \lambda(A)$ . Zum Nachweis der entgegengesetzten Ungleichung betrachten wir eine beliebige abzählbare  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung  $(E_k)_{k \geq 1}$  von  $A$ , und setzen  $E_k^* := E_k \cap A$ ,  $k \geq 1$ . Mit  $E_k$  und  $A$  liegen diese Mengen wieder in  $\mathcal{I}_{\mathbb{R}}$ ,

und wir haben  $A = \bigcup_{k \geq 1} E_k^* \subseteq \bigcup_{k \geq 1} E_k$ . Nachdem  $\lambda$  auf  $\mathcal{I}_{\mathbb{R}}$   $\sigma$ -subadditiv und monoton ist, siehe Proposition 1.1, haben wir

$$\lambda(A) \leq \sum_{k \geq 1} \lambda(E_k^*) \leq \sum_{k \geq 1} \lambda(E_k),$$

und weil die  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung  $(E_k)_{k \geq 1}$  von  $A$  beliebig war, sehen wir, dass wie gewünscht  $\lambda(A) \leq \lambda^*(A)$  gilt. ■

Wir haben also eine translationsinvariante Fortsetzung der Längenfunktion von  $\mathcal{I}_{\mathbb{R}}$  nach  $\mathcal{P}(\mathbb{R})$  gewonnen. Im Lichte von Satz 1.2 sehen wir sofort, dass  $\lambda^*$  auf  $\mathcal{P}(\mathbb{R})$  nicht  $\sigma$ -additiv sein kann. Immerhin gilt aber der grundlegende

**Satz 1.5 (Eigenschaften von  $\lambda^*$ )** *Das Lebesguesche äußere Maß  $\lambda^* : \mathcal{P}(\mathbb{R}) \rightarrow [0, \infty]$  hat folgende Eigenschaften:*

( $\lambda^*1$ )  $\lambda^*(\emptyset) = 0$ ,

( $\lambda^*2$ )  $B \subseteq A \implies \lambda^*(B) \leq \lambda^*(A)$ , und

( $\lambda^*3$ ) ist  $(A_n)_{n \geq 1}$  eine Folge von Mengen aus  $\mathcal{P}(\mathbb{R})$ , dann gilt

$$\lambda^*\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \lambda^*(A_n).$$

*Es ist also monoton und  $\sigma$ -subadditiv.*

**Beweis.** Für  $A = \emptyset$  wählen wir die Überdeckung durch die einzelne Menge  $E_1 := \emptyset \in \mathcal{I}_{\mathbb{R}}$  um zu sehen dass  $\lambda^*(\emptyset) \leq \lambda(\emptyset) = 0$ . Monotonie von  $\lambda^*$  ist auch fast offensichtlich: Falls  $B \subseteq A \subseteq X$ , so ist jede  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung von  $A$  auch eine  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung von  $B$  und somit folgt  $\lambda^*(B) \leq \lambda^*(A)$ .

Für den Beweis der  $\sigma$ -Subadditivität sei nun  $(A_n)_{n \geq 1}$  eine beliebige Folge von Mengen aus  $\mathcal{P}(X)$ . Falls  $\lambda^*(A_n) = \infty$  für ein  $n$ , so ist die Behauptung trivial. Wir können daher annehmen dass  $\lambda^*(A_n) < \infty$  für alle  $n \geq 1$ . Sei  $\varepsilon > 0$ . Für jedes  $n \geq 1$  gibt es nach Definition von  $\lambda^*$  eine abzählbare  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung  $(E_{n,k})_{k \geq 1}$  von  $A_n$  mit

$$\sum_{k \geq 1} \lambda(E_{n,k}) \leq \lambda^*(A_n) + \varepsilon \cdot 2^{-n}.$$

Gemeinsam bilden diese eine  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung  $(E_{n,k})_{n \geq 1, k \geq 1}$  von  $\bigcup_{n \geq 1} A_n$ . Daher gilt

$$\lambda^*\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \sum_{k \geq 1} \lambda(E_{n,k}) \leq \sum_{n \geq 1} \lambda^*(A_n) + \varepsilon,$$

und weil  $\varepsilon > 0$  beliebig war, folgt die  $\sigma$ -Subadditivität von  $\lambda^*$ . ■

**Bemerkung 1.5 a)** *Aus der  $\sigma$ -Subadditivität folgt ganz einfach auch die endliche Subadditivität,  $\lambda^*(\bigcup_{n=1}^N A_n) \leq \sum_{n=1}^N \lambda^*(A_n)$ , Sie brauchen dafür nur die Folge  $(A_1, \dots, A_N, \emptyset, \emptyset, \dots)$  zu betrachten.*

Wir wenden uns noch ein paar Beispielen zu. Zunächst halten wir fest:

**Proposition 1.4** *Ist  $A \subseteq \mathbb{R}$  abzählbar, dann gilt  $\lambda^*(A) = 0$ .*

**Beweis.** Sei  $A = \{a_k : k \geq 1\}$  und  $\varepsilon > 0$ . Wir können  $A$  überdecken durch die Folge  $(E_k)_{k \geq 1}$  mit  $E_k := (a_k - \frac{\varepsilon}{2^k}, a_k]$ , und sehen daher, dass

$$\lambda^*(A) \leq \sum_{k \geq 1} \lambda(E_k) = \sum_{k \geq 1} \frac{\varepsilon}{2^k} = \varepsilon.$$



Weil  $\varepsilon > 0$  beliebig war, folgt  $\lambda^*(A) = 0$ . ■

Insbesondere ist also für jede endliche Menge  $\lambda^*(A) = 0$ . In Kombination mit Proposition 1.3 und dem Satz finden wir dann zB für kompakte Intervalle, dass  $\lambda^*((a, b]) \leq \lambda^*([a, b]) \leq \lambda^*([a, b]) + \lambda^*([a, b]) = \lambda^*((a, b])$ , also  $\lambda^*([a, b]) = b - a$ . Weiters bekommen wir  $\lambda^*(\mathbb{Q}) = 0$  und erst recht  $\lambda^*(Q) = 0$  für  $Q := \mathbb{Q} \cap (-1, 1]$ . Anhand letzterer Menge können wir übrigens sehen, dass es in der Definition von  $\lambda^*$  wesentlich ist, abzählbare Überdeckungen zuzulassen, und nicht nur endliche Überdeckungen zu verwenden:

**Beispiel 1.3** *Im PS wurde gezeigt, dass jede endliche Überdeckung von  $Q = \mathbb{Q} \cap (-1, 1]$  durch Mengen  $E_1, \dots, E_N \in \mathcal{I}_{\mathbb{R}}$  automatisch das gesamte Intervall  $(-1, 1]$  überdecken muss. Somit ist*

$$\inf \left\{ \sum_{k \geq 1} \lambda(E_k) : \begin{array}{l} (E_k)_{k \geq 1} \text{ ist eine endliche} \\ \mathcal{I}_{\mathbb{R}}\text{-Überdeckung von } Q \end{array} \right\} = 2.$$

*Würden wir in der Definition nur endliche Überdeckungen zulassen, käme also eine ganz andere Mengenfunktion heraus.*

Sind nun die abzählbaren Mengen  $A \subseteq \mathbb{R}$  die einzigen mit  $\lambda^*(A) = 0$ ? Dies ist bei Weitem nicht der Fall! Wir illustrieren dies anhand einer sehr prominenten Menge, die nicht nur in der Maßtheorie von Interesse ist:

**Beispiel 1.4 (Das Cantorsche Diskontinuum)** *Ein hübsches Beispiel mit vielen interessanten Eigenschaften ist das Cantorsche Diskontinuum  $C \subseteq [0, 1]$ . Diese Menge konstruiert man schrittweise wie folgt.*

*Wir beginnen mit  $C_{(0)} := [0, \frac{1}{3}]$  und  $C_{(1)} := [\frac{2}{3}, 1]$ . Ist dann für ein  $n$ -Tupel  $(\omega_1, \dots, \omega_n)$  aus Nullen und Einsen  $\omega_k \in \{0, 1\}$  ein kompaktes Intervall  $C_{(\omega_1, \dots, \omega_n)}$  schon konstruiert, so definieren wir zwei Teilintervalle davon, nämlich*

$$\begin{aligned} C_{(\omega_1, \dots, \omega_n, 0)} &:= \text{das linke abgeschlossene Drittel von } C_{(\omega_1, \dots, \omega_n)}, \\ C_{(\omega_1, \dots, \omega_n, 1)} &:= \text{das rechte abgeschlossene Drittel von } C_{(\omega_1, \dots, \omega_n)}, \end{aligned}$$

*also zB  $C_{(0,0)} = [0, \frac{1}{9}]$  das linke abgeschlossene Drittel von  $C_{(0)}$ , und  $C_{(0,1)} = [\frac{2}{9}, \frac{1}{3}]$  das rechte abgeschlossene Drittel von  $C_{(0)}$ . Setzt man danach  $C^{(n)} := \bigcup_{\omega_1, \dots, \omega_n \in \{0,1\}} C_{(\omega_1, \dots, \omega_n)}$ ,  $n \geq 1$ , so liefert dies eine Menge die aus  $2^n$  kompakten Intervallen der Länge  $3^{-n}$  besteht. Somit ist, wegen Subadditivität von  $\lambda^*$ ,  $\lambda^*(C^{(n)}) \leq (\frac{2}{3})^n$  für jedes  $n \geq 1$ . Schliesslich definieren wir*

$$C := \bigcap_{n \geq 1} C^{(n)}.$$

*Weil  $\lambda^*(C) \leq \lambda^*(C^{(n)}) \leq (\frac{2}{3})^n$  für  $n \geq 1$  gilt, folgt  $\lambda^*(C) = 0$ .*

*Wir zeigen, dass  $C$  überabzählbar ist, indem wir eine injektive Abbildung  $\psi$  der überabzählbaren Menge  $\Omega := \{\omega = (\omega_k)_{k \geq 1} : \omega_k \in \{0, 1\}\}$  aller unendlichen 0-1-Folgen nach  $C$  angeben. Diese hilft uns generell,  $C$  besser zu verstehen.*

*Dazu halten wir jetzt eine Folge  $\omega \in \Omega$  fest. Nach Konstruktion bilden die Mengen  $C_{(\omega_1, \dots, \omega_n)}$  mit  $n \geq 1$  eine fallende Folge (dh  $C_{(\omega_1, \dots, \omega_n)} \supseteq C_{(\omega_1, \dots, \omega_{n+1})}$ ) kompakter Intervalle deren Längen  $3^{-n}$  beliebig klein werden. In deren Durchschnitt  $\bigcap_{n \geq 1} C_{(\omega_1, \dots, \omega_n)}$  kann höchstens ein Punkt  $x$  liegen. (Wären es zwei, etwa  $x \neq y$ , dann hätten wir  $x, y \in C_{(\omega_1, \dots, \omega_n)}$  für alle  $n$ . Das geht nicht, weil die Länge von  $C_{(\omega_1, \dots, \omega_n)}$  für große  $n$  den Abstand  $|x - y|$  unterschreitet.) Wir zeigen, dass es so einen Punkt auch wirklich gibt: Wählen wir  $x_n \in C_{(\omega_1, \dots, \omega_n)}$ , so bekommen wir eine Cauchyfolge (für  $k, l \geq n$  sind  $x_k, x_l \in C_{(\omega_1, \dots, \omega_n)}$  und daher  $|x_k - x_l| \leq 3^{-n}$ ), die gegen einen Punkt  $x \in \mathbb{R}$  konvergieren muss. Für jedes  $n \geq 1$  enthält das abgeschlossene Intervall  $C_{(\omega_1, \dots, \omega_n)}$  aber mit  $x_n, x_{n+1}, \dots$  auch den Grenzwert  $x$  dieser Folge. Das bedeutet gerade, dass  $x \in \bigcap_{n \geq 1} C_{(\omega_1, \dots, \omega_n)}$ . Wir bezeichnen diesen Punkt  $x$  mit  $\psi(\omega)$ . Damit ist  $\psi : \Omega \rightarrow C$  definiert. Überprüfen wir noch die Injektivität von  $\psi$ . Sind  $\omega = (\omega_k)_{k \geq 1}$  und  $\sigma = (\sigma_k)_{k \geq 1}$  zwei Elemente von*

$\Omega$  und  $\omega \neq \sigma$ , dann gibt es eine erste Stelle  $j \geq 1$  an der sie sich unterscheiden,  $\omega_j \neq \sigma_j$ . Nach Konstruktion sind dann  $C_{(\omega_1, \dots, \omega_j)}$  und  $C_{(\sigma_1, \dots, \sigma_j)}$  disjunkt. Wegen  $\psi(\omega) \in C_{(\omega_1, \dots, \omega_j)}$  und  $\psi(\sigma) \in C_{(\sigma_1, \dots, \sigma_j)}$  folgt sofort  $\psi(\omega) \neq \psi(\sigma)$ . Die Abbildung ist also injektiv.

Tatsächlich ist  $\psi : \Omega \rightarrow C$  eine Bijektion (siehe PS). Dies erlaubt es uns die Punkte  $x \in C$  vollständig durch unendliche 0-1-Folgen  $\omega \in \Omega$  zu beschreiben. Sie können die zu  $x$  gehörige Folge  $\omega$  als Adresse von  $x$  interpretieren. Je mehr ihrer Stellen  $\omega_1, \dots, \omega_n$  Sie kennen, desto genauer wissen Sie, wo  $x$  liegt (nämlich im Intervall  $C_{(\omega_1, \dots, \omega_n)}$ ).

Die Menge  $C$  besitzt allerlei interessante Eigenschaften, die im PS besprochen werden. Insbesondere ist sie kompakt, sie enthält kein Intervall, aber auch keine isolierten Punkte. Oft wird  $C$  einfach als die Cantormenge bezeichnet. (In der Topologie bezeichnet man aber gerne jede Menge, die zu  $C$  topologisch äquivalent ist, als Cantormenge.)

**Vitali-Mengen\*.** Versuchen wir uns doch an den schrecklichen Vitali-Mengen  $V$  von oben. Da es eine Vielzahl solcher Mengen gibt, können wir nicht erwarten, einen konkreten Wert für  $\lambda^*(V)$  zu finden. Interessant ist aber

**Proposition 1.5** Jede Vitali-Menge  $V$  erfüllt  $\lambda^*(V) > 0$ .

**Beweis.** Wir erinnern uns daran (siehe Satz 1.2), dass

$$(0, 1] \subseteq \biguplus_{t \in Q} (V + t).$$

Wäre nun  $\lambda^*(V) = 0$ , dann auch  $\lambda^*(V + t) = 0$  für jedes  $t \in Q$  (wegen Translationsinvarianz von  $\lambda^*$ ), und dann aufgrund von Monotonie und  $\sigma$ -Subadditivität des äußeren Maßes,

$$\lambda^*((0, 1]) = \lambda^*\left(\biguplus_{t \in Q} (0, 1] \cap (V + t)\right) \leq \sum_{t \in Q} \lambda^*((0, 1] \cap (V + t)) \leq \sum_{t \in Q} \lambda^*(V + t) = 0,$$

in Widerspruch zu  $\lambda^*((0, 1]) = 1$  (siehe Proposition 1.3). ■

Interessant ist diese Beobachtung ua aus folgendem Grund: Satz 1.2 zeigt bereits, dass  $\lambda^*$  auf  $\mathcal{P}(\mathbb{R})$  nicht  $\sigma$ -additiv sein kann. Mittels der vorangehenden Proposition können wir schnell sehen, dass es sogar noch schlimmer steht:

**Proposition 1.6**  $\lambda^*$  ist auf  $\mathcal{P}(\mathbb{R})$  nicht endlich additiv.

**Beweis.** Wir betrachten eine bestimmte Vitali-Menge  $V$ . Wegen  $\lambda^*(V) > 0$  gibt es ein  $n \geq 1$  mit  $\lambda^*(V) > \frac{1}{n}$ . Es sei  $J \subseteq Q$  eine Menge von  $3n$  Elementen,  $\#J = 3n$ . Wäre  $\lambda^*$  endlich additiv, so hätten wir

$$\lambda^*\left(\biguplus_{t \in J} (V + t)\right) = \sum_{t \in J} \lambda^*(V + t) = \sum_{t \in J} \lambda^*(V) = 3n \lambda^*(V) > 3.$$

Wir wissen aber (siehe Satz 1.2), dass  $\biguplus_{t \in Q} (V + t) \subseteq (-1, 2]$  gilt, und daher

$$\lambda^*\left(\biguplus_{t \in J} (V + t)\right) \leq \lambda^*\left(\biguplus_{t \in Q} (V + t)\right) \leq \lambda^*((-1, 2]) = 3.$$

Der Widerspruch zwingt uns, die Annahme endlicher Additivität zu verwerfen. ■

**Der  $d$ -dimensionale Fall.** Die Ideen, Konzepte und Ergebnisse dieses Abschnittes übertragen sich leicht auf den Fall von Mengen in  $\mathbb{R}^d$ . Als Grundbausteine verwenden wir dann eben  $d$ -dimensionale Quader aus  $\mathcal{I}_{\mathbb{R}}^d$ , und das  $d$ -dimensionale Volumen  $\lambda^d$ . Eine abzählbare  $\mathcal{I}_{\mathbb{R}}^d$ -Überdeckung (kurz  $\mathcal{I}_{\mathbb{R}}^d$ -Ü) von  $A \subseteq \mathbb{R}^d$ , ist natürlich eine Folge  $(E_k)_{k \geq 1}$  von Mengen aus  $\mathcal{I}_{\mathbb{R}}^d$  deren Vereinigung  $A$  enthält:  $A \subseteq \bigcup_{k \geq 1} E_k$ . Das  $d$ -dimensionale Lebesguesche äußere Maß auf  $\mathbb{R}^d$  ist die Mengenfunktion  $\lambda^{d*} : \mathcal{P}(\mathbb{R}^d) \rightarrow [0, \infty]$  mit

$$\lambda^{d*}(A) := \inf \left\{ \sum_{k \geq 1} \lambda^d(E_k) : \begin{array}{l} (E_k)_{k \geq 1} \text{ ist eine abzählbare} \\ \mathcal{I}_{\mathbb{R}}^d\text{-Überdeckung von } A \end{array} \right\}, \quad A \subseteq \mathbb{R}^d.$$

Diese Mengenfunktion ist wiederum translationsinvariant, monoton,  $\sigma$ -subadditiv, und sie setzt  $\lambda^d$  fort, es gilt also  $\lambda^{d*}|_{\mathcal{I}_{\mathbb{R}}^d} = \lambda^d$ . (Bitte überzeugen Sie sich davon, dass dies wirklich mit genau denselben Argumenten wie oben geht!) Für beliebige  $A \subseteq \mathbb{R}^d$  ist  $\lambda^{d*}(A)$  eine obere Schranke für das "richtige"  $d$ -dimensionale Volumen von  $A$ , sofern dieses überhaupt definiert werden kann.

Betrachten wir etwa den Fall  $d = 2$ , so müssen wir die Menge  $A \subseteq \mathbb{R}^2$  eben möglichst sparsam mit einer Folge halboffener Rechtecke  $(a, b] \times (c, d]$  überdecken, und die Summe ihrer Flächen gibt uns eine Schranke für  $\lambda^{2*}(A)$ . Zum Beispiel können wir das achsenparallele Streckenstück  $A := [0, 1] \times \{0\}$  für jedes  $\varepsilon > 0$  durch das einzelne Rechteck  $R_\varepsilon := (-\varepsilon, 1] \times (-\varepsilon, 0]$  mit  $\lambda^2(R_\varepsilon) = \varepsilon(1 + \varepsilon)$  überdecken, also gilt  $\lambda^{2*}(A) = 0$ . Für nicht achsenparallele Streckenstücke sollte das natürlich auch gelten, aber man braucht halt mehr als ein Rechteck um diese zu überdecken. Probieren Sie dies bitte aus (PS).

## 1.5 Lebesgue-messbare Mengen

Um endlich ernsthaft über  $\mathcal{I}_{\mathbb{R}}$  hinauszukommen, nehmen wir jetzt einmal an, dass es eine größere Familie  $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R})$  mit  $\mathcal{I}_{\mathbb{R}} \subseteq \mathcal{A}$  gibt, deren Elementen wir in konsistenter Weise eine Länge zuordnen können, sodass sich also eine Mengenfunktion  $\mu : \mathcal{A} \rightarrow [0, \infty]$  finden lässt, die  $\lambda$  fortsetzt,  $\mu|_{\mathcal{I}_{\mathbb{R}}} = \lambda$ , und die wenigstens endlich additiv ist. Wie sollen wir aber die "guten" Mengen  $A \in \mathcal{A}$  dann nur erkennen? Zur Motivation der "richtigen" Bedingung, stellen wir uns eine beschränkte Menge  $A$  in  $\mathbb{R}$  vor, die also in einem gewissen Intervall  $Q \in \mathcal{I}_{\mathbb{R}}$  enthalten ist. Wir nehmen an, dass  $A$  zu  $\mathcal{A}$  gehört. Mittels Approximation durch Mengen aus  $\mathcal{I}_{\mathbb{R}}$  von außen, dh durch Überdeckungen, erhalten wir das äußere Maß  $\lambda^*(A)$  als obere Schranke für  $\mu(A)$ . Falls wir uns auch eine untere Schranke  $\lambda_*(A)$  verschaffen, dann sehen wir jedenfalls im Fall  $\lambda_*(A) = \lambda^*(A)$ , dass es nur die eine Möglichkeit, nämlich  $\mu(A) = \lambda^*(A)$  für den Wert von  $\mu(A)$  gibt.

Tatsächlich brauchen wir aber gar keine neue Mengenfunktion einzuführen um eine untere Schranke zu bekommen. Wenn man nämlich mit  $\mu(A)$  wie gewohnt rechnen darf, dann muss doch

$$\mu(A) = \mu(Q) - \mu(Q \cap A^c) = \lambda^*(Q) - \mu(Q \cap A^c) \geq \lambda^*(Q) - \lambda^*(Q \cap A^c)$$

gelten, weil  $\lambda^*(Q \cap A^c)$  obere Schranke für  $\mu(Q \cap A^c)$  ist. Wie erhalten also eine untere Schranke ebenfalls mittels  $\lambda^*$ . Gleichheit der beiden Schranken bedeutet dann  $\lambda^*(Q) - \lambda^*(Q \cap A^c) = \lambda^*(A)$ , und wegen  $\lambda^*(A) = \lambda^*(Q \cap A)$  wird dies zu

$$\lambda^*(Q) = \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c), \quad (\diamond)$$

dh  $Q$  wird durch  $A$  additiv zerschnitten. Wollen wir nun allgemein Mengen  $A$  betrachten, die nicht unbedingt beschränkt sind, so liegt die Forderung nahe, dass beschränkte Abschnitte, also Mengen der Form  $Q \cap A$  für beliebige  $Q \in \mathcal{I}_{\mathbb{R}}$  obige Bedingung erfüllen sollten. Als Ansatz könnten wir also die Familie jener  $A \subseteq \mathbb{R}$  betrachten, die gerade die Bedingung  $(\diamond)$  für alle  $Q \in \mathcal{I}_{\mathbb{R}}$  erfüllen. Wir verwenden im folgenden aber eine formal schärfere Variante dieser Bedingung, die auf C. Carathéodory zurückgeht. Sie verlangt die Gültigkeit von  $(\diamond)$  für *alle* Mengen  $Q$ :

**Definition 1.4** Eine Menge  $A \subseteq \mathbb{R}$  heißt Lebesgue-messbar oder  $\lambda^*$ -messbar, falls

$$\lambda^*(Q) = \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c) \quad \text{für alle } Q \subseteq \mathbb{R}, \quad (\blacklozenge)$$

dh wenn  $A$  alle Mengen bezüglich  $\lambda^*$  additiv zerschneidet. Die Familie solcher Mengen werden wir mit  $\mathcal{L}_{\mathbb{R}} \subseteq \mathcal{P}(\mathbb{R})$  bezeichnen.

Die Bedeutung der Bedingung  $(\blacklozenge)$  (etwa gegenüber der Variante mit  $Q \in \mathcal{I}_{\mathbb{R}}$ ) liegt gerade darin, dass sie eine Familie von Mengen auszeichnet, auf der  $\lambda^*$  sich gut verhält, und für die dies auch relativ leicht nachzuweisen ist.

**Bemerkung 1.6** a) Da  $\lambda^*$  ohnehin  $(\lambda^*3)$  erfüllt, ist  $A$  eine  $\lambda^*$ -messbare Menge sobald

$$\lambda^*(Q) \geq \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c) \quad \text{für alle } Q \subseteq X. \quad (\heartsuit)$$

Dabei braucht man außerdem nur Mengen  $Q$  mit  $\lambda^*(Q) < \infty$  zu betrachten, da diese Ungleichung sonst automatisch erfüllt ist.

b) Ist  $\lambda^*(A) = 0$ , dann ist  $A$   $\lambda^*$ -messbar, da wegen  $(\lambda^*2)$  dann  $\lambda^*(Q \cap A) = 0$  gilt, und obige Ungleichung zu  $\lambda^*(Q) \geq \lambda^*(Q \cap A^c)$  wird, was wegen  $(\lambda^*2)$  immer stimmt. Entsprechend, wenn  $\lambda^*(A^c) = 0$ .

Um sicherzugehen, dass wir uns da nicht verzettelt haben, nun gleich mal

**Proposition 1.7 (Messbarkeit der Intervalle)** Jedes  $A \in \mathcal{I}_{\mathbb{R}}$  ist Lebesgue-messbar.

**Beweis.** Sei  $A \in \mathcal{I}_{\mathbb{R}}$  und  $Q \subseteq \mathbb{R}$ . Wir überprüfen die Bedingung  $(\heartsuit)$ . Das äußere Maß  $\lambda^*(Q)$  ist definiert mittels abzählbarer Überdeckungen von  $Q$  durch Mengen aus  $\mathcal{I}_{\mathbb{R}}$ . Sei  $(E_k)_{k \geq 1}$  eine solche. So wie  $Q$  zerschneiden wir nun auch die überdeckenden Mengen  $E_k$  mittels  $A$  und erhalten Mengen  $B_k := E_k \cap A \in \mathcal{I}_{\mathbb{R}}$  und  $E_k \cap A^c = E_k \setminus A = \bigsqcup_{j=1}^{m_k} C_{k,j}$  für geeignete disjunkte  $C_{k,1}, \dots, C_{k,m_k} \in \mathcal{I}_{\mathbb{R}}$ . Nachdem  $\lambda^*|_{\mathcal{I}_{\mathbb{R}}} = \lambda$  und  $\lambda$  auf  $\mathcal{I}_{\mathbb{R}}$  endlich additiv ist, folgt

$$\lambda^*(E_k) = \lambda^*\left(B_k \sqcup \bigsqcup_{j=1}^{m_k} C_{k,j}\right) = \lambda^*(B_k) + \sum_{j=1}^{m_k} \lambda^*(C_{k,j}) \quad \text{für } k \geq 1.$$

Nun bilden  $(B_k)_{k \geq 1}$  und  $(C_{k,j})_{k \geq 1, 1 \leq j \leq m_k}$  aber abzählbare  $\mathcal{I}_{\mathbb{R}}$ -Überdeckungen von  $Q \cap A$  bzw von  $Q \cap A^c$ , daher gilt die Ungleichung in

$$\begin{aligned} \sum_{k \geq 1} \lambda^*(E_k) &= \sum_{k \geq 1} \lambda^*(B_k) + \sum_{k \geq 1} \sum_{j=1}^{m_k} \lambda^*(C_{k,j}) \\ &\geq \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c). \end{aligned}$$

Da die  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung  $(E_k)_{k \geq 1}$  von  $Q$  aber beliebig war, sehen wir, nach Definition von  $\lambda^*(Q)$ ,

$$\lambda^*(Q) \geq \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c),$$

also ist wirklich  $(\heartsuit)$  erfüllt. ■

Na gut, das bringt uns aber noch nicht weiter. Wir sollten jetzt herausfinden, welche anderen, komplizierteren und interessanteren Mengen Lebesgue-messbar sind. Angekündigt wurde ja schon, dass  $\mathcal{L}_{\mathbb{R}}$  so ziemlich alles enthält, was Ihnen im mathematischen Alltag begegnet. Um uns davon zu überzeugen müssen wir jetzt womöglich die Bedingung  $(\heartsuit)$  für endlos viele Mengenarten nachrechnen?

Zum Glück geht es viel einfacher wenn wir die Sache ein wenig anders betrachten. Wir werden uns überlegen, welche *Struktur* die Familie  $\mathcal{L}_{\mathbb{R}}$  besitzt, und damit sehen, dass man durch die meisten natürlichen Operationen der Analysis nicht aus  $\mathcal{L}_{\mathbb{R}}$  herausfallen kann.

**Der  $d$ -dimensionale Fall.** Ganz analog und mit derselben Vorstellung geht das auch in  $\mathbb{R}^d$ , wo wir  $A \subseteq \mathbb{R}^d$  *Lebesgue-messbar* oder  $\lambda^{d*}$ -*messbar* nennen, falls  $\lambda^{d*}(Q) = \lambda^{d*}(Q \cap A) + \lambda^{d*}(Q \cap A^c)$  für alle  $Q \subseteq \mathbb{R}^d$ . Dabei muss man ggf wieder nur  $\geq$  beweisen, und zwar wenn  $\lambda^{d*}(Q) < \infty$ . Die Familie Lebesgue-messbarer Mengen bezeichnen wir mit  $\mathcal{L}_{\mathbb{R}^d}$ . Auch die Proposition überträgt sich leicht, es gilt  $\mathcal{I}_{\mathbb{R}^d} \subseteq \mathcal{L}_{\mathbb{R}^d}$ . (Wieder derselbe Beweis!)

## 1.6 Algebren und $\sigma$ -Algebren

Wir wenden uns also der Frage nach der *Struktur* relevanter Familien von Mengen zu. Dabei ist es günstig, gleich Teilmengen eines beliebigen Raumes zu erlauben,

dadurch wird die Sache nicht schwieriger. (Sie können zunächst aber gerne einfach an  $\mathbb{R}$  denken.) Es sei daher  $X$  eine beliebige nichtleere Menge. Wir interessieren uns für Teilfamilien  $\mathcal{A}$  von  $\mathcal{P}(X)$ . Damit wir mit den betreffenden Mengen auch etwas tun können, streben wir für unsere Familie die *Abgeschlossenheit unter einfachen Mengenoperationen* an, mit  $A, B$  soll etwa auch  $A \cup B$  dabei sein, damit wir auch mit dieser Menge arbeiten können. Für die Zwecke der Analysis möchte man sogar sicherstellen, dass auch abzählbar viele Operationen nicht aus  $\mathcal{A}$  herausführen. Wir wollen daher Familien von Mengen zu untersuchen, die unter endlich oder abzählbar vielen Mengenoperationen abgeschlossen sind.

**Definition 1.5** Eine Familie  $\mathcal{A} \subseteq \mathcal{P}(X)$  von Teilmengen einer Menge  $X$  ist eine Algebra (auf, in, oder über  $X$ ), falls sie folgende Eigenschaften hat:

- (A1)  $X \in \mathcal{A}$ ,
- (A2)  $A \in \mathcal{A} \implies A^c \in \mathcal{A}$ ,
- (A3)  $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$ .

Die Familie  $\mathcal{A} \subseteq \mathcal{P}(X)$  ist eine  $\sigma$ -Algebra (auf, in, oder über  $X$ ), falls die Bedingungen (A1) und (A2) erfüllt sind, und dazu noch

- (A3)'  $A_n \in \mathcal{A}, n \geq 1 \implies \bigcup_{n \geq 1} A_n \in \mathcal{A}$ .

Die Bedingungen (A1)-(A3) (bzw (A3)') stellen in weiterer Folge sicher, dass  $\mathcal{A}$  auch unter Differenzbildung und endlichen (bzw abzählbaren) Durchschnitten abgeschlossen ist. Wir halten diese und ein paar weitere einfache Eigenschaften solcher Mengensysteme fest:

**Proposition 1.8 (Algebren und  $\sigma$ -Algebren)** Sei  $\mathcal{A} \subseteq \mathcal{P}(X)$ .

a) Ist  $\mathcal{A}$  eine Algebra, so gelten

- $\emptyset \in \mathcal{A}$ ,
- $A, B \in \mathcal{A} \implies A \cap B, A \setminus B, A \triangle B \in \mathcal{A}$ ,
- $A_1, \dots, A_n \in \mathcal{A} \implies \bigcup_{k=1}^n A_k, \bigcap_{k=1}^n A_k \in \mathcal{A}$ .

b)  $\mathcal{A}$  ist eine Algebra  $\iff \mathcal{A}$  erfüllt (A1), (A2) und

- (A4)  $A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}$ .

c) Ist  $\mathcal{A}$  eine  $\sigma$ -Algebra, dann auch eine Algebra.

d)  $\mathcal{A}$  ist eine  $\sigma$ -Algebra  $\iff \mathcal{A}$  erfüllt (A1), (A2) und

- (A4)'  $A_n \in \mathcal{A}, n \geq 1 \implies \bigcap_{n \geq 1} A_n \in \mathcal{A}$ .

**Beweis.** a) Folgt aus den Identitäten  $\emptyset = X^c, A \cap B = (A^c \cup B^c)^c, A \setminus B = A \cap B^c, A \triangle B = (A \setminus B) \cup (B \setminus A)$ , von deren rechten Seiten wir jeweils wissen, dass sie in  $\mathcal{A}$  liegen. Die letzte Behauptung folgt durch Induktion nach  $n$ .

b) Wir wissen schon, dass in jeder Algebra (A4) gilt. Für die Umkehrung verwende man  $A \cup B = (A^c \cap B^c)^c$ .

c) (A3) folgt aus (A3)' und (A1)&(A2) mit  $A_1 := A, A_2 := B$  und  $A_n := \emptyset = X^c$  für  $n > 2$ .

d) Folgt aus  $\bigcap_{n \geq 1} A_n = (\bigcup_{n \geq 1} A_n^c)^c$  und  $\bigcup_{n \geq 1} A_n = (\bigcap_{n \geq 1} A_n^c)^c$ . ■

Kurz: Algebren sind die unter endlichen Mengenoperationen abgeschlossenen Mengensysteme,  $\sigma$ -Algebren die unter abzählbaren Mengenoperationen abgeschlossenen. Tatsächlich erweisen sich gerade  $\sigma$ -Algebren als idealer Lebensraum für interessante Mengenfunktionen.

Ist  $X$  oder auch nur  $\mathcal{A}$  endlich, so stimmen die Begriffe Algebra und  $\sigma$ -Algebra offenbar überein. Wenige, ganz einfache Beispiele drängen sich sofort auf, wesentlich spannendere Situationen besprechen wir demnächst.

**Beispiel 1.5 (Algebren und  $\sigma$ -Algebren)** *a)  $\mathcal{A} := \{\emptyset, X\}$  und  $\mathcal{A} := \mathcal{P}(X)$  sind in jedem Fall ( $\sigma$ -)Algebren (offenbar die kleinst- bzw größtmögliche).  
b) Für  $A \subseteq X$  ist  $\mathcal{A} := \{\emptyset, X, A, A^c\}$  die kleinste ( $\sigma$ -)Algebra die  $A$  enthält.  
c) In jedem Fall ist  $\mathcal{A} := \{A \subseteq X : A \text{ oder } A^c \text{ ist endlich}\}$  eine Algebra, diese Familie ist eine  $\sigma$ -Algebra genau wenn  $X$  endlich ist.  
d) In jedem Fall ist  $\mathcal{A} := \{A \subseteq X : A \text{ oder } A^c \text{ ist abzählbar}\}$  eine  $\sigma$ -Algebra.  
e) Die eingangs erwähnte Familie  $\mathcal{I}_{\mathbb{R}} := \{(a, b] : a, b \in \mathbb{R}, a \leq b\} \subseteq \mathcal{P}(\mathbb{R})$  halboffener Intervalle in  $\mathbb{R}$ , auf der die elementare Längenfunktion  $\lambda$  definiert wurde, ist nicht einmal eine Algebra. Nicht nur, dass  $X = \mathbb{R} \notin \mathcal{I}_{\mathbb{R}}$ , auch sind Komplemente von Intervallen aus  $\mathcal{I}_{\mathbb{R}}$  nicht in  $\mathcal{I}_{\mathbb{R}}$ .*

**Bemerkung 1.7 (Von Algebren zu  $\sigma$ -Algebren)** *Um nachzuweisen, dass eine Familie  $\mathcal{A} \subseteq \mathcal{P}(X)$  eine  $\sigma$ -Algebra ist, kann es günstig sein, erst zu zeigen, dass eine Algebra vorliegt. Danach muss man nämlich (A3)' nicht mehr ganz allgemein zeigen, sondern es reicht,  $\bigcup_{n \geq 1} A_n \in \mathcal{A}$  für disjunkte Folgen  $(A_n)_{n \geq 1}$  in  $\mathcal{A}$  zu verifizieren. (Siehe PS.)*

**Bemerkung 1.8 (Algebren bzw  $\sigma$ -Algebren auf Teilmengen)** *Ist  $\mathcal{A} \subseteq \mathcal{P}(X)$  eine Algebra oder  $\sigma$ -Algebra auf  $X$  und  $\emptyset \neq Y \subseteq X$ , dann ist  $\mathcal{A}_Y := \{Y \cap A : A \in \mathcal{A}\}$  eine Algebra oder  $\sigma$ -Algebra auf  $Y$ , die Spur von  $\mathcal{A}$  in  $Y$ . (Siehe PS.)*

Jetzt aber zu der oben getroffenen Behauptung, strukturelle Einsichten könnten es uns etwa im Fall  $X = \mathbb{R}$  ersparen, die unhandliche Messbarkeitsbedingung ( $\heartsuit$ ) für viele (Typen von) Mengen wieder nachzuprüfen. Das geht wirklich, und zwar ganz einfach, weil sich viele für die Analysis relevante Mengen mittels abzählbarer Mengenoperationen aus *Intervallen* gewinnen lassen:

**Proposition 1.9 (Jede  $\sigma$ -Algebra  $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R})$  mit  $\mathcal{I}_{\mathbb{R}} \subseteq \mathcal{A}$  enthält auch ...)** *Ist  $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R})$  eine  $\sigma$ -Algebra mit  $\mathcal{I}_{\mathbb{R}} \subseteq \mathcal{A}$ , dann enthält  $\mathcal{A}$  sämtliche Intervalle und auch alle offenen und alle abgeschlossenen Teilmengen von  $\mathbb{R}$ .*

**Beweis.** Wir haben schon im PS gesehen, dass Intervalle jeder Art mittels abzählbar vieler Mengenoperationen durch Folgen in  $\mathcal{I}_{\mathbb{R}}$  dargestellt werden können, zB

$$[a, b] = \bigcap_{n \geq 1} (a - \frac{1}{n}, b].$$

Da jede der Mengen rechts in  $\mathcal{I}_{\mathbb{R}}$  und damit in  $\mathcal{A}$  liegt, und  $\mathcal{A}$  unter abzählbaren Durchschnitten abgeschlossen ist, folgt  $[a, b] \in \mathcal{A}$ . Analog für andere Intervalle.

Ist  $G \subseteq \mathbb{R}$  eine offene Menge, so garantiert Lemma 1.1, dass  $G = \bigcup_{n \geq 1} A_n$  für eine geeignete Folge  $(A_n)_{n \geq 1}$  in  $\mathcal{I}_{\mathbb{R}}$ . Wie zuvor folgt  $G \in \mathcal{A}$ . Jede abgeschlossene Menge  $F$  ist das Komplement  $F = G^c$  einer offenen Menge  $G$ . Wie eben bemerkt ist  $G \in \mathcal{A}$ , und wegen (A2) auch  $F \in \mathcal{A}$ . ■

Das ging jetzt wirklich sehr bequem. Durch vergleichbar einfache Argumente sieht man für die meisten in der Analysis auftretenden Mengen ziemlich leicht, dass sie automatisch zu jeder  $\sigma$ -Algebra mit  $\mathcal{I}_{\mathbb{R}} \subseteq \mathcal{A}$  gehören. Damit Sie mir das schon jetzt leichter glauben, hier noch ein Beispiel.

**Beispiel 1.6** *Es sei  $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R})$  eine  $\sigma$ -Algebra mit  $\mathcal{I}_{\mathbb{R}} \subseteq \mathcal{A}$ , und  $(x_n)_{n \geq 1}$  eine beliebige Folge in  $\mathbb{R}$ . Dann liegt auch die Menge  $H$  ihrer Häufungspunkte in  $\mathcal{A}$ , man kann sie nämlich darstellen als*

$$\begin{aligned} H &= \{x \in \mathbb{R} : \forall \varepsilon > 0 \ \forall m \geq 1 \ \exists n \geq m : x \in (x_n - \varepsilon, x_n + \varepsilon)\} \\ &= \{x \in \mathbb{R} : \forall k \geq 1 \ \forall m \geq 1 \ \exists n \geq m : x \in (x_n - 1/k, x_n + 1/k)\} \\ &= \bigcap_{k \geq 1} \bigcap_{m \geq 1} \bigcup_{n \geq m} (x_n - 1/k, x_n + 1/k), \end{aligned}$$

und wir wissen ja schon, dass alle Intervalle in  $\mathcal{A}$  liegen. Also folgt  $H \in \mathcal{A}$ . (Diese Aussage folgt übrigens auch auf andere Art aus der Proposition. Sehen Sie wie?)

Das letzte Beispiel illustriert ein wichtiges Prinzip: Man kann die typischen Bedingungen der Form " $\forall \varepsilon > 0 \dots$ " aus der Analysis entsprechend durch " $\forall \varepsilon = 1/k$ ,  $k \in \mathbb{N}$ ,  $\dots$ " ersetzen, also durch abzählbar viele Forderungen. Daher sind  $\sigma$ -Algebren ausreichend robust für die Analysis.

**Der  $d$ -dimensionale Fall.** Es wird Sie nicht überraschen zu lesen, dass entsprechendes auch in  $\mathbb{R}^d$  funktioniert. Konkret: Ist  $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^d)$  eine  $\sigma$ -Algebra mit  $\mathcal{I}_{\mathbb{R}}^d \subseteq \mathcal{A}$ , dann enthält  $\mathcal{A}$  sämtliche  $d$ -dimensionalen Quader (also alle " $d$ -dimensionalen Intervalle"), wobei wir auch unbeschränkte Versionen einschließen. Mit dabei sind zB in  $\mathbb{R}^2$  insbesondere

$$\{a\} \times \{c\}, \quad (a, b) \times \{c\}, \quad (a, b) \times [c, d], \quad (a, \infty) \times [c, d], \quad (-\infty, b] \times (-\infty, d], \text{ etc}$$

$\mathcal{A}$  enthält auch alle offenen und alle abgeschlossenen Teilmengen von  $\mathbb{R}^d$ . Der Beweis verläuft genau wie oben. Man überlegt jeweils, dass man Mengen des betreffenden Typs durch eine Folge halboffener Quader aus  $\mathcal{I}_{\mathbb{R}}^d$  ausdrücken kann. (Probieren Sie das doch bitte an einigen Beispielen aus.)

## 1.7 Das Lebesgue-Maß

Mit dem Konzept der  $\sigma$ -Algebra ausgerüstet kommen wir zu unserer ursprünglichen Fragestellung zurück. Wie sieht es denn nun mit der Struktur von  $\mathcal{L}_{\mathbb{R}}$  aus? Naja, Sie können schon erraten, dass dies wohl eine  $\sigma$ -Algebra ist. Es gilt aber noch mehr: wir zeigen auch, dass sich  $\lambda^*$  auf  $\mathcal{L}_{\mathbb{R}}$  sehr gut verhält. Da wir längst wissen, dass  $\lambda^*$  die Längenfunktion fortsetzt, erlauben wir uns dann, die Einschränkung des äußeren Maßes auf die Familie der messbaren Mengen wieder mit  $\lambda$  zu bezeichnen. Das Hauptergebnis dieses Kapitels lautet nun:

**Satz 1.6 (Lebesgue- $\sigma$ -Algebra & Lebesgue-Maß auf  $\mathbb{R}$ )** *Die Familie  $\mathcal{L}_{\mathbb{R}} \subseteq \mathcal{P}(\mathbb{R})$  Lebesgue-messbarer Mengen in  $\mathbb{R}$  ist eine translationsinvariante  $\sigma$ -Algebra. Sie enthält sämtliche Intervalle, sowie alle offenen Mengen. Die durch*

$$\lambda := \lambda^*|_{\mathcal{L}_{\mathbb{R}}} : \mathcal{L}_{\mathbb{R}} \rightarrow [0, \infty]$$

*definierte Fortsetzung der Längenfunktion nach  $\mathcal{L}_{\mathbb{R}}$  ist translationsinvariant und  $\sigma$ -additiv, für jede Folge  $(A_n)_{n \geq 1}$  paarweise disjunkter Mengen  $A_n \in \mathcal{L}_{\mathbb{R}}$  gilt also*

$$\lambda \left( \biguplus_{n \geq 1} A_n \right) = \sum_{n \geq 1} \lambda(A_n).$$

*Ist  $A \in \mathcal{L}_{\mathbb{R}}$  mit  $\lambda(A) = 0$ , dann gilt  $B \in \mathcal{L}_{\mathbb{R}}$  für alle  $B \subseteq A$ .*

**Definition 1.6** *Wir nennen  $\mathcal{L}_{\mathbb{R}}$  die Lebesguesche  $\sigma$ -Algebra auf  $\mathbb{R}$ . Die Mengenfunktion  $\lambda := \lambda^*|_{\mathcal{L}_{\mathbb{R}}} : \mathcal{L}_{\mathbb{R}} \rightarrow [0, \infty]$  ist das eindimensionale Lebesguesche Maß (oder Lebesgue-Maß).*

Das Lebesguesche Maß ist also die Einschränkung von  $\lambda^*$  auf  $\mathcal{L}_{\mathbb{R}}$ . Um keine Zweifel aufkommen zu lassen, notieren wir gleich:

**Proposition 1.10** *Das Lebesguesche Maß  $\lambda : \mathcal{L}_{\mathbb{R}} \rightarrow [0, \infty]$  ist auch endlich additiv, monoton, endlich subadditiv,  $\sigma$ -subadditiv, und stetig von unten.*

**Beweis.** Wie im Beweis von Proposition 1.1. (Wir kommen darauf zurück.) ■

Mit  $\lambda$  erhalten wir eines der wichtigsten Objekte der Analysis. Das Lebesguesche Maß ist *der* natürliche Größenbegriff für eine riesige Familie von Teilmengen von  $\mathbb{R}$ . Wir halten noch ausdrücklich fest, dass es auf  $\mathcal{L}_{\mathbb{R}}$  wirklich keine Alternative zu  $\lambda$  gibt.

**Proposition 1.11 (Charakterisierungen von  $\lambda$ )** *Das Lebesguesche Maß  $\lambda$  ist die einzige  $\sigma$ -additive Mengenfunktion auf  $\mathcal{L}_{\mathbb{R}}$  welche die Längenfunktion fortsetzt, und die einzige  $\sigma$ -additive Mengenfunktion auf  $\mathcal{L}_{\mathbb{R}}$  welche translationsinvariant ist und dem Intervall  $(0, 1]$  Maß 1 zuordnet.*

**Beweis. a)** Sei  $\nu : \mathcal{L}_{\mathbb{R}} \rightarrow [0, \infty]$  irgendeine  $\sigma$ -additive Mengenfunktion mit  $\nu|_{\mathcal{I}_{\mathbb{R}}} = \lambda|_{\mathcal{I}_{\mathbb{R}}}$ . Für  $A \in \mathcal{L}_{\mathbb{R}}$  und jede  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung  $(E_k)_{k \geq 1}$  von  $A$  gilt (weil  $\nu$   $\sigma$ -additiv und damit monoton ist und  $\lambda|_{\mathcal{I}_{\mathbb{R}}}$  fortsetzt)

$$\nu(A) \leq \nu\left(\bigcup_{k \geq 1} E_k\right) \leq \sum_{k \geq 1} \nu(E_k) = \sum_{k \geq 1} \lambda(E_k),$$

und somit  $\nu \leq \lambda^*$  auf  $\mathcal{L}_{\mathbb{R}}$ . Ist nun  $Q \in \mathcal{I}_{\mathbb{R}}$ , dann folgt aus dieser Ungleichung

$$\lambda^*(Q) = \nu(Q) = \nu(Q \cap A) + \nu(Q \cap A^c) \leq \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c), \quad (1.8)$$

für  $A \in \mathcal{L}_{\mathbb{R}}$ , und nach Definition dieser Familie stimmen die Ausdrücke ganz links und ganz rechts überein, es muss also durchweg Gleichheit herrschen. Weil in (1.8) sowohl  $\nu(Q \cap A) \leq \lambda^*(Q \cap A)$  als auch  $\nu(Q \cap A^c) \leq \lambda^*(Q \cap A^c)$  gilt, muss insbesondere  $\nu(Q \cap A) = \lambda^*(Q \cap A)$  sein. Letzteres gilt also für jedes  $Q \in \mathcal{I}_{\mathbb{R}}$ . Nehmen wir  $Q = (n, n + 1]$ ,  $n \in \mathbb{Z}$ , so finden wir

$$\nu(A) = \sum_{n \in \mathbb{Z}} \nu((n, n + 1] \cap A) = \sum_{n \in \mathbb{Z}} \lambda^*((n, n + 1] \cap A) = \lambda^*(A),$$

wie behauptet.

**b)** Ist  $\nu : \mathcal{L}_{\mathbb{R}} \rightarrow [0, \infty]$  nun eine translationsinvariante  $\sigma$ -additive Mengenfunktion mit  $\nu((0, 1]) = 1$ , dann gilt nach Proposition 1.2, dass  $\nu|_{\mathcal{I}_{\mathbb{R}}} = \lambda|_{\mathcal{I}_{\mathbb{R}}}$ . Aus Schritt a) folgt dann  $\nu = \lambda$ . ■

Wenig überraschend ist nun, dass  $\mathcal{L}_{\mathbb{R}}$  eine echte Teilfamilie von  $\mathcal{P}(\mathbb{R})$  sein muss.

**Proposition 1.12 Vitali-Mengen sind nicht Lebesgue-messbar.**

**Beweis.** Wäre  $\mathcal{L}_{\mathbb{R}} = \mathcal{P}(\mathbb{R})$ , so müsste nach obigem Satz  $\lambda^*$   $\sigma$ -additiv auf  $\mathcal{P}(\mathbb{R})$  sein. Wir wissen aber seit Satz 1.2, dass dies unmöglich ist.

Konkreter ist jede Vitali-Menge  $V$  eine nicht Lebesgue-messbare Menge: Nehmen wir an, es wäre  $V \in \mathcal{L}_{\mathbb{R}}$ . Da  $\mathcal{L}_{\mathbb{R}}$  translationsinvariant ist, folgt dann  $V + t \in \mathcal{L}_{\mathbb{R}}$  für jedes  $t \in \mathbb{R}$ . Erinnern wir uns einmal mehr an (1.6), so finden wir genau wie im Beweis von Satz 1.2 (nun mit  $\lambda$  anstelle von  $\mu$ ), dass dies  $\lambda((0, 1]) \in \{0, \infty\}$  impliziert, ein Widerspruch. ■

**Bemerkung 1.9** *Leider gibt es, von der Definition der Lebesgue-Messbarkeit (und Varianten davon) abgesehen, keine "konstruktive" Möglichkeit,  $\mathcal{L}_{\mathbb{R}}$  expliziter zu beschreiben. (Gemeint ist damit nicht, dass man noch keine Möglichkeit gefunden hat, sondern dass man beweisen kann, dass es sowas nicht gibt.)*

Nun aber zurück zum zentralen Satz 1.6. Der Beweis ist (erwartungsgemäß) nicht ganz trivial, wir müssen und noch einmal gut konzentrieren. Dafür bekommen wir aber auch ein wirklich fundamentales Ergebnis.

**Beweis von Satz 1.6.** Da wir schon wissen, dass  $\mathcal{L}_{\mathbb{R}}$  und  $\lambda^*$  translationsinvariant sind, und dass  $\lambda^*$  die Längenfunktion fortsetzt, müssen wir noch zeigen, dass  $\mathcal{L}_{\mathbb{R}}$  eine  $\sigma$ -Algebra und  $\lambda^*$  dort  $\sigma$ -additiv ist. Proposition 1.9 stellt dann sicher, dass  $\mathcal{L}_{\mathbb{R}}$  sämtliche Intervalle und alle offenen Teilmengen von  $\mathbb{R}$  enthält. Die Letzte Aussage des Satzes folgt sofort aus Bemerkung 1.6 b). (Denn  $\lambda(A) = 0$  bedeutet  $\lambda^*(A) = 0$ , und daher  $\lambda^*(B) = 0$ , woraus sich dank der Bemerkung  $B \in \mathcal{L}_{\mathbb{R}}$  ergibt.)

**(i)** Wir beobachten zuerst, dass  $\mathcal{L}_{\mathbb{R}}$  eine Algebra ist: Offenkundig ist  $\mathbb{R} \in \mathcal{L}_{\mathbb{R}}$ , und da die Bedingung der  $\lambda^*$ -Messbarkeit symmetrisch in  $A$  und  $A^c$  ist, führt



die Komplementbildung nicht aus  $\mathcal{L}_{\mathbb{R}}$  heraus. Schließlich ist für  $A, B \in \mathcal{L}_{\mathbb{R}}$  auch  $A \cup B \in \mathcal{L}_{\mathbb{R}}$ , denn für alle  $Q \subseteq \mathbb{R}$  gilt

$$\begin{aligned}\lambda^*(Q) &= \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c) & (A \in \mathcal{L}_{\mathbb{R}}) \\ &= \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c \cap B) + \lambda^*(Q \cap A^c \cap B^c) & (B \in \mathcal{L}_{\mathbb{R}}) \\ &\geq \lambda^*((Q \cap A) \cup (Q \cap A^c \cap B)) + \lambda^*(Q \cap A^c \cap B^c) & ((\lambda^*3)) \\ &= \lambda^*(Q \cap (A \cup B)) + \lambda^*(Q \cap (A \cup B)^c).\end{aligned}$$

(ii) Nun zeigen wir, dass für jede *disjunkte* Folge  $(A_n)_{n \geq 1}$  in  $\mathcal{L}_{\mathbb{R}}$

$$A := \bigsqcup_{n \geq 1} A_n \in \mathcal{L}_{\mathbb{R}} \quad \text{und} \quad \lambda^*(A) = \sum_{n \geq 1} \lambda^*(A_n). \quad (1.9)$$

Dazu stellen wir fest, dass für *disjunkte*  $E, F \in \mathcal{L}_{\mathbb{R}}$  und jedes  $Q \subseteq \mathbb{R}$ ,

$$\begin{aligned}\lambda^*(Q \cap (E \uplus F)) &= \lambda^*(Q \cap (E \uplus F) \cap E) + \lambda^*(Q \cap (E \uplus F) \cap E^c) \\ &= \lambda^*(Q \cap E) + \lambda^*(Q \cap F).\end{aligned}$$

Per Induktion (man verwende  $E := \bigsqcup_{n=1}^N A_n$  und  $F := A_{N+1}$ ) folgt

$$\lambda^*\left(Q \cap \bigsqcup_{n=1}^N A_n\right) = \sum_{n=1}^N \lambda^*(Q \cap A_n) \quad \text{für } N \geq 1 \text{ und } Q \subseteq \mathbb{R}. \quad (1.10)$$

Nach Schritt (i) ist  $\bigsqcup_{n=1}^N A_n \in \mathcal{L}_{\mathbb{R}}$ , und daher für jedes  $Q \subseteq \mathbb{R}$  und  $N \geq 1$ ,

$$\begin{aligned}\lambda^*(Q) &= \lambda^*\left(Q \cap \bigsqcup_{n=1}^N A_n\right) + \lambda^*\left(Q \cap \left(\bigsqcup_{n=1}^N A_n\right)^c\right) \\ &\geq \sum_{n=1}^N \lambda^*(Q \cap A_n) + \lambda^*(Q \cap A^c).\end{aligned}$$

(Der zweite Schritt verwendet (1.10) und  $(\lambda^*2)$ .) Es folgt, für  $N \rightarrow \infty$ ,

$$\begin{aligned}\lambda^*(Q) &\geq \sum_{n \geq 1} \lambda^*(Q \cap A_n) + \lambda^*(Q \cap A^c) \\ &\geq \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c) \geq \lambda^*(Q), \quad (\text{zweimal } (\lambda^*3))\end{aligned}$$

also für alle  $Q \subseteq \mathbb{R}$ ,

$$\lambda^*(Q) = \sum_{n \geq 1} \lambda^*(Q \cap A_n) + \lambda^*(Q \cap A^c) = \lambda^*(Q \cap A) + \lambda^*(Q \cap A^c),$$

dh  $A \in \mathcal{L}_{\mathbb{R}}$ , und es ist  $\lambda^*(A) = \sum_{n \geq 1} \lambda^*(A_n)$  (wähle  $Q := A$ ).

(iii) Nach (i) und (ii) ist  $\mathcal{L}_{\mathbb{R}}$  eine Algebra, die unter *disjunkten* abzählbaren Vereinigungen abgeschlossen ist. Daher ist  $\mathcal{L}_{\mathbb{R}}$  sogar eine  $\sigma$ -Algebra (siehe Bemerkung 1.7). Außerdem ist nach (ii) das äußere Maß  $\lambda^*$  auch  $\sigma$ -additiv auf  $\mathcal{L}_{\mathbb{R}}$ . ■

Das sollten Sie jetzt wirklich feiern!

**Der  $d$ -dimensionale Fall.** Wie schon in einigen früheren Abschnitten erwähnen wir noch, dass man all dies auch in  $\mathbb{R}^d$  anstelle von  $\mathbb{R}$  analog durchführen kann. (Genauere Rechtfertigung im nächsten Kapitel.) Es gilt:

**Satz 1.7 (Lebesgue- $\sigma$ -Algebra & Lebesgue-Maß auf  $\mathbb{R}^d$ )** Die Familie  $\mathcal{L}_{\mathbb{R}^d} \subseteq \mathcal{P}(\mathbb{R}^d)$  Lebesgue-messbarer Mengen in  $\mathbb{R}^d$  ist eine translationsinvariante  $\sigma$ -Algebra. Sie enthält sämtliche Typen  $d$ -dimensionaler Quader, sowie alle offenen Mengen. Die durch

$$\lambda^d := \lambda^{d*} |_{\mathcal{L}_{\mathbb{R}^d}} : \mathcal{L}_{\mathbb{R}^d} \rightarrow [0, \infty]$$

definierte Fortsetzung der Längenfunktion nach  $\mathcal{L}_{\mathbb{R}^d}$  ist translationsinvariant und  $\sigma$ -additiv, für jede Folge  $(A_n)_{n \geq 1}$  paarweise disjunkter Mengen  $A_n \in \mathcal{L}_{\mathbb{R}^d}$  gilt also

$$\lambda^d \left( \biguplus_{n \geq 1} A_n \right) = \sum_{n \geq 1} \lambda^d(A_n).$$

Ist  $A \in \mathcal{L}_{\mathbb{R}^d}$  mit  $\lambda^d(A) = 0$ , dann gilt  $B \in \mathcal{L}_{\mathbb{R}^d}$  für alle  $B \subseteq A$ .

**Definition 1.7** Wir nennen  $\mathcal{L}_{\mathbb{R}^d}$  die Lebesguesche  $\sigma$ -Algebra auf  $\mathbb{R}^d$ . Die Mengenfunktion  $\lambda^d := \lambda^{d*} |_{\mathcal{L}_{\mathbb{R}^d}} : \mathcal{L}_{\mathbb{R}^d} \rightarrow [0, \infty]$  ist das  $d$ -dimensionale Lebesguesche Maß (oder Lebesgue-Maß).

Auch die geometrischen Charakterisierungen bekommt man ganz analog:

**Proposition 1.13 (Charakterisierungen von  $\lambda^d$ )** Das  $d$ -dimensionale Lebesgue-Maß  $\lambda^d$  ist die einzige  $\sigma$ -additive Mengenfunktion auf  $\mathcal{L}_{\mathbb{R}^d}$  welche das  $d$ -dimensionale Volumen fortsetzt, und die einzige  $\sigma$ -additive Mengenfunktion auf  $\mathcal{L}_{\mathbb{R}^d}$  welche translationsinvariant ist und dem Einheitswürfel  $(0, 1]^d$  Maß 1 zuordnet.

Wie im eindimensionalen Fall gibt es auch in  $\mathbb{R}^d$  Mengen, die nicht Lebesgue-messbar sind.

**Klarstellung und Ausblick:  $m$ -dimensionales Maß in  $\mathbb{R}^d$ .** Mit dem  $d$ -dimensionalen Lebesgue-Maß  $\lambda^d$  haben wir jetzt eine Mengenfunktion gewonnen, die jeder Menge  $A$  aus der riesigen Familie  $\mathcal{L}_{\mathbb{R}^d}$  ihre (durch wünschenswerte Grundeigenschaften eindeutig bestimmte)  $d$ -dimensionale Größe zuordnet. Konkret also: für L-messbares  $A \subseteq \mathbb{R}$  ist  $\lambda(A)$  die Länge, für L-messbares  $A \subseteq \mathbb{R}^2$  ist  $\lambda^2(A)$  die Fläche, für L-messbares  $A \subseteq \mathbb{R}^3$  ist  $\lambda^3(A)$  das Volumen dieser Menge etc. Beachten Sie aber bitte:  $\lambda^d$  wird immer nur auf Teilmengen von  $\mathbb{R}^d$  angewandt, wir können sogar für ein Streckenstück  $A \subseteq \mathbb{R}^3$  nicht ohne weiteres von seiner Länge  $\lambda(A)$  sprechen, da  $\lambda$  eben nur auf  $\mathcal{L}_{\mathbb{R}} \subseteq \mathcal{P}(\mathbb{R})$  definiert,  $A$  aber keine Teilmenge von  $\mathbb{R}$  ist. Um die Länge komplizierter Teilmengen von  $\mathbb{R}^2$  oder  $\mathbb{R}^3$  zu definieren, benötigt man jeweils wieder ein eigenes Maß, das man eben auf geeignete Teilmengen dieses höherdimensionalen Raumes anwenden kann. Genau dies leisten die *Hausdorff-Maße*. Um sie zu bekommen muss man sich allerdings noch ein wenig mehr anstrengen. Dafür wird man aber auch reich belohnt, man bekommt damit zB auch einen Dimensionsbegriff für beliebig wilde Mengen (und die Hausdorff-Dimension braucht nicht einmal eine ganze Zahl zu sein). Aus Zeit- und Platzgründen gehen wir auf diese Maße hier nicht näher ein. Sie dürfen sich aber schon darauf freuen, in anderen LVen mehr über diese Maße zu erfahren, die (gemeinsam) die Dimension einer Menge erkennen können.

Neben diesen geometrisch spannenden Maßen gibt es noch allerlei weitere, die in verschiedenen Bereichen der Mathematik eine Rolle spielen. Im letzten Kapitel etwa werden wir kurz besprechen, wie die Mathematik des Zufalls auf der Maßtheorie beruht. Insgesamt lohnt es sich daher zu bemerken, dass sich viele der oben beim Lebesgue-Maß gewonnenen Erkenntnisse ohne nennenswerten Aufwand auf weitere wichtige Szenarien übertragen. Das sehen wir uns nun an.

## 2. Ein wenig abstrakte Maßtheorie

---

Das erste Kapitel kreiste um die konkrete Frage nach einer angemessenen Verallgemeinerung der Längenfunktion bzw des  $d$ -dimensionalen Volumens auf reichhaltige Familien von Mengen. Dabei mussten wir genau beachten, mit welcher Art von Mengen jeweils gerade gearbeitet wurde. Hilfreich waren da gewisse strukturelle Beobachtungen, und es wurde bereits angedeutet, dass wir diesen Aspekt wieder aufgreifen werden. Im gegenwärtigen Kapitel besprechen wir kurz den Rahmen der abstrakten Maßtheorie. Dabei passiert zunächst kaum etwas neues, vielmehr wiederholen wir erstmal, was im vorigen Kapitel geschehen ist. Allerdings formulieren wir dabei einige der schon getroffenen Überlegungen ein wenig anders. Viele unserer früheren Argumente funktionieren nämlich nicht nur im konkreten Fall von  $\lambda$  und  $\lambda^d$ , sondern immer wenn wir Mengenfamilien einer bestimmten Struktur, und Mengenfunktionen mit gewissen Grundeigenschaften vor uns haben. Das führt uns zum Begriff eines allgemeinen Maßes und eines Maßraumes etc. Dies zu bemerken bedeutet für sich genommen schon eine zusätzliche Erkenntnis. Der Gewinn, den wir aus diesem Perspektivenwechsel ziehen, reicht aber noch weiter, denn tatsächlich gibt es noch eine Vielzahl anderer interessanter Situationen, in denen unterschiedliche Maße eine zentrale Rolle spielen. Ein paar davon werden Sie im weiteren Verlauf der Vorlesung noch kennenlernen.

### 2.1 Halbringe, Prämaße und Maße

**Halbringe.** Ausgangspunkt des ersten Kapitels war die elementare Längenfunktion  $\lambda : \mathcal{I}_{\mathbb{R}} \longrightarrow [0, \infty)$  (bzw das  $d$ -dimensionalen Volumen  $\lambda^d : \mathcal{I}_{\mathbb{R}}^d \rightarrow [0, \infty)$ ), und noch bevor wir uns diesen Mengenfunktionen widmeten, beschäftigte uns die Struktur des Definitionsbereiches (siehe Bemerkung 1.2). Auf den ersten Blick erschien dabei die Beobachtung (1.7) wohl ein wenig seltsam, aber wir haben mittlerweile gesehen, dass diese Eigenschaft (und ihre Konsequenzen) wesentlich für unsere Argumentation ist. Zum Glück gilt entsprechendes auch für  $\mathcal{I}_{\mathbb{R}}^d$ , sodass wir keine neuen Argumente benötigten. Um uns auch in Zukunft Arbeit zu ersparen, fassen wir die nützlichsten Grundeigenschaften der Mengenfamilien  $\mathcal{I}_{\mathbb{R}}$  und  $\mathcal{I}_{\mathbb{R}}^d$  in folgender Definition zusammen. Wie schon im Abschnitt über  $(\sigma)$ -Algebren ist nun  $X$  eine beliebige nichtleere Menge.

**Definition 2.1** Eine Familie  $\mathcal{H} \subseteq \mathcal{P}(X)$  ist ein Halbring oder Semiring (auf, in, oder über  $X$ ), falls sie folgende Eigenschaften hat:

- (HR1)  $\emptyset \in \mathcal{H}$ ,
- (HR2)  $A, B \in \mathcal{H} \implies A \cap B \in \mathcal{H}$ ,
- (HR3)  $A, B \in \mathcal{H} \implies$  es gibt disjunkte  $C_1, \dots, C_m \in \mathcal{H}$  sodass  $A \setminus B = \bigsqcup_{j=1}^m C_j$ .

**Beispiel 2.1 a)** Die Familie  $\mathcal{I}_{\mathbb{R}} = \{(a, b] : a, b \in \mathbb{R}, a \leq b\} \subseteq \mathcal{P}(\mathbb{R})$  halboffener Intervalle in  $\mathbb{R}$  ist ein Halbring.

**b)** Ist  $I \in \mathcal{I}_{\mathbb{R}}$  ein Intervall, dann bilden dessen halboffenen Teilintervalle einen Halbring  $\mathcal{I}_I := I \cap \mathcal{I}_{\mathbb{R}} = \{(a, b] \subseteq I : a \leq b\}$ .

**c)** Die Familie der halboffenen Rechtecke in  $\mathbb{R}^2$  bildet einen Halbring  $\mathcal{I}_{\mathbb{R}^2} := \mathcal{I}_{\mathbb{R}} \times \mathcal{I}_{\mathbb{R}} = \{(a, b] \times (c, d] : (a, b], (c, d] \in \mathcal{I}_{\mathbb{R}}\}$ . Entsprechend für  $\mathcal{I}_{\mathbb{R}^d} := \mathcal{I}_{\mathbb{R}}^d \subseteq \mathcal{P}(\mathbb{R}^d)$ , siehe folgende Proposition.

**d)** Es sei  $\mathcal{E}$  eine endliche Partition von  $X$ . Dann sind  $\mathcal{E} \cup \{\emptyset\}$  und  $\mathcal{E} \cup \{\emptyset, X\}$  Halbringe auf  $X$ .

Wie üblich (und wie bei den  $(\sigma)$ -Algebren oben) sind die in der Definition angegebenen Bedingungen möglichst sparsam gehalten, damit man sie schnell überprüfen kann. Ein paar einfache aber wichtige Konsequenzen kennen Sie schon: In Eigenschaft (HR3) darf man  $B$  durch eine endliche Vereinigung solcher Mengen ersetzen (Bemerkung 1.2 e)). Ausserdem bilden Cartesische Produkte von Mengen aus Halbringen wieder einen Halbring (auf dem Cartesischen Produkt). Das hatten wir uns schon angesehen, als wir im PS die Halbring-Eigenschaften von  $\mathcal{I}_{\mathbb{R}}^d$  besprochen haben. Damit Sie nicht Sorge haben müssen, ich würde Ihnen etwas verheimlichen, finden Sie jeweils den kompletten Beweis unten. Als Zeichen dafür, dass wir genau dieses Argument schon gesehen haben, gibt es ihn aber nur kleingedruckt.

**Proposition 2.1 (Eigenschaften von Halbringen)**

- a) Es sei  $\mathcal{H} \subseteq \mathcal{P}(X)$  ein Halbring. Sind  $m \in \mathbb{N}$  und  $A, B_1, \dots, B_m \in \mathcal{H}$ , so gibt es disjunkte  $C_1, \dots, C_m \in \mathcal{H}$  sodass

$$A \setminus (B_1 \cup \dots \cup B_m) = \bigsqcup_{j=1}^m C_j.$$

(Man beachte, dass  $B_1 \cup \dots \cup B_m$  nicht zu  $\mathcal{H}$  gehören muss.)

- b) Sind  $\mathcal{H}_1$  bzw  $\mathcal{H}_2$  Halbringe auf  $X_1$  bzw  $X_2$ , dann ist

$$\mathcal{H}_1 \times \mathcal{H}_2 := \{A_1 \times A_2 : A_i \in \mathcal{H}_i\} \subseteq \mathcal{P}(X_1 \times X_2)$$

ein Halbring auf  $X_1 \times X_2$ .

**Beweis.** a) Simple Induktion nach  $m$ . Der Induktionsbeginn  $m = 1$  ist durch (HR3) gegeben. Nehmen wir also an, die Aussage sei für  $m$  korrekt, und betrachten wir  $A, B_1, \dots, B_{m+1} \in \mathcal{H}$ , dann ist

$$A \setminus \bigcup_{j=1}^{m+1} B_j = \left( A \setminus \bigcup_{j=1}^m B_j \right) \setminus B_{m+1} = \left( \bigcup_{k=1}^m C_k \right) \setminus B_{m+1} = \bigcup_{k=1}^m (C_k \setminus B_{m+1})$$

mit disjunkten  $C_k \in \mathcal{H}$ . Nach (HR3) ist aber jede der Mengen  $C_k \setminus B_{m+1}$  wiederum eine endliche disjunkte Vereinigung von Mengen aus  $\mathcal{H}$ , woraus sich sofort die Behauptung für  $m + 1$  ergibt.

- b) Auch ganz einfach: Jedenfalls ist  $\emptyset = \emptyset \times \emptyset \in \mathcal{H}_1 \times \mathcal{H}_2$ , also (HR1) erfüllt. Mit  $A_1 \times A_2, B_1 \times B_2 \in \mathcal{H}_1 \times \mathcal{H}_2$  ist auch

$$(A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \cap B_2) \in \mathcal{H}_1 \times \mathcal{H}_2,$$

somit gilt (HR2). Schließlich ist (bitte eine Skizze machen!)

$$(A_1 \times A_2) \setminus (B_1 \times B_2) = ((A_1 \setminus B_1) \times A_2) \sqcup ((A_1 \cap B_1) \times (A_2 \setminus B_2)),$$

und  $(A_1 \setminus B_1)$  bzw  $(A_2 \setminus B_2)$  sind endliche disjunkte Vereinigungen von Mengen aus  $\mathcal{H}_1$  bzw  $\mathcal{H}_2$ . Daraus folgt (HR3). ■

**Prämaße.** Danach war es wichtig, erst einmal zu prüfen, ob die auf  $\mathcal{I}_{\mathbb{R}}$  (bzw  $\mathcal{I}_{\mathbb{R}}^d$ ) definierte einfache Mengenfunktion  $\lambda$  (bzw  $\lambda^d$ ) wenigstens dort  $\sigma$ -additiv ist. Damit erwies sich diese als ein *Prämaß* im Sinne folgender

**Definition 2.2** Sei  $\mathcal{H} \subseteq \mathcal{P}(X)$  ein Halbring auf  $X$ , und  $\mu : \mathcal{H} \rightarrow [0, \infty]$  eine Mengenfunktion mit  $\mu(\emptyset) = 0$ . Man nennt  $\mu$  ein *Prämaß* auf  $\mathcal{H}$ , falls für disjunkte Folgen  $(A_n)_{n \geq 1}$  in  $\mathcal{H}$  mit  $\bigsqcup_{n \geq 1} A_n \in \mathcal{H}$  stets gilt, dass

$$\mu \left( \bigsqcup_{n \geq 1} A_n \right) = \sum_{n \geq 1} \mu(A_n).$$

Ob eine konkret gegebene Mengenfunktion ein Prämaß ist, muss man halt jeweils überprüfen. Erst danach erledigt die nun zu beschreibende allgemeine Theorie den Rest ohne weitere Anstrengung Ihrerseits.

Grundlegende Eigenschaften des Prämaßes  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  haben wir dann in Proposition 1.1 aus der  $\sigma$ -Additivität abgeleitet. Ein kritischer Blick zeigt, dass der Beweis ohne Änderung für jedes Prämaß funktioniert. Sie finden den Beweis wieder im Kleingedruckten unten. (Er stimmt mit dem Beweis von Proposition 1.1 überein, es wurde nur  $\lambda$  durch  $\mu$  ersetzt.)

**Proposition 2.2 (Eigenschaften von Prämaßen)** Jedes Prämaß  $\mu : \mathcal{H} \rightarrow [0, \infty]$  hat folgende Eigenschaften:

a)  $\mu$  ist endlich additiv auf  $\mathcal{H}$ : für paarweise disjunkte  $A_1, \dots, A_N \in \mathcal{H}$  mit  $\biguplus_{n=1}^N A_n \in \mathcal{H}$  gilt

$$\mu \left( \biguplus_{n=1}^N A_n \right) = \sum_{n=1}^N \mu(A_n).$$

b)  $\mu$  ist monoton (isoton) auf  $\mathcal{H}$ : sind  $A, B \in \mathcal{H}$ ,  $B \subseteq A$ , so gilt  $\mu(B) \leq \mu(A)$ .

c)  $\mu$  ist endlich subadditiv auf  $\mathcal{I}_{\mathbb{R}}$ : für  $A_1, \dots, A_N \in \mathcal{H}$  mit  $\bigcup_{n=1}^N A_n \in \mathcal{H}$  gilt

$$\mu \left( \bigcup_{n=1}^N A_n \right) \leq \sum_{n=1}^N \mu(A_n).$$

d)  $\mu$  ist  $\sigma$ -subadditiv auf  $\mathcal{H}$ :

$$A_n \in \mathcal{H}, n \geq 1, \text{ und } \bigcup_{n \geq 1} A_n \in \mathcal{H} \implies \mu \left( \bigcup_{n \geq 1} A_n \right) \leq \sum_{n \geq 1} \mu(A_n).$$

e)  $\mu$  ist stetig von unten auf  $\mathcal{H}$ :

$$A_n \in \mathcal{H}, n \geq 1, \text{ und } A_n \nearrow A \in \mathcal{H} \implies \lim_{n \rightarrow \infty} \mu(A_n) = \mu(A).$$

**Beweis.** a) Definiere  $A_n := \emptyset \in \mathcal{H}$  für  $n > N$  und verwende die  $\sigma$ -Additivität:

$$\mu \left( \biguplus_{n=1}^N A_n \right) = \mu \left( \biguplus_{n \geq 1} A_n \right) = \sum_{n \geq 1} \mu(A_n) = \sum_{n=1}^N \mu(A_n).$$

b) Aus  $A, B \in \mathcal{H}$ ,  $B \subseteq A$  folgt  $A = B \uplus (A \setminus B) = B \uplus \biguplus_{k=1}^n C_k$  mit  $B, C_k \in \mathcal{H}$  disjunkt. Daher ist, nach a),  $\mu(A) = \mu(B) + \sum_{k=1}^n \mu(C_k) \geq \mu(B)$ . Beachten Sie:  $\mu(A \setminus B)$  ist hier ia nicht definiert! Wir müssen peinlich genau darauf achten, auf welche Mengen  $\mu$  angewandt werden kann!

c) Folgt aus d) mit  $A_n := \emptyset$  für  $n > N$ .

d) Nach Bemerkung 1.2 e) ist  $A'_n := A_n \setminus (A_1 \cup \dots \cup A_{n-1}) = \biguplus_{k=1}^{m_n} C_{n,k}$  mit disjunkten  $C_{n,k} \in \mathcal{H}$ . Daher gilt ( $\sigma$ -Additivität)

$$\mu(A) = \sum_{n \geq 1} \sum_{k=1}^{m_n} \mu(C_{n,k}) \leq \sum_{n \geq 1} \mu(A_n),$$

wobei der zweite Schritt wieder mit Bemerkung 1.2 e) zu rechtfertigen ist, wonach ja für jedes  $n \geq 1$ ,  $A_n \setminus A'_n = A_n \setminus (\biguplus_{k=1}^{m_n} C_{n,k}) = \biguplus_{j=1}^{l_n} D_{n,j}$  mit  $D_{n,j} \in \mathcal{H}$ , sodass

$$\mu(A_n) = \mu \left( \biguplus_{k=1}^{m_n} C_{n,k} \uplus \biguplus_{j=1}^{l_n} D_{n,j} \right) = \sum_{k=1}^{m_n} \mu(C_{n,k}) + \sum_{j=1}^{l_n} \mu(D_{n,j}) \geq \sum_{k=1}^{m_n} \mu(C_{n,k}).$$

e) Siehe PS. ■

**Bemerkung 2.1** Es ist nützlich, eine Beobachtung aus dem Argument für Aussage d) noch explizit festzuhalten: Ist  $\mathcal{H}$  ein Halbring und  $(A_n)_{n \geq 1}$  eine Folge in  $\mathcal{H}$ , dann gibt es eine disjunkte Folge  $(B_l)_{l \geq 1}$  in  $\mathcal{H}$  mit  $\bigcup_{n \geq 1} A_n = \biguplus_{l \geq 1} B_l$  derart, dass jedes  $B_l$  in einem  $A_n$  enthalten ist. (Man wähle einfach eine Aufzählung der Mengen  $C_{n,k}$  mit  $n \geq 1$  und  $1 \leq k \leq m_n$  aus dem Beweis.)

**Maße.** Prämaße sind ia aber insbesondere deshalb unpraktisch und heikel, weil ihr Definitionsbereich nicht einmal unter endlichen Mengenoperationen abgeschlossen zu sein braucht. Damit kann man noch nicht ordentlich (und entspannt) arbeiten! Vielmehr wünschen wir uns eine Funktion mit ebenso schönen Eigenschaften, aber eben auf einem größeren, robusten Definitionsbereich.

**Definition 2.3** Ein Maß ist ein auf einer  $\sigma$ -Algebra  $\mathcal{A} \subseteq \mathcal{P}(X)$  definiertes Prämaß. Ein Maßraum ist ein Tripel  $(X, \mathcal{A}, \mu)$  bestehend aus einer Menge  $X$ , einer  $\sigma$ -Algebra  $\mathcal{A}$  auf  $X$ , und einem Maß  $\mu$  auf  $\mathcal{A}$ . Man spricht von einem endlichen bzw unendlichen Maß(raum), falls  $\mu(X) < \infty$  bzw  $\mu(X) = \infty$  ist, und von einem normierten Maß(raum) wenn  $\mu(X) = 1$ . Die  $A \in \mathcal{A}$  sind die messbaren Mengen.

Beachte: Maße sind immer auf  $\sigma$ -Algebren  $\mathcal{A}$  definiert, in diesem Fall ist mit der Folge  $(A_n)$  auch  $\bigcup_{n \geq 1} A_n$  automatisch in  $\mathcal{A}$  enthalten. Weiters: Jedes Maß ist ein Prämaß, daher erfüllt es automatisch die Eigenschaften aus der vorigen Proposition auf ganz  $\mathcal{A}$ .

**Beispiel 2.2 (Einfache Maße)** a) Das triviale Maß auf  $\mathcal{A}$  ist gegeben durch  $\mu(A) := 0$  für alle  $A \in \mathcal{A}$ .

b) Wie ganz am Anfang festgehalten, ist die Zählfunktion  $A \mapsto \#A$  auf der Potenzmenge einer beliebigen Menge  $X$   $\sigma$ -additiv. Wir nennen  $\# : \mathcal{P}(X) \rightarrow [0, \infty]$  das Zählmaß auf  $X$ .  $\#$  ist ein endliches Maß genau wenn  $X$  eine endliche Menge ist.

c) Auch die Funktion  $\mu : \mathcal{P}(X) \rightarrow [0, \infty]$  mit  $\mu(A) := 0$  falls  $A$  abzählbar ist, und  $\mu(A) := \infty$  sonst, hat die Eigenschaften eines Maßes (siehe PS).

d) Sei  $x \in X$ . Dann wird durch

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

ein Maß  $\delta_x : \mathcal{P}(X) \rightarrow [0, \infty]$  definiert, das Punktmaß (oder Dirac-Maß) in  $x$ .

e) Natürlich ist das Lebesguesche Maß  $\lambda : \mathcal{L}_{\mathbb{R}} \rightarrow [0, \infty]$  nun wirklich ein Maß, genauso seine  $d$ -dimensionale Version  $\lambda^d : \mathcal{L}_{\mathbb{R}^d} \rightarrow [0, \infty]$ . Offenbar sind die  $(\mathbb{R}^d, \mathcal{L}_{\mathbb{R}^d}, \lambda^d)$ ,  $d \geq 1$ , jeweils unendliche Maßräume weil  $\lambda^d(\mathbb{R}^d) = \infty$ .

Man beachte, dass es hier nicht möglich ist, als Definitionsbereich  $\mathcal{A}$  die gesamte Potenzmenge zu verwenden.

Es gibt ein Paar sehr einfache Möglichkeiten, aus vorgegebenen Maßen neue Maße zu gewinnen:

**Bemerkung 2.2 (Einfache Operationen mit Maßen)** Sei  $\mathcal{A} \subseteq \mathcal{P}(X)$  eine  $\sigma$ -Algebra.

a) Ist  $\mu$  ein Maß auf  $\mathcal{A}$ , dann auch  $c \cdot \mu$  für jedes  $c \in [0, \infty]$ .

b) Sind  $\mu_k$ ,  $k \geq 1$ , Maße auf  $\mathcal{A}$ , dann definiert die Summe  $\mu(A) := \sum_{k \geq 1} \mu_k(A)$  ein Maß auf  $\mathcal{A}$ , kurz:  $\mu := \sum_{k \geq 1} \mu_k$ .

(Denn: Ist  $A = \biguplus_{l \geq 1} A_l$  mit  $A_l \in \mathcal{A}$ , dann gilt, weil jedes  $\mu_k$  ein Maß ist,  $\mu(A) = \sum_{k \geq 1} \mu_k(A) = \sum_{k \geq 1} \sum_{l \geq 1} \mu_k(A_l)$ . Andererseits haben wir per Definition von  $\mu$  auch  $\sum_{l \geq 1} \mu(A_l) = \sum_{l \geq 1} \sum_{k \geq 1} \mu_k(A_l)$ . Nach Satz 1.1 kommt es hier nicht auf die Reihenfolge der Summation an, also gilt wirklich  $\mu(A) = \sum_{l \geq 1} \mu(A_l)$ .)

c) Mittels a) und b) erhalten wir insbesondere die diskreten Maßräume: Es sei  $(X, \mathcal{A})$  ein messbarer Raum,  $(x_j)_{j \geq 1}$  eine Folge in  $X$ , und  $(m_j)_{j \geq 1}$  eine Folge in  $[0, \infty]$ . Dann ist  $\mu := \sum_{j \geq 1} m_j \cdot \delta_{x_j}$  ein diskretes Maß auf  $\mathcal{A}$ . Man interpretiert dieses oft so, dass der Punkt  $x_j$  mit Masse  $m_j$  versehen wird. Es ist dann  $\mu(A)$  die Summe aller Massen, die eben in  $A$  liegen.

d) Ist  $\mathcal{B} \subseteq \mathcal{A}$  eine kleinere  $\sigma$ -Algebra, und  $\mu$  ein Maß auf  $\mathcal{A}$ , dann ist natürlich die Einschränkung  $\mu|_{\mathcal{B}}$  ein Maß auf  $\mathcal{B}$ .

e) Ist  $\mu$  ein Maß auf  $\mathcal{A}$ , und  $Y \in \mathcal{A}$ , dann ist  $\mu|_{Y \cap \mathcal{A}}$  ein Maß auf der Spur- $\sigma$ -Algebra  $Y \cap \mathcal{A}$  (siehe Bemerkung 1.8).

## 2.2 Äußere Maße und Fortsetzungssatz

Die meisten der oben genannten Beispiele von Maßen sind ganz elementar zu bekommen. Um das Lebesguesche Maß  $\lambda$  zu erhalten mussten wir uns aber ziemlich anstrengen. Der große Schritt bestand darin, das L-B-Prämaß vom Halbring der Intervalle auf eine diesen umfassende  $\sigma$ -Algebra fortzusetzen. Dadurch haben wir die zentrale Existenzaussage des letzten Kapitels erhalten.

**Äußere Maße.** In diesem Abschnitt wiederholen wir nochmals die Vorgehensweise, und formulieren sie so, dass wir dieselben Ideen auch in anderen Situationen nützen können. Erster wesentlicher Schritt über das Prämaß hinaus war die Konstruktion des Lebesgueschen äußeren Maßes  $\lambda^*$ . Als dessen wichtigste Eigenschaften hatten wir  $(\lambda^*1)$ - $(\lambda^*3)$  hervorgehoben. Tatsächlich nennt man jede auf der gesamten Potenzmenge von  $X$  definierte Mengenfunktion mit analogem Verhalten ein äußeres Maß:

**Definition 2.4** *Ein äußeres Maß auf  $X$  ist eine für alle Teilmengen von  $X$  definierte Mengenfunktion  $\mu^* : \mathcal{P}(X) \rightarrow [0, \infty]$  mit folgenden Eigenschaften:*

$$(\mu^*1) \quad \mu^*(\emptyset) = 0,$$

$$(\mu^*2) \quad B \subseteq A \implies \mu^*(B) \leq \mu^*(A), \text{ und}$$

$$(\mu^*3) \quad \text{ist } (A_n)_{n \geq 1} \text{ eine Folge von Mengen aus } \mathcal{P}(X), \text{ dann gilt}$$

$$\mu^* \left( \bigcup_{n \geq 1} A_n \right) \leq \sum_{n \geq 1} \mu^*(A_n).$$

Ein äußeres Maß  $\mu^*$  ist also monoton (oder isotone) und  $\sigma$ -subadditiv.

Äußere Maße kann man, wie im Falle von  $\lambda^*$ , insbesondere mittels äußerer Approximation aus anderen Mengenfunktionen  $\mu : \mathcal{E} \rightarrow [0, \infty]$  mit kleinerem Definitionsbereich  $\mathcal{E} \subseteq \mathcal{P}(X)$  gewinnen, wobei wir immer  $\emptyset \in \mathcal{E}$  voraussetzen. Dazu verwendet man *abzählbare  $\mathcal{E}$ -Überdeckungen* von  $A \subseteq X$ , dh Folgen  $(E_k)_{k \geq 1}$  von Mengen aus  $\mathcal{E}$  deren Vereinigung  $A$  enthält:  $A \subseteq \bigcup_{k \geq 1} E_k$  (wobei ruhig einige  $E_k$  leer sein dürfen), und sieht die Summe der Maße einer approximierenden Überdeckung als Approximation der Größe von  $A$  an.

**Proposition 2.3 (Das von  $\mu$  erzeugte äußere Maß)** *(a) Sei  $\mathcal{E} \subseteq \mathcal{P}(X)$  mit  $\emptyset \in \mathcal{E}$ , und  $\mu : \mathcal{E} \rightarrow [0, \infty]$  eine Mengenfunktion mit  $\mu(\emptyset) = 0$ . Dann ist die durch*

$$\mu^*(A) := \inf \left\{ \sum_{k \geq 1} \mu(E_k) : \begin{array}{l} (E_k)_{k \geq 1} \text{ ist eine abzählbare} \\ \mathcal{E}\text{-Überdeckung von } A \end{array} \right\} \quad (2.1)$$

*definierte Mengenfunktion  $\mu^* : \mathcal{P}(X) \rightarrow [0, \infty]$  ein äußeres Maß. (Wobei wir der Konvention  $\inf \emptyset := \infty$  folgen.)*

*(b) Nehmen wir zusätzlich an, dass  $\mathcal{E} = \mathcal{H}$  ein Halbring und  $\mu$  ein Prämaß auf  $\mathcal{H}$  ist, so gilt außerdem*

$$\mu^*|_{\mathcal{H}} = \mu$$

*dh  $\mu^*$  ist dann eine Fortsetzung der Mengenfunktion  $\mu : \mathcal{H} \rightarrow [0, \infty]$  auf die ganze Potenzmenge  $\mathcal{P}(X)$ .*

In dieser Situation nennt man  $\mu^*$  das *von  $\mu : \mathcal{E} \rightarrow [0, \infty]$  erzeugte äußere Maß*. (Der Definitionsbereich  $\mathcal{E}$  von  $\mu$  kann dabei wichtig sein.) Auch das haben wir im konkreten Kontext der Längenfunktion schon gesehen. Weil sich das Argument (von der Notation abgesehen) nicht ändert, kommt es zum Kleingedruckten.

**Beweis.** (a) Für  $A = \emptyset$  wählen wir die Überdeckung durch die einzelne Menge  $E_1 := \emptyset \in \mathcal{E}$  um zu sehen dass  $\mu^*(\emptyset) \leq \mu(\emptyset) = 0$ . Monotonie von  $\mu^*$  ist auch fast offensichtlich: Falls  $B \subseteq A \subseteq X$ , so ist jede  $\mathcal{E}$ -Überdeckung von  $A$  auch eine  $\mathcal{E}$ -Überdeckung von  $B$  und somit  $\mu^*(B) \leq \mu^*(A)$ .

Für den Beweis der  $\sigma$ -Subadditivität sei nun  $(A_n)_{n \geq 1}$  eine beliebige Folge von Mengen aus  $\mathcal{P}(X)$ . Falls  $\mu^*(A_n) = \infty$  für ein  $n$ , so ist die Behauptung trivial. Wir können daher annehmen dass  $\mu^*(A_n) < \infty$  für alle  $n \geq 1$ . Sei  $\varepsilon > 0$ . Für jedes  $n \geq 1$  gibt es nach Definition von  $\mu^*$  eine abzählbare  $\mathcal{E}$ -Überdeckung  $(E_{n,k})_{k \geq 1}$  von  $A_n$  mit

$$\sum_{k \geq 1} \mu(E_{n,k}) \leq \mu^*(A_n) + \varepsilon \cdot 2^{-n}.$$

Gemeinsam bilden diese eine  $\mathcal{E}$ -Überdeckung  $(E_{n,k})_{n \geq 1, k \geq 1}$  von  $\bigcup_{n \geq 1} A_n$ . Daher gilt

$$\mu^* \left( \bigcup_{n \geq 1} A_n \right) \leq \sum_{n \geq 1} \sum_{k \geq 1} \mu(E_{n,k}) \leq \sum_{n \geq 1} \mu^*(A_n) + \varepsilon,$$

und weil  $\varepsilon > 0$  beliebig war, folgt die  $\sigma$ -Subadditivität von  $\mu^*$ .

(b) Sei  $A \in \mathcal{H}$  fest. Da  $(A, \emptyset, \emptyset, \dots)$  eine  $\mathcal{H}$ -Überdeckung von  $A$  ist, gilt jedenfalls  $\mu^*(A) \leq \mu(A)$ . Zum Nachweis der entgegengesetzten Ungleichung betrachten wir eine abzählbare  $\mathcal{H}$ -Überdeckung  $(E_k)_{k \geq 1}$  von  $A$ , und setzen  $E_k^* := E_k \cap A$ ,  $k \geq 1$ . Mit  $E_k$  und  $A$  liegen diese Mengen wieder in  $\mathcal{H}$ , und wir haben  $A = \bigcup_{k \geq 1} E_k^* \subseteq \bigcup_{k \geq 1} E_k$ . Nachdem  $\mu$  auf  $\mathcal{H}$  ein Prämaß ist, folgt wegen  $\sigma$ -Subadditivität und Monotonie auf  $\mathcal{H}$ , siehe Proposition 2.2, Aussagen b) und d),

$$\mu(A) \leq \sum_{k \geq 1} \mu(E_k^*) \leq \sum_{k \geq 1} \mu(E_k),$$

und weil die  $\mathcal{H}$ -Überdeckung  $(E_k)_{k \geq 1}$  von  $A$  beliebig war, sehen wir, dass

$$\mu(A) \leq \inf \left\{ \sum_{k \geq 1} \mu(E_k) : (E_k)_{k \geq 1} \dots \right\} = \mu^*(A),$$

wie gewünscht. ■

Viele wichtige äußere Maße entstehen auf diesem Weg durch Überdeckungen. Es gibt aber auch andere relevante Beispiele.

**$\mu^*$ -messbare Mengen.** Der nächste Schritt zum Lebesgueschen Maß war dann die Definition  $\lambda^*$ -messbarer Mengen. Die oben angegebene Fassung erweist sich sogar für beliebige äußere Maße als passend. Wir vereinbaren daher:

**Definition 2.5** *Es sei  $\mu^* : \mathcal{P}(X) \rightarrow [0, \infty]$  ein äußeres Maß. Eine Menge  $A \subseteq X$  heißt  $\mu^*$ -messbar, falls*

$$\mu^*(Q) = \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \quad \text{für alle } Q \subseteq X, \quad (2.2)$$

*dh wenn  $A$  alle Mengen bezüglich  $\mu^*$  additiv zerschneidet. Wir schreiben*

$$\mathcal{A}(\mu^*) := \{A \subseteq X : A \text{ ist } \mu^*\text{-messbar}\}. \quad (2.3)$$

Analog zu Bemerkung 1.6 finden wir dann:

**Bemerkung 2.3 a)** *Da  $\mu^*$  ohnehin  $(\mu^*3)$  erfüllt, ist  $A$  eine  $\mu^*$ -messbare Menge sobald*

$$\mu^*(Q) \geq \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \quad \text{für alle } Q \subseteq X.$$

*Dabei braucht man außerdem nur Mengen  $Q$  mit  $\mu^*(Q) < \infty$  zu betrachten, da diese Ungleichung sonst automatisch erfüllt ist.*

**b)** *Ist  $\mu^*(A) = 0$ , dann ist  $A$   $\mu^*$ -messbar, da wegen  $(\mu^*2)$  dann  $\mu^*(Q \cap A) = 0$  gilt, und obige Ungleichung zu  $\mu^*(Q) \geq \mu^*(Q \cap A^c)$  wird, was wegen  $(\mu^*2)$  immer stimmt. Entsprechend, wenn  $\mu^*(A^c) = 0$ .*

Und parallel zu Proposition 1.7 sehen wir, dass wenigstens jene Mengen, die wir als Ausgangspunkt gewählt haben, dieser Definition genügen.

**Proposition 2.4 (Messbarkeit des Halbringes)** *Ist  $\mathcal{H} \subseteq \mathcal{P}(X)$  ein Halbring,  $\mu : \mathcal{H} \rightarrow [0, \infty]$  ein Prämaß, und  $\mu^*$  das davon erzeugte äußere Maß, dann ist  $\mathcal{H} \subseteq \mathcal{A}(\mu^*)$ .*

**Beweis.** Sei  $A \in \mathcal{H}$ . Wir zeigen, dass  $A$  dann  $\mu^*$ -messbar ist. Dazu betrachten wir irgendein  $Q \subseteq X$  mit  $\mu^*(Q) < \infty$ . Das äußere Maß  $\mu^*(Q)$  ist definiert mittels abzählbarer Überdeckungen von  $Q$  durch Mengen aus  $\mathcal{H}$ . Sei  $(E_n)_{n \geq 1}$  eine solche (wegen  $\mu^*(Q) < \infty$  gibt es zumindest eine). Da  $\mathcal{H}$  ein Halbring ist, sind für jedes  $n \geq 1$  die Mengen  $B_n := E_n \cap A \in \mathcal{H}$  und  $E_n \cap A^c = E_n \setminus A = \bigsqcup_{k=1}^{m_n} C_{n,k}$  für geeignete disjunkte  $C_{n,1}, \dots, C_{n,m_n} \in \mathcal{H}$ . Nachdem  $\mu^*|_{\mathcal{H}} = \mu$  und  $\mu$  ein Prämaß ist, folgt

$$\mu^*(E_n) = \mu^*(B_n) + \sum_{k=1}^{m_n} \mu^*(C_{n,k}) \quad \text{für } n \geq 1.$$

Nun bilden  $(B_n)_{n \geq 1}$  und  $(C_{n,k})_{n \geq 1, 1 \leq k \leq m_n}$  aber abzählbare  $\mathcal{H}$ -Überdeckungen von  $Q \cap A$  bzw von  $Q \cap A^c$ , daher gilt

$$\begin{aligned} \sum_{n \geq 1} \mu^*(E_n) &= \sum_{n \geq 1} \mu^*(B_n) + \sum_{n \geq 1} \sum_{k=1}^{m_n} \mu^*(C_{n,k}) \\ &\geq \mu^*(Q \cap A) + \mu^*(Q \cap A^c). \end{aligned}$$



Da die  $\mathcal{H}$ -Überdeckung  $(E_n)_{n \geq 1}$  von  $Q$  aber beliebig war, sehen wir, dass

$$\mu^*(Q) = \inf \left\{ \sum_{n \geq 1} \mu^*(E_n) : (E_n)_{n \geq 1} \dots \right\} \geq \mu^*(Q \cap A) + \mu^*(Q \cap A^c),$$

also ist wirklich  $A \in \mathcal{A}(\mu^*)$ . ■

Der entscheidende Schritt um letztlich Satz 1.6 und damit das Lebesgue-Maß zu erhalten bestand darin, die Familie  $\mathcal{L}_{\mathbb{R}}$  Lebesgue-messbarer Mengen als  $\sigma$ -Algebra zu erkennen, und zu zeigen, dass  $\lambda^*$  auf dieser immer noch  $\sigma$ -additiv ist. Wer mutig genug ist, das Argument aus dem Beweis dieses Satzes abstrakt anzusehen, wird reich belohnt: auch dort kommt es nicht darauf an, dass wir über Teilmengen von  $\mathbb{R}$  und Längensummen von Überdeckungen durch Intervalle sprechen. Entscheidend ist nur, dass die betrachtete Mengenfunktion die Eigenschaften eines äußeren Maßes hat. Wir bekommen also mit demselben Beweis den bemerkenswerten

**Satz 2.1 (Satz von Carathéodory)** *Ist  $\mu^* : \mathcal{P}(X) \rightarrow [0, \infty]$  ein äußeres Maß, dann ist die Familie  $\mathcal{A}(\mu^*)$  eine  $\sigma$ -Algebra, und die Einschränkung  $\mu^*|_{\mathcal{A}(\mu^*)}$  ist  $\sigma$ -additiv, dh ein Maß.*

**Beweis.** (i) Wir bemerken zuerst, dass  $\mathcal{A}(\mu^*)$  eine Algebra ist: Offenkundig ist  $X \in \mathcal{A}(\mu^*)$ , und da die Bedingung der  $\mu^*$ -Messbarkeit symmetrisch in  $A$  und  $A^c$  ist, führt die Komplementbildung nicht aus  $\mathcal{A}(\mu^*)$  heraus. Schließlich ist für  $A, B \in \mathcal{A}(\mu^*)$  auch  $A \cup B \in \mathcal{A}(\mu^*)$ , denn für alle  $Q \subseteq X$  gilt

$$\begin{aligned} \mu^*(Q) &= \mu^*(Q \cap A) + \mu^*(Q \cap A^c) & (A \in \mathcal{A}(\mu^*)) \\ &= \mu^*(Q \cap A) + \mu^*(Q \cap A^c \cap B) + \mu^*(Q \cap A^c \cap B^c) & (B \in \mathcal{A}(\mu^*)) \\ &\geq \mu^*((Q \cap A) \cup (Q \cap A^c \cap B)) + \mu^*(Q \cap A^c \cap B^c) & ((\mu^*3)) \\ &= \mu^*(Q \cap (A \cup B)) + \mu^*(Q \cap (A \cup B)^c). \end{aligned}$$

(ii) Nun zeigen wir, dass für jede *disjunkte* Folge  $(A_n)_{n \geq 1}$  in  $\mathcal{A}(\mu^*)$

$$A := \biguplus_{n \geq 1} A_n \in \mathcal{A}(\mu^*) \quad \text{und} \quad \mu^*(A) = \sum_{n \geq 1} \mu^*(A_n). \quad (2.4)$$

Dazu stellen wir fest, dass für *disjunkte*  $E, F \in \mathcal{A}(\mu^*)$  und jedes  $Q \subseteq X$ ,

$$\mu^*(Q \cap (E \uplus F)) = \mu^*(Q \cap E) + \mu^*(Q \cap F)$$

(man schneide  $Q' = Q \cap (E \uplus F)$  mit  $E$  und  $E^c$ ). Per Induktion folgt

$$\mu^*\left(Q \cap \biguplus_{n=1}^N A_n\right) = \sum_{n=1}^N \mu^*(Q \cap A_n) \quad \text{für } N \geq 1 \text{ und } Q \subseteq X. \quad (2.5)$$

Nach Schritt (i) ist  $\biguplus_{n=1}^N A_n \in \mathcal{A}(\mu^*)$ , und daher für jedes  $Q \subseteq X$  und  $N \geq 1$ ,

$$\begin{aligned} \mu^*(Q) &= \mu^*\left(Q \cap \biguplus_{n=1}^N A_n\right) + \mu^*\left(Q \cap \left(\biguplus_{n=1}^N A_n\right)^c\right) \\ &\geq \sum_{n=1}^N \mu^*(Q \cap A_n) + \mu^*(Q \cap A^c). \end{aligned}$$

(Der zweite Schritt verwendet (2.5) und  $(\mu^*2)$ .) Es folgt, für  $N \rightarrow \infty$ ,

$$\begin{aligned} \mu^*(Q) &\geq \sum_{n \geq 1} \mu^*(Q \cap A_n) + \mu^*(Q \cap A^c) \\ &\geq \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \geq \mu^*(Q), \end{aligned} \quad (\text{zweimal } (\mu^*3))$$

also für alle  $Q \subseteq X$ ,

$$\mu^*(Q) = \sum_{n \geq 1} \mu^*(Q \cap A_n) + \mu^*(Q \cap A^c) = \mu^*(Q \cap A) + \mu^*(Q \cap A^c),$$

dh  $A \in \mathcal{A}(\mu^*)$ , und es ist  $\mu^*(A) = \sum_{n \geq 1} \mu^*(A_n)$  (wähle  $Q := A$ ).

(iii) Nach (i) und (ii) ist  $\mathcal{A}(\mu^*)$  eine Algebra, die unter *disjunkten* abzählbaren Vereinigungen abgeschlossen ist. Daher ist  $\mathcal{A}(\mu^*)$  sogar eine  $\sigma$ -Algebra (vgl. Bemerkung 1.7). Außerdem ist nach (ii) das äußere Maß  $\mu^*$ , welches sowieso  $(\mu^*1)$  erfüllt,  $\sigma$ -additiv auf  $\mathcal{A}(\mu^*)$ , also ein Maß auf dieser  $\sigma$ -Algebra. ■

**Der allgemeine Fortsetzungssatz.** Wir müssen noch klären, wie groß die  $\sigma$ -Algebra  $\mathcal{A}(\mu^*)$  der  $\mu^*$ -messbaren Mengen allgemein ist. Für beliebige äußere Maße kann sie durchaus recht mickrig ausfallen. Gehen wir von einem Prämaß auf einem Halbring aus, so ist  $\mathcal{A}(\mu^*)$  aber reichhaltig genug. Indem wir uns davon überzeugen (wiederum mit einem schon bekannten Argument, vgl. Proposition 1.7), bekommen wir den fundamentalen

**Satz 2.2 (Fortsetzungssatz)** Ist  $\mathcal{H} \subseteq \mathcal{P}(X)$  ein Halbring,  $\mu : \mathcal{H} \rightarrow [0, \infty]$  ein Prämaß, und  $\mu^*$  das davon erzeugte äußere Maß, dann ist  $\mathcal{H} \subseteq \mathcal{A}(\mu^*)$  und  $\mu^*|_{\mathcal{A}(\mu^*)}$  ist eine Fortsetzung von  $\mu$  zu einem Maß auf  $\mathcal{A}(\mu^*)$ .

**Beweis.** Nach Proposition 2.3 b) ist  $\mu^*|_{\mathcal{H}} = \mu$ ,  $\mu^*$  also eine Fortsetzung von  $\mu$ . Gemäß Proposition 2.4 gilt  $\mathcal{H} \subseteq \mathcal{A}(\mu^*)$ , und nach Satz 2.1 ist  $\mu^*|_{\mathcal{A}(\mu^*)}$  ein Maß. ■

Insbesondere kann man also jedes Prämaß zu einem Maß fortsetzen. Dadurch wird die Bezeichnung *Prämaß* nachträglich gerechtfertigt.

**Eindeutigkeit der Fortsetzung.** Der Fortsetzungssatz garantiert die Existenz vieler Maße. Sind die solcherart aus "natürlichen" Prämaßen gewonnenen Maße aber wieder "natürliche" Objekte, oder können wir dabei willkürlich zwischen verschiedenen Fortsetzungen wählen? Erst wenn sich die Fortsetzung als *eindeutig* herausstellt können wir sie gefahrlos wieder mit  $\mu$  bezeichnen und von der Fortsetzung des Prämaßes nach  $\mathcal{A}(\mu^*)$  sprechen. Beim Lebesgue-Maß ist dies der Fall (Proposition 1.11). Allgemein ist das aber nicht richtig, was man anhand eines ärgerlich einfachen Beispiels sieht:

**Beispiel 2.3** Es sei  $X$  beliebig,  $\mathcal{A} := \{\emptyset, X\}$  und  $\mathcal{H} := \{\emptyset\}$ , ein (langweiliger) Halbring. Das Prämaß  $\mu : \mathcal{H} \rightarrow [0, \infty]$  mit  $\mu(\emptyset) := 0$  hat überabzählbar viele verschiedene Fortsetzungen nach  $\mathcal{A}(\mu^*) = \mathcal{A}$ : Für jedes  $r \in [0, \infty]$  ist das Maß  $\mu_r : \mathcal{A} \rightarrow [0, \infty]$  mit  $\mu_r(X) := r$  eine solche.

In diesem Beispiel ist die Fortsetzung  $\mu$  eindeutig bestimmt sobald man  $\mu(X)$  kennt. Um letzteres festzulegen, sollte  $X \in \mathcal{H}$  gelten. Dann reicht die Kenntnis von  $\mu(X)$  sogar allgemein aus, sofern dieser Wert nur endlich ist. Diese Aussage ist zwar auch oft nützlich, wir geben uns aber nicht damit zufrieden, weil ja nicht einmal das Lebesguesche Maß auf  $(\mathbb{R}, \mathcal{L}_{\mathbb{R}})$  endlich ist. Am Ende des Beweises von Proposition 1.11 haben wir uns zu Nutze gemacht, dass wir  $\mathbb{R}$  als abzählbare disjunkte Vereinigung der Mengen  $(n, n+1]$ ,  $n \in \mathbb{Z}$ , schreiben können, wobei diese jeweils endliche Länge besitzen. Diese Möglichkeit, einen Raum in abzählbar viele Teile endlichen Maßes zu zerlegen ist oft sehr vorteilhaft. Man gelangt daher zu

**Definition 2.6** Sei  $\mathcal{E} \subseteq \mathcal{P}(X)$ . Eine Mengenfunktion  $\mu : \mathcal{E} \rightarrow [0, \infty]$  ist  $\sigma$ -endlich, falls es eine Folge  $(E_n)_{n \geq 1}$  in  $\mathcal{E}$  gibt mit  $\mu(E_n) < \infty$  für jedes  $n \geq 1$ , und  $X = \bigcup_{n \geq 1} E_n$ . (Das ist offensichtlich erfüllt falls  $X \in \mathcal{E}$  und  $\mu(X) < \infty$ .)

**Beispiel 2.4 a)** Das Prämaß  $\lambda : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  ist  $\sigma$ -endlich, wir können zB die abzählbar vielen Mengen  $E_m := (m, m+1]$ ,  $m \in \mathbb{Z}$  nehmen.

**b)** Ebenso für  $\lambda^d : \mathcal{I}_{\mathbb{R}^d} \rightarrow [0, \infty)$ . Zur Abwechslung verwenden wir diesmal  $E_n := (-n, n]^d$ ,  $n \geq 1$ .

**c)** Die Einschränkung des Lebesgue-Maßes  $\lambda$  auf  $\{(-\infty, b] : b \in \mathbb{R}\}$  ist nicht  $\sigma$ -endlich, weil jede dieser Mengen unendliches Maß hat.

**d)** Das Zählmaß  $\# : \mathcal{P}(X) \rightarrow [0, \infty]$  ist  $\sigma$ -endlich genau wenn  $X$  abzählbar ist.

**Bemerkung 2.4** Im Falle eines Halbringes  $\mathcal{E} = \mathcal{H}$  bedeutet (wie schon früher gesehen, Bemerkung 2.1) die Existenz einer Folge  $(E_n)$  wie in obiger Definition, dass es auch eine disjunkte Folge  $(F_n)_{n \geq 1}$  in  $\mathcal{H}$  gibt mit  $X = \bigsqcup_{n \geq 1} F_n$  und  $\mu(F_n) < \infty$ .

Damit können wir den Fortsetzungssatz ergänzen durch

**Satz 2.3 (Eindeutigkeit der Fortsetzung)** Sei  $\mathcal{H} \subseteq \mathcal{P}(X)$  ein Halbring,  $\mu : \mathcal{H} \rightarrow [0, \infty]$  ein  $\sigma$ -endliches Prämaß. Dann gibt es nur eine Fortsetzung von  $\mu$  zu einem Maß auf  $\mathcal{A}(\mu^*)$ .

Unter dieser Annahme funktioniert der Beweis von Proposition 1.11 a) wieder. (Wir müssen nur ganz am Ende einen schon bekannten Kunstgriff einfügen.)

**Beweis.** Sei  $\nu : \mathcal{A}(\mu^*) \rightarrow [0, \infty]$  irgendeine  $\sigma$ -additive Mengenfunktion mit  $\nu|_{\mathcal{H}} = \mu^*|_{\mathcal{H}}$ . Für  $A \in \mathcal{A}(\mu^*)$  und jede  $\mathcal{H}$ -Überdeckung  $(E_k)_{k \geq 1}$  von  $A$  gilt (weil  $\nu$   $\sigma$ -additiv und damit monoton ist und  $\mu$  fortsetzt)

$$\nu(A) \leq \nu\left(\bigcup_{k \geq 1} E_k\right) \leq \sum_{k \geq 1} \nu(E_k) = \sum_{k \geq 1} \mu(E_k),$$

und somit  $\nu \leq \mu^*$  auf  $\mathcal{A}(\mu^*)$ . Ist nun  $Q \in \mathcal{H}$  mit  $\mu(Q) < \infty$ , dann folgt aus dieser Ungleichung

$$\mu^*(Q) = \nu(Q) = \nu(Q \cap A) + \nu(Q \cap A^c) \leq \mu^*(Q \cap A) + \mu^*(Q \cap A^c), \quad (2.6)$$

für  $A \in \mathcal{A}(\mu^*)$ , und nach Definition dieser Familie stimmen die Ausdrücke ganz links und ganz rechts überein, es muss also durchweg Gleichheit herrschen. Weil hier sowohl  $\nu(Q \cap A) \leq \mu^*(Q \cap A)$  als auch  $\nu(Q \cap A^c) \leq \mu^*(Q \cap A^c)$  gilt, muss insbesondere ist  $\nu(Q \cap A) = \mu^*(Q \cap A)$  sein. Letzteres gilt also für jedes  $Q \in \mathcal{H}$  mit  $\mu(Q) < \infty$ .

Sei  $(F_n)_{n \geq 1}$  wie in der vorigen Bemerkung. Indem wir oben  $Q = F_n$  wählen, finden wir

$$\nu(A) = \sum_{n \geq 1} \nu(F_n \cap A) = \sum_{n \geq 1} \mu^*(F_n \cap A) = \mu^*(A),$$

wie behauptet. ■

Zum Ende dieses Abschnittes tragen wir noch eine bisher übergangene Grundeigenschaft von (Prä)maßen nach.

**Bemerkung 2.5 (Stetigkeit von oben)** *Aus gutem Grund findet sich in Proposition 2.2 zwar die Stetigkeit von unten, Eigenschaft e), aber nicht die analoge Aussage, dass aus  $A_n \in \mathcal{H}$ ,  $n \geq 1$ , und  $A_n \searrow A \in \mathcal{H}$  immer  $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)$  folgt. Diese ist nämlich nicht einmal für Maße korrekt. Zum Beispiel können wir  $(\mathbb{R}, \mathcal{L}_{\mathbb{R}}, \lambda)$  betrachten, mit  $A_n := (n, \infty)$ ,  $n \geq 1$ . Dann gilt  $A_n \searrow \emptyset$  aber  $\lambda(A_n) = \infty$  während  $\lambda(\emptyset) = 0$ .*

*Richtig, und zwar für alle Prämaße, ist allerdings folgende Version:*

$$A_n \in \mathcal{H}, n \geq 1, \text{ mit } \mu(A_1) < \infty \text{ und } A_n \searrow A \in \mathcal{H} \implies \lim_{n \rightarrow \infty} \mu(A_n) = \mu(A).$$

*Denn: in der genannten Situation gilt  $A_1 \setminus A_n \nearrow A_1 \setminus A$  und die bereits bekannte Stetigkeit des Prämaßes von unten garantiert, dass  $\lim_{n \rightarrow \infty} \mu(A_1 \setminus A_n) = \mu(A_1 \setminus A)$ . Wegen  $\mu(A_1) < \infty$  können wir jetzt  $\mu(A_n) = \mu(A_1) - \mu(A_1 \setminus A_n)$  und  $\mu(A) = \mu(A_1) - \mu(A_1 \setminus A)$  verwenden, um  $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)$  zu bekommen.*

*Nun ja, wenn Sie aufgepasst haben, dann stört Sie hier sicher, dass ich einfach mit  $\mu(A_1 \setminus A_n)$  rechne, diese Mengen liegen ja nicht unbedingt im Halbring  $\mathcal{H}$ . Allerdings: Aufgrund des Fortsetzungssatzes dürfen wir jetzt einfach anstelle von  $\mu : \mathcal{H} \rightarrow [0, \infty]$  gleich die Fortsetzung  $\mu : \mathcal{A}(\mu^*) \rightarrow [0, \infty]$  verwenden, und in  $\mathcal{A}(\mu^*)$  sind die  $A_1 \setminus A_n$  natürlich mit dabei.*

## 2.3 Anwendung: Lebesgue-Stieltjes Maße auf $\mathbb{R}$

Obige Rückschau auf das erste Kapitel “durch die abstrakte Brille” ermöglicht es uns nun ohne wesentlichen Mehraufwand einen Überblick über alle Maße auf  $\mathbb{R}$  zu geben, welche Intervalle messen können, und *beschränkten Mengen endliche Größe verleihen*. Falls nämlich  $\mu : \mathcal{A} \rightarrow [0, \infty]$  von dieser Art ist, dann muss die Einschränkung auf  $\mathcal{I}_{\mathbb{R}} \subseteq \mathcal{A}$  ein Prämaß sein, das nur endliche Werte annimmt, und daher folgender Definition genügt.

**Definition 2.7** *Ein Lebesgue-Stieltjessches Prämaß (L-S-Prämaß) ist einfach ein Prämaß  $\mu : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$ . (Die einzige Einschränkung besteht darin, dass der Wert  $\infty$  nicht angenommen wird.)*

Wir bemerken zuerst, dass man jedes L-S-Prämaß  $\mu : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  durch eine simple reelle Funktion eindeutig beschreiben kann. Das ist hilfreich, weil uns reelle Funktionen doch besser vertraut sind als Maße. Wir definieren  $F_\mu : \mathbb{R} \rightarrow \mathbb{R}$  durch

$$F_\mu(t) := \begin{cases} \mu((0, t]) & \text{falls } t \geq 0, \\ -\mu((t, 0]) & \text{falls } t < 0. \end{cases}$$

Dann gilt für  $F = F_\mu$  (bitte nachprüfen!)

$$\mu((a, b]) = F(b) - F(a) \quad \text{für } a \leq b \text{ in } \mathbb{R}. \quad (2.7)$$

Mittels Proposition 2.2 b) sehen wir sofort, dass  $F$  *monoton wachsend* ist. Umgekehrt führt jede wachsende reelle Funktion  $F$  mittels (2.7) zu einer nicht-negativen Mengenfunktion auf  $\mathcal{I}_{\mathbb{R}}$ . Bekommen wir damit vielleicht automatisch lauter Prämaße? Nun, ganz so einfach ist das nicht. So wie die Monotonieeigenschaft für Prämaße aus Proposition 2.2 b) die Monotonie von  $F$  vorgibt, erzwingt die Stetigkeitseigenschaft aus Bemerkung 2.5 eine entsprechende Stetigkeitseigenschaft von  $F$ . Für Folgen in  $\mathbb{R}$  mit  $t_n \searrow t$ ,  $t > 0$ , gilt ja  $(0, t_n] \searrow (0, t]$ , und wegen  $\mu((0, t_1]) < \infty$  folgt dann  $F(t_n) \searrow F(t)$  (und entsprechend wenn  $t < 0$ ). Somit ist  $F$  *rechtsseitig stetig*. Das ist nun aber alles, was man für ein Prämaß braucht:

**Satz 2.4 (Lebesgue-Stieltjesche Prämaße)** *Ist  $F : \mathbb{R} \rightarrow \mathbb{R}$  wachsend und rechtsseitig stetig, so ist  $\mu_F : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$  mit  $\mu_F((a, b]) := F(b) - F(a)$  ein Prämaß.*

Ist  $F : \mathbb{R} \rightarrow \mathbb{R}$  wachsend und rechtsseitig stetig, so spricht man allgemein von einer *Verteilungsfunktion (VF)*. Hängt diese wie in (2.7) mit  $\mu$  zusammen, so ist  $F$  eine *Verteilungsfunktion von  $\mu$* . Mit  $F$  ist, für jedes  $c \in \mathbb{R}$ , auch  $F + c$  eine VF von  $\mu$ . Ist  $\mu$  *endlich* in dem Sinne, dass die Gesamtmasse  $\sum_{n \in \mathbb{Z}} \mu((n, n+1]) < \infty$ , so verwendet man als Standard-Variante meist  $F := F_\mu + c$  mit  $c := -\lim_{t \rightarrow -\infty} F_\mu(t)$ , sodass dann eben  $\lim_{t \rightarrow -\infty} F(t) = 0$  gilt. Dies bedeutet, dass  $F(t) = \mu((-\infty, t])$  für alle  $t \in \mathbb{R}$  (sobald  $\mu$  auf die einzig sinnvolle Weise mittels  $\mu((-\infty, t]) := \sum_{n \geq 0} \mu((t-n-1, t-n])$  auf solche Intervalle erweitert wird). Gilt zusätzlich  $\lim_{t \rightarrow \infty} F(t) = 1$  (und damit  $\mu(\mathbb{R}) = \lim_{t \rightarrow \infty} (F(t) - F(-t)) = 1$  sobald definiert), so ist die VF  $F$  *normiert*.

In deutlichem Kontrast zur Längenfunktion  $\lambda$ , die der stetigen Verteilungsfunktion  $F(t) := t$  entspricht, beschreiben Lebesgue-Stieltjes Prämaße  $\mu_F$  mit unstetigem  $F$  Situationen in denen Masse in einzelnen Punkten sitzt.

**Beispiel 2.5 (Punktmaß und Sprungfunktion)** *Seien  $x \in \mathbb{R}$  und  $c \in [0, \infty)$  fest, und  $F : \mathbb{R} \rightarrow \mathbb{R}$  die Verteilungsfunktion mit  $F(t) := c \cdot 1_{[x, \infty)}(t)$ . Dann finden wir*

$$\mu_F((a, b]) = c \cdot (1_{[x, \infty)}(b) - 1_{[x, \infty)}(a)) = \begin{cases} c & \text{falls } a < x \leq b \\ 0 & \text{sonst} \end{cases} = c \cdot \delta_x((a, b]),$$

*dh das Prämaß  $\mu_F : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty]$  ist die Einschränkung auf  $\mathcal{I}_{\mathbb{R}}$  des in  $x$  konzentrierten Maßes  $c \cdot \delta_x$ . Allgemein entspricht eine Sprungstelle einer VF  $F$  stets einem Punkt mit positiver Masse.*

**Beispiel 2.6 (VF und Prämaße aus Dichten)** *Sei  $f : \mathbb{R} \rightarrow [0, \infty)$  auf jedem kompakten Intervall  $[a, b]$  R-integrierbar, und  $F : \mathbb{R} \rightarrow \mathbb{R}$  gegeben durch*

$$F(t) := \begin{cases} \int_0^t f(s) ds & \text{falls } t \geq 0 \\ -\int_t^0 f(s) ds & \text{sonst.} \end{cases}$$

*Dann ist  $F$  eine stetige Verteilungsfunktion, und  $\mu_F((a, b]) = \int_a^b f(s) ds$ . Man spricht von der VF bzw dem Prämaß mit Dichte  $f$ . Aus der elementaren Stochastik ist die Idee, verschiedene Bereiche des Raumes mittels einer Dichte unterschiedlich zu gewichten ja hinlänglich bekannt.*

Wir beweisen nun endlich den Satz. Auch dieses Argument kennen Sie bereits, denn wir haben den Ausgangspunkt für die Konstruktion des Lebesgue-Maßes, den Satz 1.4 genauso bewiesen. Anstelle von  $\lambda((a, b]) = b - a$  verwenden wir jetzt aber  $\mu_F((a, b]) = F(b) - F(a)$ . Die ersten beiden Schritte im Beweis erfordern keine weitere Änderung, man muss nur beachten, dass wegen der Monotonie von  $F$  für  $a \leq b$  stets  $F(b) - F(a) \geq 0$  ist. Deshalb sind sie auch kleingedruckt. Der Dritte

Schritt verläuft ebenfalls analog zum Fall  $\mu_F = \lambda$ , da er aber ein wenig subtiler ist und auf der weniger trivialen Voraussetzung der rechtsseitigen Stetigkeit beruht, gestehen wir ihm die normale Schriftgröße zu.

**Beweis.** Zu zeigen ist: Für jede disjunkte Folge  $((a_n, b_n])_{n \geq 1}$  mit  $\bigcup_{n \geq 1} (a_n, b_n] = (a, b] \in \mathcal{I}_{\mathbb{R}}$  gilt

$$\sum_{n \geq 1} (F(b_n) - F(a_n)) = F(b) - F(a).$$

(i) Wir zeigen zuerst: Für jede disjunkte Folge  $((a_n, b_n])_{n \geq 1}$  in  $\mathcal{I}_{\mathbb{R}}$  mit  $\bigcup_{n \geq 1} (a_n, b_n] \subseteq (a, b] \in \mathcal{I}_{\mathbb{R}}$  gilt

$$\sum_{n \geq 1} (F(b_n) - F(a_n)) \leq F(b) - F(a).$$

Betrachte endlich viele dieser Intervalle,  $(a_1, b_1], \dots, (a_N, b_N]$ . Da disjunkt, können wir sie so ordnen, dass

$$a \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq \dots \leq a_N \leq b_N \leq b.$$

Dann ist, wenn wir die allen Paaren aufeinanderfolgender Punkte entsprechenden (nichtnegativen) Differenzen anschreiben,

$$\begin{aligned} \sum_{n=1}^N (F(b_n) - F(a_n)) &\leq \\ &\leq (F(a_1) - F(a)) + (F(b_1) - F(a_1)) + (F(a_2) - F(b_1)) + \dots \\ &\quad \dots + (F(b_N) - F(a_N)) + (F(b) - F(b_N)) \\ &= F(b) - F(a), \end{aligned}$$

wie behauptet. Weil das für jedes  $N \geq 1$  gilt, folgt die entsprechende Abschätzung auch für den Grenzwert der unendlichen Reihe:

$$\sum_{n \geq 1} (F(b_n) - F(a_n)) = \lim_{N \rightarrow \infty} \sum_{n=1}^N (F(b_n) - F(a_n)) \leq F(b) - F(a).$$

(ii) Nun zeigen wir: Für jede endliche Überdeckung  $\bigcup_{n=1}^N (a_n, b_n] \supseteq (a, b]$  gilt

$$F(b) - F(a) \leq \sum_{n=1}^N (F(b_n) - F(a_n)).$$

Für  $N = 1$  ist die Behauptung trivial. Um sie induktiv für beliebiges  $N$  nachzuweisen, nehmen wir an sie gelte für  $N - 1$ . Wir ordnen die Intervalle so, dass  $b \in (a_N, b_N]$ . Falls auch  $a \in (a_N, b_N]$ , dann gilt sogar  $F(b) - F(a) \leq F(b_N) - F(a_N)$ . Andernfalls wird das Teilintervall  $(a, b] \setminus (a_N, b_N] = (a, a_N]$  von den  $N - 1$  verbleibenden Intervallen überdeckt,  $(a, a_N] \subseteq \bigcup_{n=1}^{N-1} (a_n, b_n]$ , und unsere Induktionsvoraussetzung garantiert, dass  $\sum_{n=1}^{N-1} (F(b_n) - F(a_n)) \geq F(a_N) - F(a)$ . Insgesamt folgt also  $F(b) - F(a) = (F(b) - F(a_N)) + (F(a_N) - F(a)) \leq (F(b_N) - F(a_N)) + \sum_{n=1}^{N-1} (F(b_n) - F(a_n))$ , wie gefordert.

(iii) Schließlich der Fall unendlicher Überdeckungen:  $\bigcup_{n \geq 1} (a_n, b_n] \supseteq (a, b]$ . Wegen der rechtsseitigen Stetigkeit von  $F$  können wir, für beliebiges  $\varepsilon > 0$ , Punkte  $\alpha \in (a, b]$  mit  $F(\alpha) \leq F(a) + \varepsilon/2$  und Punkte  $\beta_n > b_n$  mit  $F(\beta_n) \leq F(b_n) + 2^{-(n+1)}\varepsilon$ ,  $n \geq 1$ , wählen. Dann ist

$$[\alpha, b] \subseteq \bigcup_{n \geq 1} (a_n, \beta_n),$$

das kompakte Intervall auf der linken Seite wird also von den offenen Mengen rechts überdeckt. Nach dem Satz von Heine und Borel enthält diese offene Überdeckung eine endliche Teilüberdeckung, insbesondere gibt es ein  $N \geq 1$  sodass

$$(\alpha, b] \subseteq \bigcup_{n=1}^N (a_n, \beta_n].$$

Damit befinden wir uns im vorher betrachteten Fall endlicher Überdeckungen, und wissen daher, dass

$$\begin{aligned} F(b) - F(a) &\leq F(b) - F(\alpha) + \varepsilon \leq \sum_{n=1}^N (F(\beta_n) - F(a_n)) + \frac{\varepsilon}{2} \\ &\leq \sum_{n=1}^N (F(b_n) - F(a_n)) + \frac{\varepsilon}{2} \sum_{n=1}^N 2^{-n} + \frac{\varepsilon}{2} \leq \sum_{n \geq 1} (F(b_n) - F(a_n)) + \varepsilon, \end{aligned}$$

und da  $\varepsilon > 0$  beliebig war, folgt

$$F(b) - F(a) \leq \sum_{n \geq 1} (F(b_n) - F(a_n)),$$

wie gewünscht. ■

Mit den Sätzen über die Fortsetzung eines Prämaßes zu einem Maß liefert das:

**Satz 2.5 (Verteilungsfunktionen und Lebesgue-Stieltjesche Maße)** *Zu jeder Verteilungsfunktion  $F$  gibt es genau ein L-S-Maß auf der  $\sigma$ -Algebra  $\mathcal{A}(\mu_F^*) \supseteq \mathcal{I}_{\mathbb{R}}$ .*

**Beweis.** Nach Satz 2.4 bestimmt  $F$  das L-S-Prämaß  $\mu_F : \mathcal{I}_{\mathbb{R}} \rightarrow [0, \infty)$ . Gemäß Satz 2.2 besitzt dieses eine Fortsetzung zu einem Maß auf  $\mathcal{A}(\mu_F^*) \supseteq \mathcal{I}_{\mathbb{R}}$ . Weil jedes L-S-Prämaß auf  $\mathcal{I}_{\mathbb{R}}$   $\sigma$ -endlich ist, zeigt Satz 2.3 Eindeutigkeit der Fortsetzung. ■

**Definition 2.8** *Wir bezeichnen das Maß aus dem Satz, wie das Prämaß oben, mit  $\mu_F$  und erhalten so das Lebesgue-Stieltjes Maß (L-S-Maß)  $\mu_F : \mathcal{A}(\mu_F^*) \rightarrow [0, \infty]$ .*

Wenden wir den Satz auf die Prämaße aus Beispiel 2.6 an, so bekommen wir

**Beispiel 2.7 (L-S-Maße aus R-integrierbaren Dichten)** *Ist  $f : \mathbb{R} \rightarrow [0, \infty)$  auf jedem kompakten Intervall  $[a, b]$  R-integrierbar, so gibt es genau ein L-S-Maß  $\mu$  mit*

$$\mu((a, b]) = \int_a^b f(s) ds, \quad \text{für } a < b \text{ in } \mathbb{R}.$$

*Man nennt  $\mu$  das Maß mit Dichte  $f$ , später werden wir es mit  $\mu = f \odot \lambda$  bezeichnen. Ob ein gegebenes L-S-Maß von dieser Art ist, können Sie übrigens oft an der Verteilungsfunktion  $F_\mu$  erkennen. Ist diese zB stetig differenzierbar, so gilt ja*

$$\mu((a, b]) = F_\mu(b) - F_\mu(a) = \int_a^b F'_\mu(s) ds, \quad \text{für } a < b \text{ in } \mathbb{R}.$$

*Also besitzt  $\mu$  in diesem Fall gerade die Dichte  $f := F'_\mu$ .*

Mit dem Satz haben wir, wie angekündigt, eine vollständige Beschreibung sämtlicher Maße gewonnen, die Intervalle aus  $\mathcal{I}_{\mathbb{R}}$  messen können, und ihnen jeweils endliche Masse zuordnen. Ein Aspekt bleibt noch ein wenig unangenehm: die  $\sigma$ -Algebra  $\mathcal{A}(\mu_F^*)$  der zugehörigen  $\mu_F^*$ -messbaren Mengen hängt von  $F$  ab. Wesentlich ist aber, dass sie in jedem Fall  $\mathcal{I}_{\mathbb{R}}$  und damit (Proposition 1.9) sämtliche Intervalle und auch alle offenen und alle abgeschlossenen Teilmengen von  $\mathbb{R}$  enthält. Dennoch wäre es praktisch, einfach eine gemeinsame  $\sigma$ -Algebra für alle  $\mu_F$  verwenden zu können. Wie das geht, besprechen wir im nächsten Abschnitt.

Zuvor wollen wir noch andeuten, dass der Zoo der L-S-Maße auf  $\mathbb{R}$  noch viel exotischere Kreaturen beherbergt, als die einfachen Beispiele oben. (Weil es letztlich “nur” um Verteilungsfunktionen geht, unterschätzt man sehr leicht die Vielfalt der so entstehenden Maße.)

**Beispiel 2.8 (Ein Maß auf der Cantor-Menge)** *Wir erinnern uns an das Cantorsche Diskontinuum  $C \subseteq [0, 1]$  aus Beispiel 1.4. Wie besprochen ist dieses für das Lebesgue-Maß unsichtbar, weil  $\lambda(C) = 0$ . Die Menge ist aber sehr interessant, und hat eine bemerkenswerte Struktur. Wir stellen uns nun die Frage, ob wir ein L-S-Maß  $\mu = \mu_F$  finden können, das auf dieser Menge konzentriert ist in dem Sinne, dass  $\mu(C) = 1$  und  $\mu(C^c) = 0$ , und das auch Teilmengen von  $C$  sinnvoll messen kann, etwa indem für die linke Hälfte der Cantor-Menge gerade  $\mu((0, \frac{1}{2}] \cap C) = \frac{1}{2}$  gilt.*

*Dazu nehmen wir mal an, dass so ein  $\mu$  existiert, und versuchen zu verstehen, wie dann eine zugehörige VF mit  $F(t) := \mu((-\infty, t])$  aussehen muss. Wegen  $C \subseteq [0, 1]$  gilt jedenfalls  $F(t) = 0$  für  $t < 0$  und  $F(t) = 1$  für  $t \geq 1$ . Betrachten wir das offene Intervall  $D_{(\phi)} := [0, 1] \setminus (C_{(0)} \cup C_{(1)}) = (\frac{1}{3}, \frac{2}{3})$  welches im ersten Schritt*

der Konstruktion von  $C$  weggesehritten wird. Ist  $t \in D_{(\phi)}$ , dann besteht  $(-\infty, t] \cap C$  genau aus der linken Hälfte von  $C$ , und es sollte daher  $F(t) = \frac{1}{2}$  gelten. Analoges gilt wenn wir die im nächsten Konstruktionsschritt entfernten offenen Intervalle  $D_{(0)} := C_{(0)} \setminus (C_{(00)} \cup C_{(01)}) = (\frac{1}{9}, \frac{2}{9})$  und  $D_{(1)} := C_{(1)} \setminus (C_{(10)} \cup C_{(11)}) = (\frac{7}{9}, \frac{8}{9})$  betrachten: Für  $t \in D_{(0)}$  besteht  $(-\infty, t] \cap C$  genau aus dem linken Viertel von  $C$ , und es sollte daher  $F(t) = \frac{1}{4}$  sein, während für  $t \in D_{(1)}$  die Menge  $(-\infty, t] \cap C$  sich aus den drei linken Viertel von  $C$  zusammensetzt, sodass wir  $F(t) = \frac{3}{4}$  bekommen.

Allgemein sei für  $n \geq 1$  und  $\omega_1, \dots, \omega_n \in \{0, 1\}$  nun  $D_{(\omega_1, \dots, \omega_n)} := C_{(\omega_1, \dots, \omega_n)} \setminus (C_{(\omega_1, \dots, \omega_n, 0)} \cup C_{(\omega_1, \dots, \omega_n, 1)})$  das mittig aus  $C_{(\omega_1, \dots, \omega_n)}$  entfernte offene Intervall, dann ist für alle  $t \in D_{(\omega_1, \dots, \omega_n)}$  die Menge  $(-\infty, t] \cap C$  jeweils dieselbe, also muss  $F$  auf  $D_{(\omega_1, \dots, \omega_n)}$  konstant sein. Man sieht, dass konkret alle Zahlen aus  $\{\frac{k}{2^n} : n \geq 1, 0 \leq k \leq 2^n\}$  als Werte angenommen werden. Auf diese Weise können wir  $F$  auf der offenen Menge  $D := C^c$  bestimmen. Die Funktion ist dort natürlich wachsend.

Im PS überlegen wir, dass  $F$  auf genau eine (einfache) Weise zu einer VF  $F : \mathbb{R} \rightarrow \mathbb{R}$  fortgesetzt werden kann, und dass das zugehörige L-S-Maß  $\mu = \mu_F$  die gewünschten Eigenschaften besitzt. Wir sehen also mithilfe der Verteilungsfunktion, dass es genau ein L-S-Maß gibt, welches Teilmengen von  $C$  in der gewünschten natürlichen Weise misst!

Bemerkenswert ist übrigens auch die so gewonnene Cantor-Funktion  $F$ . Sie ist stetig und wächst auf  $[0, 1]$  von  $F(0) = 0$  bis  $F(1) = 1$ . Dabei ist sie aber fast überall konstant! Genauer: Die Vereinigung  $[0, 1] \setminus C$  aller oben verwendeten offenen Intervalle  $D_{(\omega_1, \dots, \omega_n)}$  in  $[0, 1]$  hat Lebesgue-Maß  $\lambda([0, 1] \setminus C) = 1$ , und auf jedem dieser Intervalle ist  $F$  konstant. Die Funktion ist also auf einer Menge vollen Maßes differenzierbar, und hat dort Ableitung  $F'(t) = 0$ . Trotzdem schafft sie es, echt zu wachsen!

## 2.4 Erzeugte $\sigma$ -Algebren

Neben ein paar wenigen elementaren Beispielen haben wir bisher nur über die  $\sigma$ -Algebren  $\mathcal{A}(\mu^*)$  der zu einem äußeren Maß  $\mu^*$  messbaren Mengen gesprochen, die in der Regel keine konstruktive Beschreibung gestatten. Letzteres ist aber nicht schlimm, wir brauchen meist nur zu wissen, dass die für uns relevanten Mengen dort enthalten sind. Sobald wir es mit mehr als einem Maß zu tun haben, sollten diese Mengenfunktionen aber möglichst auf einer gemeinsamen  $\sigma$ -Algebra definiert sein. Überhaupt wäre es schön, eine ausreichend große  $\sigma$ -Algebra zu haben, die unabhängig von der Wahl eines Maßes ein natürliches Objekt ist. Zum Glück geht das recht einfach:

**Proposition 2.5 (Durchschnitte und erzeugte  $\sigma$ -Algebra)** .

- a) Ist  $(\mathcal{A}_i)_{i \in \mathcal{I}}$  eine beliebige Kollektion von  $\sigma$ -Algebren auf  $X$  mit  $\mathcal{I} \neq \emptyset$ , so ist auch  $\bigcap_{i \in \mathcal{I}} \mathcal{A}_i$  eine  $\sigma$ -Algebra auf  $X$ .
- b) Sei  $\mathcal{E} \subseteq \mathcal{P}(X)$  eine beliebige Mengenfamilie. Dann gibt es eine eindeutig bestimmte kleinste  $\sigma$ -Algebra  $\sigma(\mathcal{E})$ , welche  $\mathcal{E}$  enthält, nämlich

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathcal{A} \subseteq \mathcal{P}(X), \mathcal{E} \subseteq \mathcal{A} \text{ und} \\ \mathcal{A} \text{ ist eine } \sigma\text{-Algebra}}} \mathcal{A}.$$

- c) Ist  $\mathcal{A} \subseteq \mathcal{P}(X)$  eine  $\sigma$ -Algebra, dann folgt  $\sigma(\mathcal{A}) = \mathcal{A}$ .
- d) Für beliebige  $\mathcal{E}, \mathcal{F} \subseteq \mathcal{P}(X)$  gilt: ist  $\mathcal{E} \subseteq \mathcal{F}$ , dann folgt  $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{F})$ .
- e) Für beliebige  $\mathcal{E}, \mathcal{F} \subseteq \mathcal{P}(X)$  gilt: ist  $\mathcal{E} \subseteq \sigma(\mathcal{F})$  und  $\mathcal{F} \subseteq \sigma(\mathcal{E})$ , dann folgt  $\sigma(\mathcal{E}) = \sigma(\mathcal{F})$ .

"Kleinste" bedeutet hier, dass jede  $\sigma$ -Algebra  $\mathcal{A}$  welche  $\mathcal{E}$  enthält auch  $\sigma(\mathcal{E})$  enthält. Ist  $\mathcal{E}$  selbst schon eine  $\sigma$ -Algebra  $\mathcal{A}$ , dann ist natürlich  $\sigma(\mathcal{A}) = \mathcal{A}$ .

**Definition 2.9** Die Familie  $\sigma(\mathcal{E})$  ist die von  $\mathcal{E}$  erzeugte  $\sigma$ -Algebra. Ist  $\mathcal{A}$  eine  $\sigma$ -Algebra, so nennt man jede Familie  $\mathcal{E} \subseteq \mathcal{A}$  mit  $\sigma(\mathcal{E}) = \mathcal{A}$  ein Erzeugendensystem (kurz einen Erzeuger) von  $\mathcal{A}$ .

**Beweis.** a) Den Nachweis der Eigenschaften (A1)-(A3)' darf man beinahe als trivial bezeichnen. ZB für (A2): ist  $A \in \bigcap_{\iota \in \mathcal{J}} \mathcal{A}_\iota$ , dann also  $A \in \mathcal{A}_\iota$  für jedes  $\iota \in \mathcal{J}$ , und da dies lauter ( $\sigma$ -)Algebren sind, folgt jeweils, dass auch  $A^c \in \mathcal{A}_\iota$ . Somit ist  $A^c \in \bigcap_{\iota \in \mathcal{J}} \mathcal{A}_\iota$ . Undsowweiterundsowweiter...

b) Zunächst sind die Definitionen sinnvoll, da die Familien deren Durchschnitte betrachtet werden stets nicht leer sind: die maximale  $\sigma$ -Algebra  $\mathcal{A} := \mathcal{P}(X)$  ist immer dabei. Nach a) ist  $\sigma(\mathcal{E})$  eine  $\sigma$ -Algebra, und aufgrund der Definition ist sie in jeder  $\mathcal{E}$  umfassenden  $\sigma$ -Algebra enthalten.

c) Ist klar.

d) Wenn  $\mathcal{E} \subseteq \mathcal{F}$ , dann ist jede  $\mathcal{F}$  umfassende  $\sigma$ -Algebra  $\mathcal{A}$  auch eine  $\mathcal{E}$  umfassende  $\sigma$ -Algebra.

e) Falls  $\mathcal{E} \subseteq \sigma(\mathcal{F})$ , so folgt  $\sigma(\mathcal{E}) \subseteq \sigma(\sigma(\mathcal{F})) = \sigma(\mathcal{F})$ , und umgekehrt. ■

**Beispiel 2.9** a) Für  $\mathcal{E} = \emptyset$  und für  $\mathcal{E} = \{\emptyset\}$  ist jeweils  $\sigma(\mathcal{E}) = \{\emptyset, X\}$ .

b) Für  $A \subseteq X$  und  $\mathcal{E} = \{A\}$  ist  $\sigma(\mathcal{E}) = \{\emptyset, X, A, A^c\}$ .

c) Für  $\mathcal{E} = \{A \subseteq X : \#A < \infty\}$  ist  $\sigma(\mathcal{E}) = \{A \subseteq X : A \text{ oder } A^c \text{ abzählbar}\}$ .

Für die Zwecke der Analysis brauchen wir auf  $\mathbb{R}$  bzw  $\mathbb{R}^d$  jedenfalls eine  $\sigma$ -Algebra welche alle Intervalle (bzw  $d$ -dimensionalen Quader) und (damit, siehe Proposition 1.9) sämtliche offenen und abgeschlossenen Mengen enthält. Sehr häufig ist die kleinste unter diesen die beste Wahl.

**Definition 2.10** Die Borel- $\sigma$ -Algebra auf  $\mathbb{R}^d$  ist  $\mathcal{B}_{\mathbb{R}^d} := \sigma(\mathcal{I}_{\mathbb{R}^d}^d) \subseteq \mathcal{P}(\mathbb{R}^d)$ , ihre Elemente sind die Borel-Mengen von  $\mathbb{R}^d$ .

Das ist also die kleinste  $\sigma$ -Algebra auf  $\mathbb{R}^d$ , welche die wichtigsten Typen von Mengen enthält. Sie erweist sich als sehr reichhaltig und für sehr viele Gelegenheiten als ausreichend groß. Wie schon  $\mathcal{L}_{\mathbb{R}^d}$  erlaubt aber auch  $\mathcal{B}_{\mathbb{R}^d}$  keine konstruktive Beschreibung.

Da bekanntlich  $\mathcal{I}_{\mathbb{R}^d}^d \subseteq \mathcal{L}_{\mathbb{R}^d}$  ist, folgt (Teil d) obiger Proposition) sofort  $\mathcal{B}_{\mathbb{R}^d} \subseteq \mathcal{L}_{\mathbb{R}^d}$ . Entsprechend gilt für alle L-S-Maße  $\mu_F$  auf  $\mathbb{R}$  dass  $\mathcal{B}_{\mathbb{R}^d} \subseteq \mathcal{A}(\mu_F^*)$ . Wir können also gefahrlos das Lebesguesche Maß und jedes L-S-Maß auf Borel-Mengen anwenden, bzw diese Mengenfunktionen auf jene Teil- $\sigma$ -Algebra einschränken. Auf diesem Weg bekommt man dann zB das (hier wieder mit  $\lambda$  bzw  $\lambda^d$  bezeichnete)  $d$ -dimensionale Lebesgue-Borel-Maß  $\lambda^d : \mathcal{B}_{\mathbb{R}^d} \rightarrow [0, \infty]$ . Obwohl das genaugenommen nun wirklich eine andere Mengenfunktion als das Lebesgue-Maß  $\lambda^d : \mathcal{L}_{\mathbb{R}^d} \rightarrow [0, \infty]$  ist, spricht man aber oft einfach wieder vom Lebesgue-Maß oder halt Lebesgue-Maß auf  $\mathcal{B}_{\mathbb{R}^d}$ .

Wir verzichten auf den Beweis des folgenden Satzes, der zeigt, dass sich  $\mathcal{B}_{\mathbb{R}^d}$  und  $\mathcal{L}_{\mathbb{R}^d}$  nur durch Teilmengen von Borel-messbaren  $\lambda^d$ -Nullmengen unterscheiden.

**Satz 2.6 (Beziehung zwischen  $\mathcal{B}_{\mathbb{R}^d}$  und  $\mathcal{L}_{\mathbb{R}^d}$ )** Für jedes  $d \geq 1$  ist  $\mathcal{B}_{\mathbb{R}^d} \neq \mathcal{L}_{\mathbb{R}^d}$ . Eine Menge  $A \subseteq \mathbb{R}^d$  liegt genau dann in  $\mathcal{L}_{\mathbb{R}^d}$  wenn sie von der Form  $A = B \uplus C$  ist mit  $B \in \mathcal{B}_{\mathbb{R}^d}$  und  $C \subseteq N \in \mathcal{B}_{\mathbb{R}^d}$ , wobei  $\lambda^d(N) = 0$ .

Nachdem man sich bei der Beschreibung relevanter  $\sigma$ -Algebren sonst schwer tut, bieten Erzeuger sehr häufig die beste Möglichkeit, um mit ihnen umzugehen. Oft versteht man  $\sigma$ -Algebren und Maße am besten über Erzeuger. Deshalb ist es auch nützlich wenn man mehrere Erzeuger kennt, dann kann man nämlich mit jenem arbeiten, der gerade am angenehmsten ist.

**Proposition 2.6 (Verschiedene Erzeuger für  $\mathcal{B}_{\mathbb{R}}$ )** Folgende Familien in  $\mathcal{P}(\mathbb{R})$  erzeugen jeweils die Borel- $\sigma$ -Algebra  $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{I}_{\mathbb{R}})$ .



- a)  $\mathcal{I} := \{I \subseteq \mathbb{R} : I \text{ ist irgendein Intervall in } \mathbb{R}\},$
- b)  $\mathcal{I}_o := \{I \subseteq \mathbb{R} : I \text{ ist ein offenes Intervall in } \mathbb{R}\},$
- c)  $\mathcal{I}_a := \{I \subseteq \mathbb{R} : I \text{ ist ein abgeschlossenes Intervall in } \mathbb{R}\},$
- d)  $\mathcal{I}_k := \{I \subseteq \mathbb{R} : I \text{ ist ein kompaktes Intervall in } \mathbb{R}\},$
- e)  $\mathcal{I}_* := \{(-\infty, b] : b \in \mathbb{R}\},$
- f)  $\mathcal{I}_{**} := \{(-\infty, b] : b \in \mathbb{Q}\},$
- g)  $\mathcal{O}_{\mathbb{R}} := \{M \subseteq \mathbb{R} : M \text{ ist eine offene Teilmenge von } \mathbb{R}\},$
- h)  $\mathcal{F}_{\mathbb{R}} := \{M \subseteq \mathbb{R} : M \text{ ist eine abgeschlossene Teilmenge von } \mathbb{R}\},$
- i)  $\mathcal{K}_{\mathbb{R}} := \{M \subseteq \mathbb{R} : M \text{ ist eine kompakte Teilmenge von } \mathbb{R}\}.$

**Beweis.** Man verwendet jeweils Teil e) obiger Proposition. Wir illustrieren dies anhand von  $\mathcal{O}_{\mathbb{R}}$ . Zunächst ist  $\mathcal{I}_{\mathbb{R}} \subseteq \sigma(\mathcal{O}_{\mathbb{R}})$ , weil jedes Element von  $\mathcal{I}_{\mathbb{R}}$  dargestellt werden kann als  $(a, b] = \bigcap_{n \geq 1} (a, b + \frac{1}{n})$ , und die Menge rechts in  $\sigma(\mathcal{O}_{\mathbb{R}})$  liegt. Umgekehrt ist  $\mathcal{O}_{\mathbb{R}} \subseteq \sigma(\mathcal{I}_{\mathbb{R}})$ , wissen wir doch schon (Lemma 1.1), dass es für beliebiges offenes  $G \subseteq \mathbb{R}$  eine Folge  $(A_n)_{n \geq 1}$  in  $\mathcal{I}_{\mathbb{R}}$  gibt mit  $G = \bigcup_{n \geq 1} A_n$ , woraus  $G \in \sigma(\mathcal{I}_{\mathbb{R}})$  folgt. ■

Gerade die Charakterisierungen von  $\mathcal{B}_{\mathbb{R}}$  über große Familien topologisch einfacher (insbesondere offener) Mengen erweisen sich als besonders nützlich, sie stellen reibungsloses Zusammenspiel mit Stetigkeits- und Konvergenzargumenten sicher. Entsprechendes gilt auch wenn wir ausreichend große  $\sigma$ -Algebren in anderen Räumen suchen. Die  $\sigma$ -Algebra  $\mathcal{B}_{\mathbb{R}}$  ist ein Spezialfall folgenden Konzeptes:

**Definition 2.11** Ist  $(X, \mathcal{O}_X)$  ein topologischer Raum (zB  $(X, d)$  ein metrischer Raum, oder speziell  $X$  eine Teilmenge eines euklidischen Raumes  $\mathbb{R}^d$ , und  $\mathcal{O}_X$  die Familie der offenen Teilmengen von  $X$ ), dann ist  $\mathcal{B}_X := \sigma(\mathcal{O}_X)$  die Borel- $\sigma$ -Algebra von  $(X, \mathcal{O}_X)$ . Die  $A \in \mathcal{B}_X$  sind die Borel-(messbaren-)Mengen in  $X$ .

Die Erfahrung zeigt, dass diese Definition die meisten in der mathematischen Praxis auftretenden Mengen erfasst. Borel- $\sigma$ -Algebren sind die wichtigsten konkreten  $\sigma$ -Algebren der Maßtheorie. Für die Analysis sind natürlich die Borel- $\sigma$ -Algebren der euklidischen Räume  $\mathbb{R}^d$  von größter Bedeutung. Analog zum eindimensionalen Fall finden wir:

**Proposition 2.7 (Verschiedene Erzeuger für  $\mathcal{B}_{\mathbb{R}^d}$ )** Folgende Teilfamilien von  $\mathcal{P}(\mathbb{R}^d)$  erzeugen jeweils die Borel- $\sigma$ -Algebra  $\mathcal{B}_{\mathbb{R}^d}$ :

$$\mathcal{I}_{\mathbb{R}^d} = \mathcal{I}_{\mathbb{R}}^d, \mathcal{I}_o^d, \mathcal{I}_a^d, \mathcal{I}_k^d, \mathcal{I}_*^d, \mathcal{I}_{**}^d, \mathcal{O}_{\mathbb{R}^d}, \mathcal{K}_{\mathbb{R}^d} = \{A \subseteq \mathbb{R}^d : A \text{ ist kompakt}\}.$$

**Beispiel 2.10** Wenn man  $\overline{\mathbb{R}} = [-\infty, \infty]$  mit einer passenden Metrik versieht (so dass Konvergenz  $a_n \rightarrow \infty$  das bedeutet was es bedeuten soll), dann findet man  $\mathcal{B}_{\overline{\mathbb{R}}} = \{B \cup C : B \in \mathcal{B}_{\mathbb{R}}, C \subseteq \{-\infty, \infty\}\}$ , und diese  $\sigma$ -Algebra wird zB von  $\{[-\infty, b] : b \in \mathbb{R}\}$  erzeugt.

**Der allgemeine Eindeutigkeitssatz für Maße.** Wir erwähnen noch ein Resultat der allgemeinen Maßtheorie, welches den Grundsatz “ $\sigma$ -Algebren und Maße versteht man oft am besten über Erzeuger” illustriert. Eines der zentralen Ergebnisse dieses Kapitels ist der Satz 2.3 über die Eindeutigkeit der Fortsetzung. Er zeigt insbesondere, dass (im  $\sigma$ -endlichen Fall) ein Maß auf  $\mathcal{A}$  schon vollständig bestimmt ist, wenn man seine Einschränkung auf einen erzeugenden Halbring kennt:

**Proposition 2.8 (Eindeutigkeit über Halbringe)** Es seien  $\mu$  und  $\nu$  Maße auf  $(X, \mathcal{A})$ , die auf einem Halbring  $\mathcal{H}$  mit  $\sigma(\mathcal{H}) = \mathcal{A}$  übereinstimmen,  $\mu|_{\mathcal{H}} = \nu|_{\mathcal{H}}$ . Falls die Maße  $\sigma$ -endlich auf  $\mathcal{H}$  sind, dann ist  $\mu = \nu$ .

**Beweis.** Das folgt sofort aus Satz 2.3 weil  $\mu$  und  $\nu$  Fortsetzungen ein- und desselben  $\sigma$ -endlichen Prämaßes  $\mu|_{\mathcal{H}} = \nu|_{\mathcal{H}}$  nach  $\mathcal{A}$  mit  $\mathcal{A} = \sigma(\mathcal{H}) \subseteq \mathcal{A}(\mu|_{\mathcal{H}}^*)$  (vgl Proposition 2.4) sind. ■

Dies ist ein Spezialfall des folgenden Eindeutigkeitsresultates, das zu den wichtigsten Standardresultaten der Maßtheorie gehört. (Auf den Beweis verzichten wir aber, da muss ich Sie auf weiterführende LVen bzw die Literatur verweisen.) Um ein Maß eindeutig festzulegen darf man den Halbring durch einen anderen Erzeuger  $\mathcal{E} \subseteq \mathcal{P}(X)$  von  $\mathcal{A}$  ersetzen, wenn dieser nur *(durch)schnittstabil* ist, dh wenn für  $A, B \in \mathcal{E}$  stets auch  $A \cap B \in \mathcal{E}$  ist.

**Satz 2.7 (Eindeutigkeitssatz für Maße)** *Es seien  $\mu$  und  $\nu$  Maße auf  $(X, \mathcal{A})$ , die auf einem durchschnittstabilen Erzeuger  $\mathcal{E}$  der  $\sigma$ -Algebra  $\mathcal{A}$  übereinstimmen,  $\mu|_{\mathcal{E}} = \nu|_{\mathcal{E}}$ . Falls  $\mu(X) = \nu(X) < \infty$ , oder falls die Maße  $\sigma$ -endlich auf  $\mathcal{E}$  sind, dann ist  $\mu = \nu$ .*

Auf die Durchschnittstabilität des Erzeugers kann man aber auch im endlichen Fall nicht verzichten:

**Beispiel 2.11** *Sei  $X := \{1, 2, 3, 4\}$ , dann erzeugt  $\mathcal{E} := \{\{1, 2\}, \{2, 3\}\}$  die  $\sigma$ -Algebra  $\mathcal{A} = \mathcal{P}(X)$ . Die beiden endlichen Maße  $\mu := \delta_1 + \delta_3$  und  $\nu := \delta_2 + \delta_4$  auf  $\mathcal{A}$  stimmen für alle  $A \in \mathcal{E} \cup \{X\}$  überein, obwohl  $\mu \neq \nu$ .*

Weiters übersieht man leicht, dass im Satz die  $\sigma$ -Endlichkeit auf  $\mathcal{E}$  vorausgesetzt wird. Es reicht nicht, dass  $\mu$  und  $\nu$  einfach  $\sigma$ -endliche Maße (dh  $\sigma$ -endlich auf  $\mathcal{A}$ ) sind!

**Beispiel 2.12** *Es seien  $(X, \mathcal{A}) := (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  und  $\mu := \lambda$  während  $\nu := 2\lambda$ . Dann sind  $\mu$  und  $\nu$  zwei  $\sigma$ -endliche Maße auf  $(X, \mathcal{A})$ . Offenbar gilt  $\nu = 2\mu \neq \mu$ . Dies obwohl  $\mu$  und  $\nu$  auf dem durchschnittstabilen Erzeuger  $\mathcal{E} = \mathcal{I}_* := \{(-\infty, b] : b \in \mathbb{R}\}$  von  $\mathcal{A}$  übereinstimmen.*

## 2.5 Messbare Abbildungen

Hat man eine Abbildung  $f : X \rightarrow Y$  eines Maßraumes  $(X, \mathcal{A}, \mu)$  in eine Menge  $Y$ , so liegt es nahe, die durch das Maß  $\mu$  auf  $X$  gegebene "Masseverteilung" mittels  $f$  nach  $Y$  zu transportieren. Die Größe einer Menge  $B \subseteq Y$  sollte das Maß aller jener Punkte aus  $X$  sein, die in  $B$  landen. Wir interessieren uns daher für die *Urbildoperation*, die (unabhängig davon, ob  $f$  injektiv ist) stets eine Abbildung  $f^{-1} : \mathcal{P}(Y) \rightarrow \mathcal{P}(X)$  liefert. Bekanntlich verhält sich diese sehr gutmütig, indem sie mit Mengenoperationen kommutiert: es ist immer  $f^{-1}B^c = (f^{-1}B)^c$  und für beliebige Familien  $(B_i)_{i \in \mathcal{I}}$  in  $\mathcal{P}(Y)$  gelten stets  $f^{-1}(\bigcup_{i \in \mathcal{I}} B_i) = \bigcup_{i \in \mathcal{I}} f^{-1}B_i$  und  $f^{-1}(\bigcap_{i \in \mathcal{I}} B_i) = \bigcap_{i \in \mathcal{I}} f^{-1}B_i$ . (Überzeugen Sie sich bitte davon!)

**Definition 2.12** *Sind  $(X, \mathcal{A})$  und  $(Y, \mathcal{B})$  messbare Räume dh jeweils ein Paar bestehend aus einer Menge und einer  $\sigma$ -Algebra auf derselben, dann heißt eine Abbildung  $f : X \rightarrow Y$  messbar bzgl  $\mathcal{A}$  und  $\mathcal{B}$ , kurz:  $\mathcal{A}$ - $\mathcal{B}$ -messbar, falls*

$$f^{-1}\mathcal{B} := \{f^{-1}B : B \in \mathcal{B}\} \subseteq \mathcal{A}.$$

Wenn klar ist, um welche  $\sigma$ -Algebren es geht, so spricht man einfach von messbaren Abbildungen. Das gilt insbesondere wenn  $Y = \mathbb{R}$  oder  $\overline{\mathbb{R}}$ ,  $\mathbb{R}^d$ ,  $\mathbb{C}$  ist, in solchen Fällen ist mit messbar generell  $\mathcal{A}$ - $\mathcal{B}_Y$ -messbar (Borel- $\sigma$ -Algebra auf  $Y$ ) gemeint, sofern nicht explizit etwas anderes gesagt wird. Um implizit auszudrücken, dass unter einer erst zu spezifizierenden  $\sigma$ -Algebra die Borelsche gemeint ist, nennt man  $f$  kurz Borel-messbar.

**Bemerkung 2.6** Für  $f : X \rightarrow Y$  mit  $Y \subseteq Z$ , wobei etwa  $Y$  und  $Z$  beide eine Borel- $\sigma$ -Algebra besitzen, könnten wir  $f$  auch als Abbildung  $f : X \rightarrow Z$  ansehen, und es stellt sich die Frage, welche  $\sigma$ -Algebra wir dann nehmen sollen. Es ist dann aber  $\mathcal{B}_Y = Y \cap \mathcal{B}_Z$  und für jedes  $C \subseteq Z$  gilt  $f^{-1}C = f^{-1}(Y \cap C)$ . Daher sehen wir, dass  $f^{-1}\mathcal{B}_Z = f^{-1}(Y \cap \mathcal{B}_Z) = f^{-1}\mathcal{B}_Y$ , sodass  $f$  genau dann  $\mathcal{A}$ - $\mathcal{B}_Y$ -messbar ist, wenn es  $\mathcal{A}$ - $\mathcal{B}_Z$ -messbar ist.

Insbesondere:  $f : X \rightarrow \mathbb{R} \subseteq \overline{\mathbb{R}}$  ist  $\mathcal{A}$ - $\mathcal{B}_{\mathbb{R}}$ -messbar genau wenn  $\mathcal{A}$ - $\mathcal{B}_{\overline{\mathbb{R}}}$ -messbar.

Sehr praktisch ist folgende Schreibweise für Urbilder unter  $f : X \rightarrow Y$ ,

$$\{f \in B\} := \{x \in X : f(x) \in B\} = f^{-1}B \quad \text{für } B \subseteq Y, \quad (2.8)$$

und, speziell wenn  $f : X \rightarrow \overline{\mathbb{R}}$ ,

$$\{f \leq b\} := \{x \in X : f(x) \leq b\} = f^{-1}[-\infty, b], \quad \{f = b\} := f^{-1}\{b\}, \quad \text{etc.}$$

Nun zur eingangs angekündigten Möglichkeit, Maße durch entsprechende Abbildungen zu transportieren:

**Proposition 2.9 (Zusammensetzungen und Bildmaße)** (i) Sind  $f : X \rightarrow Y$  und  $g : Y \rightarrow Z$  jeweils  $\mathcal{A}$ - $\mathcal{B}$ -messbar bzw  $\mathcal{B}$ - $\mathcal{C}$ -messbar, so ist  $g \circ f : X \rightarrow Z$  stets  $\mathcal{A}$ - $\mathcal{C}$ -messbar.

(ii) Sind  $(X, \mathcal{A})$  und  $(Y, \mathcal{B})$  messbare Räume und  $f : X \rightarrow Y$  eine  $(\mathcal{A}, \mathcal{B})$ -messbare Abbildung, dann ist für jedes Maß  $\mu : \mathcal{A} \rightarrow [0, \infty]$  die Mengenfunktion

$$\mu \circ f^{-1} : \mathcal{B} \rightarrow [0, \infty] \quad (\text{dh } \mu \circ f^{-1}(B) := \mu(f^{-1}B), B \in \mathcal{B})$$

ein Maß auf  $\mathcal{B}$ , das Bild(maß) von  $\mu$  unter  $f$ .

(iii) Ist  $(Z, \mathcal{C})$  ein weiterer messbarer Raum und ist  $g : Y \rightarrow Z$  eine  $(\mathcal{B}, \mathcal{C})$ -messbare Abbildung, dann folgt  $(\mu \circ f^{-1}) \circ g^{-1} = \mu \circ (g \circ f)^{-1}$ .

**Beweis.** (i) Das stimmt wegen  $(g \circ f)^{-1}\mathcal{C} = f^{-1}(g^{-1}\mathcal{C}) \subseteq f^{-1}\mathcal{B} \subseteq \mathcal{A}$ .

(ii) Für jedes  $B \in \mathcal{B}$  ist  $\mu \circ f^{-1}(B)$  definiert, weil  $f$   $(\mathcal{A}, \mathcal{B})$ -messbar ist. Offenbar ist  $\mu \circ f^{-1}(\emptyset) = \mu(f^{-1}\emptyset) = \mu(\emptyset) = 0$ . Zum Nachweis der  $\sigma$ -Additivität sei  $(B_n)_{n \geq 1}$  eine disjunkte Folge in  $\mathcal{B}$ . Dann sind auch die  $f^{-1}B_n$  paarweise disjunkt, und daher

$$\mu \circ f^{-1} \left( \biguplus_{n \geq 1} B_n \right) = \mu \left( \biguplus_{n \geq 1} f^{-1}B_n \right) = \sum_{n \geq 1} \mu(f^{-1}B_n) = \sum_{n \geq 1} \mu \circ f^{-1}(B_n).$$

(iii) Es ist ja  $f^{-1} \circ g^{-1} = (g \circ f)^{-1}$ . ■

Wie erkennt man nun aber messbare Abbildungen? In einfachen Situationen kann man direkt die Definition anwenden.

**Beispiel 2.13 (Grundlegende Beispiele messbarer Abbildungen)** .

a) Jede konstante Abbildung  $f : X \rightarrow Y$  ist messbar, weil in diesem Fall  $f^{-1}\mathcal{B} = \{\emptyset, X\} \subseteq \mathcal{A}$ .

b) Die Indikatorfunktion  $f = 1_A : X \rightarrow \mathbb{R}$  von  $A \subseteq X$  ist messbar (dh  $\mathcal{A}$ - $\mathcal{B}_{\mathbb{R}}$ -messbar) genau wenn  $A \in \mathcal{A}$  (siehe PS).

Meistens ist das aber kaum möglich, weil wir bei interessanten  $\sigma$ -Algebren die messbaren Mengen gar nicht explizit kennen. Zum Glück reicht es dann, die Mengen aus einem Erzeuger zu betrachten:

**Proposition 2.10 (Messbarkeit erkennen)** Eine Abbildung  $f : X \rightarrow Y$  ist  $\mathcal{A}$ - $\mathcal{B}$ -messbar sobald  $f^{-1}\mathcal{E} \subseteq \mathcal{A}$  für einen Erzeuger  $\mathcal{E} \subseteq \mathcal{P}(Y)$  von  $\mathcal{B}$ .

Der Beweis dieser Aussage wird ganz einfach, wenn wir uns zuvor allgemein überlegen, wie sich die Urbildoperation mit dem Erzeugen von  $\sigma$ -Algebren verträgt. Geklärt wird dies in folgendem Lemma (im PS besprochen).

**Lemma 2.1 (Urbilder und Erzeuger)** Sei  $f : X \rightarrow Y$  und  $\mathcal{E} \subseteq \mathcal{P}(Y)$  beliebig. Dann gilt

$$f^{-1}(\sigma(\mathcal{E})) = \sigma(f^{-1}(\mathcal{E})).$$

(Wobei wir rechts natürlich die von  $f^{-1}(\mathcal{E}) \subseteq \mathcal{P}(X)$  auf  $X$  erzeugte  $\sigma$ -Algebra meinen. Es ist nicht üblich, die Grundmenge in dieser Notation anzuführen.)

**Beweis.** Offenbar ist  $\sigma(f^{-1}(\mathcal{E})) \subseteq \sigma(f^{-1}(\sigma(\mathcal{E}))) = f^{-1}(\sigma(\mathcal{E}))$ . Zu zeigen bleibt also  $f^{-1}(\sigma(\mathcal{E})) \subseteq \sigma(f^{-1}(\mathcal{E}))$ , dh

$$\forall A \in \sigma(\mathcal{E}) : f^{-1}A \in \sigma(f^{-1}(\mathcal{E})).$$

Wir verwenden ein als *Prinzip der guten Mengen* bekanntes Argument: wegen

$$\mathcal{E} \subseteq \mathcal{G} := \{A \in \sigma(\mathcal{E}) : f^{-1}A \in \sigma(f^{-1}(\mathcal{E}))\} \subseteq \sigma(\mathcal{E})$$

folgt unsere Behauptung  $\mathcal{G} = \sigma(\mathcal{E})$  sofort, wenn wir zeigen, dass die Familie  $\mathcal{G}$  der in unserem Sinne guten Mengen die Struktur einer  $\sigma$ -Algebra hat.

(A1) Ist klar wegen  $f^{-1}X = X' \in \sigma(f^{-1}(\mathcal{E}))$ . Für (A2) sei  $A \in \mathcal{G}$ , dann ist mit  $f^{-1}A \in \sigma(f^{-1}(\mathcal{E}))$  auch  $f^{-1}(A^c) = (f^{-1}A)^c \in \sigma(f^{-1}(\mathcal{E}))$ . Entsprechend folgt (A3): Für  $A_n \in \mathcal{G}$ ,  $n \geq 1$ , ist mit den  $f^{-1}A_n \in \sigma(f^{-1}(\mathcal{E}))$ ,  $n \geq 1$ , auch  $f^{-1}(\bigcup_{n \geq 1} A_n) = \bigcup_{n \geq 1} f^{-1}A_n \in \sigma(f^{-1}(\mathcal{E}))$ , also  $\bigcup_{n \geq 1} A_n \in \mathcal{G}$ . ■

Wie angekündigt ergibt sich dann leicht der

**Beweis von Proposition 2.10.**  $f^{-1}\mathcal{B} = f^{-1}\sigma(\mathcal{E}) = \sigma(f^{-1}\mathcal{E}) \subseteq \sigma(\mathcal{A}) = \mathcal{A}$ . ■

Damit können wir nun mit relativ geringem Aufwand in vielen wichtigen Situationen sehen, dass zahlreiche Typen von Abbildungen automatisch messbar sind. (Beachten Sie, dass es dabei sehr nützlich ist, verschiedene Erzeuger zu kennen.)

**Beispiel 2.14 (Stetige Abbildungen)** Sind  $X, Y$  offene Mengen in  $\mathbb{R}^d$  bzw  $\mathbb{R}^m$  (oder, allgemeiner, irgendwelche metrischen oder gar topologischen Räume), dann sind stetige Abbildung  $f : X \rightarrow Y$  immer  $\mathcal{B}_X$ - $\mathcal{B}_Y$ -messbar.

Denn: bekanntlich ist  $f$  stetig genau wenn das Urbild jeder offenen Menge offen ist, dh wenn  $f^{-1}\mathcal{O}_Y \subseteq \mathcal{O}_X$ . Also haben wir  $f^{-1}\mathcal{O}_Y \subseteq \mathcal{O}_X \subseteq \sigma(\mathcal{O}_X) = \mathcal{B}_X$ . Da  $\mathcal{O}_Y$  die  $\sigma$ -Algebra  $\mathcal{B}_Y$  erzeugt, folgt die Behauptung.

**Beispiel 2.15 (Reelle Funktionen)** a) Ist  $(X, \mathcal{A})$  beliebig, so finden wir

$$f : X \rightarrow \overline{\mathbb{R}} \quad \text{ist } (\mathcal{A}-\mathcal{B}_{\overline{\mathbb{R}}})\text{-messbar} \Leftrightarrow \{f \leq b\} \in \mathcal{A} \quad \text{für jedes } b \in \mathbb{R}.$$

Denn  $\{f \leq b\} = f^{-1}[-\infty, b]$ , und  $\mathcal{E} := \{[-\infty, b] : b \in \mathbb{R}\}$  erzeugt  $\mathcal{B}_{\overline{\mathbb{R}}}$ . Alternativ kann man auch überprüfen ob alle Mengen der Form  $\{f < b\}$  in  $\mathcal{A}$  liegen, weil auch  $\{[-\infty, b) : b \in \mathbb{R}\}$  die  $\sigma$ -Algebra  $\mathcal{B}_{\overline{\mathbb{R}}}$  erzeugt.

b) Als Spezialfall halten wir fest, dass jede monotone Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  auch  $\mathcal{B}_{\mathbb{R}}$ - $\mathcal{B}_{\mathbb{R}}$ -messbar ist. (Wir wissen ja, dass  $\mathcal{B}_{\mathbb{R}}$  von der Familie aller Intervalle erzeugt wird, und Monotonie bedeutet, dass das Urbild jedes Intervalles wieder ein Intervall ist. Überprüfen Sie das bitte.)

**Beispiel 2.16 ( $\mathbb{R}^d$ -wertige Funktionen)** Ist  $(X, \mathcal{A})$  beliebig und  $f : X \rightarrow \mathbb{R}^d$ , können wir  $f$  als Tupel  $f = (f_1, \dots, f_d)$  reeller Funktionen  $f_j : X \rightarrow \mathbb{R}$  auffassen. Da deren Messbarkeit eben besprochen wurde, ist folgende Aussage nützlich:

$$\begin{aligned} f = (f_1, \dots, f_d) : X \rightarrow \mathbb{R}^d \text{ ist } (\mathcal{A}-\mathcal{B}_{\mathbb{R}^d})\text{-messbar} \\ \Leftrightarrow \text{jedes } f_j : X \rightarrow \mathbb{R} \text{ ist } (\mathcal{A}-\mathcal{B}_{\mathbb{R}})\text{-messbar.} \end{aligned}$$

Um das zu zeigen, drücken wir die Komponenten  $f_j$  von  $f$  mittels der kanonischen Projektionen  $\pi_j : \mathbb{R}^d \rightarrow \mathbb{R}$  (mit  $\pi_j(x_1, \dots, x_d) = x_j$ ) als  $f_j = \pi_j \circ f$  aus. Weil diese stetig, und daher nach dem Beispiel oben  $\mathcal{B}_{\mathbb{R}^d}$ - $\mathcal{B}_{\mathbb{R}}$ -messbar sind, erkennen wir, dass aus der Messbarkeit von  $f$  sofort jene der  $f_j$  folgt (Proposition 2.9 a)).

Falls umgekehrt alle  $f_j$  messbar sind, überprüfen wir, dass  $f^{-1}\mathcal{I}_{\mathbb{R}}^d \subseteq \mathcal{A}$  gilt, woraus Messbarkeit von  $f$  folgt, weil  $\sigma(\mathcal{I}_{\mathbb{R}}^d) = \mathcal{B}_{\mathbb{R}^d}$ . Dazu schreiben wir Elemente von  $\mathcal{I}_{\mathbb{R}}^d$  mithilfe der  $\pi_j$  als

$$I_1 \times \dots \times I_d = \bigcap_{j=1}^d \pi_j^{-1} I_j \in \mathcal{I}_{\mathbb{R}}^d \quad \text{für } I_1, \dots, I_d \in \mathcal{I}_{\mathbb{R}},$$

sodass

$$\begin{aligned} f^{-1}(I_1 \times \dots \times I_d) &= f^{-1}\left(\bigcap_{j=1}^d \pi_j^{-1} I_j\right) = \bigcap_{j=1}^d f^{-1}(\pi_j^{-1} I_j) \\ &= \bigcap_{j=1}^d (\pi_j \circ f)^{-1} I_j = \bigcap_{j=1}^d f_j^{-1} I_j, \end{aligned}$$

wobei  $f_j^{-1} I_j \in \mathcal{A}$  wegen Messbarkeit der  $f_j$ . Also gilt  $f^{-1}(I) \in \mathcal{A}$  für alle  $I = I_1 \times \dots \times I_d \in \mathcal{I}_{\mathbb{R}}^d$ .

Hat man erst einmal die Messbarkeit reeller Funktionen überprüft, so kann kaum noch etwas passieren. Man darf mit diesen dann nämlich entspannt rechnen, ohne diese Eigenschaft wieder zu verlieren. Dabei ist es praktisch, gleich Funktionen  $f, g : X \rightarrow \overline{\mathbb{R}}$  zu betrachten, weil bei Grenzübergängen auch unendliche Werte auftreten können. Interessiert man sich dann zB für  $f + g$ , so muss man aber natürlich Situationen ausschliessen, in denen etwa Punkte  $x$  mit  $f(x) = \infty = -g(x)$  vorkommen, dort ist  $f + g$  ja nicht definiert.

**Proposition 2.11 (Rechnen mit messbaren Funktionen)** *Es seien  $(X, \mathcal{A})$  ein messbarer Raum,  $A \in \mathcal{A}$ , und  $f, g : X \rightarrow \overline{\mathbb{R}}$  zwei  $(\mathcal{A}, \mathcal{B}_{\overline{\mathbb{R}}})$ -messbare Funktionen. Dann gilt:*

- a) Die Funktionen  $1_A f$ ,  $f + g$ ,  $f \cdot g$ ,  $f/g$ ,  $\max(f, g)$ ,  $\min(f, g)$  sind messbar (falls definiert).
- b) Die Mengen  $\{f \leq g\}$ ,  $\{f < g\}$ ,  $\{f = g\}$  sind messbar.

**Beweis.** a) Wir beginnen mit  $1_A f$ , siehe Proseminar

Nun sehen wir uns nun den Fall  $f + g$  genau an. In den anderen Fällen kann man ähnlich vorgehen. Zuerst setzen wir voraus, dass  $f$  und  $g$  nur reelle Werte annehmen, und drücken  $f + g$  als Komposition zweier simpler Funktionen aus: Nach Beispiel 2.16 ist die Funktion

$$F : X \rightarrow \mathbb{R}^2 \quad \text{mit} \quad F(x) := (f(x), g(x))$$

$\mathcal{A}$ - $\mathcal{B}_{\mathbb{R}^2}$ -messbar. Weiters ist die stetige Funktion

$$G : \mathbb{R}^2 \rightarrow \mathbb{R} \quad \text{mit} \quad G(s, t) := s + t$$

automatisch  $\mathcal{B}_{\mathbb{R}^2}$ - $\mathcal{B}_{\mathbb{R}}$ -messbar (Beispiel 2.14). Insgesamt, wegen Proposition 2.9 a), ist dann auch  $f + g = G \circ F$  wirklich  $\mathcal{A}$ - $\mathcal{B}_{\mathbb{R}}$ -messbar.

Im allgemeinen Fall  $f, g : X \rightarrow \overline{\mathbb{R}}$  sei  $A := \{f \in \mathbb{R}\} \cap \{g \in \mathbb{R}\} \in \mathcal{A}$ . Es sind auch  $\{f = -\infty\}, \{g = -\infty\} \in \mathcal{A}$ . Die Funktionen  $1_A f$  und  $1_A g$  sind reellwertig und messbar, nach dem vorangehenden Argument, ist also  $1_A(f + g)$  messbar. Deshalb ist für jedes  $b \in \mathbb{R}$ ,

$$\{f + g \leq b\} = \{f = -\infty\} \cup \{g = -\infty\} \cup (A \cap \{1_A(f + g) \leq b\}) \in \mathcal{A},$$

also  $f + g$  messbar nach Beispiel 2.15.

b) Es ist  $\{f \leq g\} = \{f - g \leq 0\} \in \mathcal{A}$ , da  $f - g$  messbar. Entsprechend für  $\{f < g\}$ . Schließlich ist  $\{f = g\} = \{f \leq g\} \cap \{g \leq f\} \in \mathcal{A}$ . ■

**Beispiel 2.17 (Positiv- und Negativteil von  $f$ )** Als wichtige Anwendung sehen wir uns, für  $f : X \rightarrow \overline{\mathbb{R}}$  den Positiv- und Negativteil von  $f$  an dh die minimalen nichtnegativen Funktionen  $f^+ := \max(f, 0)$  und  $f^- := \max(-f, 0)$  mit

$$f = f^+ - f^-.$$

Der Proposition zeigt: Ist  $f$  messbar, dann auch  $f^+, f^-$  und  $|f| = f^+ + f^-$ .

Es kommt aber noch viel besser: Messbarkeit von Funktionen  $f : X \rightarrow \overline{\mathbb{R}}$  ist tatsächlich eine ungewöhnlich robuste Eigenschaft! Sie verträgt sich auch mit Grenzübergängen, und erlaubt es daher, ohne Verrenkungen Analysis zu betreiben. Während etwa die Grenzfunktion einer punktweise konvergenten Folge stetiger reeller Funktionen keineswegs stetig zu sein braucht, hält die Messbarkeit solche Grenzübergänge ohne weiteres aus. Am allerbesten aber ist, dass man das auch ganz leicht beweisen kann.

**Proposition 2.12 (Messbarkeit punktwieser Suprema und Limiten)** Es seien  $(X, \mathcal{A})$  ein messbarer Raum,  $A \in \mathcal{A}$ , und  $f, f_n (n \geq 1) : X \rightarrow \overline{\mathbb{R}}$  messbare Funktionen. Dann sind auch die (punktweise definierten) Funktionen

$$\sup_{n \geq 1} f_n, \inf_{n \geq 1} f_n, \overline{\lim}_{n \rightarrow \infty} f_n, \underline{\lim}_{n \rightarrow \infty} f_n$$

messbar. Falls  $(f_n)_{n \geq 1}$  punktweise konvergiert, so ist auch  $\lim_{n \rightarrow \infty} f_n$  messbar.

**Beweis.** Für  $b \in \mathbb{R}$  ist

$$\begin{aligned} \left\{ \sup_{n \geq 1} f_n \leq b \right\} &= \left\{ x \in X : \sup_{n \geq 1} f_n(x) \leq b \right\} = \{ x \in X : f_n(x) \leq b \text{ für alle } n \geq 1 \} \\ &= \{ x \in X : x \in \{ f_n \leq b \} \text{ für alle } n \geq 1 \} = \bigcap_{n \geq 1} \{ f_n \leq b \} \in \mathcal{A}, \end{aligned}$$

also  $\sup_{n \geq 1} f_n$  messbar laut Beispiel 2.15. Entsprechend für  $\inf_{n \geq 1} f_n$ . Weiters ist (weil die  $g_n := \sup_{k \geq n} f_k$  eine fallende Folge bilden)

$$\overline{\lim}_{n \rightarrow \infty} f_n = \limsup_{n \rightarrow \infty} f_n := \lim_{n \rightarrow \infty} \left( \sup_{k \geq n} f_k \right) = \inf_{n \geq 1} \left( \sup_{k \geq n} f_k \right).$$

Aus der eben gezeigten Messbarkeit punktwieser Suprema und Infima folgt somit jene von  $\overline{\lim}_{n \rightarrow \infty} f_n$ . Analog für  $\underline{\lim}_{n \rightarrow \infty} f_n$ . Falls der punktweise Grenzwert  $\lim_{n \rightarrow \infty} f_n$  existiert, stimmt er natürlich mit  $\lim_{n \rightarrow \infty} f_n$  überein. ■

# 3. Das Lebesgue-Integral

Interpretiert man das Integral  $\int_a^b f(x) dx$  einer (zB stetigen) Funktion  $f : [a, b] \rightarrow [0, \infty)$  als Fläche unter dem Graphen, so wird sofort klar, dass die Frage nach einer geeigneten Definition des Integrals wilderer Funktionen engstens mit der Maßtheorie zusammenhängt. Wie dort (auf Ebene der Mengen) ist hier (auf Ebene der Funktionen) für gewisse einfache Situationen klar, wie man den gewünschten Größenbegriff zu verstehen hat, und man bemüht sich dann, dies zu einer nützlichen Definition für eine große Familie von Mengen bzw Funktionen auszudehnen.

## 3.1 Die Grundidee - Einfache Funktionen

**Als Aufwärmübung erinnern wir uns zuerst an das Riemann-Integral.** Die Grundbausteine dort sind die Indikatorfunktionen  $1_I$  von Teilintervallen  $I \subseteq [a, b]$ , bzw ihre endlichen Linearkombinationen, die *Intervall-Treppenfunktionen*  $g = \sum_{k=1}^n \alpha_k 1_{I_k}$  wobei  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$  sind und die Teilintervalle  $I_1, \dots, I_n$  eine Partition  $\mathcal{Z}$  von  $[a, b]$  bilden. Der Inhalt der Fläche unter dem Graphen ist offensichtlich

$$\int_a^b g(x) dx := \sum_{k=1}^n \alpha_k \lambda(I_k).$$

Zu einer Partition  $\mathcal{Z}$  wie oben seien  $|\mathcal{Z}| := \max_{1 \leq k \leq n} \lambda(I_k) = \max_{I \in \mathcal{Z}} \lambda(I)$  die Länge des größten Zerlegungsintervalles, und weiters

$$g(f, \mathcal{Z}) := \sum_{k=1}^n \left( \inf_{I_k} f \right) 1_{I_k} \quad \text{und} \quad G(f, \mathcal{Z}) := \sum_{k=1}^n \left( \sup_{I_k} f \right) 1_{I_k}$$

die kleinste bzw größte Intervall-Treppenfunktion zu  $\mathcal{Z}$  mit

$$g(f, \mathcal{Z}) \leq f \leq G(f, \mathcal{Z}).$$

Deren Integrale sind die *Unter-* bzw *Obersumme* zu  $f$  bzgl  $\mathcal{Z}$ ,

$$U(f, \mathcal{Z}) := \sum_{k=1}^n \left( \inf_{I_k} f \right) \lambda(1_{I_k}) = \int_a^b g(f, \mathcal{Z})(x) dx$$

und

$$O(f, \mathcal{Z}) := \sum_{k=1}^n \left( \sup_{I_k} f \right) \lambda(1_{I_k}) = \int_a^b G(f, \mathcal{Z})(x) dx.$$

In Darboux' Zugang zum R-Integral wählt man eine Folge  $(\mathcal{Z}_N)_{N \geq 1}$  immer feiner werdender Zerlegungen mit  $\lim_{N \rightarrow \infty} |\mathcal{Z}_N| = 0$ , und setzt

$$g_N := g(f, \mathcal{Z}_N), \quad G_N := G(f, \mathcal{Z}_N), \quad N \geq 1.$$

Dann können die  $g_N, G_N$  unser  $f$  evtl immer besser approximieren, und es gilt

$$g_1 \leq g_2 \leq \dots \leq f \leq \dots \leq G_2 \leq G_1.$$

Aufgrund dieser Monotonie ist klar, dass  $\lim_{N \rightarrow \infty} U(f, \mathcal{Z}_N) \leq \lim_{N \rightarrow \infty} O(f, \mathcal{Z}_N)$  punktweise existieren, und bekanntlich heißt  $f$  *Riemann-integrierbar* wenn diese Grenzwerte übereinstimmen, ihr gemeinsamer Wert ist dann das *Riemann-Integral*

$\int_a^b f(x) dx$  von  $f$ .

**Allerdings** funktioniert das nur für relativ gutmütige Funktionen, schon einfache Gegenbeispiele zeigen, dass im allgemeinen von guter Approximation durch die  $g_N$  und  $G_N$  nicht die Rede sein kann: Für  $f := 1_{\mathbb{Q} \cap [a,b]}$  etwa ist ja stets  $U(f, \mathcal{Z}_N) = 0$  und  $O(f, \mathcal{Z}_N) = b - a$ , der Fall ist aus Sicht der Riemannschen Theorie also hoffnungslos. Die Auswahl an Bausteinen (nur die Intervall-Treppenfunktion) ist viel zu klein um solche Funktionen gut (etwa in dem Sinne, dass  $g_N \nearrow f$  punktweise) approximieren zu können.

**Andererseits** ist  $f = 1_{\mathbb{Q} \cap [a,b]}$  als Indikatorfunktion  $1_M$  doch auch wieder ziemlich einfach, ganz allgemein könnte man doch den Inhalt der Fläche unter  $1_M$  definieren als Höhe (= 1) mal Länge der Grundmenge  $M$ , sofern die zweite Größe vernünftig definiert ist. Nachdem wir diese Frage schon ausführlich behandelt haben, sehen wir, dass dieser Definitionsvorschlag für jede Menge  $M$  aus  $\mathcal{L}_{[a,b]}$  Sinn macht, insbesondere auch für unser  $f$ , wo wir  $1 \cdot \lambda(\mathbb{Q} \cap [a,b]) = 0$  erhalten. Die Maßtheorie erlaubt uns also, einen ganz naheliegenden Integralbegriff für eine viel größere Familie von Indikatorfunktionen (und dann auch deren endlichen Linearkombinationen) einzuführen. Diese Familie ist so groß, dass wir im weiteren so gut wie alle in der Praxis auftretenden Funktionen gut approximieren können.

**Als Bausteine der Lebesgueschen Integrationstheorie** dienen uns nun folgende besonders simple messbare Funktionen.

**Definition 3.1** Sei  $(X, \mathcal{A})$  ein messbarer Raum. Eine einfache Funktion (oder Treppenfunktion) ist eine messbare Funktion  $f : X \rightarrow \mathbb{R}$  welche nur endlich viele Werte annimmt.  $\mathcal{T}$  bezeichne die Familie aller einfachen Funktionen,  $\mathcal{T}^+ \subseteq \mathcal{T}$  jene der nichtnegativen.

**Proposition 3.1 (Einfache Funktionen)** Sei  $(X, \mathcal{A})$  ein messbarer Raum.

**a)**  $f : X \rightarrow \mathbb{R}$  ist genau dann eine einfache Funktion, wenn  $n \in \mathbb{N}$ ,  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ , und  $A_1, \dots, A_n \in \mathcal{A}$  existieren mit

$$f = \sum_{k=1}^n \alpha_k 1_{A_k}. \quad (3.1)$$

Dabei können die  $A_k$  als Partition von  $X$  gewählt werden (Normaldarstellung).

**b)**  $\mathcal{T}$  ist ein reeller Vektorraum. Sind  $f, g$  einfache Funktionen, dann auch  $|f|$ ,  $f \cdot g$ ,  $f/g$  (falls definiert),  $\min(f, g)$ , und  $\max(f, g)$ . (Insbesondere also auch  $f^+$ ,  $f^-$ , und  $f 1_A$  für  $A \in \mathcal{A}$ .)

**Beweis.** **a)** Für  $f \in \mathcal{T}$  seien  $\alpha_1, \dots, \alpha_n$  die unterschiedlichen Werte von  $f$  und  $A_k := \{f = \alpha_k\}$ . Dann bilden die  $A_k$ ,  $1 \leq k \leq n$ , eine Partition von  $X$ , und da  $f$  messbar ist, sind die  $A_k$  in  $\mathcal{A}$ . Evident gilt (3.1) und wir haben eine Normaldarstellung gefunden.

Ist umgekehrt  $f$  eine Funktion der Form (3.1), dann ist  $f$  messbar (da jedes  $\alpha_k 1_{A_k}$  eine messbare Funktion ist), und als Werte nimmt  $f$  Summen der Form  $\alpha_{k_1} + \dots + \alpha_{k_j}$  (mit  $k_1 < \dots < k_j$ ) an. Dafür gibt es aber nur endlich viele Möglichkeiten.

**b)** Seien  $f, g \in \mathcal{T}$ ,  $f = \sum_{k=1}^n \alpha_k 1_{A_k}$  und  $g = \sum_{j=1}^m \beta_j 1_{B_j}$  in Normaldarstellung. Da  $f$  auf den zu  $g$  gehörigen Partitions Mengen  $B_j$  nicht konstant zu sein braucht (und entsprechend für  $g$  und  $A_k$ ), gehen wir zur *gemeinsamen Verfeinerung* der beiden Partitionen über, dh zu  $\{A_k \cap B_j : 1 \leq k \leq n, 1 \leq j \leq m\}$ . Dann finden



wir

$$\begin{aligned} f + g &= \sum_{k=1}^n \alpha_k \sum_{j=1}^m 1_{A_k \cap B_j} + \sum_{j=1}^m \beta_j \sum_{k=1}^n 1_{A_k \cap B_j} \\ &= \sum_{k=1}^n \sum_{j=1}^m (\alpha_k + \beta_j) 1_{A_k \cap B_j}, \end{aligned}$$

klarerweise die Normaldarstellung einer Treppenfunktion. Also ist  $f + g \in \mathcal{T}$ . Banales niederschreiben der Definition zeigt, dass auch  $a \cdot f \in \mathcal{T}$  für jedes  $a \in \mathbb{R}$ , also ist  $\mathcal{T}$  ein Vektorraum über  $\mathbb{R}$ . Ganz analog bekommt man zB

$$f \cdot g = \sum_{k=1}^n \sum_{j=1}^m (\alpha_k \cdot \beta_j) 1_{A_k \cap B_j}.$$

Die verbleibenden Aussagen können Sie leicht selbst beweisen. ■

Wir überzeugen uns nun davon, dass wir jede messbare Funktion durch einfache Funktionen punktweise approximieren können. Das Hauptsätzchen über messbare Funktionen lautet so:

**Proposition 3.2 (Approximation nichtnegativer messbarer Funktionen)**  
Es sei  $(X, \mathcal{A})$  ein messbarer Raum, und  $f : X \rightarrow [0, \infty]$  messbar. Dann gibt es eine Folge  $(f_n)_{n \geq 1}$  in  $\mathcal{T}^+$  mit  $f_n \nearrow f$  (punktweise), d.h. für alle  $x \in X$  ist  $0 \leq f_1(x) \leq f_2(x) \leq \dots$  und  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ .

Den einfachen Beweis bereiten wir durch ein noch einfacheres Lemma vor, es ist nämlich nützlich dieses explizit zur Verfügung zu haben. Ausserdem können Sie es ganz leicht selbst beweisen (was Sie auch wirklich tun sollten).

**Lemma 3.1** Für jedes  $n \geq 1$  ist die Funktion  $d^{(n)} : [0, \infty] \rightarrow [0, \infty]$  mit

$$d^{(n)}(t) := \begin{cases} 0 & \text{für } t = 0, \\ (j-1)2^{-n} & \text{für } (j-1)2^{-n} < t \leq j2^{-n} \leq n, \\ n & \text{für } t > n. \end{cases}$$

Borel-messbar mit endlich vielen Werten und  $0 \leq d^{(n)} \leq n$ . Es gilt

$$d^{(n)}(t) \nearrow t \quad \text{für jedes } t \in [0, \infty]. \quad (3.2)$$

Damit bekommen wir sehr leicht den

**Beweis von Proposition 3.2.** Es sei  $f_n := d^{(n)} \circ f$ ,  $n \in \mathbb{N}$ . Als Komposition messbarer Funktionen ist  $f_n$  wieder messbar, und weil  $d^{(n)}$  nur endlich viele Werte annimmt, gilt dies auch für  $f_n$ , sodass also  $f_n \in \mathcal{T}^+$ . Aus (3.2) folgt nun sofort dass  $f_n(x) = d^{(n)}(f(x)) \nearrow f(x)$  für jedes  $x \in X$ . ■

## 3.2 Definition des Integrals

In diesem Abschnitt definieren wir für eine sehr grosse Familie messbarer Funktionen  $f : X \rightarrow \mathbb{R}$  auf einem (beliebigen) festen Maßraum  $(X, \mathcal{A}, \mu)$  das (*Lebesgue-*) *Integral von  $f$  über  $X$  bezüglich  $\mu$* , geschrieben als

$$\int_X f \, d\mu = \int_X f(x) \, d\mu(x) = \int_X f(x) \, \mu(dx) \in \overline{\mathbb{R}}.$$

(Wenn klar ist um welchen Maßraum es sich handelt schreibt man auch kurz  $\int f \, d\mu$  oder einfach  $\int f$ .) Die Definition erfolgt in drei Schritten:

- A) für Treppenfunktionen  $f \in \mathcal{T}^+$
- B) für messbare  $f : X \rightarrow [0, \infty]$
- C) für messbare  $f : X \rightarrow \overline{\mathbb{R}}$  mit  $\int f^+$  oder  $\int f^-$  endlich

wobei der zweite Schritt der eigentlich interessante und spannende ist.

**A) Integrale von Treppenfunktionen.** Wie angekündigt, vereinbart man:

**Definition 3.2** Für  $f \in \mathcal{T}^+$  mit Normaldarstellung  $f = \sum_{k=1}^n \alpha_k 1_{A_k}$  ist

$$\int_X f d\mu := \sum_{k=1}^n \alpha_k \cdot \mu(A_k) \in [0, \infty].$$

**Lemma 3.2** Dies ist tatsächlich eine Definition, dh der Wert hängt nicht von der Wahl der Normaldarstellung ab.

**Beweis.** Sei  $f = \sum_{j=1}^m \beta_j 1_{B_j}$  eine weitere Normaldarstellung. Indem wir wie gewohnt zur gemeinsamen Verfeinerung der beiden Partitionen übergehen erhalten wir

$$\sum_{k=1}^n \alpha_k \mu(A_k) = \sum_{k=1}^n \sum_{j=1}^m \alpha_k \cdot \mu(A_k \cap B_j)$$

und

$$\sum_{j=1}^m \beta_j \mu(B_j) = \sum_{j=1}^m \sum_{k=1}^n \beta_j \cdot \mu(A_k \cap B_j).$$

Zu zeigen ist, dass diese Summen übereinstimmen. Relevant sind dabei nur Paare von Indices  $k, j$  mit  $\mu(A_k \cap B_j) > 0$ . In diesem Fall ist  $A_k \cap B_j \neq \emptyset$  und für jedes  $x \in A_k \cap B_j$  gilt  $\alpha_k = f(x) = \beta_j$ . ■

Der so gewonnene simple Integralbegriff für Treppenfunktionen besitzt vernünftige Grundeigenschaften:

**Proposition 3.3 (Linearität und Monotonie)** Seien  $f, g \in \mathcal{T}^+$ ,  $a \geq 0$ . Dann gilt:

$$a) \int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

$$b) \int_X (a \cdot f) d\mu = a \cdot \int_X f d\mu.$$

$$c) f \geq g \implies \int_X f d\mu \geq \int_X g d\mu.$$

**Beweis.** a) Wie gesehen gibt es Normaldarstellungen von  $f, g, f + g$  in der Form

$$f = \sum_{k=1}^n \sum_{j=1}^m \alpha_k \cdot 1_{A_k \cap B_j}, \quad g = \sum_{k=1}^n \sum_{j=1}^m \beta_j \cdot 1_{A_k \cap B_j}, \quad f + g = \sum_{k=1}^n \sum_{j=1}^m (\alpha_k + \beta_j) \cdot 1_{A_k \cap B_j}.$$

Daher ist einfach

$$\begin{aligned} \int_X (f + g) d\mu &= \sum_{k=1}^n \sum_{j=1}^m (\alpha_k + \beta_j) \cdot \mu(A_k \cap B_j) \\ &= \sum_{k=1}^n \sum_{j=1}^m \alpha_k \cdot \mu(A_k \cap B_j) + \sum_{k=1}^n \sum_{j=1}^m \beta_j \cdot \mu(A_k \cap B_j) \\ &= \int_X f d\mu + \int_X g d\mu. \end{aligned}$$

b) Noch viel einfacher.

c) Wir haben  $f = g + (f - g)$  mit  $f - g \in \mathcal{T}^+$ , daher nach a)

$$\int_X f d\mu = \int_X g d\mu + \int_X (f - g) d\mu \geq \int_X g d\mu.$$

■

**B) Integrale nichtnegativer messbarer Funktionen.** Mittels Proposition 3.2 erklären wir nun:

**Definition 3.3** Für messbares  $f : X \rightarrow [0, \infty]$  sei  $(f_n)_{n \geq 1}$  eine Folge in  $\mathcal{T}^+$  mit  $f_n \nearrow f$ . Dann ist

$$\int_X f d\mu := \lim_{n \rightarrow \infty} \int_X f_n d\mu \in [0, \infty].$$

**Lemma 3.3** Dies ist tatsächlich eine Definition, dh der Wert hängt nicht von der Wahl der Folge  $(f_n)$  ab.

**Beweis.** (i) Wir zeigen zuerst: Sind  $g_1 \leq g_2 \leq \dots$  in  $\mathcal{T}^+$ , und  $f \in \mathcal{T}^+$  mit  $f \leq \lim_{n \rightarrow \infty} g_n$  (punktweise), so folgt

$$\int_X f d\mu \leq \lim_{n \rightarrow \infty} \int_X g_n d\mu.$$

Dazu sei  $f = \sum_{k=1}^m \alpha_k 1_{A_k}$  in Normaldarstellung, und  $\varepsilon > 0$ . Definieren wir  $B_n := \{(1 + \varepsilon)g_n \geq f\} \in \mathcal{A}$ ,  $n \geq 1$ , so gilt  $B_n \nearrow X$ . Insbesondere gilt für jedes  $k \in \{1, \dots, m\}$ , dass  $A_k \cap B_n \nearrow A_k$  und somit (Stetigkeit des Maßes von unten)

$$\lim_{n \rightarrow \infty} \mu(A_k \cap B_n) = \mu(A_k).$$

Daher folgt

$$\begin{aligned} \int_X f d\mu &= \sum_{k=1}^m \alpha_k \cdot \mu(A_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^m \alpha_k \cdot \mu(A_k \cap B_n) \\ &= \lim_{n \rightarrow \infty} \int_X \left( \sum_{k=1}^m \alpha_k \cdot 1_{A_k} \right) \cdot 1_{B_n} d\mu \\ &= \lim_{n \rightarrow \infty} \int_X f \cdot 1_{B_n} d\mu \\ &\leq \lim_{n \rightarrow \infty} \int_X (1 + \varepsilon)g_n d\mu \\ &= (1 + \varepsilon) \lim_{n \rightarrow \infty} \int_X g_n d\mu, \end{aligned}$$

und, da  $\varepsilon > 0$  beliebig war, unsere Behauptung.

(ii) Nun sei  $(g_k)_{k \geq 1}$  eine weitere Folge in  $\mathcal{T}^+$  mit  $g_k \nearrow f$ . Zu zeigen ist

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \lim_{k \rightarrow \infty} \int_X g_k d\mu.$$

Für jedes  $k$  ist  $f_k \leq \lim_{n \rightarrow \infty} f_n = \lim_{n \rightarrow \infty} g_n$ , und nach (i) daher

$$\int_X f_k d\mu \leq \lim_{n \rightarrow \infty} \int_X g_n d\mu.$$

Somit gilt auch

$$\lim_{k \rightarrow \infty} \int_X f_k d\mu \leq \lim_{n \rightarrow \infty} \int_X g_n d\mu,$$

und entsprechend wenn wir die Rollen von  $(f_n)$  und  $(g_k)$  vertauschen. ■

**Bemerkung 3.1** Das Lemma - und damit die Definition - stützt sich ganz wesentlich auf die Stetigkeit des Maßes, bzw die  $\sigma$ -Additivität. Für nicht  $\sigma$ -additive Mengenfunktionen liefert obiger Ansatz tatsächlich keine Definition!

Wir überprüfen wieder die Grundeigenschaften des nun wesentlich erweiterten Integralkonzeptes.

**Proposition 3.4 (Linearität und Monotonie)** Sind  $f, g : X \rightarrow [0, \infty]$  messbar, und  $a \geq 0$ , dann gelten

$$a) \int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

$$b) \int_X (a \cdot f) d\mu = a \cdot \int_X f d\mu.$$

$$c) f \geq g \implies \int_X f d\mu \geq \int_X g d\mu.$$

**Beweis.** a)+b) Seien  $(f_n)$  und  $(g_n)$  Folgen in  $\mathcal{T}^+$  mit  $f_n \nearrow f$  und  $g_n \nearrow g$ . Dann gilt  $af_n + g_n \in \mathcal{T}^+$  für alle  $n$  und  $af_n + g_n \nearrow af + g$ . Also ist

$$\begin{aligned} \int_X (af + g) d\mu &= \lim_{n \rightarrow \infty} \int_X (af_n + g_n) d\mu \\ &= a \lim_{n \rightarrow \infty} \int_X f_n d\mu + \lim_{n \rightarrow \infty} \int_X g_n d\mu = a \int_X f d\mu + \int_X g d\mu. \end{aligned}$$

c) Wie zuvor. ■

**C) (Quasi-) integrierbare Funktionen.** Der Rest ist harmlos:

**Definition 3.4** Ist  $(X, \mathcal{A}, \mu)$  ein Maßraum, dann heißt eine Funktion  $f : X \rightarrow \overline{\mathbb{R}}$  integrierbar bzw quasi-integrierbar falls sie messbar ist mit

$$\int_X f^+ d\mu < \infty \text{ und } \int_X f^- d\mu < \infty \quad \text{bzw} \quad \int_X f^+ d\mu < \infty \text{ oder } \int_X f^- d\mu < \infty.$$

In beiden Fällen definiert man

$$\int_X f d\mu := \int_X f^+ d\mu - \int_X f^- d\mu \in \overline{\mathbb{R}}.$$

Falls  $\int_X f^+ d\mu = \infty$  und  $\int_X f^- d\mu = \infty$ , so bleibt  $\int_X f d\mu$  undefiniert. Die Familie aller integrierbaren reellen Funktionen wird mit  $\mathcal{L}_1(X, \mathcal{A}, \mu) = \mathcal{L}_1(\mu)$  bezeichnet. Schließlich vereinbaren wir noch: Für  $Y \in \mathcal{A}$  sei

$$\int_Y f d\mu := \int_X 1_Y f d\mu,$$

das Integral von  $f$  (bzgl  $\mu$ ) über  $Y$ , wann immer die rechte Seite definiert ist.

**Bemerkung 3.2** Eine Funktion  $f : X \rightarrow \overline{\mathbb{R}}$  ist genau dann integrierbar, wenn sie messbar ist und  $\int_X |f| d\mu < \infty$ . (Denn:  $|f| = f^+ + f^-$ .)

Natürlich übertragen sich die bekannten Grundeigenschaften des Integrals:

**Proposition 3.5 (Linearität und Monotonie)** Für Funktionen  $f, g : X \rightarrow \overline{\mathbb{R}}$  gilt:

$$a) f \text{ (quasi-)integrierbar und } a \in \mathbb{R} \implies a \cdot f \text{ (quasi-)integrierbar und } \int_X (a \cdot f) d\mu = a \cdot \int_X f d\mu.$$

$$b) f, g \text{ integrierbar und } f + g \text{ definiert} \implies$$

$$f + g \text{ integrierbar und } \int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

$$c) f, g \text{ (quasi-)integrierbar und } f \geq g \implies \int_X f d\mu \geq \int_X g d\mu.$$

$$d) f \text{ (quasi-)integrierbar} \implies \left| \int_X f d\mu \right| \leq \int_X |f| d\mu.$$

**Beweis.** Das schaffen Sie selbst ! ■

**D) Nullmengen.** Schon vom Riemann -Integral wissen wir, dass kleine Änderungen an einer Funktion deren Integral nicht verändern, zB wenn man die Funktion an nur endlich vielen Punkten modifiziert. Entsprechendes sollte wohl auch in unserem Rahmen gelten. Tatsächlich sind es genau die  $(\mu\text{-})$ Nullmengen (die  $A \in \mathcal{A}$  mit  $\mu(A) = 0$ ) über die das Integral großzügig hinwegsieht. Wir führen eine praktische Sprechweise ein:

**Definition 3.5** Es sei  $(X, \mathcal{A}, \mu)$  ein Maßraum. Man sagt, eine Eigenschaft gelte  $(\mu\text{-})$  fast überall (f.ü.) wenn sie außerhalb einer  $(\mu\text{-})$ Nullmenge gilt. ZB bedeuten " $f = g$  f.ü.", dass  $\{f \neq g\} \subseteq N \in \mathcal{A}$  mit  $\mu(N) = 0$  und " $f > g$  f.ü.", dass  $\{f \leq g\} \subseteq N \in \mathcal{A}$  mit  $\mu(N) = 0$ .

Damit können wir obige Behauptung präzisieren:

**Proposition 3.6 (Integral und Nullmengen)** Sind  $f, g : X \rightarrow [0, \infty]$  messbar, so gelten

$$a) \int_X f d\mu = 0 \iff f = 0 \text{ f.ü.}$$

$$b) \int_X f d\mu < \infty \implies f < \infty \text{ f.ü.}$$

$$c) f \geq g \text{ f.ü.} \implies \int_X f d\mu \geq \int_X g d\mu.$$

$$d) f = g \text{ f.ü.} \implies \int_X f d\mu = \int_X g d\mu.$$

**Beweis. a)** Sei  $A := \{f > 0\} \in \mathcal{A}$ . Wir nehmen an, dass  $f = 0$  f.ü., also  $\mu(A) = 0$ , und zeigen, dass dann für jedes  $g \in \mathcal{T}^+$  mit  $g \leq f$  das Integral verschwindet. Da  $\int_X f d\mu = \lim_{n \rightarrow \infty} \int_X f_n d\mu$  für eine Folge solcher  $g = f_n$ , folgt  $\int_X f d\mu = 0$ .

Ist  $g = \sum_{k=0}^n \alpha_k 1_{A_k}$  eine Normaldarstellung von  $g$ , dann ist für jedes  $k$  mit  $\mu(A_k) > 0$  notwendigerweise  $\alpha_k = 0$ , da sonst  $f \geq g = \alpha_k > 0$  auf  $A_k \subseteq A$  im Widerspruch zu  $\mu(A) = 0$  stünde. Daher ist tatsächlich  $\int_X g d\mu = 0$ .

Sei umgekehrt  $\int_X f d\mu = 0$ . Wir setzen  $A_n := \{f > 1/n\} \in \mathcal{A}$ ,  $n \geq 1$ , sodass  $A_n \nearrow A$ , also auch  $\mu(A_n) \nearrow \mu(A)$ . Nun ist

$$0 = \int_X f d\mu = \int_X 1_{A_n} f d\mu + \int_X 1_{A_n^c} f d\mu \geq \int_X 1_{A_n} f d\mu,$$

und, da  $1_{A_n} f \geq \frac{1}{n} 1_{A_n} \in \mathcal{T}^+$ ,

$$\int_X 1_{A_n} f d\mu \geq \frac{1}{n} \int_X 1_{A_n} d\mu = \frac{\mu(A_n)}{n},$$

also  $\mu(A_n) = 0$  für alle  $n$  und somit  $\mu(A) = 0$ .

**b)** Sei  $A := \{f = \infty\} \in \mathcal{A}$ , und  $f_n := n \cdot 1_A \in \mathcal{T}^+$ ,  $n \geq 1$ . Wegen  $f_n \leq f$  gilt

$$n \cdot \mu(A) = n \int_X 1_A d\mu = \int_X f_n d\mu \leq \int_X f d\mu < \infty$$

für alle  $n$ , und daher  $\mu(A) = 0$ .

c) Sei  $A := \{f < g\} \in \mathcal{A}$ , sodass  $\mu(A) = 0$ . Dann ist  $1_A f = 0 = 1_A g$  f.ü. und nach a) somit

$$\int_X 1_A f d\mu = \int_X 1_A g d\mu = 0.$$

Andererseits ist  $1_{A^c} f \geq 1_{A^c} g$  auf ganz  $X$ , und deshalb

$$\int_X 1_{A^c} f d\mu = \int_X 1_{A^c} g d\mu + \underbrace{\int_X 1_{A^c} (f - g) d\mu}_{\geq 0} \geq \int_X 1_{A^c} g d\mu.$$

Insgesamt erhalten wir daher

$$\begin{aligned} \int_X f d\mu &= \int_X 1_A f d\mu + \int_X 1_{A^c} f d\mu \\ &\geq \int_X 1_A g d\mu + \int_X 1_{A^c} g d\mu = \int_X g d\mu. \end{aligned}$$

d) Folgt sofort aus c). ■

### 3.3 Riemann-Integral und Unendliche Reihen

**A) Riemann-Integral.** Nachdem wir recht allgemein erklärt haben, was wir unter dem Integral einer auf einem Maßraum definierten messbaren Funktion verstehen wollen, sehen wir uns genauer an, was dies im wichtigsten Spezialfall, dem des Lebesgueschen Maßes  $\lambda$  auf einem Intervall, konkret bedeutet. Für (hinreichend gutmütige) Funktionen auf einem Intervall stünde uns ja auch der Riemannsche Integralbegriff zur Verfügung. Es seien  $a \leq b$  und  $f : [a, b] \rightarrow \mathbb{R}$  eine beschränkte Funktion. Wie hängen nun Existenz und (gegebenenfalls) die Werte des Riemann-Integrals  $\int_a^b f(x) dx$  und des Lebesgue-Integrals  $\int_{[a,b]} f d\lambda$  zusammen?

Das Integral  $\int_{[a,b]} f d\lambda$  ist das im vorigen Abschnitt definierte, bezogen auf den Maßraum  $([a, b], \mathcal{B}_{[a,b]}, \lambda)$  oder  $([a, b], \mathcal{L}_{[a,b]}, \lambda)$ . Wir erinnern uns daran, dass die Lebesguesche  $\sigma$ -Algebra  $\mathcal{L}_{[a,b]}$  etwas größer als  $\mathcal{B}_{[a,b]}$  ist, und die elementargeometrische Längenfunktion eben sogar zu einem eindeutig bestimmten Maß auf  $\mathcal{L}_{[a,b]}$  fortgesetzt werden kann. Eine auf  $([a, b], \mathcal{L}_{[a,b]}, \lambda)$  integrierbare Funktion nennen wir kurz *Lebesgue-integrierbar* auf  $[a, b]$ .

Der folgende Satz liefert eine sehr befriedigende Antwort auf unsere Frage.

**Satz 3.1 (Riemann versus Lebesgue)** *Es sei  $f : [a, b] \rightarrow \mathbb{R}$  eine beschränkte Funktion. Ist  $f$  R-integrierbar, dann auch L-integrierbar und es gilt*

$$\int_a^b f(x) dx = \int_{[a,b]} f d\lambda.$$

Diese Aussage bestätigt also unsere schon früher getroffene Behauptung, dass wir einen stärkeren, den Riemannschen umfassenden Integralbegriff entwickelt haben. Wegen der Übereinstimmung im Falle R-integrierbarer Funktionen brauchen wir - sobald der Satz bewiesen ist - in der Bezeichnung nicht mehr zwischen den beiden Konzepten zu unterscheiden: Sie können einen Ausdruck der Form  $\int_a^b f(x) dx$  zukünftig als L-Integral interpretieren, und man verwendet diese Notation eben auch wenn  $f$  nur L- und nicht unbedingt R-integrierbar ist.

**Beweis.** (i) Wir verwenden die zu Beginn des Kapitels eingeführte Notation. Natürlich sind Intervall-Treppenfunktionen spezielle einfache Funktionen (sie sind Borel-messbar), und per Definition stimmen ihre R- und L-Integrale überein. Insbesondere ist für jede Zerlegung  $\mathcal{Z}$  von  $[a, b]$  in Intervalle daher  $U(f, \mathcal{Z}) = \int_{[a,b]} g(f, \mathcal{Z}) d\lambda$  und  $O(f, \mathcal{Z}) = \int_{[a,b]} G(f, \mathcal{Z}) d\lambda$ .

Sei  $(\mathcal{Z}_N)_{N \geq 1}$  eine Folge feiner werdender Zerlegungen mit  $\lim_{N \rightarrow \infty} |\mathcal{Z}_N| = 0$ , und  $g_N := g(f, \mathcal{Z}_N)$ ,  $G_N := G(f, \mathcal{Z}_N)$ ,  $N \geq 1$ . Dann sind  $g_N, G_N \in \mathcal{T}^+$  für  $N \geq 1$ , und es gilt  $g_1 \leq g_2 \leq \dots \leq f \leq \dots \leq G_2 \leq G_1$ . Aufgrund der Monotonie existieren die (punktweisen) Grenzwerte

$$g := \lim_{N \rightarrow \infty} g_N \quad \text{und} \quad G := \lim_{N \rightarrow \infty} G_N,$$

und mit den  $g_N, G_N$  sind auch diese Funktionen Borel-messbar (siehe Proposition 2.12), mit

$$g \leq f \leq G$$

sowie (nach Definition des Integrals nichtnegativer messbarer Funktionen und, für  $G$ , Linearität - ergänzen Sie bitte die Details)

$$\int_{[a,b]} g d\lambda = \lim_{N \rightarrow \infty} U(f, \mathcal{Z}_N). \quad (3.3)$$

Dazu erinnern wir an die Definition des Integrals, wir haben ja  $g_N \in \mathcal{T}$  und  $g_N \nearrow g$ . Allerdings muss hier nichts  $\geq 0$  sein. Weil aber  $f$  beschränkt ist, dh  $|f| \leq \kappa$  für eine geeignete konstante  $\kappa \in (0, \infty)$ , können wir unsere Funktionen einfach in den positiven Bereich heben, und  $g_N + \kappa \nearrow g + \kappa$  mit  $g_N + \kappa \in \mathcal{T}^+$  verwenden. Das ergibt

$$\int_{[a,b]} g_N d\lambda + \kappa(b-a) = \int_{[a,b]} (g_N + \kappa) d\lambda \nearrow \int_{[a,b]} (g + \kappa) d\lambda = \int_{[a,b]} g d\lambda + \kappa(b-a),$$

also (3.3). Analog bekommen wir (wegen  $\kappa - G_N \nearrow \kappa - G$ )

$$\int_{[a,b]} G d\lambda = \lim_{N \rightarrow \infty} O(f, \mathcal{Z}_N).$$

Somit finden wir (wenn wir an die Wiederholung zu Beginn des Kapitels denken):

$$\begin{aligned} f \text{ ist R-integrierbar} &\iff \int_{[a,b]} g d\lambda = \int_{[a,b]} G d\lambda \\ &\quad \left( \text{dann} = \int_a^b f(x) dx \right) \\ &\iff g = G \text{ f.ü.} \end{aligned} \quad (3.4)$$

wobei die zweite Äquivalenz wegen  $g \leq G$  aus Proposition 3.6 folgt.

(ii) Nun nehmen wir  $f$  als R-integrierbar an. Der entscheidende Schritt ist der Beweis der Messbarkeit von  $f$ , dann ist ja  $\int_{[a,b]} f d\lambda$  definiert, und wegen  $g \leq f \leq G$  gilt

$$\int_{[a,b]} g d\lambda \leq \int_{[a,b]} f d\lambda \leq \int_{[a,b]} G d\lambda.$$

Mit (3.4) sehen wir dann sofort, dass wirklich  $\int_a^b f(x) dx = \int_{[a,b]} f d\lambda$ .

Nun folgt aus der R-Integrierbarkeit zwar nicht zwangsläufig, dass  $f$  Borel-messbar ist. (Obwohl, wie schon betont, fast jede im mathematischen Alltag auftretende Funktion dies erfüllt.) Wir zeigen, dass  $f$  aber in jedem Fall messbar bzgl der etwas größeren Lebesgueschen  $\sigma$ -Algebra  $\mathcal{L}_{[a,b]}$  ist. Dazu weisen wir nach:

$$\text{für jedes } c \in \mathbb{R} \text{ ist } \{f \leq c\} \in \mathcal{L}_{[a,b]}.$$

Sei  $c \in \mathbb{R}$  fest, dann gilt, wegen  $g \leq f \leq G$ ,

$$B_1 := \{G \leq c\} \subseteq \{f \leq c\} \subseteq \{g \leq c\} =: B_2.$$

Da  $g$  und  $G$  aber Borel-messbar sind (s.o.), liegen  $B_1$  und  $B_2$  in  $\mathcal{B}_{[a,b]}$ . Verwenden wir den zweiten Teil von (3.4), so sehen wir

$$\lambda(B_2 \setminus B_1) = \lambda(\underbrace{\{g \leq c\} \setminus \{G \leq c\}}_{\subseteq \{g \neq G\}}) = 0.$$

Nach Satz 1.6 gilt damit  $B \in \mathcal{L}_{\mathbb{R}}$  für alle  $B \subseteq B_2 \setminus B_1$ , also ist insbesondere  $\{f \leq c\} \setminus \{G \leq c\} \in \mathcal{L}_{\mathbb{R}}$ . Das bedeutet aber, dass auch

$$\{f \leq c\} = \{G \leq c\} \cup (\{f \leq c\} \setminus \{G \leq c\})$$

zu  $\mathcal{L}_{[a,b]}$  gehört. ■

Die Lebesguesche Theorie ermöglicht es übrigens auch, eine schöne und einleuchtende Charakterisierung der R-Integrierbarkeit anzugeben. Wir verzichten in der VO darauf, den Beweis im Detail zu besprechen, eine Skizze finden Sie hier.

**Satz 3.2 (Lebesgue-Kriterium für R-Integrierbarkeit)** *Es sei  $f : [a, b] \rightarrow \mathbb{R}$  eine beschränkte Funktion. Dann ist  $f$  R-integrierbar genau wenn  $f$  im Sinne des Lebesgue-Maßes fast überall stetig ist.*

**Beweis.** Wir verwenden Schritt (i) des vorigen Beweises. Definieren wir

$$M := \{x \in [a, b] : x \text{ ist ein Teilungspunkt für ein } \mathcal{Z}_N, N \geq 1\},$$

so ist  $M$  abzählbar und damit in  $\mathcal{B}_{[a,b]}$  mit  $\lambda(M) = 0$ . Das Verhalten von  $f$  auf  $M$  ist daher für unsere Zwecke irrelevant, es gilt ja

$$f \text{ stetig f.ü. auf } [a, b] \iff f \text{ stetig f.ü. auf } [a, b] \setminus M.$$

Erinnern wir uns an die Definition von  $g(x)$  und  $G(x)$  als Grenzwerte der Folgen  $(g_N(x))$  und  $(G_N(x))$  welche inf und sup von  $f$  in dem kleinen  $x$  enthaltenden Zerlegungsintervall angeben, so sieht man (überzeugen Sie sich bitte):

$$\text{Für } x \in [a, b] \setminus M \text{ gilt } g(x) = G(x) \text{ genau wenn } f \text{ stetig in } x \text{ ist.} \quad (3.5)$$

Daraus folgt unsere Behauptung. ■

**B) Summen und Unendliche Reihen.** Das allgemeine Integral umfasst aber auch gewöhnliche Summen und viele unendliche Reihen. Sei  $X = \{1, \dots, n\}$  oder  $\mathbb{N}^+$ ,  $\mathcal{A} = \mathcal{P}(X)$ ,  $\mu = \# = \sum_{x \in X} \delta_x = \text{Zählmaß}$ . Auf diesem messbaren Raum  $(X, \mathcal{A})$  sind alle Funktionen messbar, und die  $f : X \rightarrow [0, \infty)$  entsprechen den Folgen  $(a_k)_{k=1}^n$  bzw  $(a_k)_{k \geq 1}$  in  $[0, \infty)$  via  $a_k = f(k)$  bzw  $f = \sum_k a_k 1_{\{k\}}$ .

a) Sei  $X = \{1, \dots, n\}$ , dann ist jede Funktion einfach und  $\int_X f d\mu = \sum_{k=1}^n a_k$ . Endliche Summen sind also immer spezielle Integrale.

b) Sei ab jetzt  $X = \mathbb{N}^+$ . Für  $f \geq 0$  bzw  $(a_k) \geq 0$  sei  $f_n := \sum_{k=1}^n a_k 1_{\{k\}}$ ,  $n \geq 1$ , dann gilt  $f_n \nearrow f$  und  $f_n \in \mathcal{T}^+$ . Also ist

$$\int_X f d\mu = \lim_{n \rightarrow \infty} \int_X f_n d\mu = \lim_{n \rightarrow \infty} \sum_{k=1}^n a_k = \sum_{k \geq 1} a_k. \quad (3.6)$$

Unendliche Reihen nichtnegativer reeller Zahlen sind also spezielle Integrale.

c) Allgemeines  $f$  ist nach Bemerkung 3.2 genau dann integrierbar wenn  $\int |f| d\mu = \sum_{k \geq 1} |a_k| < \infty$ , dh wenn die Reihe *absolut* konvergiert. In diesem Fall gilt dann



$\int_X f \, d\mu = \int_X f^+ \, d\mu - \int_X f^- \, d\mu = \sum_{k \geq 1} a_k^+ - \sum_{k \geq 1} a_k^- = \sum_{k \geq 1} a_k$ . Absolut konvergente Reihen reeller Zahlen sind also immer spezielle Integrale.

d) Sind  $\sum_{k \geq 1} a_k^+ = \sum_{k \geq 1} a_k^- = \infty$ , so kann die Reihe zwar noch konvergieren, ihr Grenzwert (der dann bekanntlich von der Reihenfolge der Summation abhängt) ist aber *kein Integral* im oben definierten Sinne. ZB passiert dies bei  $\sum_{k \geq 1} (-1)^k/k$ .

## 3.4 Konvergenzsätze und Transformationsformel

**Fragestellung.** Wir wenden uns nun Aussagen zu, die man als Hauptsätze der allgemeinen Integrationstheorie bezeichnen kann. Ihre Bedeutung für die Analysis ist kaum zu überschätzen. Hauptsächlich befassen wir uns dabei mit Konvergenzsätzen für das Integral, dh Ergebnissen die erklären, wann wir aus der Kenntnis der Integrale  $\int_X f_n \, d\mu$  einer (punktweise oder wenigstens f.ü.) konvergenten Funktionenfolge  $(f_n)_{n \geq 1}$  auf das Integral  $\int_X f \, d\mu$  der Grenzfunktion  $f := \lim_{n \rightarrow \infty} f_n$  schließen können. Im allgemeinen kann man nicht davon ausgehen, dass sicher die erhoffte Beziehung

$$\int_X f \, d\mu = \lim_{n \rightarrow \infty} \int_X f_n \, d\mu \quad (3.7)$$

besteht. Es ist wichtig sich klar zu machen, dass dies schon in sehr simplen Beispielen grausam scheitern kann:

**Beispiel 3.1** Auf dem Maßraum  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \lambda)$  betrachten wir die Treppenfunktionen  $f_n := -1_{[n, \infty)}$ . Für diese Folge einfacher Funktionen gilt ja  $f_n \nearrow f := 0$ , aber  $\int f_n \, d\mu = -\infty$  für alle  $n$ , während  $\int f \, d\mu = 0$ .

**Beispiel 3.2** Auf demselben Maßraum betrachten wir die Treppenfunktionen  $f_n := (1/n) \cdot 1_{[0, n]}$ ,  $n \geq 1$ . Offensichtlich gilt  $\int_X f_n \, d\lambda = 1$  für alle  $n$ , andererseits konvergiert die Folge der  $f_n$  sogar gleichmäßig gegen die Grenzfunktion  $f = 0$  mit  $\int_X f \, d\lambda = 0$ .

Wir müssen uns also nach zusätzlichen Bedingungen umsehen, welche die erwünschte Relation (3.7) sicherstellen. Die Ergebnisse dieses Abschnittes stellen uns entsprechende Aussagen zur Verfügung. Auf ihnen beruht die Leistungsfähigkeit der Lebesgueschen Integrationstheorie.

**Monotone Konvergenz.** Es sei  $(X, \mathcal{A}, \mu)$  ein fester Maßraum. Erinnern wir uns zunächst daran, dass wir schon mit der Definition des Integrals einer messbaren Funktion  $f \geq 0$  als  $\int_X f \, d\mu := \lim_{n \rightarrow \infty} \int_X f_n \, d\mu$ , wo  $(f_n)$  eine Folge von Treppenfunktionen mit  $0 \leq f_n \nearrow f$  sein sollte, eine Konvergenzeigenschaft dieser Art fest eingebaut haben. Grundlegend ist nun die Einsicht, dass entsprechendes auch für monotone Folgen beliebiger messbarer Funktionen  $f_n \geq 0$  mit  $f_n \nearrow f$  gilt:

**Satz 3.3 (Satz von der monotonen Konvergenz, Version 1)** Es sei  $(f_n)$  eine wachsende Folge messbarer nichtnegativer Funktionen und  $f := \lim_{n \rightarrow \infty} f_n$ . (Also  $0 \leq f_1(x) \leq \dots \leq f_n(x) \nearrow f(x)$  für alle  $x \in X$ .) Dann gilt

$$\lim_{n \rightarrow \infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

**Beweis. (i)** Wegen  $f_n \nearrow f$  (punktweise) ist auch  $f$  messbar. Weiters gilt (nach Proposition 3.4)

$$0 \leq \int_X f_n \, d\mu \leq \int_X f_{n+1} \, d\mu \leq \int_X f \, d\mu \quad \text{für alle } n \geq 1,$$

also existiert  $\lim_{n \rightarrow \infty} \int_X f_n d\mu$  (in  $[0, \infty]$ ) und es gilt

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu \leq \int_X f d\mu.$$

(ii) Es bleibt zu zeigen, dass auch

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu \geq \int_X f d\mu. \quad (3.8)$$

Dazu verschaffen wir uns eine Folge  $(h_n)$  in  $\mathcal{T}^+$  mit

$$h_n \leq f_n \text{ für alle } n \geq 1 \quad \text{und} \quad h_n \nearrow f.$$

Aus diesen beiden Eigenschaften folgen sofort

$$\int_X h_n d\mu \leq \int_X f_n d\mu \text{ für alle } n \geq 1 \quad \text{und} \quad \int_X h_n d\mu \nearrow \int_X f d\mu,$$

und somit auch unsere Behauptung (3.8).

(iii) Zur Konstruktion der  $h_n$ : Für jedes  $f_n$ ,  $n \geq 1$ , gibt es eine Folge  $(g_{n,k})_{k \geq 1}$  in  $\mathcal{T}^+$  mit  $g_{n,k} \nearrow f_n$  (für  $k \rightarrow \infty$ ), insbesondere also mit  $\int_X g_{n,k} d\mu \nearrow \int_X f_n d\mu$ . Da wir  $g_{n,k}$  und  $g_{m,j}$  für  $n \neq m$  im allgemeinen nicht vergleichen können, betrachten wir

$$h_n := \max(g_{1,n}, g_{2,n}, \dots, g_{n,n}), \quad n \geq 1.$$

Nach Proposition 3.1 ist  $h_n \in \mathcal{T}^+$ , und aus der Definition folgen (ganz leicht)  $h_n \leq f_n$  und  $h_n \leq h_{n+1}$ .

Wir müssen noch nachweisen, dass  $h_n \nearrow f$ . Sei  $x \in X$  fest. Im Falle  $f(x) = 0$ , ist die Aussage trivial (warum?). Nehmen wir also  $f(x) > 0$  an. Wegen  $h_n \leq f_n \leq f$  reicht es zu zeigen, dass es für jedes  $\varepsilon > 0$  ein  $M \in \mathbb{N}$  gibt mit  $h_n(x) > f(x) - \varepsilon$  für  $n \geq M$ .

Da  $f_n(x) \nearrow f(x)$  gibt es ein  $N$  mit  $f_N(x) > f(x) - \varepsilon/2$ . Weiters, da  $g_{N,k}(x) \nearrow f_N(x)$ , gibt es auch ein  $K$  mit  $g_{N,K}(x) > f(x) - \varepsilon$ . Wählen wir  $M := \max(N, K)$ , so gilt für  $n \geq M$  stets

$$\begin{aligned} h_n(x) &= \max(g_{1,n}(x), g_{2,n}(x), \dots, g_{N,n}(x), \dots, g_{n,n}(x)) \\ &\geq g_{N,n}(x) \geq g_{N,K}(x) > f(x) - \varepsilon, \end{aligned}$$

wie gefordert. ■

Um ihn bequemer anwenden zu können, halten wir eine etwas flexiblere Fassung des Satzes von der monotonen Konvergenz fest. Einerseits verlangen wir nur noch, dass Monotonie und Konvergenz fast überall gelten sollen. Andererseits schwächen wir die Forderung  $f_n \geq 0$  ab. Sie wird ersetzt durch die Bedingung, dass die kleinste der Funktionen, nämlich  $f_1$ , nicht allzu negativ ist.

**Satz 3.4 (Satz von der monotonen Konvergenz, Version 2)** Seien  $f, f_n : X \rightarrow \overline{\mathbb{R}}$ ,  $n \geq 1$ , messbare Funktionen. Falls

$$f_n \leq f_{n+1} \text{ f.ü. } \forall n \geq 1 \quad \text{und} \quad \lim_{n \rightarrow \infty} f_n = f \text{ f.ü.} \quad \text{und} \quad \int_X f_1^- d\mu < \infty,$$

dann gilt

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu.$$

**Beweis.** Zunächst garantiert  $\int_X f_1^- d\mu < \infty$ , dass  $f_1^- < \infty$  f.ü.. Definieren wir

$$A := \left\{ \lim_{n \rightarrow \infty} f_n = f \right\} \cap \bigcap_{n \geq 1} \{f_n \leq f_{n+1}\} \cap \{f_1^- < \infty\} \in \mathcal{A},$$

so gilt daher  $\mu(A^c) = 0$  (weshalb?). Wir setzen  $g_n := 1_A(f_n + f_1^-)$ ,  $n \geq 1$ , dann sind die  $g_n$  messbar, nichtnegativ, und es gilt, auf ganz  $X$ ,

$$g_n \nearrow g := 1_A(f + f_1^-).$$

Satz 3.3 zeigt somit

$$\lim_{n \rightarrow \infty} \int_A (f_n + f_1^-) d\mu = \lim_{n \rightarrow \infty} \int_X g_n d\mu = \int_X g d\mu = \int_A (f + f_1^-) d\mu,$$

und wegen  $\int_A (f_n + f_1^-) d\mu = \int_A f_n d\mu + \int_A f_1^- d\mu = \int_X f_n d\mu + \int_X f_1^- d\mu$  (und entsprechend mit  $f$  statt  $f_n$ ) erhalten wir

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu.$$

■

**Fatou und Lebesgue.** Da uns nicht alle interessanten Funktionenfolgen den Gefallen tun, monoton zu sein, sollten wir auch überlegen, was man ohne Monotonie anzunehmen noch sagen kann. Im nächsten Resultat setzen wir nicht einmal voraus, dass die Folge  $(f_n)$  konvergiert, sondern betrachten von vornherein nur  $\underline{\lim}_{n \rightarrow \infty} f_n$  und  $\overline{\lim}_{n \rightarrow \infty} f_n$ .

**Satz 3.5 (Das Lemma von Fatou)** *Es seien  $g, f_n : X \rightarrow \overline{\mathbb{R}}$ ,  $n \geq 1$ , messbare Funktionen.*

a) Falls  $\int_X g^- d\mu < \infty$  und  $f_n \geq g$  f.ü. für  $n \geq 1$ , so gilt

$$\int_X \underline{\lim}_{n \rightarrow \infty} f_n d\mu \leq \underline{\lim}_{n \rightarrow \infty} \int_X f_n d\mu.$$

b) Falls  $\int_X g^+ d\mu < \infty$  und  $f_n \leq g$  f.ü. für  $n \geq 1$ , so gilt

$$\int_X \overline{\lim}_{n \rightarrow \infty} f_n d\mu \geq \overline{\lim}_{n \rightarrow \infty} \int_X f_n d\mu.$$

c) Falls  $g$  integrierbar und  $|f_n| \leq g$  f.ü. für  $n \geq 1$ , so gilt

$$\int_X \underline{\lim}_{n \rightarrow \infty} f_n d\mu \leq \underline{\lim}_{n \rightarrow \infty} \int_X f_n d\mu \leq \overline{\lim}_{n \rightarrow \infty} \int_X f_n d\mu \leq \int_X \overline{\lim}_{n \rightarrow \infty} f_n d\mu.$$

**Beweis.** a) Wir erinnern uns, dass  $\underline{\lim}_{n \rightarrow \infty} f_n = \lim_{n \rightarrow \infty} h_n$ , wobei die  $h_n := \inf_{k \geq n} f_k$ ,  $n \geq 1$ , eine monotone Folge bilden:  $h_n \leq h_{n+1}$ . Wegen  $h_1 = \inf_{k \geq 1} f_k \geq g$  f.ü. und  $\int_X g^- d\mu < \infty$  gilt  $\int_X h_1^- d\mu < \infty$ . Wir können daher den vorangehenden Satz auf die Folge  $(h_n)$  anwenden, und erhalten

$$\lim_{n \rightarrow \infty} \int_X h_n d\mu = \int_X \lim_{n \rightarrow \infty} h_n d\mu = \int_X \underline{\lim}_{n \rightarrow \infty} f_n d\mu. \quad (3.9)$$

Da offenbar  $h_n \leq f_n$  f.ü. für alle  $n \geq 1$ , folgt  $\int_X h_n d\mu \leq \int_X f_n d\mu$ , und damit

$$\lim_{n \rightarrow \infty} \int_X h_n d\mu \leq \underline{\lim}_{n \rightarrow \infty} \int_X f_n d\mu. \quad (3.10)$$

Die Behauptung folgt unmittelbar aus (3.9) und (3.10).

b) Betrachte  $(-f_n)_{n \geq 1}$  und verwende a).

c) Es gilt  $-g \leq f_n \leq g$ , sodass wir a) und b) anwenden können. ■

**Beispiel 3.3** Die Ungleichungen im Lemma von Fatou versteht und merkt man sich am besten, wenn man sich ein ganz einfaches Beispiel einprägt: Es seien  $X := (0, 1]$ ,  $\mathcal{A} := \mathcal{B}_{(0,1]}$  und  $\mu = \lambda$ . Betrachte die Folge von Funktionen  $f_n$  mit

$$f_n := \begin{cases} 1_{(0, \frac{1}{2}]} & \text{falls } n \text{ gerade} \\ 1_{(\frac{1}{2}, 1]} & \text{falls } n \text{ ungerade} \end{cases}$$

Dann gilt jedenfalls  $\lim_{n \rightarrow \infty} f_n = 0$  auf  $X$ , also auch  $\int_X \lim_{n \rightarrow \infty} f_n d\mu = 0$ . Andererseits ist für jedes  $n \geq 1$ ,  $\int_X f_n d\mu = \frac{1}{2}$ , und daher  $\lim_{n \rightarrow \infty} \int_X f_n d\mu = \frac{1}{2}$ . Entsprechend für  $\overline{\lim}_{n \rightarrow \infty} f_n = 1_X$ .

Nun brauchen wir kaum noch was zu tun, um den zweiten berühmten Konvergenzsatz der Lebesgueschen Integrationstheorie zu bekommen.

**Satz 3.6 (Satz von der dominierten Konvergenz)** Es seien  $f, f_n : X \rightarrow \overline{\mathbb{R}}$ ,  $n \geq 1$ , messbare Funktionen mit  $\lim_{n \rightarrow \infty} f_n = f$  f.ü.. Falls es eine integrierbare Funktion  $g : X \rightarrow \overline{\mathbb{R}}$  gibt mit  $|f_n| \leq g$  f.ü. für  $n \geq 1$ , so folgt

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu.$$

**Beweis.** Da die Voraussetzungen von Teil c) des Fatou'schen Lemmas erfüllt sind, gilt

$$\int_X \lim_{n \rightarrow \infty} f_n d\mu \leq \lim_{n \rightarrow \infty} \int_X f_n d\mu \leq \overline{\lim}_{n \rightarrow \infty} \int_X f_n d\mu \leq \int_X \overline{\lim}_{n \rightarrow \infty} f_n d\mu.$$

Nach Voraussetzung stimmen aber  $\lim_{n \rightarrow \infty} f_n$  und  $\overline{\lim}_{n \rightarrow \infty} f_n$  f.ü. mit  $f$  überein, daher sind die beiden äußeren Integrale beide gleich  $\int_X f d\mu$ . Dann folgt aber klarerweise auch, dass

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \overline{\lim}_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu,$$

und damit unsere Behauptung. ■

Im wichtigen Spezialfall endlicher Maßräume ist jede konstante reelle Funktion  $g$  integrierbar, und wir erhalten insbesondere

**Folgerung 3.1** Es seien  $f, f_n : X \rightarrow \overline{\mathbb{R}}$  messbare Funktionen mit  $\lim_{n \rightarrow \infty} f_n = f$  f.ü.. Ist  $\mu(X) < \infty$  und gibt es eine Konstante  $M \in \mathbb{R}$  mit  $|f_n| \leq M$  f.ü. für  $n \geq 1$ , so folgt

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu.$$

**Die Transformationsformel.** Zum Ende dieses Abschnittes besprechen wir noch kurz ein anderes Grundprinzip, das sowohl für die Theorie als auch für die konkrete Berechnung von Integralen von Bedeutung ist. Erinnern Sie sich bitte an die Möglichkeit, mittels messbarer Abbildungen Maße auf andere Räume zu transportieren. Das Zusammenspiel zwischen Integration und diesem Übergang zum Bildmaß gestaltet sich erfreulich problemlos.

**Satz 3.7 (Allgemeine Transformationsformel)** Es sei  $T : X \rightarrow Y$  eine  $\mathcal{A}$ - $\mathcal{B}$ -messbare Abbildung,  $\mu$  ein Maß auf  $\mathcal{A}$ , und  $\mu \circ T^{-1}$  sein Bild unter  $T$ . Dann gilt für jedes messbare  $g : Y \rightarrow \overline{\mathbb{R}}$ ,

$$\int_Y g d(\mu \circ T^{-1}) = \int_X (g \circ T) d\mu$$

sobald eines dieser Integrale definiert ist.

**Beweis.** Wir betrachten zuerst den Fall  $g = \sum_{j=1}^m \beta_j 1_{B_j} \in \mathcal{T}^+$ , und finden

$$\begin{aligned} \int_Y g d(\mu \circ T^{-1}) &= \sum_{j=1}^m \beta_j \mu(T^{-1}B_j) = \sum_{j=1}^m \beta_j \int_X 1_{T^{-1}B_j} d\mu \\ &= \sum_{j=1}^m \beta_j \int_X 1_{B_j} \circ T d\mu = \int_X \left( \sum_{j=1}^m \beta_j 1_{B_j} \right) \circ T d\mu. \end{aligned}$$

Für messbares  $g : Y \rightarrow [0, \infty]$  wählen wir einfache Funktionen  $g_n \in \mathcal{T}^+$  mit  $g_n \nearrow g$ . Dann gilt auch  $g_n \circ T \nearrow g \circ T$ , und somit

$$\int_Y g d(\mu \circ T^{-1}) = \lim_{n \rightarrow \infty} \int_Y g_n d(\mu \circ T^{-1}) = \lim_{n \rightarrow \infty} \int_X (g_n \circ T) d\mu = \int_X (g \circ T) d\mu.$$

Der Rest ist ganz einfach. ■

### 3.5 Mehr Beispiele und Anwendungen

**A) Unbeschränkte Intervalle und unbeschränkte Funktionen.** Wir wenden uns wieder der Integration auf dem Raum  $(\mathbb{R}, \mathcal{L}_{\mathbb{R}}, \lambda)$  zu. (Erinnerung: Um eine Funktion  $f$  die auf einer Lebesgue-messbaren Menge  $A$  (hier meist ein Intervall) definiert und messbar ist zu integrieren, können wir entweder den Raum einschränken, und zB über  $([0, 1], \mathcal{L}_{[0,1]}, \lambda)$  sprechen, oder (äquivalent dazu)  $f$  nach  $\mathbb{R}$  fortsetzen, mit  $f(x) := 0$  für  $x \notin A$ .) Bisher haben wir für beschränkte Funktionen auf kompakten Intervallen gesehen, dass man das Integral einfach als Riemann-Integral ausrechnen kann. Wie sieht es aber nun mit unbeschränkten Intervallen und unbeschränkten Funktionen aus, die in der Definition des Lebesgue-Integrals ja von vornherein zugelassen sind?

In der Riemannschen Theorie spricht man ggf von uneigentlichen R-Integralen. Die typische Situation sieht dabei so aus: Wir studieren eine messbare Funktion  $f : (a, b) \rightarrow [0, \infty)$ , wobei  $-\infty \leq a < b \leq \infty$  ist, und nehmen an, dass  $f$  auf jedem der Teilintervalle  $[a_n, b_n]$  R-integrierbar ist, mit  $a_n \searrow a$  und  $b_n \nearrow b$ , sodass  $[a_n, b_n] \nearrow (a, b)$ . Falls dann der Grenzwert

$$\int_a^b f(x) dx := \lim_{n \rightarrow \infty} \int_{a_n}^{b_n} f(x) dx$$

endlich ist (wegen  $f \geq 0$  ist die Folge monoton), so nennt man ihn das *uneigentliche Riemann-Integral von  $f$  über  $(a, b)$* . Das kennen Sie ja schon.

**Beispiel 3.4** Im Sinne dieser Definition ist zB

$$\int_0^\infty e^{-x} dx = \lim_{n \rightarrow \infty} \int_0^n e^{-x} dx = 1 \text{ und } \int_0^1 \log x dx = \lim_{n \rightarrow \infty} \int_{1/n}^1 \log x dx = -1.$$

Beachten wir nun, dass  $1_{[a_n, b_n]} \nearrow 1_{(a, b)}$ , dann finden wir (unabhängig von etwaiger R-Integrierbarkeit auf den  $[a_n, b_n]$ ) mithilfe des Satzes von der Monotonen Konvergenz entsprechend

$$\int_{(a, b)} f d\lambda = \lim_{n \rightarrow \infty} \int_{(a_n, b_n)} f d\lambda,$$

sogar wenn der Grenzwert  $\infty$  sein sollte. Dabei gilt rechts  $\lim_{n \rightarrow \infty} \int_{(a_n, b_n)} f d\lambda = \lim_{n \rightarrow \infty} \int_{a_n}^{b_n} f(x) dx$  sofern diese Riemann-Integrale existieren. Lebesgue-Integrale

unbeschränkter Funktionen bzw über unbeschränkte Intervalle kann man also genau wie im Riemannschen Fall berechnen, und uneigentliche R-Integrale sind ganz normale Lebesgue-Integrale - jedenfalls wenn  $f \geq 0$ .

Bei Funktionen mit wechselndem Vorzeichen ist allerdings Vorsicht geboten, diese können uneigentlich R-integrierbar sein, obwohl das L-Integral nicht existiert.

**Beispiel 3.5** Betrachte die Funktion  $f : [0, \infty) \rightarrow \mathbb{R}$  mit  $f(x) := \frac{1}{n}$  falls  $x \in [2n - 2, 2n - 1)$  und  $f(x) := -\frac{1}{n}$  falls  $x \in [2n - 1, 2n)$ ,  $n \geq 1$ . Dann ist  $\int_0^\infty f(x) dx := \lim_{b \rightarrow \infty} \int_0^b f(x) dx = 0$  weil  $\left| \int_0^b f(x) dx \right| \leq \frac{1}{n}$  wenn  $b \geq 2n$ . Da aber  $\int_{[0, \infty)} f^+ d\lambda = \int_{[0, \infty)} f^- d\lambda = \sum_{n \geq 1} \frac{1}{n} = \infty$  gilt, ist diese Funktion auf  $[0, \infty)$  nicht einmal quasi-integrierbar, und daher  $\int_{[0, \infty)} f d\lambda$  nicht definiert! Man sollte diesen Grenzwert daher nicht mit  $\int_{[0, \infty)} f d\lambda$  oder  $\int_0^\infty f d\lambda$  bezeichnen. Entsprechendes gilt für uneigentliche R-Integrale unbeschränkter Funktionen. Um zu sehen, ob man gefahrlos über L-Integrale solcher Funktionen sprechen kann, muss man eben zuerst überprüfen, ob  $\int_{[0, \infty)} f^+ d\lambda$  oder  $\int_{[0, \infty)} f^- d\lambda$  endlich ist.

**B) Integrale bezüglich diskreter Maße.** Sehen wir uns als weiteres konkretes Beispiel den Fall diskreter Maße (vgl Bemerkung 2.2) an. Hier ist  $(X, \mathcal{A})$  ein beliebiger messbarer Raum,  $(x_j)_{j \geq 1}$  eine Folge unterschiedlicher ( $x_i \neq x_j$  falls  $i \neq j$ ) Punkten in  $X$ , und  $(m_j)_{j \geq 1}$  eine Folge von "Gewichten" in  $[0, \infty]$ . Durch  $\nu := \sum_{j \geq 1} m_j \cdot \delta_{x_j}$  wird ein diskretes Maß auf  $\mathcal{A}$  definiert, das dem Punkt  $x_j$  die Masse  $m_j$  zuordnet. Nehmen wir an, dass jede einpunktige Menge  $\{x\}$  in  $\mathcal{A}$  liegt, so sind auch die  $A_n := \{x_j : 1 \leq j \leq n\}$  und  $A := \{x_j : j \geq 1\}$  messbare Mengen, und natürlich gilt  $A_n \nearrow A$ . Das Maß  $\nu$  ist auf  $A$  konzentriert in dem Sinne, dass  $\nu(A^c) = 0$ .

Ist nun  $f : X \rightarrow [0, \infty]$  messbar, so sehen wir, dass  $1_{A_n} f \nearrow 1_A f$  konvergiert, wobei  $1_A f = f$  f.ü. Damit folgt

$$\int_X f d\nu = \int_X 1_A f d\nu = \lim_{n \rightarrow \infty} \int_X 1_{A_n} f d\nu = \lim_{n \rightarrow \infty} \sum_{j=1}^n f(x_j) m_j = \sum_{j \geq 1} f(x_j) m_j. \quad (3.11)$$

Identifizieren wir  $f$  wieder mit der Folge der Funktionswerte  $a_j := f(x_j)$ , so beschreibt das Integral bezüglich  $\nu$  also die Operation, welche einer Folge  $(a_j)_{j \geq 1}$  in  $[0, \infty]$  die durch die  $m_j$  gewichtete Summe zuordnet,

$$(a_j)_{j \geq 1} \mapsto \sum_{j \geq 1} a_j m_j.$$

Insbesondere wenn  $\nu$  normiert ist, also  $\sum_{j \geq 1} m_j = 1$ , spricht man auch von einem *gewichteten Mittel*. Eine messbare Funktion  $f : X \rightarrow \mathbb{R}$  ist  $\nu$ -integrierbar genau wenn  $\sum_{j \geq 1} |f(x_j)| m_j = \sum_{j \geq 1} |a_j| m_j < \infty$ , und in diesem Fall ist natürlich wieder  $\int_X f d\mu = \sum_{j \geq 1} f(x_j) m_j = \sum_{j \geq 1} a_j m_j$ .

**Beispiel 3.6** Auf  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  sei  $\nu := \sum_{j \geq 0} 2^{-j} \delta_j$ . Für  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit  $f(x) := x + 1$  finden wir dann

$$\int_X f d\nu = \sum_{j \geq 0} (j+1) 2^{-j} = 4.$$

Bei Maßen auf  $\mathcal{B}_{\mathbb{R}}$  welche manchen Punkten positive Masse verleihen, kann dann eben auch das Integral über einzelne Punkte positiv sein, hier etwa  $\int_{\{0\}} f d\nu = f(0)\nu(\{0\}) = 1$ . Daher ist es wichtig, nicht leichtfertig  $\int_a^b$  zu schreiben, sondern genau zu spezifizieren, welche der Endpunkte dazugehören sollen. Im vorliegenden Fall ist etwa  $\int_{[0,1)} f d\nu = 1$ , aber  $\int_{(0,1)} f d\nu = 0$ . Wie sollte man da  $\int_0^1 f d\nu$  interpretieren?

**C) Maße mit Dichten.** Erinnern Sie sich bitte daran, dass wir im konkreten Fall von  $\lambda$  schon früher neue Maße durch Integration erhalten haben, siehe Bsp

2.7. Damals stand uns aber nur das Riemann-Integral zur Verfügung. Mithilfe des Satzes von der Monotonen Konvergenz sehen wir sehr leicht, dass wir ganz allgemein durch die Integration nichtnegativer messbarer Funktionen über Teilmengen immer ein Maß erhalten.

**Proposition 3.7 (Integration liefert Maße)** Ist  $(X, \mathcal{A}, \mu)$  ein Maßraum und  $h \geq 0$  messbar, so wird durch

$$h \odot \mu(A) := \int_A h \, d\mu = \int_X 1_A h \, d\mu, \quad A \in \mathcal{A}, \quad (3.12)$$

ein Maß  $h \odot \mu$  auf  $(X, \mathcal{A})$  definiert. Für dieses gilt

$$\mu(A) = 0 \implies h \odot \mu(A) = 0. \quad (3.13)$$

**Beweis.** Klarerweise ist  $h \odot \mu(\emptyset) = \int_X 0 \, d\mu = 0$ . Zum Nachweis der  $\sigma$ -Additivität sei  $(A_n)_{n \geq 1}$  eine disjunkte Folge in  $\mathcal{A}$ , und  $A := \biguplus_{n \geq 1} A_n$ . Dann gilt  $1_A h = \sum_{n \geq 1} 1_{A_n} h$ , also  $\sum_{n=1}^N 1_{A_n} h \nearrow 1_A h$ , und somit

$$\begin{aligned} h \odot \mu(A) &= \int_X 1_A h \, d\mu = \lim_{N \rightarrow \infty} \int_X \left( \sum_{n=1}^N 1_{A_n} h \right) d\mu \\ &= \lim_{N \rightarrow \infty} \sum_{n=1}^N \int_X 1_{A_n} h \, d\mu = \sum_{n \geq 1} h \odot \mu(A_n). \end{aligned}$$

Wir wenden uns (3.13) zu und nehmen  $\mu(A) = 0$  an. Dann ist  $\{1_A h > 0\} \subseteq A$ , also auch  $\mu(\{1_A h > 0\}) = 0$ , und daher  $h \odot \mu(A) = \int_X 1_A h \, d\mu = 0$ . ■

Diese Beobachtung stellt einen der wichtigsten Wege, neue Maße zu gewinnen, bereit. Auch die dort festgehaltene Eigenschaft (3.13) so definierter Maße verdient eine eigene Bezeichnung.

**Definition 3.6** Das in (3.12) definierte Maß  $h \odot \mu$  heißt das Maß mit Dichte  $h$  (bezüglich  $\mu$ ).

**Beispiel 3.7 a)** Offensichtlich ist für  $c \in [0, \infty]$  stets  $(c1_X) \odot \mu = c\mu$ .

**b)** Für  $B \in \mathcal{A}$  ist  $\nu := 1_B \odot \mu$  das auf  $B$  eingeschränkte Maß  $\mu$ , also  $\nu(A) = 1_B \odot \mu(A) = \mu(A \cap B)$  für  $A \in \mathcal{A}$ .

**c)** Das diskrete Maß  $\nu = \sum_{j \geq 1} m_j \delta_{x_j}$  besitzt die Dichte  $h = \sum_{j \geq 1} m_j 1_{\{x_j\}}$  bezüglich  $\mu = \sum_{j \geq 1} \delta_{x_j}$  (das "Zählmaß auf  $\{x_j : j \geq 1\}$ "). Es ist ja  $\nu(A) = \sum_{j \geq 1} m_j 1_A(x_j) = \sum_{j \geq 1} m_j 1_A(x_j) \mu(\{x_j\}) = \int_A h \, d\mu$ .

**Bemerkung 3.3 a)** Da es beim Integrieren auf Nullmengen nicht ankommt, ist die Dichte eines Maßes ggf nur bis auf Nullmengen eindeutig bestimmt:  $g$  ist eine Dichte von  $h \odot \mu$  genau wenn  $g = h$  f.ü.

**b)** Ist  $\nu = h \odot \mu$ , so wird die Dichte  $h$  auch mit  $d\nu/d\mu$  bezeichnet. In Hinblick auf Bsp 2.7 kann man diese nämlich als verallgemeinerte Ableitung ansehen. (Mehr dazu im nächsten Unterabschnitt.)

Integrale bezüglich eines Maßes der Form  $h \odot \mu$  können leicht als Integrale bezüglich  $\mu$  umgeschrieben werden.

**Proposition 3.8 (Integration bezüglich  $h \odot \mu$ )** Ist  $(X, \mathcal{A}, \mu)$  ein Maßraum und sind  $g, h \geq 0$  messbar, so gilt

$$\int_X g \, d(h \odot \mu) = \int_X gh \, d\mu = \int_X g(x)h(x) \, d\mu(x). \quad (3.14)$$

**Beweis.** Siehe PS. ■

**Beispiel 3.8 a)** Für  $\nu := 1_B \odot \mu$  aus dem Beispiel b) oben ist das ganz einfach,  $\int_X g d(1_B \odot \mu) = \int_X g 1_B d\mu = \int_B g d\mu$ .

**b)** Für das diskrete Maß  $\nu = \sum_{j \geq 1} m_j \delta_{x_j}$  aus dem Beispiel c) oben haben wir diese Formel auch schon nachgerechnet. Einerseits steht da nämlich, dass für  $g \geq 0$  und  $h, \mu$  wie dort beschrieben,  $\int_X g d\nu = \int_X g d(h \odot \mu)$ . Gemäß (3.11) in Abschnitt B oben gilt aber andererseits für  $g \geq 0$ ,

$$\int_X g d\nu = \sum_{j \geq 1} g(x_j) m_j = \sum_{j \geq 1} g(x_j) h(x_j) = \sum_{j \geq 1} g(x_j) h(x_j) \mu(\{x_j\}) = \int_X gh d\mu.$$

**D) Der Hauptsatz der Differential- und Integralrechnung und Absolute Stetigkeit.** Maße mit Dichten sind nun aufs Engste mit einem zentralen Thema der Analysis verknüpft, dem Zusammenspiel zwischen Integration und Differentiation. Um dies zu besprechen, konzentrieren wir uns auf ein kompaktes Intervall  $[a, b]$  in  $\mathbb{R}$ , und das Lebesguesche Maß  $\lambda$  auf  $\mathcal{L}_{[a, b]}$ . Ist dann  $f : [a, b] \rightarrow [0, \infty)$  L-integrierbar, so ist die Verteilungsfunktion  $F$  des Maßes  $f \odot \lambda$  gegeben durch das *bestimmte (Lebesgue-) Integral*

$$F(t) := f \odot \lambda([a, t]) = \int_a^t f(x) dx, \quad t \in [a, b].$$

Jetzt wissen Sie natürlich, dass dies im speziellen Fall einer stetigen Funktion  $f$  eine *Stammfunktion* des Integranden liefert,  $F' = f$ , und dafür reicht bereits das Riemann-Integral aus. Wie ist das aber nun für allgemeine Lebesgue-integrierbare Funktionen  $f$ ? Damit wir nicht zuviel erwarten, nehmen wir einmal an es gelte  $F' = f$ . Wenn wir nun  $f$  auf einer Lebesgue-Nullmenge  $A \subseteq [a, b]$  abändern, bekommen wir eine neue L-integrierbare Funktion  $\tilde{f}$ . Das Integral erkennt den Unterschied aber nicht, dh wir haben  $\tilde{F}(t) := \int_a^t \tilde{f}(x) dx = F(t)$  für jedes  $t$ . Somit folgt  $\tilde{F}'(t) = F'(t) = f(t) \neq \tilde{f}(t)$  falls  $t \in A$ . Kurz: weil das bestimmte Integral  $F$  den Integranden  $f$  nur bis auf eine Nullmenge festlegt, können wir bestenfalls hoffen, dass seine Ableitung *fast überall* mit  $f$  übereinstimmt. Das ist aber tatsächlich der Fall:

**Satz 3.8 (Hauptsatz der Differential- und Integralrechnung, Teil I)** Ist  $f : [a, b] \rightarrow \mathbb{R}$  L-integrierbar, und  $F(t) := \int_a^t f d\lambda$ , dann ist  $F$  f.ü. differenzierbar mit

$$F' = f \quad \text{f.ü. auf } [a, b].$$

Für den Beweis dieses schönen und fundamentalen Satzes fehlt hier leider die Zeit, Sie werden ihn in anderen LVen kennenlernen. Hier sind wir also von  $f$  ausgegangen, und haben zuerst integriert, um dann durch Differentiation von  $F$  die ursprüngliche Funktion f.ü. wiederzufinden. Vorausgesetzt wurde nur L-Integrierbarkeit (offensichtlich notwendig damit der erste Schritt Sinn ergibt), und nach obiger Bemerkung ist die f.ü.-Aussage der Schlussfolgerung bestmöglich.

Die Gegenrichtung erweist sich als etwas subtiler, deshalb behandeln wir sie getrennt: Für Teil II des Hauptsatzes fragen wir uns was passiert, wenn wir von einer Funktion  $F$  ausgehend zuerst differenzieren, und dann versuchen durch Integration der Ableitung  $F'$  zu  $F$  zurückzukehren, indem  $F(t) = F(a) + \int_a^t F' d\lambda$ . Bekanntlich funktioniert dies in der besonders gutmütigen Situation einer stetig differenzierbaren Funktion  $F$ . Allgemein gibt es wiederum eine offenkundige notwendige Voraussetzung, nämlich jene, dass  $F'$  wenigstens f.ü. existiert (damit das Integral von  $F'$  eindeutig festgelegt ist). Allerdings reicht dies noch nicht aus, und so enttäuschend Ihnen das zunächst erscheinen mag, so interessant ist doch der Grund dafür: Es gibt eben Funktionen  $F : [a, b] \rightarrow \mathbb{R}$  die zwar fast überall differenzierbar sind, deren Ableitung das Wachstum der Funktion aber überhaupt nicht widerspiegelt! Tatsächlich habes Sie so eine Funktion schon gesehen: Die monotone Cantor-Funktion  $F : [0, 1] \rightarrow [0, 1]$  aus Beispiel 2.8 ist stetig und hat



f.ü. Ableitung  $F'(t) = 0$ . Trotzdem schafft sie es, echt zu wachsen indem  $F(0) = 0$  und  $F(1) = 1$  gilt.

Also ist Differenzierbarkeit f.ü. keine hinreichende Bedingung dafür, dass  $F$  ein unbestimmtes Integral ist. Tatsächlich ist letzteres eine separate Eigenschaft, die man als starke Form der Stetigkeit ansehen kann. Weil eine stetige Funktion  $F$  auf dem Kompaktum  $[a, b]$  automatisch gleichmäßig stetig ist, wissen wir, dass  $F : [a, b] \rightarrow \mathbb{R}$  genau dann stetig ist, wenn es zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  gibt derart, dass  $|F(\beta) - F(\alpha)| < \varepsilon$  sobald  $\lambda([\alpha, \beta]) < \delta$ . Schärfer ist die Forderung in der folgenden Definition, dass nämlich die Summe der Zuwächse über viele kleine Intervalle  $[\alpha_i, \beta_i]$  durch  $\varepsilon$  beschränkt bleibt wenn insgesamt  $\sum_{i=1}^n \lambda([\alpha_i, \beta_i]) < \delta$ . (Der Unterschied besteht darin, dass wir hier im gesamten Intervall  $[a, b]$  eine große Zahl besonders gefährlicher Punkte  $\alpha_i$  kombinieren dürfen, und  $F$  nicht nur in einer Umgebung eines einzelnen Punktes  $\alpha$  betrachten.)

**Definition 3.7** Eine Funktion  $F : [a, b] \rightarrow \mathbb{R}$  nennt man absolutstetig wenn gilt:

$$\begin{aligned} &\text{für jedes } \varepsilon > 0 \text{ gibt es ein } \delta > 0 \text{ sodass für beliebige } n \in \mathbb{N} \\ &\text{und } a \leq \alpha_1 < \beta_1 \leq \dots \leq \alpha_n < \beta_n \leq b \text{ mit } \sum_{i=1}^n (\beta_i - \alpha_i) < \delta \\ &\text{stets } \sum_{i=1}^n |F(\beta_i) - F(\alpha_i)| < \varepsilon \text{ ist.} \end{aligned} \quad (3.15)$$

Dies erweist sich als die richtige Bedingung:

**Satz 3.9 (Hauptsatz der Differential- und Integralrechnung, Teil II)** Eine reelle Funktion  $F : [a, b] \rightarrow \mathbb{R}$  ist genau dann ein unbestimmtes  $L$ -Integral, wenn sie absolutstetig ist. In diesem Fall ist  $F$  f.ü. differenzierbar, und es gilt

$$F(t) - F(a) = \int_a^t F' d\lambda, \quad t \in [a, b]. \quad (3.16)$$

Wir verzichten hier wiederum darauf, den Satz komplett zu beweisen. Damit Sie die Bedingung aber nicht allzu abschreckend empfinden, zeigen wir, dass sie für monotonen  $F$  jedenfalls notwendig ist:

**Beweis der Notwendigkeit für monotonen  $F$ .** Es sei  $f : [a, b] \rightarrow [0, \infty)$   $L$ -integrierbar, und  $F(t) := \int_a^t f d\lambda$ . Wir zeigen, dass  $F$  absolutstetig ist. Wegen  $\{f > n\} \searrow \emptyset$  gilt  $f_n := 1_{\{f > n\}} f \rightarrow 0$ , und da  $|f_n| \leq f$  bekommen wir mit dominierter Konvergenz

$$\lim_{n \rightarrow \infty} \int_{\{f > n\}} f d\lambda = 0.$$

Ist nun  $\varepsilon > 0$  gegeben, so gibt es daher ein  $N \geq 1$  mit  $\int_{\{f > N\}} f d\lambda < \frac{\varepsilon}{2}$ . Setzen wir  $\delta := \frac{\varepsilon}{2N}$ , so finden wir für beliebige  $A \in \mathcal{L}_{[a, b]}$  mit  $\lambda(A) < \delta$ , dass

$$\int_A f d\lambda = \int_{A \cap \{f > N\}} f d\lambda + \int_{A \cap \{f \leq N\}} f d\lambda \leq \int_{\{f > N\}} f d\lambda + \int_A N d\lambda < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.$$

Im Falle  $A = \biguplus_{i=1}^n (\alpha_i, \beta_i]$  ist dabei  $\int_A f d\lambda = \sum_{i=1}^n \int_{\alpha_i}^{\beta_i} f d\lambda = \sum_{i=1}^n |F(\beta_i) - F(\alpha_i)|$  und  $\lambda(A) = \sum_{i=1}^n (\beta_i - \alpha_i)$ . ■

**Bemerkung 3.4 a)** Da der Hauptsatz für stetig differenzierbares  $F$  funktioniert, wissen Sie sofort, dass solche  $F$  auch absolutstetig sind. Versuchen Sie doch bitte, diese Aussage ohne Verwendung des Hauptsatzes zu beweisen.

**b)** Wenn  $F$  nicht gerade stetig differenzierbar ist, kann es recht schwierig werden, herauszufinden ob diese Funktion absolutstetig ist.

**Beispiel 3.9** Die oben erwähnte Cantor-Funktion aus Beispiel 2.8 ist natürlich nicht absolut stetig. Können Sie das ohne Verwendung des Hauptsatzes sehen?

Im Beweis haben wir uns auf die Situation der Verteilungsfunktion eines Maßes mit Dichte konzentriert (weil  $f \geq 0$  bzw  $F$  wachsend). Der schwierige Teil des Satzes ist es,  $f$  zu finden, falls  $F$  absolutstetig ist. Bei wachsendem  $F$  bedeutet dies, zu dem LS-Maß  $\mu_F$  mit Verteilungsfunktion  $F$  eine Dichte zu finden. Als Ausblick erwähnen wir ein prominentes Resultat der Maßtheorie, welches in allgemeinen Maßräumen die Existenz von Dichten garantiert, sobald eine geeignete Form der Absolutstetigkeit vorliegt.

**Definition 3.8** Ein Maß  $\nu$  auf  $(X, \mathcal{A})$  nennt man absolutstetig bezüglich  $\mu$ , geschrieben als  $\nu \ll \mu$ , wenn die Implikation  $\mu(A) = 0 \implies \nu(A) = 0$  gilt.

Offensichtlich erfüllt jedes Maß  $\nu = h \odot \mu$  das eine Dichte bezüglich  $\mu$  besitzt diese Bedingung,  $\nu \ll \mu$ , weil  $\nu(A) = \int_A h d\mu = 0$  sobald  $\mu(A) = 0$ . Man kann zeigen, dass im Fall  $(X, \mathcal{A}, \mu) = ([a, b], \mathcal{L}_{[a, b]}, \lambda)$  die VF  $F$  von  $\nu$  absolutstetig ist, genau wenn  $\nu \ll \lambda$  gilt. Daher ist der schwierige Teil von Hauptsatz II im wesentlichen enthalten in

**Satz 3.10 (Satz von Radon-Nikodym)** Ist  $\mu$  ein  $\sigma$ -endliches Maß auf  $(X, \mathcal{A})$ , und  $\nu$  ein Maß mit  $\nu \ll \mu$ , dann gibt es eine quasi-integrierbare Funktion  $f$  mit  $\nu = f \odot \mu$ . Dabei ist  $f$  bis auf  $\mu$ -Nullmengen eindeutig bestimmt.

Dieser Satz ist spektakulär, weil die Voraussetzung  $\nu \ll \mu$  so schwach zu sein scheint. Wir fordern nur, dass sich  $\nu$  nicht auf einer  $\mu$ -Nullmenge verstecken darf. Das reicht schon, um die Existenz einer Dichte sicherzustellen! Neben seiner Bedeutung in der reellen Analysis, die uns ja zum Begriff der absoluten Stetigkeit geführt hat, ist dieser Satz übrigens auch in anderen Bereichen der Analysis, und in der fortgeschrittenen Wahrscheinlichkeitstheorie wichtig, und Sie haben später siche noch Gelegenheit, mehr darüber zu erfahren.

**E) Die Transformationsformel im reellen Fall.** Die allgemeine Transformationsformel für Integrale (Satz 3.7) kann flexibel verwendet werden, wenn man das dort auftretende Bildmaß versteht. Wir sehen uns das gleich mal im Fall reeller Funktionen an.

**Satz 3.11 (Bild von  $\lambda$  unter einem  $\mathcal{C}^1$ -Diffeomorphismus)** Es sei  $T : [a, b] \rightarrow [\alpha, \beta]$  ein  $\mathcal{C}^1$ -Diffeomorphismus, dann ist das Bild des (auf  $[a, b]$  eingeschränkten) Lebesgue-Maßes  $\lambda$  das Maß auf  $[\alpha, \beta]$  mit Dichte  $1/|T'| \circ T^{-1}$ ,

$$\lambda \circ T^{-1} = \frac{1}{|T'| \circ T^{-1}} \odot \lambda. \quad (3.17)$$

Allgemeiner gilt, für messbares  $h \geq 0$ , dass das Bild von  $h \odot \lambda$  unter  $T$  die Dichte  $(h/|T'|) \circ T^{-1}$  besitzt,

$$(h \odot \lambda) \circ T^{-1} = \frac{h \circ T^{-1}}{|T'| \circ T^{-1}} \odot \lambda. \quad (3.18)$$

**Beweis.** (i) Wir nehmen an, dass  $T$  wachsend ist (sonst analog), und bestimmen das Bildmaß  $\lambda \circ T^{-1}$  indem wir die Verteilungsfunktion  $F$  mit  $F(s) = \lambda \circ T^{-1}((\alpha, s])$ ,  $s \in (\alpha, \beta]$ , betrachten. (Ausserhalb  $(\alpha, \beta]$  ist die VF natürlich konstant,  $F(s) = 0$  für  $s \leq \alpha$  und  $F(s) = 1$  für  $s > \beta$ .) Offenbar gilt  $F(s) = \lambda(T^{-1}(\alpha, s]) = \lambda((a, T^{-1}s]) = T^{-1}s - a$ , und daher ist  $F$  stetig differenzierbar mit  $F'(s) = 1/T'(T^{-1}(s))$ . Nach dem klassischen Hauptsatz der Differential-Integral-Rechnung gilt also

$$\lambda \circ T^{-1}((\alpha, s]) = F(s) = \int_{\alpha}^s \frac{ds}{T'(T^{-1}(s))} = \int_{(\alpha, s]} \frac{d\lambda}{T' \circ T^{-1}} \quad \text{für } s \in (\alpha, \beta].$$

Das L-S-Maß  $\lambda \circ T^{-1}$  stimmt also auf  $\mathcal{I}_{\mathbb{R}}$  mit dem L-S-Maß  $\frac{1}{T' \circ T^{-1}} \odot \lambda$  überein, und ist deshalb überhaupt dasselbe Maß.

(ii) Den Fall von  $(h \odot \lambda) \circ T^{-1}$  können Sie vermutlich jetzt schon selbst klären. ■

Aus der Analysis kennen Sie natürlich längst die klassische Substitutionsregel, die sich oft als Schlüssel zur Berechnung konkreter Integrale erweist. Obwohl sie auf den ersten Blick ein wenig anders aussieht, ist sie in der allgemeinen Transformationsformel enthalten. Wir überzeugen uns kurz davon, und erweitern sie auch gleich auf alle Lebesgue-integrierbaren Funktionen.

**Satz 3.12 (Substitutionsregel für eindimensionale L-Integrale)** *Es sei  $\psi : [\alpha, \beta] \rightarrow [a, b]$  ein  $\mathcal{C}^1$ -Diffeomorphismus, und  $f : [a, b] \rightarrow \mathbb{R}$  Lebesgue-messbar. Dann gilt*

$$\int_a^b f(x) dx = \int_\alpha^\beta f(\psi(s)) |\psi'(s)| ds \quad (3.19)$$

*sobald eines dieser Integrale definiert ist.*

**Beweis.** Der Übersichtlichkeit halber nehmen wir an, dass  $\psi$  wachsend ist. Durch  $T := \psi^{-1}$  wird ein  $\mathcal{C}^1$ -Diffeomorphismus  $T : [a, b] \rightarrow [\alpha, \beta]$  definiert. Betrachte  $g := f \circ \psi$ , dann ist  $f = g \circ T$ , und somit  $\int_a^b f(x) dx = \int_{[a,b]} (g \circ T) d\lambda$ . Andererseits haben wir  $\psi' = 1/T' \circ T^{-1}$  und daher wegen (3.17),

$$\begin{aligned} \int_\alpha^\beta f(\psi(s)) \psi'(s) ds &= \int_{[\alpha,\beta]} \frac{g}{T' \circ T^{-1}} d\lambda \\ &= \int_{[\alpha,\beta]} g d\left(\frac{1}{T' \circ T^{-1}} \odot \lambda\right) = \int_{[\alpha,\beta]} g d(\lambda \circ T^{-1}). \end{aligned}$$

Satz 3.7 garantiert nun, dass wirklich  $\int_{[a,b]} (g \circ T) d\lambda = \int_{[\alpha,\beta]} g d(\lambda \circ T^{-1})$  gilt. ■

## 3.6 Räume integrierbarer Funktionen

In vielen Bereichen der ("reinen" wie auch der "angewandten") höheren Analysis erweist es sich als außerordentlich nützlich, wenn man relevante Funktionen als Elemente eines geeigneten Funktionenraumes auffassen kann. Dieser soll neben einer (meist naheliegenden) algebraischen Struktur auch eine dazu passende topologische Struktur tragen, sodass man etwa über Konvergenz von Funktionenfolgen sprechen kann. Damit wird er den leistungsfähigen Hilfsmitteln der Funktionalanalysis zugänglich. Die wichtigsten Beispiele dafür liefern *normierte Räume* in denen man sich - sofern sie *vollständig* sind - bestens aufgehoben weiß.

Zumindest einen nützlichen Größenbegriff für Funktionen kennen Sie längst: auf dem Raum  $\mathcal{C}[0, 1]$  stetiger Funktionen  $f : [0, 1] \rightarrow \mathbb{R}$  ist die *Supremumsnorm* durch  $\|f\|_{\mathcal{C}[0,1]} := \sup_{x \in [0,1]} |f(x)|$  gegeben. Damit verbunden ist automatisch eine Abstandsfunktion, mittels  $d_{\mathcal{C}}(f, g) := \|f - g\|_{\mathcal{C}[0,1]}$  für  $f, g \in \mathcal{C}[0, 1]$  wird die *uniforme Metrik* definiert. Konvergenz  $f_n \rightarrow f$  im Sinne derselben, also  $d_{\mathcal{C}}(f_n, f) = \|f_n - f\|_{\mathcal{C}[0,1]} \rightarrow 0$  bedeutet gerade *gleichmäßige Konvergenz* der Funktionenfolge  $(f_n)$  gegen  $f$ . Über den Nutzen dieses Konzeptes wissen Sie längst bescheid. Es gibt aber noch eine Reihe anderer, ebenso wichtiger Konvergenzbegriffe für Funktionen, die man durch geeignete Normen beschreiben kann.

**A) Der Raum  $\mathcal{L}^1$ .** Das Integral führt uns nun zu einem recht natürlichen neuen Abstandsbegriff. Es sei  $(X, \mathcal{A}, \mu)$  ein beliebiger Maßraum. Wir schreiben

$$\mathcal{L}^1 = \mathcal{L}^1(\mu) = \mathcal{L}^1(X, \mathcal{A}, \mu) := \{f : X \rightarrow \mathbb{R} \text{ integrierbar}\}.$$

(auch die Schreibweise  $\mathcal{L}_1 = \mathcal{L}_1(\mu) = \mathcal{L}_1(X, \mathcal{A}, \mu)$  ist gebräuchlich). Falls  $X \subseteq \mathbb{R}$  ein Intervall und  $\mu = \lambda$  ist, dann schreiben wir auch einfach  $\mathcal{L}^1(X)$ , also zB

$\mathcal{L}^1[0, 1] = \mathcal{L}^1([0, 1], \mathcal{L}_{[0,1]}, \lambda)$ . Aus der Linearität des Integrals folgt sofort, dass  $\mathcal{L}^1$  ein linearer Raum über  $\mathbb{R}$  ist, wobei  $f + g$  und  $cf$  mit  $c \in \mathbb{R}$  natürlich wie schon zuvor punktweise definiert sind (also  $(f + cg)(x) = f(x) + cg(x)$ ). Das Nullelement ist dann natürlich die konstante Funktion 0. Anstatt wie in  $d_{\mathcal{L}}(f, g)$  den maximalen Abstand der Funktionswerte zu betrachten, sehen wir uns jetzt

$$d_{\mathcal{L}^1}(f, g) := \int_X |f - g| \, d\mu \quad \text{für } f, g \in \mathcal{L}^1 \quad (3.20)$$

an. Ist  $\mu$  normiert, dann ist dies gerade der durchschnittliche oder *mittlere Abstand* zwischen  $f(x)$  und  $g(x)$ , anschaulich der Inhalt der Fläche zwischen den beiden Funktionsgraphen. Ähnlich wie oben können wir diesen auch ausdrücken als

$$d_{\mathcal{L}^1}(f, g) = \|f - g\|_1 \quad \text{wobei} \quad \|f\|_1 := \int_X |f| \, d\mu \quad \text{für } f \in \mathcal{L}^1. \quad (3.21)$$

Das Funktional  $\|\cdot\|_1 : \mathcal{L}^1 \rightarrow [0, \infty)$  misst die mittlere Größe von  $|f(x)|$ . Nachdem schon von Normen die Rede war, überprüfen wir gleich, wie es diesbezüglich bei  $\|\cdot\|_1$  aussieht. Leider ist dies ja keine Norm, sondern nur eine *Halbnorm*, dh ein Funktional welches (HN1)-(HN3) unten erfüllt. Aus  $\|f\|_1 = 0$  folgt aber nicht zwingend  $f = 0$ .

**Proposition 3.9 (Halbnorm-Eigenschaften von  $\|\cdot\|_1$ )** Für  $f, g \in \mathcal{L}^1$  und  $c \in \mathbb{R}$  gelten

$$f = 0 \implies \|f\|_1 = 0, \quad (\text{HN1})$$

und

$$\|cf\|_1 = |c| \|f\|_1, \quad (\text{HN2})$$

sowie

$$\|f + g\|_1 \leq \|f\|_1 + \|g\|_1. \quad (\text{HN3})$$

Es ist

$$\|f\|_1 = 0 \iff f \in \mathcal{N} := \{g \in \mathcal{L}^1 : g = 0 \text{ f.ü.}\}.$$

**Beweis.** Die ersten beiden Aussagen beweisen sich praktisch selbst, für die dritte verwende man die gewöhnliche Dreiecksungleichung  $|f + g| \leq |f| + |g|$  um  $\|f + g\|_1 = \int |f + g| \, d\mu \leq \int (|f| + |g|) \, d\mu = \|f\|_1 + \|g\|_1$  zu bekommen. Die letzte Behauptung folgt sofort aus Proposition 3.6 a). ■

**Beispiel 3.10 a)** Im Fall  $(X, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{L}_{\mathbb{R}}, \lambda)$  ist  $f := 1_{\{\pi\}}$  nicht die Nullfunktion, erfüllt aber  $\|f\|_1 = 0$ . Hier kann  $\|\cdot\|_1$  nicht zwischen diesem  $f$  und 0 unterscheiden.

**b)** Wenn  $(X, \mathcal{A}, \mu) = (\{1, \dots, d\}, \mathcal{P}(\{1, \dots, d\}), \#)$ , dann ist  $g = 0$  f.ü. genau wenn  $g = 0$ , weil  $\emptyset$  die einzige Nullmenge dieses Raumes ist. In diesem einfachen Spezialfall ist  $\|\cdot\|_1$  also sogar eine Norm. Das wussten Sie aber schon, der Raum  $\mathcal{L}^1$  besteht in diesem Fall nämlich genau aus den  $d$ -Tupeln der Funktionswerte  $t := (t_1, \dots, t_d) = (f(1), \dots, f(d))$ , sodass wir  $\mathcal{L}^1$  und  $\mathbb{R}^d$  identifizieren können, und dann ist  $\|t\|_1 = \sum_{j=1}^d |t_j|$ , bekanntlich eine Norm auf  $\mathbb{R}^d$ .

Aufgrund des besprochenen Defektes ist nun  $d_{\mathcal{L}^1}$  im allgemeinen auch keine Metrik, sondern nur eine *Halbmetrik* ( $f = g \implies d_{\mathcal{L}^1}(f, g) = 0$ ,  $d_{\mathcal{L}^1}(f, g) = d_{\mathcal{L}^1}(g, f)$  und  $d_{\mathcal{L}^1}(f, h) \leq d_{\mathcal{L}^1}(f, g) + d_{\mathcal{L}^1}(g, h)$ ), aber der Abstands begriff ist so naheliegend, dass wir trotzdem über Konvergenz etc sprechen möchten.

**Definition 3.9** Sind  $f, f_n \in \mathcal{L}^1$  ( $n \geq 1$ ), so sagen wir,  $(f_n)$  konvergiert in  $\mathcal{L}^1$  (oder im Mittel) gegen  $f$ , geschrieben  $f_n \xrightarrow{\mathcal{L}^1} f$  oder  $f_n \longrightarrow f$  in  $\mathcal{L}^1$ , falls  $\lim_{n \rightarrow \infty} \|f_n - f\|_1 = \lim_{n \rightarrow \infty} d_{\mathcal{L}^1}(f_n, f) = 0$ .

Das einzig ungewohnte an diesem Konvergenzbegriff ist, dass die Grenzfunktion ggf nicht eindeutig bestimmt zu sein braucht, sondern nur bis auf eine Nullmenge festgelegt ist.

**Beispiel 3.11** Im Beispiel a) oben gilt  $zB$  mit  $f_n = 0$  dass  $f_n \xrightarrow{\mathcal{L}^1} 0$  und  $f_n \xrightarrow{\mathcal{L}^1} f$  obwohl  $f \neq 0$ .

Das macht aber nicht zuviel aus, wir nehmen das vorerst hin, besprechen weiter unten aber einen Ausweg. Es ist nun wichtig sich klarzumachen, dass wir dadurch ein neues Konzept eingeführt haben, es besteht kein ganz simpler Zusammenhang zwischen Konvergenz in  $\mathcal{L}^1$  und anderen Konvergenzarten.

**Beispiel 3.12 a)** Eine punktweise gegen ein  $f \in \mathcal{L}^1$  konvergente Folge  $(f_n)_{n \geq 1}$  in  $\mathcal{L}^1$  braucht nicht im Mittel gegen  $f$  zu konvergieren: Man betrachte etwa den Raum  $((0, 1], \mathcal{L}_{(0,1]}, \lambda)$  und die Funktionen  $f_n := n \mathbf{1}_{(0, 1/n]}$ ,  $n \geq 1$ , und  $f := 0$ .  
**b)** Es ist auch möglich, dass  $f_n \rightarrow f$  in  $\mathcal{L}^1$  obwohl die Folge  $(f_n(x))_{n \geq 1}$  für kein  $x \in X$  konvergiert. In obigem Maßraum wähle man  $zB$  die Folge der  $f_n = \mathbf{1}_{A_n}$ , wobei die  $A_n$  immer kleiner werdende Intervalle seien, die aber jeden Punkt unendlich oft überdecken, etwa  $(0, 1]$ ,  $(0, 1/2]$ ,  $(1/2, 1]$ ,  $(0, 1/4]$ ,  $(1/4, 2/4]$ ,  $(2/4, 3/4]$ ,  $(3/4, 1]$ ,  $(0, 1/8]$ ,  $(1/8, 2/8]$ , ... etc.

Konvergenz im Mittel wird Ihnen noch oft begegnen, für den Moment deute ich nur eine konkrete Situation an, wo sie der "richtige" Konvergenzbegriff ist.

**Beispiel 3.13** Wir stellen uns vor, dass auf  $(X, \mathcal{A}, \mu)$  endliche Maße  $\nu_n, \nu$  mit Dichten  $h_n, h$  gegeben sind ( $n \geq 1$ ). Da ist es naheliegend, anstelle der Maße mit den vertrauteren Objekten  $h_n, h$  zu arbeiten, das sind ja einfach Funktionen. Wollen wir nun verstehen, ob  $\nu_n \rightarrow \nu$  in dem (punktweisen) Sinne konvergiert, dass  $\nu_n(A) \rightarrow \nu(A)$  für jedes  $A \in \mathcal{A}$ , dann hilft es nichts, zu prüfen ob  $h_n \rightarrow h$  f.ü. gilt. (Daraus folgt nämlich nicht einmal  $\nu_n(X) \rightarrow \nu(X)$ , siehe Bsp a) oben.) Falls aber  $h_n \xrightarrow{\mathcal{L}^1} h$ , dann finden wir für beliebiges  $A \in \mathcal{A}$ ,

$$|\nu_n(A) - \nu(A)| = \left| \int_A (h_n - h) d\mu \right| \leq \int_A |h_n - h| d\mu \leq \|h_n - h\|_1 \rightarrow 0,$$

also folgt  $\nu_n \rightarrow \nu$ .

Was den (halb)normierten Raum  $(\mathcal{L}^1, \|\cdot\|_1)$  besonders schön und gutmütig macht, ist aber die grundlegend wichtige Tatsache, dass er *vollständig* ist.  $(f_n)$  ist eine *Cauchy-Folge* in  $\mathcal{L}^1$  falls es zu jedem  $\varepsilon > 0$  ein  $N_\varepsilon$  gibt mit  $\|f_n - f_m\|_1 < \varepsilon$  für  $m, n \geq N_\varepsilon$ .

**Satz 3.13 (Vollständigkeit von  $\mathcal{L}^1$ . Satz von Riesz-Fischer)**  $(\mathcal{L}^1, \|\cdot\|_1)$  ist ein vollständiger Raum, dh zu jeder Cauchy-Folge  $(f_n)_{n \geq 1}$  in  $\mathcal{L}^1$  gibt es ein  $f \in \mathcal{L}^1$  mit  $\|f_n - f\|_1 \rightarrow 0$ .

**Beweis.** \* Wir müssen zu der vorgegebenen Cauchy-Folge  $(f_n)_{n \geq 1}$  in  $\mathcal{L}^1$  eine Grenzfunktion  $f$  finden, die natürlich nicht explizit angegeben werden kann. Daher wird (wie üblich) die Existenz dieses Grenzbildes mithilfe eines Vollständigkeitsargumentes gezeigt. Da wir die Vollständigkeit von  $\mathcal{L}^1$  natürlich noch nicht heranziehen können, wollen wir jene von  $\mathbb{R}$  nutzen, und Konvergenz der reellen Funktionswerte  $f_n(x)$  erreichen, wodurch wir eine Grenzfunktion definieren könnten. Allerdings: das Beispiel oben zeigt, dass man  $\mathcal{L}^1$ -Konvergenz nicht leicht mit punktwiser Konvergenz in Verbindung bringen kann. Der wesentliche Punkt im folgenden Argument ist die Einsicht, dass man wenigstens immer eine *Teilfolge* bekommt, die f.ü. konvergiert, und damit zur Definition von  $f$  genutzt werden kann.

(i) Wir zeigen: Es gibt eine Teilfolge  $(f_{n_k})_{k \geq 0}$  ( $n_k \nearrow \infty$ ) die f.ü. gegen ein messbares  $f$  konvergiert. Wir beginnen mit  $n_0 := 1$ . Falls Indices  $n_0, \dots, n_{k-1}$  gegeben sind, wählen wir  $n_k > n_{k-1}$  so groß, dass

$$\|f_n - f_m\|_1 < 2^{-k} \quad \text{für } n, m \geq n_k.$$

Insbesondere gilt dann  $\|f_{n_{k+1}} - f_{n_k}\|_1 < 2^{-k}$  für  $k \geq 1$ . Wir betrachten die (messbaren) Funktionen  $g_k := f_{n_{k+1}} - f_{n_k}$ ,  $k \geq 1$ , und  $g := \sum_{k \geq 1} |g_k|$ . Dann konvergiert die wachsende Folge der nichtnegativen messbaren Funktionen  $\sum_{j=1}^k |g_j|$  gegen  $g$ , und wegen

$$\left\| \sum_{j=1}^k |g_j| \right\|_1 \leq \sum_{j=1}^k \|g_j\|_1 \leq \sum_{j=1}^k 2^{-j} \leq 1$$

folgt (monotone Konvergenz)

$$\int g d\mu = \lim_{k \rightarrow \infty} \int \sum_{j=1}^k |g_j| d\mu = \lim_{k \rightarrow \infty} \left\| \sum_{j=1}^k |g_j| \right\|_1 \leq 1.$$

Insbesondere ist also  $g$  integrierbar und damit f.ü. endlich, dh es gibt eine  $\mu$ -Nullmenge  $A \in \mathcal{A}$ ,  $\mu(A) = 0$ , sodass  $g < \infty$  auf  $A^c$ , dh auf  $A^c$  ist  $\sum_{k \geq 1} |g_k|$  konvergent, was auch die Konvergenz der Folge  $(\sum_{j=1}^k g_j)_{k \geq 1} = (f_{n_{k+1}} - f_{n_1})_{k \geq 1}$  auf  $A^c$  impliziert. Dann folgt auch Konvergenz von  $(f_{n_k})_{k \geq 1}$  auf  $A^c$ . Um eine auf ganz  $X$  definierte messbare reellwertige Grenzfunktion  $f$  mit  $f_{n_k} \rightarrow f$  f.ü. hinzuschreiben, können wir nun zB  $f := 1_{A^c} \overline{\lim}_{k \rightarrow \infty} f_{n_k}$  setzen.

(ii) Wir prüfen jetzt nach, dass die eben gewonnene Funktion  $f$  tatsächlich sogar zu  $\mathcal{L}^1$  gehört. Dazu sei  $h := 1_{A^c}(|f_{n_1}| + g)$ , das liegt in  $\mathcal{L}^1$  weil  $1_{A^c} f_{n_1}, 1_{A^c} g \in \mathcal{L}^1$  (für  $g$  siehe oben). Wegen

$$|1_{A^c} f_{n_k}| = 1_{A^c} \left| f_{n_1} + \sum_{j=1}^{k-1} g_j \right| \leq 1_{A^c} |f_{n_1}| + 1_{A^c} \sum_{j=1}^{k-1} |g_j| \leq h$$

folgt  $|f| \leq h$ , und damit  $\|f\|_1 \leq \|h\|_1 < \infty$ , wie gefordert.

(iii) Wir können nun über  $\mathcal{L}^1$ -Konvergenz gegen  $f$  sprechen, und halten fest, dass diese zumindest entlang unserer Teilfolge stattfindet. Setzen wir nämlich  $h_k := |f_{n_k} - f|$ ,  $k \geq 1$ , dann gilt ja  $h_k \rightarrow 0$  f.ü. Wegen  $|h_k| \leq 2h$  und der Integrierbarkeit von  $h$  folgt (dominierte Konvergenz), dass  $\int |h_k| d\mu \rightarrow 0$ , dh  $\|f_{n_k} - f\|_1 \rightarrow 0$ .

Schließlich haben wir uns davon zu überzeugen, dass tatsächlich die volle Folge  $(f_n)_{n \geq 1}$  in  $\mathcal{L}^1$  gegen  $f$  konvergiert. Das ergibt sich jetzt aber ganz leicht aus der Cauchy-Eigenschaft: Sei  $\varepsilon > 0$ , dann gilt  $\|f_n - f_m\|_1 < \varepsilon/2$  für  $m, n \geq N_{\varepsilon/2}$ . Da es aber ein  $n_k \geq N_{\varepsilon/2}$  mit  $\|f_{n_k} - f\|_1 < \varepsilon/2$  gibt, folgt

$$\|f_n - f\|_1 \leq \|f_n - f_{n_k}\|_1 + \|f_{n_k} - f\|_1 < \varepsilon \quad \text{für } n \geq N_{\varepsilon/2}.$$

Also gilt  $f_n \rightarrow f$  in  $\mathcal{L}^1$ , womit der Nachweis der Vollständigkeit erbracht ist. ■

**B) Nullmengen vergessen. Der Raum  $L^1$ .** Wie bereits festgehalten, bedeutet die fehlende Eindeutigkeit von Grenzwerten in  $(\mathcal{L}^1, \|\cdot\|_1)$  kein ernsthaftes Problem, da ja immerhin alles bis auf Nullmengen festgelegt ist. Man kann diesen Mangel aber überhaupt beheben, indem man f.ü. übereinstimmende Funktionen  $f, g \in \mathcal{L}^1$  miteinander identifiziert (dh als äquivalent bezeichnet,  $f \sim g$ ), und die so erhaltenen Äquivalenzklassen als Punkte eines neuen Raumes ansieht. Dies verträgt sich mit der linearen Struktur der Räume  $\mathcal{L}^1$ , da

$$f \sim g \Leftrightarrow f - g \in \mathcal{N} := \{h \in \mathcal{L}^1 : h = 0 \text{ f.ü.}\},$$

wobei (wie man sehr leicht sieht)  $\mathcal{N}$  ein linearer Teilraum von  $\mathcal{L}^1$  ist. Dh die Äquivalenzklassen sind gerade die *Nebenklassen*  $[f] = f + \mathcal{N}$ , und diese bilden den *Quotientenraum*

$$L^1 = L^1(\mu) = L^1(X, \mathcal{A}, \mu) := \mathcal{L}^1 / \mathcal{N} = \{[f] = f + \mathcal{N} : f \in \mathcal{L}^1\},$$

der automatisch eine lineare Struktur trägt,  $c[f] + [g] = cf + c\mathcal{N} + g + \mathcal{N} = cf + g + \mathcal{N} = [cf + g]$  weil  $c\mathcal{N} + \mathcal{N} = \mathcal{N}$ , da  $\mathcal{N}$  eben ein Untervektorraum ist. Nun ist aber  $\mathcal{N} = \{f : \|f\|_1 = 0\}$ ,  $\|f\|_1$  hängt nur von der Äquivalenzklasse von  $f$  ab, und wir können daher gefahrlos

$$\|[f]\|_1 := \|f\|_1$$

für beliebiges  $[f] \in L^1$  definieren. Das liefert jedenfalls eine Halbnorm auf  $L^1$ . Da wir aber gerade den Nullraum  $\mathcal{N}$  von  $\|\cdot\|_1$  zu einem einzigen Punkt  $[0]$  zusammenfassen, gilt nun  $\|[f]\|_1 = 0$  genau wenn  $[f] = [0]$ , somit haben wir nun wirklich eine Norm vorliegen. Konvergenz  $[f_n] \rightarrow [f]$  in derselben bedeutet eben, dass (bei einer - und damit jeder - Wahl von Repräsentanten  $f_n$  und  $f$ )  $\|f_n - f\|_1 \rightarrow 0$  konvergiert, weshalb sich die oben festgestellte Vollständigkeit auf  $(L^1, \|\cdot\|_1)$  überträgt. Ein vollständiger normierter Raum wird übrigens als *Banachraum* bezeichnet. Wir erhalten also

**Satz 3.14**  $(L^1, \|\cdot\|_1)$  ist ein Banachraum.

Tatsächlich verzichtet man "in der Praxis" oft darauf, genau zwischen  $\mathcal{L}^1$  und  $L^1$  bzw zwischen  $f$  und  $[f]$  in der Notation zu unterscheiden, eben weil man für viele Zwecke f.ü. übereinstimmende Funktionen ohne weiteres identifizieren kann. Man kann zB Elemente von  $L^1$  integrieren, indem man (unzweideutig) eben  $\int [f] d\mu := \int f d\mu$  für einen beliebigen Repräsentanten  $f$  setzt.

**C) Es geht nicht ohne Lebesgue.** Na gut, sagen Sie, dann ist das eben ein hübscher Raum. Wäre ja noch schöner, wenn es nach dem riesigen Aufwand, den wir für Lebesgue-Maß und -Integral getrieben haben nicht auch ein paar Vorteile gäbe. Ganz überzeugt sind Sie aber vielleicht noch nicht. Falls Sie etwa primär an Situationen interessiert sind, in denen es um stetige Funktionen, zB auf  $[0, 1]$  geht, so bemerken Sie, dass man den Abstand  $d_{\mathcal{L}^1}(f, g)$  für  $f, g \in \mathcal{C}[0, 1]$  dort genauso schon mit dem Riemann-Integral hätte definieren können.

Das stimmt auch, und auf  $\mathcal{C}[0, 1]$  eingeschränkt ist  $\|\cdot\|_1$  sogar eine Norm! (Überzeugen Sie sich bitte davon.) Allerdings ist der normierte Raum  $(\mathcal{C}[0, 1], \|\cdot\|_1)$  nicht vollständig:

**Beispiel 3.14** In  $\mathcal{C}[0, 1]$  betrachten wir die Funktionenfolge  $(f_n)_{n \geq 2}$  mit

$$f_n(x) := \begin{cases} 0 & \text{für } x \in [0, \frac{1}{2} - \frac{1}{2^n}], \\ 2^n x - (2^{n-1} - 1) & \text{für } x \in [\frac{1}{2} - \frac{1}{2^n}, \frac{1}{2}], \\ 1 & \text{für } x \in [\frac{1}{2}, 1]. \end{cases}$$

(Hier wächst  $f_n$  innerhalb des kurzen Intervalles  $[\frac{1}{2} - \frac{1}{2^n}, \frac{1}{2}]$  affin von 0 auf 1, die entsprechenden Stücke der Graphen werden mit  $n$  aber immer steiler.) Es ist leicht zu sehen, dass  $(f_n)$  eine Cauchy-Folge ist, konkret ist stets  $\|f_n - f_m\|_1 < 2^{-\min(n, m)}$ . Eine Skizze legt nahe, dass  $(f_n)$  gegen  $f := 1_{[\frac{1}{2}, 1]}$  konvergieren möchte. Tatsächlich gilt  $\|f_n - f\|_1 < 2^{-n}$  für alle  $n$ , und daher  $f_n \rightarrow f$  in  $\mathcal{L}^1[0, 1] = \mathcal{L}^1([0, 1], \mathcal{L}_{[0, 1]}, \lambda)$ . Nun liegt  $f$  aber nicht in dem Teilraum  $\mathcal{C}[0, 1]$  von  $\mathcal{L}^1$ , und auch keine Funktion  $g$  mit  $g = f$  f.ü. ist stetig. Dies sind aber schon alle Grenzwerte unter  $\|\cdot\|_1$  von  $(f_n)$  (siehe oben), und da keiner in  $\mathcal{C}[0, 1]$  liegt, sehen wir, dass  $(f_n)$  in  $(\mathcal{C}[0, 1], \|\cdot\|_1)$  nicht konvergiert.

Möchte man, von den stetigen Funktionen auf  $[0, 1]$  ausgehend, in einem bzgl  $\|\cdot\|_1$  vollständigen Raum arbeiten, so muss man wenigstens die Grenzfunktion  $f$  aus dem Beispiel, und überhaupt die Grenzfunktionen aller  $\|\cdot\|_1$ -Cauchy-Folgen aus  $\mathcal{C}[0, 1]$  hinzunehmen. (Das liefert dann die Vervollständigung von  $\mathcal{C}[0, 1]$  bzgl  $\|\cdot\|_1$ , dh den kleinsten vollständigen Raum der  $\mathcal{C}[0, 1]$  enthält). Nun gut, wie kompliziert können diese Grenzfunktionen schon sein? Im Beispiel ist  $f$  ja recht harmlos. Dieser Eindruck führt allerdings in die Irre: Tatsächlich ist jede Funktion aus  $\mathcal{L}^1$  Limes einer  $\|\cdot\|_1$ -Cauchy-Folge aus  $\mathcal{C}[0, 1]$ . Um  $\mathcal{C}[0, 1]$  zu einem vollständigen Raum zu machen, muss man also genau die Lebesgue-integrierbaren Funktionen hinzunehmen! (Genauer: man muss aus jeder Klasse  $[f] \in L^1$  einen Repräsentanten nehmen, oder am besten gleich  $L^1$  als den erweiternden Raum betrachten.) Beachten Sie bitte: die ursprüngliche Zielsetzung hatte nicht von vorneherein mit Maßtheorie zu tun, es stellt sich aber heraus, dass die Lebesguesche Theorie genau den richtigen Raum liefert, der das Gewünschte leistet. (Weitere ähnliche Situationen gibt es auch anderswo in der höheren Analysis, darauf können wir hier aber nicht eingehen.) Nun aber endlich zur Rechtfertigung dieser Lobrede:

**Satz 3.15** Für jede Funktion  $f \in \mathcal{L}^1[0, 1]$  gibt es eine Folge  $(f_n)$  in  $\mathcal{C}[0, 1]$  mit  $f_n \rightarrow f$  in  $\mathcal{L}^1$ . Anders gesagt:  $\mathcal{C}[0, 1]$  liegt dicht in  $(\mathcal{L}^1[0, 1], \|\cdot\|_1)$ .

Den Beweis des Satzes bereiten wir durch folgende Beobachtung vor.

**Lemma 3.4** Eine Menge  $A \subseteq [0, 1]$  ist genau dann Lebesgue-messbar,  $A \in \mathcal{L}_{[0, 1]}$ , wenn es zu jedem  $\delta > 0$  eine kompakte Teilmenge  $K \subseteq A$  und eine offene Obermenge  $G \supseteq A$  gibt mit  $\lambda(G \setminus K) < \delta$ .

**Beweis.** Wir zeigen, dass jedes  $A \in \mathcal{L}_{[0, 1]}$  die genannte Eigenschaft hat. Die Umkehrung ist eine Übungsaufgabe. Seien also  $A \in \mathcal{L}_{[0, 1]}$  und  $\delta > 0$ .

Wegen  $\lambda(A) = \lambda^*(A)$  gibt es eine abzählbare  $\mathcal{I}_{\mathbb{R}}$ -Überdeckung  $\{(a_k, b_k)\}_{k \geq 1}$  von  $A$  mit  $\sum_{k \geq 1} (b_k - a_k) < \lambda(A) + \frac{\delta}{4}$ . Setzen wir  $\beta_k := b_k + 2^{-k} \frac{\delta}{4}$ , dann ist  $\{(a_k, \beta_k)\}_{k \geq 1}$  eine (größere) offene Überdeckung von  $A$  mit  $\sum_{k \geq 1} (\beta_k - a_k) < \lambda(A) + \frac{\delta}{2}$ . Die offene Menge  $G := \bigcup_{k \geq 1} (a_k, \beta_k)$  enthält also  $A$  und es gilt  $\lambda(G) \leq \sum_{k \geq 1} (\beta_k - a_k) < \lambda(A) + \frac{\delta}{2}$ .

Indem wir dieses Teilresultat auf  $[0, 1] \setminus A$  anwenden, bekommen wir auch eine offene Menge  $U \supseteq [0, 1] \cap A^c$  mit  $\lambda(U) < 1 - \lambda(A) + \frac{\delta}{2}$ . Dann ist aber  $K := [0, 1] \cap U^c$  eine beschränkte und abgeschlossene, also kompakte Menge mit  $K \subseteq A$  und  $\lambda(K) > \lambda(A) - \frac{\delta}{2}$ . Damit leisten  $K$  und  $G$  das Gewünschte. ■

Nun sind wir schon gerüstet für den

**Beweis von Satz 3.15.** (i) Wir beweisen zuerst: Ist  $A \in \mathcal{L}_{[0,1]}$  und  $\delta > 0$ , dann gibt es eine stetige Funktion  $h \in \mathcal{C}[0, 1]$  mit  $\|1_A - h\|_1 < \delta$ . Dazu seien  $K$  und  $G$  wie im Lemma. Betrachte die abgeschlossene Menge  $F := G^c$ , diese schneidet  $K$  nicht. Daher hat die kompakte Menge  $K$  positiven Abstand zu  $F$ ,  $\text{dist}(F, K) = \inf\{|x - y| : x \in F, y \in K\} > 0$ . Wir setzen

$$h(x) := 1 - \min\left(1, \frac{\text{dist}(x, K)}{\text{dist}(F, K)}\right), \quad x \in [0, 1].$$

Weil der Abstand  $\text{dist}(x, K) = \inf\{|x - y| : y \in K\}$  von  $x$  zu  $K$  stetig von  $x$  abhängt, ist  $h \in \mathcal{C}[0, 1]$ . Weiters ist  $0 \leq h \leq 1$ . Indem wir einsetzen, finden wir ausserdem dass  $h(x) = 1$  für  $x \in K$  und  $h(x) = 0$  für  $x \in F$  gelten. Die Funktion  $h$  liegt also zwischen  $1_K$  und  $1_G$  indem  $1_K \leq h \leq 1_G$ . Für  $1_A$  gilt natürlich auch  $1_K \leq 1_A \leq 1_G$  (wegen  $K \subseteq A \subseteq G$ ). Damit bekommen wir wie gewünscht

$$\|1_A - h\|_1 \leq \|1_G - 1_K\|_1 = \lambda(G \setminus K) < \delta.$$

(ii) Im allgemeinen Fall reicht es  $f \geq 0$  zu betrachten. (Ansonsten verwende man  $f^+$  und  $f^-$ .) Sei  $\varepsilon > 0$ . Wir müssen zeigen, dass es ein  $g \in \mathcal{C}[0, 1]$  gibt mit  $\|f - g\|_1 < \varepsilon$ . Nach Definition des Integrals von  $f$  gibt es ein  $f_n \in \mathcal{T}^+$  mit  $f_n \leq f$  und  $\|f - f_n\|_1 = \int f \, d\mu - \int f_n \, d\mu < \frac{\varepsilon}{2}$ . Also reicht es,  $g \in \mathcal{C}[0, 1]$  zu finden mit  $\|f_n - g\|_1 < \frac{\varepsilon}{2}$ .

Wir haben  $f_n = \sum_{j=1}^m \alpha_j 1_{A_j}$  mit geeigneten  $\alpha_j \geq 0$  und  $A_j \in \mathcal{L}_{[0,1]}$ , und nehmen an, dass nicht alle  $\alpha_j$  gleich 0 sind. (Sonst geht es mit  $g = 0$ .) Setze  $\delta := \varepsilon / (2m \max(\alpha_1, \dots, \alpha_m))$ , dann gibt es nach Schritt (i) zu jedem  $j$  ein  $h_j \in \mathcal{C}[0, 1]$  mit  $\|1_{A_j} - h_j\|_1 < \delta$ . Die Funktion  $g := \sum_{j=1}^m \alpha_j h_j$  ist natürlich wieder stetig, und wir sehen, dass

$$\|f_n - g\|_1 = \left\| \sum_{j=1}^m \alpha_j (1_{A_j} - h_j) \right\|_1 \leq \sum_{j=1}^m \alpha_j \|1_{A_j} - h_j\|_1 < \sum_{j=1}^m \alpha_j \delta \leq \frac{\varepsilon}{2},$$

wie gefordert. ■

**D) Ausblick\*: Die Räume  $\mathcal{L}^p$  bzw  $L^p$ .** Wie bereits erwähnt (Beispiel 3.10), bekommen wir im Spezialfall einer endlichen Menge  $X = \{1, \dots, d\}$  mit Zählmaß  $\#$  gerade  $\mathcal{L}^1 \cong \mathbb{R}^d$  und  $\|t\|_1 = \|(t_1, \dots, t_d)\|_1 = \sum_{j=1}^d |t_j|$ . Im Fall des linearen Raumes  $\mathbb{R}^d$  kennen Sie natürlich auch schon andere Normen, etwa die *Euklidische Norm*  $\|t\|_2 = \|(t_1, \dots, t_d)\|_2 := (\sum_{j=1}^d t_j^2)^{1/2}$ , allgemeiner (für  $p \in [1, \infty)$ ) die *p-Norm*  $\|t\|_p = \|(t_1, \dots, t_d)\|_p := (\sum_{j=1}^d |t_j|^p)^{1/p}$ , und die *Maximumsnorm*  $\|t\|_\infty = \|(t_1, \dots, t_d)\|_\infty := \max_{1 \leq j \leq d} |t_j|$ . Erstere ist zurecht besonders prominent, da sie aufs Engste mit dem Innenprodukt  $\langle t, s \rangle := \sum_{j=1}^d t_j s_j$  verknüpft ist,  $\|t\|_2 = \langle t, t \rangle^{1/2}$ , und somit die Euklidische Geometrie des  $\mathbb{R}^d$  widerspiegelt. Insbesondere können wir da über Orthogonalität sprechen. Mitunter ist aber  $\|t\|_1$  oder  $\|t\|_\infty$  einfacher zu handhaben, weshalb es sehr nützlich ist zu wissen, dass auf einem endlichdimensionalen Raum alle Normen äquivalent sind (und damit alle denselben Konvergenzbegriff beschreiben). Dies ändert sich aber dramatisch, sobald wir uns unendlichdimensionalen Vektorräumen zuwenden. Genau das tun wir jetzt, wobei wir die genannten Normen von  $\mathbb{R}^d$  auf geeignete Räume messbarer Funktionen verallgemeinern.

Dazu betrachten wir den reellen Vektorraum  $\mathcal{L}^0 = \mathcal{L}^0(\mu) = \mathcal{L}^0(X, \mathcal{A}, \mu) := \{f : X \rightarrow \mathbb{R} \text{ messbar}\}$ . Um Analoga zu obigen Normen am  $\mathbb{R}^d$  auf Teilräumen von



$\mathcal{L}^0$  zu definieren, bemerken wir zunächst, dass für jedes  $f \in \mathcal{L}^0$  und  $p \in [1, \infty)$  auch  $|f|^p \in \mathcal{L}^0$  ist (da für jedes  $b \in \mathbb{R}$  ja  $\{|f|^p \leq b\} = \{-b^{1/p} \leq f \leq b^{1/p}\} \in \mathcal{A}$  ist). Somit macht folgende Vereinbarung Sinn:

**Definition 3.10** Für  $f \in \mathcal{L}^0$  und  $p \in [1, \infty)$  sei

$$\|f\|_p := \left( \int |f|^p d\mu \right)^{1/p} \in [0, \infty],$$

und wir bezeichnen mit  $\mathcal{L}^p = \mathcal{L}^p(\mu) = \mathcal{L}^p(X, \mathcal{A}, \mu) := \{f \in \mathcal{L}^0 : \|f\|_p < \infty\}$  den Raum der  $p$ -fach integrierbaren Funktionen.

**Beispiel 3.15** Für  $p = 1$  wird einfach  $|f|$  integriert, bei größerem  $p$  werden große Werte von  $|f|$  stärker, und kleine Werte geringer gewichtet. Für  $\alpha \in \mathbb{R}$  gilt bekanntlich (siehe PS)

$$\int_0^1 x^\alpha dx < \infty \iff \alpha > -1 \quad \text{und} \quad \int_1^\infty x^\alpha dx < \infty \iff \alpha < -1.$$

Ist  $f(x) := x^\alpha$  für ein  $\alpha \in \mathbb{R}$ , so sehen wir also unmittelbar, dass

$$1_{(0,1)} f \in \mathcal{L}^p(\lambda) \iff \alpha > -\frac{1}{p} \quad \text{und} \quad 1_{(1,\infty)} f \in \mathcal{L}^p(\lambda) \iff \alpha < -\frac{1}{p}.$$

Insbesondere sind die Räume  $\mathcal{L}^p[0,1]$ ,  $p \in [1, \infty)$ , alle verschieden. In endlichen Maßräumen kommt es für die Integrierbarkeit nur auf große Werte von  $|f|$  an. In unendlichen Maßräumen können auch kleine Werte Schwierigkeiten machen, wenn Sie eben auf großen Mengen auftreten.

Auf  $\mathcal{L}^p$  eingeschränkt ist die Funktion  $\|\cdot\|_p$  wieder beinahe eine Norm:

**Satz 3.16 (Die Räume  $(\mathcal{L}^p, \|\cdot\|_p)$ )** Für jedes  $p \in [1, \infty)$  ist  $\mathcal{L}^p$  ein linearer Raum, und  $\|\cdot\|_p : \mathcal{L}^p \rightarrow [0, \infty)$  ist eine Halbnorm mit  $\|f\|_p = 0 \iff f = 0$  f.ü.  $\|\cdot\|_p$

Um das nachzuweisen muss man nun ein wenig mehr tun als im Fall  $p = 1$  oben. Die Dreiecksungleichung beruht auf den klassischen Ungleichungen von Hölder und Minkowski (die wir übergehen). Analog zum Fall  $p = 1$  deklariert man dann:

**Definition 3.11** Eine Folge  $(f_n)_{n \geq 1}$  in  $\mathcal{L}^p$ ,  $p \in [1, \infty)$ , konvergiert im  $p$ -ten Mittel oder in  $\mathcal{L}^p$  gegen  $f \in \mathcal{L}^p$ , falls

$$\|f_n - f\|_p \longrightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Entsprechend nennt man  $(f_n)_{n \geq 1}$  eine Cauchy-Folge in  $\mathcal{L}^p$  falls es zu jedem  $\varepsilon > 0$  ein  $N_\varepsilon$  gibt mit

$$\|f_n - f_m\|_p < \varepsilon \quad \text{für } m, n \geq N_\varepsilon.$$

Konvergenz  $f_n \rightarrow f$  in  $\mathcal{L}^p$  bedeutet also  $\|f_n - f\|_p^p = \int |f_n - f|^p d\mu \rightarrow 0$ , dh man betrachtet nicht die Abstände von Funktionswerten  $|f_n(x) - f(x)|$  an festgehaltenen Stellen  $x$ , sondern interessiert sich nur für einen mittleren Abstand zwischen  $f_n$  und  $f$ . Dabei kann, indem man einen Exponenten  $p > 1$  wählt, größeren Abständen größeres Gewicht gegeben werden. Wie jeder durch eine (Halb-)Norm gegebene Konvergenzbegriff, ist auch dieser mit der linearen Struktur verträglich, dh Addition und Skalarmultiplikation sind stetig.

**Proposition 3.10 (Konvergenz in  $\mathcal{L}^p$ )** Für  $p \in [1, \infty)$ , gelten

- a) falls  $c \in \mathbb{R}$  und  $f_n \rightarrow f$ ,  $g_n \rightarrow g$  in  $\mathcal{L}^p$ , so folgt  $cf_n + g_n \rightarrow cf + g$  in  $\mathcal{L}^p$ ,
- b) aus  $f_n \rightarrow f$  und  $f_n \rightarrow g$  in  $\mathcal{L}^p$  folgt  $f = g$  f.ü.

**Beweis.** a) Dies ergibt sich unmittelbar aus den Halbnorm-Eigenschaften.  
b) Es ist ja  $0 \leq \|f - g\|_p \leq \|f - f_n\|_p + \|f_n - g\|_p \rightarrow 0$ , also  $\|f - g\|_p = 0$ , und somit  $|f - g| = 0$  f.ü. ■

Schliesslich gilt (mit ähnlichem Beweis wie im Fall  $p = 1$ )

**Satz 3.17 (Vollständigkeit von  $\mathcal{L}^p$ . Satz von Riesz-Fischer)** Für jedes  $p \in [1, \infty)$  ist  $(\mathcal{L}^p, \|\cdot\|_p)$  ein vollständiger Raum, dh zu jeder Cauchy-Folge  $(f_n)_{n \geq 1}$  in  $\mathcal{L}^p$  gibt es ein  $f \in \mathcal{L}^p$  mit  $\|f - f_n\|_p \rightarrow 0$ .

Parallel zu Satz 3.15 zeigt sich übrigens auch, dass für jedes  $p \in [1, \infty)$  der Teilraum  $\mathcal{C}[0, 1]$  dicht in  $(\mathcal{L}^p[0, 1], \|\cdot\|_p)$  liegt.

Die Halbnormen  $\|\cdot\|_p, p \in [1, \infty)$ , verallgemeinern die eingangs erwähnten Normen  $\|t\|_p$  auf  $\mathbb{R}^d$ . Auch für die Maximumsnorm  $\|t\|_\infty$  kann man eine entsprechende Version in  $\mathcal{L}^0$  definieren:

**Definition 3.12** Für  $f \in \mathcal{L}^0$  setzt man

$$\|f\|_\infty := \inf \{ \alpha \in [0, \infty] : |f| \leq \alpha \text{ f.ü.} \} \in [0, \infty]$$

(das wesentliche oder essentielle Supremum von  $|f|$ ,  $\text{ess sup } |f|$ ), und bezeichnet mit  $\mathcal{L}^\infty = \mathcal{L}^\infty(X, \mathcal{A}, \mu) := \{f \in \mathcal{L}^0 : \|f\|_\infty < \infty\}$  den Raum der wesentlich beschränkten Funktionen.

Den vorangegangenen ähnliche (aber etwas einfachere) Argumente zeigen, dass die für  $\mathcal{L}^p, p \in [1, \infty)$ , getroffenen Aussagen auch im Falle  $p = \infty$  gelten. (Wir verzichten auf den Beweis.)

**Satz 3.18 (Der Raum  $(\mathcal{L}^\infty, \|\cdot\|_\infty)$ )**  $\mathcal{L}^\infty$  ist ein linearer Raum, und  $\|\cdot\|_\infty : \mathcal{L}^\infty \rightarrow [0, \infty)$  ist eine vollständige Halbnorm mit  $\|f\|_\infty = 0$  genau wenn  $f = 0$  f.ü.

Wie zuvor kann man die Sache noch weiter veredeln, indem man f.ü. übereinstimmende Funktionen  $f, g \in \mathcal{L}^p$  miteinander identifiziert, also  $f \sim g \Leftrightarrow f - g \in \mathcal{N}^p$  setzt, wobei  $\mathcal{N}^p := \{h \in \mathcal{L}^p : h = 0 \text{ f.ü.}\}$  wieder ein linearer Teilraum von  $\mathcal{L}^p$  ist. Dh die Äquivalenzklassen sind gerade die Nebenklassen  $[f] = f + \mathcal{N}^p$ , und diese bilden die Quotientenräume

$$L^p = L^p(\mu) = L^p(X, \mathcal{A}, \mu) := \mathcal{L}^p / \mathcal{N}^p = \{[f] = f + \mathcal{N}^p : f \in \mathcal{L}^p\}, \quad p \in [1, \infty],$$

die automatisch eine lineare Struktur tragen, und auf denen durch  $\|[f]\|_p := \|f\|_p$  eine Norm gegeben ist. Das funktioniert alles genau wie im Fall  $p = 1$ , insbesondere erhalten wir

**Satz 3.19 (Die Räume  $(L^p, \|\cdot\|_p)$ )** Für  $p \in [1, \infty]$  ist  $(L^p, \|\cdot\|_p)$  ein Banachraum.

Stammt die Norm  $\|\cdot\|$  eines Banachraumes speziell von einem Innenprodukt  $\langle \cdot, \cdot \rangle$  ab, dh  $\|f\| = \langle f, f \rangle^{1/2}$ , so spricht man von einem Hilbertraum. Bei den  $L^p$ -Räumen ist dies genau für  $p = 2$  der Fall, mit

$$\langle f, g \rangle := \int f g d\mu, \quad f, g \in L^2. \quad (3.22)$$

Das hebt  $(L^2, \|\cdot\|_2)$  als einzigen Hilbertraum unter den  $L^p$ -Räumen hervor. Zwei Funktionen  $f, g \in L^2$  sind zueinander orthogonal wenn  $\langle f, g \rangle = \int f g d\mu = 0$ . In  $\mathbb{R}^d$  bewährt es sich, Elemente dieses Raumes besonders übersichtlich in einer Orthonormalbasis darzustellen. Im unendlichdimensionalen Fall verwendet man anstelle einer algebraischen Orthonormalbasis lieber ein vollständiges Orthonormalsystem  $(g_n)_{n \geq 0}$ , dh eine Folge zueinander orthogonaler ( $\langle g_n, g_m \rangle = 0$  für  $n \neq m$ ) und normierter ( $\langle g_n, g_n \rangle = 1$ ) Elemente mit der Eigenschaft, dass jedes Element  $f$  des Hilbertraumes als  $f = \sum_{n \geq 0} \langle f, g_n \rangle g_n$  dargestellt werden kann. Die unendliche Reihe ist dabei als Grenzwert der Partialsummen  $\sum_{n=0}^N \langle f, g_n \rangle g_n$  bezüglich der Norm zu verstehen.

Wir skizzieren ganz kurz eine Situation, welche die Bedeutung dieser Konzepte illustriert. (In einschlägigen LVen werden Sie noch viel mehr darüber erfahren.)

**Beispiel 3.16 (Fourierreihen)** In verschiedenen konkreten Fragen zB aus der Physik von Schwingungen und Wellen stellt sich die Frage, ob bzw wie eine Funktion  $f : [0, 2\pi] \rightarrow \mathbb{R}$  als Überlagerung (dh Summe) von Funktionen der Form

$$x \mapsto \frac{1}{\sqrt{2\pi}}, \frac{\sin(nx)}{\sqrt{\pi}}, \frac{\cos(nx)}{\sqrt{\pi}} \quad \text{mit } m, n \in \mathbb{N}^+, \quad (3.23)$$

dargestellt werden kann, dh man möchte wissen, ob es Konstanten  $a_n, b_n \in \mathbb{R}$  gibt, sodass

$$f(x) = a_0 + \sum_{n \geq 1} a_n \cos(nx) + b_n \sin(nx). \quad (3.24)$$

Eine nach obiger Diskussion naheliegende Wahl wären da

$$a_n := \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx) dx \quad \text{und} \quad b_n := \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(nx) dx$$

(um eine Summe der Form  $\sum_{n \geq 0} \langle f, g_n \rangle g_n$  zu bekommen). Man spricht dann von der Fourierreihe von  $f$  und die  $a_n, b_n$  sind die Fourierkoeffizienten der Funktion. Dabei kann man die Reihe in (3.24) zunächst hinsichtlich punktweiser Konvergenz studieren, stößt damit aber schnell auf erhebliche Schwierigkeiten. So gibt es etwa stetige Funktionen  $f$ , für welche die Reihe nicht in jedem Punkt konvergiert. Im Sinne punktweiser Konvergenz lassen sich stetige Funktionen also nicht als Superpositionen trigonometrischer Funktionen auffassen.

Dramatisch anders sieht es aber aus, wenn wir den Grenzübergang in (3.24) als Konvergenz bzgl  $\|\cdot\|_2$  interpretieren. Da wird die Angelegenheit mit einem mal viel einfacher und übersichtlicher, erweist sich die Familie in (3.23) doch als vollständiges Orthonormalsystem in  $(L^2[0, 2\pi], \|\cdot\|_2)$ , sodass nicht nur jede stetige, sondern sogar jede quadratintegrierbare Funktion  $f : [0, 2\pi] \rightarrow \mathbb{R}$  wie in (3.24) durch ihre Fourierreihe dargestellt werden kann! Der neue Konvergenzbegriff löst also das Rätsel, wie  $f$  allgemein mit seiner Fourierreihe zusammenhängt.

## 3.7 Produkträume und Satz von Fubini

Maße auf Produkträumen zählen zu den wichtigsten Objekten der Maßtheorie. Dabei geht es nicht nur um Produkte endlich vieler Räume, wie im Fall des  $d$ -dimensionalen Lebesguemaßes, sondern oft auch um unendliche Produkte, die insbesondere für die Stochastik unverzichtbar sind. Wir konzentrieren uns jetzt aber auf den grundlegenden Fall zweier Räume. Per Induktion bekommt man entsprechende Aussagen für endlich viele Faktoren.

**A) Das Produkt zweier Maßräume.** Im folgenden seien  $(X, \mathcal{A}, \mu)$  und  $(Y, \mathcal{B}, \nu)$  feste Maßräume. In diesem Abschnitt konstruieren und studieren wir den *Produktraum*  $(X \times Y, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$ . Dabei müssen wir uns auf dem Cartesischen Produkt  $X \times Y$  der Mengen  $X$  und  $Y$  zuerst eine passende  $\sigma$ -Algebra verschaffen. Da gibt es aber eine sehr naheliegende Wahl: sind  $A \in \mathcal{A}$  und  $B \in \mathcal{B}$  messbare Mengen in  $X$  bzw  $Y$ , dann soll halt wenigstens deren Cartesisches Produkt  $A \times B \subseteq X \times Y$  in unserem Produktraum messbar sein. Also definieren wir  $\mathcal{A} \otimes \mathcal{B}$  als die kleinste  $\sigma$ -Algebra auf  $X \times Y$ , die solche Mengen enthält:

**Definition 3.13** Sind  $(X, \mathcal{A})$  und  $(Y, \mathcal{B})$  messbare Räume, so ist die Produkt- $\sigma$ -Algebra von  $\mathcal{A}$  und  $\mathcal{B}$  gegeben durch

$$\mathcal{A} \otimes \mathcal{B} := \sigma(\mathcal{A} * \mathcal{B}) \subseteq \mathcal{P}(X \times Y),$$

wobei wir  $\mathcal{A} * \mathcal{B} := \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$  schreiben.

Im PS haben wir schon überlegt, dass Cartesische Produkte von Halbringen wieder Halbringe (auf dem Produktraum) sind. Insbesondere ist die erzeugende Familie  $\mathcal{A} * \mathcal{B} \subseteq \mathcal{P}(X \times Y)$  in dieser Definition ein Halbring. Damit befinden wir uns schon mal auf vertrautem Terrain.

Wir können die Definition auch noch anderes formulieren. Dazu betrachten wir die kanonischen Projektionen  $\pi_X : X \times Y \rightarrow X$ ,  $\pi_X(x, y) := x$ , und  $\pi_Y : X \times Y \rightarrow Y$ ,  $\pi_Y(x, y) := y$ , welche ja die Verbindung zwischen  $X \times Y$  und den Faktorräumen  $X$  und  $Y$  herstellen. Für  $A \in \mathcal{A}$  und  $B \in \mathcal{B}$  ist  $A \times B = \pi_X^{-1}A \cap \pi_Y^{-1}B$ . Daher enthält eine  $\sigma$ -Algebra auf  $X \times Y$  alle derartigen Mengen genau dann wenn sie die Vereinigung  $\pi_X^{-1}\mathcal{A} \cup \pi_Y^{-1}\mathcal{B}$  umfasst (machen Sie sich das bitte klar). Letzteres bedeutet aber gerade Messbarkeit beider Projektionen. Also ist  $\mathcal{A} \otimes \mathcal{B}$  die kleinste  $\sigma$ -Algebra auf  $X \times Y$  bezüglich derer  $\pi_X$  und  $\pi_Y$  messbar sind.

**Beispiel 3.17 a)** Betrachte  $(X, \mathcal{A}) = (Y, \mathcal{B}) := (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ . Wir überlegen uns, dass  $\mathcal{B}_{\mathbb{R}^2} = \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$ : Da die Projektionen  $\pi_X$  und  $\pi_Y$  stetig, und somit  $\mathcal{B}_{\mathbb{R}^2}$ - $\mathcal{B}_{\mathbb{R}}$ -messbar sind, sehen wir, dass  $\mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}} \subseteq \mathcal{B}_{\mathbb{R}^2}$  gilt. Andererseits enthält  $\mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$  die Familie  $\mathcal{I}_{\mathbb{R}}^2$ , und gemäß Proposition 2.7 ist diese ein Erzeuger von  $\mathcal{B}_{\mathbb{R}^2}$ , sodass auch  $\mathcal{B}_{\mathbb{R}^2} \subseteq \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$  gilt.

**b)** Betrachte  $(X, \mathcal{A}) := (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  und  $(Y, \mathcal{B}) := (\{1, 2\}, \mathcal{P}(\{1, 2\}))$ . Dann besteht  $X \times Y$  aus den beiden disjunkten Geraden  $\mathbb{R} \times \{1\}$  und  $\mathbb{R} \times \{2\}$  in  $\mathbb{R}^2$ . In diesem Fall ist  $\mathcal{A} \otimes \mathcal{B} = \{(A \times \{1\}) \cup (B \times \{2\}) : A, B \in \mathcal{B}_{\mathbb{R}}\}$ , dh man kann einfach innerhalb jeder Geraden eine Borel-Menge nehmen.

**c)** Betrachte  $(X, \mathcal{A}) = (Y, \mathcal{B}) := (\mathbb{N}, \mathcal{P}(\mathbb{N}))$ , dann ist  $\mathcal{A} \otimes \mathcal{B} = \mathcal{P}(\mathbb{N}^2)$ .

Mengen aus  $\mathcal{A} \otimes \mathcal{B}$  können zwar recht kompliziert sein, verhalten sich aber doch gutmütig. So lassen sie sich etwa in Scheibchen zerschneiden, die alle wieder messbar sind. Dazu nennen wir, für  $Q \subseteq X \times Y$  und Koordinatenwerte  $x \in X$ ,  $y \in Y$  die Mengen

$$Q_x := \{y : (x, y) \in Q\} \subseteq Y \quad \text{und} \quad Q^y := \{x : (x, y) \in Q\} \subseteq X$$

den (vertikalen)  $x$ -Schnitt bzw den (horizontalen)  $y$ -Schnitt von  $Q$ .

**Lemma 3.5 (Messbarkeit von Schnitten  $Q_x, Q^y$ )** Für jedes  $Q \in \mathcal{A} \otimes \mathcal{B}$  und alle  $x \in X$ ,  $y \in Y$ , sind  $Q_x \in \mathcal{B}$  und  $Q^y \in \mathcal{A}$ . Insbesondere sind  $\nu(Q_x)$  und  $\mu(Q^y)$  definiert.

**Beweis.** Sei  $\mathcal{M} := \{Q \subseteq X \times Y : Q_x \in \mathcal{B} \text{ für alle } x \in X\} \subseteq \mathcal{P}(X \times Y)$ , dann ist  $\mathcal{A} * \mathcal{B} \subseteq \mathcal{M}$  weil

$$(A \times B)_x = \begin{cases} B & \text{falls } x \in A, \\ \emptyset & \text{sonst.} \end{cases} \quad (3.25)$$

Andererseits ist  $\mathcal{M}$  eine  $\sigma$ -Algebra auf  $X \times Y$ . (Dies folgt, weil  $\emptyset_x = \emptyset$ ,  $(Q^c)_x = (Q_x)^c$  und  $(\bigcup_{n \geq 1} Q_n)_x = \bigcup_{n \geq 1} (Q_n)_x$ .) Also haben wir  $\mathcal{A} \otimes \mathcal{B} \subseteq \mathcal{M}$  wie gewünscht. (Analog für die  $Q^y$ .) ■

**Bemerkung 3.5** Aufgrund von (3.25) bemerken wir, dass die Funktion  $g_{A \times B} : X \rightarrow [0, \infty)$  mit  $g_{A \times B}(x) := \nu((A \times B)_x)$  einfach als

$$g_{A \times B} = \nu(B)1_A$$

geschrieben werden kann, woraus wir sofort sehen, dass das Integral über die Maße der  $x$ -Schnitte von  $A \times B$  gerade das Produkt der beiden Faktormäße liefert,

$$\int_X \nu((A \times B)_x) d\mu(x) = \mu(A)\nu(B) \quad \text{für } A \times B \in \mathcal{A} * \mathcal{B}. \quad (3.26)$$

Ein Produktmaß ist nun ein Maß auf der Produkt- $\sigma$ -Algebra, welches anschaulich gesprochen Rechtecken das Produkt ihrer Seitenlängen zuordnet.

**Definition 3.14** Sind  $(X, \mathcal{A}, \mu)$  und  $(Y, \mathcal{B}, \nu)$  Maßräume, so nennt man ein Maß  $\mu \otimes \nu$  auf  $\mathcal{A} \otimes \mathcal{B}$  ein Produktmaß wenn

$$\mu \otimes \nu(A \times B) = \mu(A) \cdot \nu(B) \quad \text{für } A \in \mathcal{A}, B \in \mathcal{B}. \quad (3.27)$$

Das Standard-Werkzeug um die Existenz wichtiger Maße nachzuweisen, der Fortsetzungssatz zeigt nun, dass es so etwas immer gibt. Die Notation  $\mu \otimes \nu$  suggeriert dabei, dass es *nur ein* Produktmaß gibt. Das stimmt nicht immer, aber wenn die beiden Faktorräume wenigstens  $\sigma$ -endlich sind, dann ist Eindeutigkeit gewährleistet, und die Notation ist ungefährlich.

**Satz 3.20 (Existenz und Eindeutigkeit des Produktmaßes)** Sind  $(X, \mathcal{A}, \mu)$  und  $(Y, \mathcal{B}, \nu)$  zwei Maßräume, dann existiert ein Produktmaß  $\mu \otimes \nu$ . Dieses ist eindeutig bestimmt und  $\sigma$ -endlich falls  $\mu$  und  $\nu$   $\sigma$ -endlich sind.

**Beweis.** (i) Definiere zunächst  $\rho : \mathcal{A} * \mathcal{B} \rightarrow [0, \infty]$  mittels

$$\rho(A \times B) := \mu(A) \cdot \nu(B) \quad \text{für } A \in \mathcal{A}, B \in \mathcal{B}.$$

Wenn wir zeigen, dass diese Mengenfunktion ein Prämaß ist, können wir Satz 2.2 anwenden, und sehen dass sich  $\rho$  zu einem Maß  $\mu \otimes \nu : \mathcal{A} \otimes \mathcal{B} \rightarrow [0, \infty]$  fortsetzen lässt.

(ii)\* Zur Prämaß-Eigenschaft von  $\rho$ : Sei  $A \times B = \bigsqcup_{n \geq 1} A_n \times B_n$  mit  $A, A_n \in \mathcal{A}$  und  $B, B_n \in \mathcal{B}$ . Mithilfe von (3.26) und monotoner Konvergenz finden wir dann

$$\begin{aligned} \rho(A \times B) &= \int_X \nu((A \times B)_x) d\mu(x) = \int_X \nu\left(\bigsqcup_{n \geq 1} (A_n \times B_n)_x\right) d\mu(x) \\ &= \int_X \sum_{n \geq 1} \nu((A_n \times B_n)_x) d\mu(x) = \sum_{n \geq 1} \int_X \nu((A_n \times B_n)_x) d\mu(x) \\ &= \sum_{n \geq 1} \rho(A_n \times B_n). \end{aligned}$$

(iii) Die Eindeutigkeit im  $\sigma$ -endlichen Fall folgt aus Satz 2.3 sobald wir bemerken, dass obiges Prämaß  $\rho$  selbst auf  $\mathcal{A} * \mathcal{B}$   $\sigma$ -endlich ist. (Übung!) ■

**Beispiel 3.18 a)** In Teil a) der vorigen Beispiels wählen wir  $\mu = \nu := \lambda$ , dann ist offenbar  $\lambda^2$  ein Produktmaß auf  $\mathcal{B}_{\mathbb{R}^2} = \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$ . Der Satz zeigt, dass es das einzige Produktmaß ist,  $\lambda^2 = \lambda \otimes \lambda$ .

**b)** In Teil b) der vorigen Beispiels wählen wir  $\mu := \lambda$  und  $\nu := \#$ , dann ist  $\mu \otimes \nu((A \times \{1\}) \cup (B \times \{2\})) = \lambda(A) + \lambda(B)$ . (Prüfen Sie das bitte nach.)

**c)** In Teil c) der vorigen Beispiels wählen wir  $\mu = \nu := \#$ , dann ist auch  $\mu \otimes \nu = \#$ , dh wir bekommen wieder das Zählmaß (aber nun auf  $\mathbb{N}^2$ ).

Wir setzen nun voraus, dass  $\mu$  und  $\nu$  zwei  $\sigma$ -endliche Maße sind. In dieser Situation ist es auch generell (vgl (3.26)) möglich, das Maß einer Menge im Produktraum “scheibchenweise” zu berechnen. Damit können wir  $\mu \otimes \nu(Q)$  nun für beliebige  $Q \in \mathcal{A} \otimes \mathcal{B}$  beschreiben:

**Satz 3.21 (Form des Produktmaßes)** Sind  $\mu$  und  $\nu$   $\sigma$ -endlich, so erlaubt das eindeutig bestimmte und  $\sigma$ -endliche Produktmaß  $\mu \otimes \nu$  die Darstellung

$$\mu \otimes \nu(Q) = \int_X \nu(Q_x) d\mu(x) = \int_Y \mu(Q^y) d\nu(y) \quad \text{für } Q \in \mathcal{A} \otimes \mathcal{B}. \quad (3.28)$$

Diese Aussage enthält implizit die Feststellung, dass die beiden Integranden  $x \mapsto \nu(Q_x)$  und  $y \mapsto \mu(Q^y)$  messbare Funktionen sind. (Sonst wären die Integrale nicht definiert.) Tatsächlich ist dies ein wesentlicher Schritt im Beweis des Satzes:

**Lemma 3.6 (Messbarkeit der Maße von Schnitten)** Sind  $\mu, \nu$   $\sigma$ -endlich, und  $Q \in \mathcal{A} \otimes \mathcal{B}$ , dann ist

$$\begin{aligned} x &\longmapsto \nu(Q_x) && \mathcal{A}\text{-}\mathcal{B}_{\mathbb{R}}\text{-messbar und} \\ y &\longmapsto \mu(Q^y) && \mathcal{B}\text{-}\mathcal{B}_{\mathbb{R}}\text{-messbar.} \end{aligned}$$

Den Nachweis dieser Beobachtung ersparen wir uns hier. (Der folgende Beweis des Satzes enthält übrigens ein alternatives Argument für den Nachweis der Existenz eines Produktmaßes.)

**Beweis von Satz 3.21.** Wir beweisen die Aussage nur für endliche Maße. Der  $\sigma$ -endliche Fall erfordert dann keine neuen Ideen, Sie können sich zur Übung selbst daran versuchen.

Für jedes  $Q \in \mathcal{A} \otimes \mathcal{B}$  ist, wie eben gezeigt,  $g_Q : X \rightarrow [0, \infty)$  mit  $g_Q(x) := \nu(Q_x)$  eine  $\mathcal{A}$ - $\mathcal{B}_{\mathbb{R}}$ -messbare Funktion. Deshalb ist durch

$$\rho(Q) := \int_X \nu(Q_x) d\mu(x) = \int_X g_Q d\mu$$

eine Mengenfunktion auf  $\mathcal{A} \otimes \mathcal{B}$  definiert. Tatsächlich ist  $\rho$  ein Maß: Zunächst ist  $\rho(\emptyset) = \int_X \nu(\emptyset_x) d\mu(x) = 0$  klar wegen  $\emptyset_x = \emptyset$ . Sind  $Q_j \in \mathcal{A} \otimes \mathcal{B}$ ,  $j \geq 1$ , paarweise disjunkte Mengen, und  $Q := \bigsqcup_{j \geq 1} Q_j$ , so gilt  $g_Q = \sum_{j \geq 1} g_{Q_j}$ . Daher folgt  $\rho(Q) = \int_X g_Q d\mu = \lim_{m \rightarrow \infty} \sum_{j=1}^m \int_X g_{Q_j} d\mu = \sum_{j \geq 1} \int_X g_{Q_j} d\mu = \sum_{j \geq 1} \rho(Q_j)$  mit monotoner Konvergenz, dh  $\rho$  ist  $\sigma$ -additiv. Dass es sich um ein Produktmaß handelt folgt sofort aus (3.26). Wegen der Eindeutigkeitsaussage des vorigen Satzes folgt somit  $\rho = \mu \otimes \nu$ . Analog für  $\tilde{\rho}(Q) := \int_Y \mu(Q^y) d\nu(y)$ . ■

Jetzt können wir endlich bestätigen, dass für messbares  $f : \mathbb{R} \rightarrow [0, \infty]$  das Integral  $\int_{\mathbb{R}} f d\lambda$  tatsächlich der Fläche (nämlich dem zweidimensionalen Lebesgue-Maß) der *Ordinatenmenge* von  $f$ ,  $O(f) := \{(x, y) \in \mathbb{R}^2 : 0 \leq y < f(x)\}$  entspricht.

**Folgerung 3.2 (Das Integral ist die Fläche unter dem Graphen)** Für messbares  $f : \mathbb{R} \rightarrow [0, \infty)$  gilt

$$\int_{\mathbb{R}} f d\lambda = \lambda^2(O(f)).$$

**Beweis.** Sei  $Q := O(f) \subseteq \mathbb{R}^2$ , dann ist  $Q \in \mathcal{B}_{\mathbb{R}^2}$ : Betrachte  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  mit  $\varphi(x, y) := f(x) - y$ , dann ist  $\varphi$  eine  $\mathcal{B}_{\mathbb{R}^2}$ - $\mathcal{B}_{\mathbb{R}}$ -messbare Abbildung (da  $(x, y) \mapsto (f(x), y)$   $\mathcal{B}_{\mathbb{R}^2}$ - $\mathcal{B}_{\mathbb{R}^2}$ -messbar ist, siehe Bsp 2.16, und  $(s, t) \mapsto s - t$  stetig und somit  $\mathcal{B}_{\mathbb{R}^2}$ - $\mathcal{B}_{\mathbb{R}}$ -messbar ist). Wegen  $O(f) = \{\varphi > 0\}$  folgt die Behauptung.

Für jedes  $x \in \mathbb{R}$  ist nun  $Q_x = [0, f(x))$ , und daher  $\lambda(Q_x) = f(x)$ . Aufgrund unseres Satzes finden wir also  $\lambda^2(O(f)) = \lambda \otimes \lambda(Q) = \int_{\mathbb{R}} \lambda(Q_x) d\lambda(x) = \int_{\mathbb{R}} f(x) d\lambda(x)$ . ■

Analog können wir nun offiziell die Fläche  $\lambda^2(Q)$  einer beliebigen Borel-Menge  $Q \in \mathcal{B}_{\mathbb{R}^2}$  scheibchenweise bestimmen, mittels

$$\lambda^2(Q) = \int_{\mathbb{R}} \lambda(Q_x) d\lambda(x) = \int_{\mathbb{R}} \lambda(Q_x) dx. \quad (3.29)$$

**Beispiel 3.19** Für die Kreisscheibe  $Q := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}$  mit Radius  $r \geq 0$  ist

$$Q_x = \begin{cases} [-\sqrt{r^2 - x^2}, \sqrt{r^2 - x^2}] & \text{falls } |x| \leq r, \\ \emptyset & \text{sonst,} \end{cases}$$

und deshalb  $\lambda(Q_x) = 1_{[-r, r]}(x) \cdot 2\sqrt{r^2 - x^2}$ ,  $x \in \mathbb{R}$ . Somit (wenig überraschend)

$$\lambda^2(Q) = \int_{\mathbb{R}} \lambda(Q_x) dx = 2 \int_{-r}^r \sqrt{r^2 - x^2} dx = 2r^2 \int_{-1}^1 \sqrt{1 - t^2} dt = \pi r^2.$$

**B) Integration auf Produkträumen.** Es seien  $(X, \mathcal{A}, \mu)$  und  $(Y, \mathcal{B}, \nu)$  zwei  $\sigma$ -endliche Maßräume. Im Rahmen des vorigen Abschnittes bemühen wir uns nun, die Integration messbarer Funktionen auf Produkträumen in den Griff zu bekommen. Dazu übertragen wir das Konzept der Schnitte von Mengen auf Funktionen, indem wir für  $f : X \times Y \rightarrow \overline{\mathbb{R}}$  und  $x \in X$ ,  $y \in Y$ , die Funktionen

$$f_x : Y \rightarrow \overline{\mathbb{R}} \text{ mit } f_x(y) := f(x, y) \quad \text{und} \quad f^y : X \rightarrow \overline{\mathbb{R}} \text{ mit } f^y(x) := f(x, y)$$

als (vertikalen)  $x$ -Schnitt bzw (horizontalen)  $y$ -Schnitt von  $f$  bezeichnen.

**Bemerkung 3.6** Für  $Q \subseteq X \times Y$  und  $x \in X$  gilt  $(1_Q)_x = 1_{Q_x}$ . Weiters finden wir  $(cf + g)_x = cf_x + g_x$  und  $(f^\pm)_x = (f_x)^\pm$ .

Parallel zu Lemma 3.5 bekommen wir:

**Proposition 3.11 (Messbarkeit von Schnitten  $f_x, f^y$ )** Ist  $f : X \times Y \rightarrow \overline{\mathbb{R}}$   $\mathcal{A} \otimes \mathcal{B}$ -messbar, so sind alle Schnitte  $f_x, x \in X$ ,  $\mathcal{B}$ -messbar, und alle Schnitte  $f^y, y \in Y$ ,  $\mathcal{A}$ -messbar.

**Beweis.** Zu zeigen ist, dass für  $x \in X$  und  $t \in \mathbb{R}$  stets  $\{f_x \leq t\} \in \mathcal{B}$  gilt. Nun ist

$$\{f_x \leq t\} = \{y : (x, y) \in \{f \leq t\}\} = Q_x$$

mit  $Q := \{f \leq t\} \in \mathcal{A} \otimes \mathcal{B}$ , und die Behauptung folgt nun aus Lemma 3.5. ■

Wiederum sind auch Integrale von Schnitten messbar, und liefern, ein weiteres mal integriert, das Integral von  $f$  bzgl  $\mu \otimes \nu$ .

**Satz 3.22 (Satz von Fubini)** Es seien  $\mu$  und  $\nu$   $\sigma$ -endliche Maße.

a) Für  $\mathcal{A} \otimes \mathcal{B}$ -messbares  $f : X \times Y \rightarrow [0, \infty]$  gilt

$$\int_{X \times Y} f \, d\mu \otimes \nu = \int_X \int_Y f \, d\nu \, d\mu = \int_Y \int_X f \, d\mu \, d\nu. \quad (3.30)$$

b) Für  $\mu \otimes \nu$ -integrierbares  $f : X \times Y \rightarrow \overline{\mathbb{R}}$  gilt (3.30) ebenfalls.

c) Ist für  $\mathcal{A} \otimes \mathcal{B}$ -messbares  $f : X \times Y \rightarrow \overline{\mathbb{R}}$  eines der Integrale

$$\int_{X \times Y} |f| \, d\mu \otimes \nu, \quad \int_X \int_Y |f| \, d\nu \, d\mu, \quad \int_Y \int_X |f| \, d\mu \, d\nu$$

endlich, dann ist  $f$  eine  $\mu \otimes \nu$ -integrierbare Funktion, und somit gilt (3.30).

Dabei ist  $\int_X \int_Y f \, d\nu \, d\mu$  die Kurzschreibweise für  $\int_X \left( \int_Y f_x(y) \, d\nu(y) \right) d\mu(x) = \int_X \left( \int_Y f(x, y) \, d\nu(y) \right) d\mu(x)$ , etc.

**Bemerkung 3.7** Implizit in der Aussage des Satzes enthalten ist die Behauptung, dass die angeführten Integrale auch definiert sind. Insbesondere beinhaltet a) die Feststellung, dass

$$x \mapsto \int_Y f_x \, d\nu \text{ } \mathcal{A}\text{-messbar} \quad \text{und} \quad y \mapsto \int_X f^y \, d\mu \text{ } \mathcal{B}\text{-messbar ist,}$$

und b) enthält die Beobachtungen, dass

$$\begin{aligned} &\text{für } \mu\text{-fast alle } x \in X \text{ die Funktion } f_x \text{ } \nu\text{-integrierbar, und} \\ &\text{für } \nu\text{-fast alle } y \in Y \text{ die Funktion } f^y \text{ } \mu\text{-integrierbar ist,} \end{aligned}$$

und weiters, wenn wir für dass

$$x \mapsto \int_Y f_x \, d\nu \text{ } \mu\text{-integrierbar,} \quad \text{und} \quad y \mapsto \int_X f^y \, d\mu \text{ } \nu\text{-integrierbar ist.}$$

**Beweis. a)** Für Indikatorfunktionen  $f = 1_Q$  mit  $Q \in \mathcal{A} \otimes \mathcal{B}$  ist, nach Bemerkung 3.6,  $\int_Y f_x \, d\nu = \int_Y 1_{Q_x} \, d\nu = \nu(Q_x)$ , was nach Lemma 3.6  $\mathcal{A}$ - $\mathcal{B}_{\mathbb{R}}$ -messbar von  $x$  abhängt. Satz 3.21 zeigt nun, dass  $\int_{X \times Y} f \, d\mu \otimes \nu = \mu \otimes \nu(Q) = \int_X \nu(Q_x) \, d\mu(x) = \int_X \int_Y f \, d\nu \, d\mu$ . Genauso mit vertauschten Rollen von  $x$  und  $y$ . Somit gilt a) für Indikatoren.

Für Treppenfunktionen  $f = \sum_{k=1}^n \alpha_k 1_{Q_k} \in \mathcal{T}^+$  mit  $Q_k \in \mathcal{A} \otimes \mathcal{B}$  haben wir (Bemerkung 3.6)  $f_x = \sum_{k=1}^n \alpha_k 1_{(Q_k)_x}$  und nach obiger Aussage zu Indikatoren folgt die  $\mathcal{B}$ -Messbarkeit von  $f_x$ . Wir dürfen also integrieren, und bekommen  $F(x) :=$

$\int_Y f_x d\nu = \sum_{k=1}^n \alpha_k \nu((Q_k)_x)$ . Da die Summanden  $\mathcal{A}$ -messbar von  $x$  abhängen, gilt entsprechendes für  $F(x)$ . Somit darf  $F$  integriert werden, und es ist (wegen Satz 3.21)

$$\begin{aligned} \int_X \int_Y f d\nu d\mu &= \int_X F d\mu = \sum_{k=1}^n \alpha_k \int_X \nu((Q_k)_x) d\mu(x) \\ &= \sum_{k=1}^n \alpha_k \mu \otimes \nu(Q) = \int_{X \times Y} f d\mu \otimes \nu. \end{aligned}$$

Genauso mit vertauschten Rollen von  $x$  und  $y$ . Also gilt a) für  $f \in \mathcal{T}^+(\mathcal{A} \otimes \mathcal{B})$ .

Sei jetzt  $f : X \times Y \rightarrow [0, \infty]$  eine beliebige  $\mathcal{A} \otimes \mathcal{B}$ -messbare Funktion, dann gibt es  $f_n \in \mathcal{T}^+(\mathcal{A} \otimes \mathcal{B})$  mit  $f_n \nearrow f$ . Insbesondere gilt auch für jedes  $x \in X$  dann  $(f_n)_x \nearrow f_x$ , weshalb wir (monotone Konvergenz) sehen, dass

$$F_n(x) := \int_Y (f_n)_x d\nu \nearrow F(x) := \int_Y f_x d\nu \quad \text{für alle } x \in X,$$

wobei mit den  $F_n$  auch  $F$   $\mathcal{A}$ -messbar ist. Unter Verwendung des vorigen Schrittes erhalten wir

$$\begin{aligned} \int_X \int_Y f d\nu d\mu &= \int_X F d\mu = \lim_{n \rightarrow \infty} \int_X F_n d\mu \\ &= \lim_{n \rightarrow \infty} \int_X \int_Y f_n d\nu d\mu \\ &= \lim_{n \rightarrow \infty} \int_{X \times Y} f_n d\mu \otimes \nu \\ &= \int_{X \times Y} f d\mu \otimes \nu. \end{aligned}$$

Dasselbe mit vertauschten Rollen schließt den Nachweis von Aussage a) ab.

b) Ist  $f : X \times Y \rightarrow \mathbb{R}$   $\mu \otimes \nu$ -integrierbar, so auch  $|f|$ , und Teil a) auf  $|f|$  angewandt zeigt, dass

$$\int_{X \times Y} |f| d\mu \otimes \nu = \int_X \int_Y |f| d\nu d\mu < \infty.$$

Damit muss aber die  $\mathcal{A}$ -messbare Funktion  $x \mapsto \int_Y |f_x| d\nu$  auf  $X$   $\mu$ -f.ü. endlich sein, dh es gibt ein  $X_f \in \mathcal{A}$  mit  $\mu(X_f^c) = 0$  sodass  $\int_Y |f_x| d\nu < \infty$  (also  $f_x$   $\nu$ -integrierbar) für  $x \in X_f$ . Entsprechend für die  $f_y$ . Unsere Behauptung folgt nun in der übliche Weise indem wir die Zerlegung  $f = f^+ - f^-$  betrachten.

c) Die letzte Behauptung des Satzes ergibt sich leicht aus a) und b). ■

Die etwas längliche Formulierung des Satzes ist durchaus gerechtfertigt, man sollte keine der Voraussetzungen fallen lassen. Es reicht nicht, einfach eines der Doppelintegrale auszurechnen!

**Beispiel 3.20** Wie schon Cauchy bemerkt hat, ist

$$\int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} dy dx = \frac{\pi}{4} = - \int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} dx dy,$$

was man leicht nachprüfen kann wenn man den Integranden als  $\frac{\partial^2}{\partial x \partial y} \arctan \frac{x}{y}$  erkennt. Das widerspricht unserem Satz nicht, die Funktion ist einfach nicht  $\lambda^2$ -integrierbar, dh es ist

$$\int_0^1 \int_0^1 \left| \frac{x^2 - y^2}{(x^2 + y^2)^2} \right| dy dx = \int_0^1 \int_0^1 \left| \frac{x^2 - y^2}{(x^2 + y^2)^2} \right| dx dy = \infty.$$

Wendet man den Satz von Fubini aber korrekt an, so kann man nun allerlei Integrale von Funktionen  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  bestimmen. Häufig sind die Integranden nur auf einer Teilmenge  $Q \in \mathcal{B}_{\mathbb{R}^2}$  gegeben, dann kommt eben die Indikatorfunktion  $1_Q$  dazu.



**Beispiel 3.21** Sei  $Q := \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq x^2\}$ . (Diese Menge ist abgeschlossen, also in  $\mathcal{B}_{\mathbb{R}^2}$ .) Gesucht ist das Integral der Funktion  $f : Q \rightarrow [0, \infty)$  mit  $f(x, y) := xe^y$  für  $(x, y) \in Q$ . Wir setzen  $f$  auf die ganze Ebene fort, indem wir allgemeiner  $f(x, y) := 1_Q(x, y) \cdot xe^y$  für  $(x, y) \in \mathbb{R}^2$  definieren. Dann ist einfach  $\int_{\mathbb{R}^2} f d\lambda^2 = \int_{\mathbb{R}^2} f(x, y) d\lambda^2(x, y)$  zu bestimmen. Die Funktion ist Borel-messbar (können Sie das begründen?) und wegen  $f \geq 0$  gilt sicher (3.30). Wir brauchen also nur eines der Doppelintegrale auszurechnen, zB

$$\int_Q f d\lambda^2 = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_Q(x, y) \cdot xe^y dy dx.$$

Weil  $f$  außerhalb  $Q$  ohnehin verschwindet, brauchen wir jeweils nur über Teilbereiche zu integrieren, das führt uns konkret zu

$$\int_0^1 \int_0^{x^2} xe^y dy dx.$$

Vielleicht ist Ihnen das sowieso klar, sicherheitshalber machen wir das aber ruhig mal ganz ausführlich und formal. Zuerst ist also für jedes feste  $x$  das innere Integral  $\int_{\mathbb{R}} 1_Q(x, y) \cdot xe^y dy$  zu bestimmen. Dabei ist es günstig, die Indikatorfunktion entsprechend zu faktorisieren. Wegen  $Q_x = \{y \in \mathbb{R} : 0 \leq y \leq x^2\} = [0, x^2]$  für  $x \in [0, 1]$  und  $Q_x = \emptyset$  sonst, ist

$$1_Q(x, y) = 1_{[0,1]}(x) \cdot 1_{[0,x^2]}(y) \quad \text{für } (x, y) \in \mathbb{R}^2.$$

Daher bekommen wir

$$\begin{aligned} \int_{\mathbb{R}} 1_Q(x, y) \cdot xe^y dy &= 1_{[0,1]}(x) \cdot x \int_{\mathbb{R}} 1_{[0,x^2]}(y) \cdot e^y dy \\ &= 1_{[0,1]}(x) \cdot x \int_0^{x^2} e^y dy = 1_{[0,1]}(x) \cdot x(e^{x^2} - 1). \end{aligned}$$

Und letztlich

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} 1_Q(x, y) \cdot xe^y dy dx &= \int_{\mathbb{R}} 1_{[0,1]}(x) \cdot x(e^{x^2} - 1) dx \\ &= \int_0^1 x(e^{x^2} - 1) dx \\ &= \left[ \frac{e^{x^2}}{2} - \frac{x^2}{2} \right] \Big|_{x=0}^1 = \frac{e}{2} - 1. \end{aligned}$$

Berechnen Sie doch bitte auch das andere Doppelintegral  $\int_{\mathbb{R}} \int_{\mathbb{R}} 1_Q(x, y) \cdot xe^y dx dy$ .

Manchmal sind Integrale über mehrdimensionale Bereiche auch nützlich, um bestimmte eindimensionale Integrale zu berechnen. Hier ist ein sehr prominentes Beispiel.

**Beispiel 3.22** Die Funktion mit  $f(x) := e^{-x^2}$ ,  $x \in \mathbb{R}$ , besitzt keine elementare Stammfunktion. (Das ist ein Satz aus der Differentiellen Algebra.) Man kann es daher nicht ganz direkt bestimmen. Seit Euler wissen wir aber, dass

$$I := \int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}. \quad (3.31)$$

Ein Argument, das auf Laplace zurückgeht verwendet Fubini: Zunächst ist

$$\begin{aligned} \int_0^\infty \int_0^\infty ye^{-(1+x^2)y^2} dy dx &= \int_0^\infty \left[ \frac{-e^{-(1+x^2)y^2}}{2(1+x^2)} \right] \Big|_{y=0}^\infty dx \\ &= \int_0^\infty \frac{1}{2(1+x^2)} dx = \frac{\pi}{4}. \end{aligned}$$

Nach Fubini stimmt dieses Doppelintegral aber überein mit

$$\begin{aligned}\int_0^\infty \int_0^\infty y e^{-(1+x^2)y^2} dx dy &= \int_0^\infty \left( \int_0^\infty e^{-x^2 y^2} dx \right) y e^{-y^2} dy \\ &= \int_0^\infty \left( \int_0^\infty e^{-t^2} dt \right) e^{-y^2} dy = I^2.\end{aligned}$$

(Im zweiten Schritt substituieren wir innen  $t = xy$  bei festem  $y$ .) Also folgt  $I^2 = \frac{\pi}{4}$ .

**Bemerkung 3.8 (Klage über das Riemann-Integral)** Auch was den Produktraum  $\mathbb{R}^2$  mit dem natürlichsten Maß  $\lambda^2$  betrifft, ist das Lebesgue-Integral dem Riemannschen in der Systematik weit überlegen. So gibt es zB Funktionen  $f : [0, 1]^2 \rightarrow \mathbb{R}$  die (zweidimensional) R-integrierbar sind, für die aber keines der Doppelintegrale  $\int_0^1 \int_0^1 f(x, y) dx dy$  und  $\int_0^1 \int_0^1 f(x, y) dy dx$  im Riemannschen Sinne existiert (und das nicht einmal dann, wenn man ein beliebiges anderes Cartesisches Koordinatensystem wählen darf). So ein  $f$  ist aber sicher  $\lambda^2$ -integrierbar, und erfüllt (3.30) wenn man damit eben L-Integrale meint.

**Bemerkung 3.9 (Fubini und Doppelreihen)** Zum Ende des Abschnittes wenden wir den Satz von Fubini noch in der Situation von Teil c) der Beispiele 3.17 und 3.18 an: Dabei sei  $f : \mathbb{N}^2 \rightarrow [0, \infty]$  wie früher identifiziert mit der Doppelfolge  $(b_{k,l})_{k \geq 0, l \geq 0}$  der Werte  $b_{k,l} := f(k, l)$ . Dann ist (Aufgabe!)

$$\int_{\mathbb{N}^2} f d\mu \otimes \nu = \sup \left\{ \sum_{(k,l) \in M} b_{k,l} : M \subseteq \mathbb{N} \times \mathbb{N} \text{ endlich} \right\} \in [0, \infty],$$

und die Doppelintegrale werden zu

$$\int_X \int_Y f d\nu d\mu = \sum_{k \geq 0} \sum_{l \geq 0} b_{k,l} \quad \text{und} \quad \int_Y \int_X f d\mu d\nu = \sum_{l \geq 0} \sum_{k \geq 0} b_{k,l}$$

Wir erkennen in Satz 1.1 d) also einen Spezialfall der Aussage a) bei Fubini.

**C) Produkte endlich vieler Faktoren und weitere Beispiele. ...**

## 3.8 Die Transformationsformel in $\mathbb{R}^d$

Neben dem Satz von Fubini ist die Transformationsformel bzw Substitutionsregel in  $\mathbb{R}^d$  das wichtigste Werkzeug im Umgang mit Integralen von Funktionen mehrerer Variablen (bezüglich  $\lambda^d$ ). Das ist kaum verwunderlich, spielt die Technik der Variablensubstitution doch schon in der eindimensionalen Integration eine bedeutende Rolle. Ziel dieses Abschnittes ist es also, Proposition 3.12 auf  $d$ -dimensionale Situationen zu übertragen. Grundlage dafür ist erneut die allgemeine Transformationsformel (Satz 3.7). Um diese anzuwenden müssen wir in den relevanten Situationen halt wieder das Bildmaß  $\lambda^d \circ T^{-1}$  bestimmen.

**A) Das Bild von  $\lambda^d$  und die Substitutionsregel.** Als Aufwärmübung dazu sehen wir uns zuerst an, wie  $\lambda^d$  durch eine invertierbare lineare Abbildung transformiert wird. Das Ergebnis kennen Sie eigentlich schon aus der linearen Algebra: Ist  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  ein linearer Automorphismus, so gibt  $|\det T|$ , der Betrag der Determinante, nämlich gerade das  $d$ -dimensionale Volumen des Bildes  $T((0, 1]^d)$  des  $d$ -dimensionalen Einheitswürfels an. Anwendung von  $T$  ändert also

die Größe der Menge um den Faktor  $|\det T|$ . Beim Übergang von  $\lambda^d$  zum Bildmaß  $\lambda^d \circ T^{-1}$  müssen wir aber  $T^{-1}$  anwenden, und bekommen daher den Faktor  $|\det(T^{-1})| = 1/|\det T|$ . Natürlich spricht man in der Linearen Algebra nur über einen elementaren Größenbegriff für sehr einfache Mengen. Wir überzeugen uns nun formal davon, dass diese Argumentation wirklich die richtige Antwort für das Lebesgue-Maß liefert.

**Proposition 3.12 (Bild von  $\lambda^d$  unter linearen Automorphismen)** *Ist  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  ein linearer Automorphismus, so gilt*

$$\lambda^d \circ T^{-1} = \frac{1}{|\det T|} \lambda^d = \frac{1}{|\det T|} \odot \lambda^d \quad \text{auf } \mathcal{L}_{\mathbb{R}^d}, \quad (3.32)$$

bzw. äquivalent dazu,

$$\lambda^d \circ T = |\det T| \lambda^d = |\det T| \odot \lambda^d \quad \text{auf } \mathcal{L}_{\mathbb{R}^d}. \quad (3.33)$$

Die Aussagen bleibt auch für bijektive affine Abbildungen  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  korrekt.

**Bemerkung 3.10 a)** Weil  $|\det T|$  hier eine Konstante ist, können wir diese in (3.32) und (3.33) sowohl als skalaren Faktor, als auch als Dichte schreiben.

**b)** Die äußerlich einfachere und auch in der Literatur oft bevorzugte Variante (3.33) ist hier nachgereiht, damit Sie nicht vergessen, dass  $\mu \circ T^{-1}$  zwar für jede messbare Abbildung,  $\mu \circ T$  aber nur im speziellen Fall einer invertierbaren Abbildung  $T$  (dh wenn eben  $T = S^{-1}$  für ein messbares  $S$ ) ohne Probleme ein Maß liefert.

**Beweis. (i)** Wir bemerken zuerst, dass das Bildmaß  $\lambda^d \circ T^{-1}$  wieder translationsinvariant ist: Für  $S : \mathbb{R}^d \rightarrow \mathbb{R}^d$  mit  $Sx := x + a$  (mit festem  $a \in \mathbb{R}^d$ ) und jede Menge  $A \in \mathcal{L}_{\mathbb{R}^d}$  ist (wegen Linearität von  $T^{-1}$  und Translationsinvarianz von  $\lambda^d$ )

$$(\lambda^d \circ T^{-1}) \circ S^{-1}(A) = \lambda^d \circ T^{-1}(A - a) = \lambda^d(T^{-1}A - T^{-1}a) = \lambda^d \circ T^{-1}(A).$$

**(ii)** Weil  $T$  ein Homöomorphismus ist, sehen wir, dass  $T^{-1}(0, 1)^d$  eine nichtleere offene Menge ist, und somit  $\lambda^d \circ T^{-1}((0, 1)^d) > 0$ . Aus dieser topologischen Eigenschaft von  $T$  folgt weiters, dass  $T^{-1}[0, 1]^d$  kompakt ist, sodass  $\lambda^d \circ T^{-1}([0, 1]^d) < \infty$  sein muss. Insgesamt ist also  $0 < \lambda^d \circ T^{-1}((0, 1)^d) \leq \lambda^d \circ T^{-1}([0, 1]^d) < \infty$ .

**(iii)** Wir wissen bereits, dass jedes translationsinvariante Maß  $\mu$  auf  $\mathcal{L}_{\mathbb{R}^d}$  mit  $0 < \mu([0, 1]^d) < \infty$  ein konstantes Vielfaches von  $\lambda^d$  sein muss (Proposition 1.13). Aufgrund von (i) und (ii) gibt somit es eine Konstante  $c_T \in (0, \infty)$  derart, dass  $\lambda^d \circ T^{-1} = c_T \lambda^d$  auf  $\mathcal{L}_{\mathbb{R}^d}$ . Es bleibt zu zeigen, dass stets  $c_T = 1/|\det T|$  gilt. Setzen wir eine beliebige Menge  $A \in \mathcal{L}_{\mathbb{R}^d}$  positiven endlichen Maßes ein, so können wir hoffen, mittels  $c_T = \lambda^d(T^{-1}A)/\lambda^d(A)$  die Konstante bestimmen.

**(iv)** Im Spezialfall einer orthogonalen Abbildung  $T$  lässt sich der letzte Vorschlag ganz leicht umsetzen: wählen wir für  $A$  die abgeschlossene Einheitskugel in  $\mathbb{R}^d$ , dann ist, weil Längen unverändert bleiben,  $T^{-1}A = A$ . Wir erhalten  $c_T = \lambda^d(T^{-1}A)/\lambda^d(A) = 1 = 1/|\det T|$ . Damit ist die Behauptung für alle orthogonalen  $T$  bewiesen.

**(v)** Ein weiterer simpler Spezialfall ist jener eines Automorphismus der durch eine Diagonalmatrix

$$\begin{pmatrix} \rho_1 & & 0 \\ & \ddots & \\ 0 & & \rho_d \end{pmatrix}$$

gegeben ist. Wählen wir hier nämlich den Einheitswürfel  $A := [0, 1]^d$ , so ist offenbar  $T^{-1}A = \prod_{j=1}^d [0, \rho_j^{-1}]$  (wobei wir  $[0, r]$  als  $[r, 0]$  interpretieren, falls  $r < 0$ ). Dies zeigt, dass  $c_T = \lambda^d(T^{-1}A)/\lambda^d(A) = \prod_{j=1}^d \rho_j^{-1} = 1/\prod_{j=1}^d \rho_j = 1/|\det T|$ . Die Behauptung der Proposition ist also auch für alle diagonalen Automorphismen korrekt.

(vi) Den allgemeinen Fall können wir nun auf (iv) und (v) zurückführen, zB indem wir folgende algebraische Tatsache verwenden: Jeder lineare Automorphismus  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  kann als Zusammensetzung  $T = O_2 \circ S \circ O_1$  mit orthogonalen  $O_1, O_2$  und diagonalem  $S$  dargestellt werden. Aufgrund der Multiplikativität der Determinante haben wir dann  $|\det T| = |\det O_1 \det S \det O_2| = c_{O_1} c_S c_{O_2} = c_S$ , und nach (iv) und (v) gilt wie gewünscht

$$\lambda^d \circ T^{-1} = \lambda^d \circ O_1^{-1} \circ S^{-1} \circ O_2^{-1} = c_{O_1} \lambda^d \circ S^{-1} \circ O_2^{-1} = c_{O_1} c_S \lambda^d \circ O_2^{-1} = c_{O_1} c_S c_{O_2} \lambda^d.$$

(vii) Ist schließlich  $T$  eine bijektive affine Abbildung auf  $\mathbb{R}^d$ , so ist diese die Zusammensetzung eines linearen Automorphismus und einer Translation. Aus dem eben bewiesenen und der Translationsinvarianz von  $\lambda^d$  folgt damit sofort die letzte Behauptung. ■

**Bemerkung 3.11 (Lokale Version der Proposition)** *Weil jeder lineare Automorphismus bijektiv mit messbarer Umkehrabbildung ist, geschieht die Transformation von  $\lambda^d$  nach  $|\det T|^{-1} \lambda^d$  dabei in folgendem Sinne lokal: Ist  $T : A \rightarrow TA$  die Einschränkung eines solchen Automorphismus auf eine beliebige (zB sehr kleine) Menge  $A \in \mathcal{L}_{\mathbb{R}^d}$ , dann ist  $TA \in \mathcal{L}_{\mathbb{R}^d}$  und (3.32) gilt weiterhin indem*

$$\lambda^d \circ T^{-1} = \frac{1}{|\det T|} \lambda^d = \frac{1}{|\det T|} \odot \lambda^d \quad \text{auf } \mathcal{L}_{TA}. \quad (3.34)$$

Dh die Einschränkung von  $\lambda^d$  auf  $A$  wird durch  $T$  auf die Einschränkung von  $|\det T|^{-1} \lambda^d$  auf  $TA$  abgebildet.

**Beispiel 3.23** *So wird  $1_{[0,1]^2} \odot \lambda^2$ , das zweidimensionale Lebesgue-Maß auf dem Einheitsquadrat, durch  $\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$  auf  $1_{[0,3] \times [-2,0]} \odot \frac{1}{6} \lambda^2$ , das normierte zweidimensionale Lebesgue-Maß auf  $T[0,1]^2 = [0,3] \times [-2,0]$  abgebildet, und  $\begin{pmatrix} 1/2 & 0 \\ 0 & 1/5 \end{pmatrix}$  bildet  $1_{[0,1] \times \mathbb{R}} \odot \lambda^2$ , das Lebesgue-Maß auf dem Streifen  $[0,1] \times \mathbb{R}$ , auf  $1_{[0,1/2] \times \mathbb{R}} \odot 10 \lambda^2$ , das um den Faktor 10 verdichtete Lebesgue-Maß auf dem Streifen  $[0,1/2] \times \mathbb{R}$  ab.*

Eine wesentlich allgemeinere Version obigen Ergebnisses erlaubt es uns, die affine Abbildung  $T$  aus der Proposition durch einen Homöomorphismus  $T : X \rightarrow Y$  offener Teilmengen zu ersetzen, wenn dieser lokal sehr gut durch invertierbare affine Abbildungen approximiert werden kann indem  $T$  differenzierbar mit invertierbarer Ableitung  $DT$  ist. (Wir schreiben dann  $DT(x)$  für die Ableitung an der Stelle  $x$ , und  $DT(x)(v)$  oder  $DT(x) \cdot v$  für das Bild von  $v \in \mathbb{R}^d$  unter  $DT(x)$ .) In diesem Fall kann man ja hoffen, dass für eine kleine Umgebung  $A$  von  $x$ , das Bild unter  $T$  von  $1_A \odot \lambda^d$  auf der Umgebung  $TA$  von  $y = Tx$  gut approximiert wird durch das Bild desselben Maßes unter der approximierenden affinen Abbildung, und das wäre laut (3.34) ja gerade  $\frac{1}{|\det DT|(x)} \odot \lambda^d = \frac{1}{|\det DT|(T^{-1}y)} \odot \lambda^d$  auf  $TA$ . Der zentrale Satz dieses Abschnittes bestätigt diese Vermutung.

**Satz 3.23 (Bild von  $\lambda^d$  unter einem  $\mathcal{C}^1$ -Diffeomorphismus)** *Seien  $X, Y \subseteq \mathbb{R}^d$  offen und  $T : X \rightarrow Y$  ein  $\mathcal{C}^1$ -Diffeomorphismus. Dann ist*

$$\lambda^d \circ T^{-1} = \frac{1}{|\det DT| \circ T^{-1}} \odot \lambda^d \quad \text{auf } \mathcal{L}_Y, \quad (3.35)$$

bzw, äquivalent dazu mit  $\psi := T^{-1}$ ,

$$\lambda^d \circ \psi = |\det D\psi| \odot \lambda^d \quad \text{auf } \mathcal{L}_Y. \quad (3.36)$$

(Siehe nochmals Bemerkung 3.10 b.) Den Beweis dieses Satzes schieben wir noch ein wenig auf, und besprechen zuerst die angekündigte Folgerung: Die Transformationsformel in  $\mathbb{R}^d$  ergibt sich nun wie im eindimensionalen Fall sofort aus Satz 3.7. (Weil der Beweis ganz analog verläuft, kommt er zum Kleingedruckten.)

**Satz 3.24 (Substitutionsregel für  $d$ -dimensionale Lebesgue-Integrale)** Es seien  $X, Y \subseteq \mathbb{R}^d$  offen,  $\psi : Y \rightarrow X$  ein  $\mathcal{C}^1$ -Diffeomorphismus und  $f : X \rightarrow \mathbb{R}$  Lebesgue-messbar. Dann gilt

$$\int_X f(x) d\lambda^d(x) = \int_Y f(\psi(y)) |\det D\psi(y)| d\lambda^d(y) \quad (3.37)$$

so bald eines dieser Integrale definiert ist.

Nur scheinbar allgemeiner ist die Formulierung, dass für  $Z \subseteq Y$  mit  $Z \in \mathcal{L}_{\mathbb{R}^d}$ ,

$$\int_{\psi(Z)} f(x) d\lambda^d(x) = \int_Z f(\psi(z)) |\det D\psi(z)| d\lambda^d(z), \quad (3.38)$$

die man sofort aus (3.37) bekommt indem man einfach  $1_{\psi(Z)}f$  statt  $f$  betrachtet.

**Beweis des Satzes.** Durch  $T := \psi^{-1}$  wird ein  $\mathcal{C}^1$ -Diffeomorphismus  $T : X \rightarrow Y$  definiert. Betrachte  $g := f \circ \psi$ , dann ist  $f = g \circ T$ , und somit  $\int_X f(x) dx = \int_X (g \circ T) d\lambda^d$ . Andererseits haben wir  $\psi' = 1/T' \circ T^{-1}$  und daher wegen (3.35),

$$\begin{aligned} \int_Y f(\psi(y)) |\det D\psi| dy &= \int_Y \frac{g}{|\det DT| \circ T^{-1}} d\lambda^d \\ &= \int_Y g d\left(\frac{1}{|\det DT| \circ T^{-1}} \odot \lambda^d\right) = \int_Y g d(\lambda^d \circ T^{-1}). \end{aligned}$$

Satz 3.7 garantiert nun, dass  $\int_X (g \circ T) d\lambda^d = \int_Y g d(\lambda^d \circ T^{-1})$  gilt, wie in (3.37) behauptet. ■

**B) Anwendungen der  $d$ -dimensionalen Transformationsformel.** Möchte man Satz 3.24 anwenden um ein konkretes Integral  $\int_X f(x) d\lambda^d(x)$  mit  $X \subseteq \mathbb{R}^d$  offen und  $f : X \rightarrow \mathbb{R}$  Lebesgue-messbar auszurechnen, so sucht man in der Regel zuerst einen potenziell vorteilhaften Koordinatenwechsel, dh einen  $\mathcal{C}^1$ -Diffeomorphismus  $\psi : Y \rightarrow X$  für welchen  $f(\psi(y))$  ein (wenigstens für die Zwecke des Integrierens) einfacherer Ausdruck ist als  $f(x)$ . Dann muss man aber noch die *Funktional-* (oder *Jacobi-*) *determinante*  $|\det D\psi(y)|$  bestimmen, um zu sehen ob der neue Integrand  $f(\psi(y)) |\det D\psi(y)|$  insgesamt gut zu handhaben bleibt. Ist  $\psi$  durch die eindimensionalen Komponentenfunktionen  $\psi_j : Y \rightarrow \mathbb{R}$ ,  $j \in \{1, \dots, d\}$ , gegeben, so ist natürlich

$$\det D\psi = \begin{vmatrix} \frac{\partial \psi_1}{\partial y_1} & \dots & \frac{\partial \psi_1}{\partial y_d} \\ \vdots & & \vdots \\ \frac{\partial \psi_d}{\partial y_1} & \dots & \frac{\partial \psi_d}{\partial y_d} \end{vmatrix}.$$

Kandidaten für nützliche Koordinatenwechsel ergeben sich manchmal ganz natürlich aufgrund von Symmetrieeigenschaften der Funktion  $f$ . Dies ist insbesondere im folgenden Beispiel der Fall.

**Polarkoordinaten in der Ebene.** Bekanntlich lassen sich Punkte  $(x, y) \in \mathbb{R}^2$  der Ebene auch durch Angabe ihres Abstandes  $r$  vom Ursprung  $(0, 0)$  und des Winkels  $\varphi$  zwischen der  $x$ -Achse und der Strecke von  $(0, 0)$  nach  $(x, y)$  beschreiben. Um dies geeignet zu formalisieren, setzen wir  $Y := (0, \infty) \times (0, 2\pi)$  und definieren  $\psi(r, \varphi) := (r \cos \varphi, r \sin \varphi)$  um die zugehörigen Cartesischen Koordinaten zu bekommen. Dies liefert uns eine Bijektion zwischen zwei offenen Mengen,  $\psi : Y \rightarrow X$  mit  $X := \mathbb{R}^2 \setminus ([0, \infty) \times 0)$ . Dass hierbei  $X$  nicht ganz  $\mathbb{R}^2$  ist, stört für die Belange der Integration nicht, weil wir mit dem Strahl  $[0, \infty) \times 0$  nur eine Nullmenge entfernt haben. Dafür ist  $\psi$  zwischen diesen Mengen aber ein  $\mathcal{C}^1$ -Diffeomorphismus mit

$$D\psi(r, \varphi) = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix},$$

und somit

$$|\det D\psi|(r, \varphi) = r \cos^2 \varphi + r \sin^2 \varphi = r.$$

Von einer in Cartesischen Koordinaten  $(x, y)$  gegebenen Funktion  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  ausgehend, können wir diese jetzt “in Polarkoordinaten ausdrücken” indem wir zu  $\tilde{f} := f \circ \psi$  übergehen. Es ist ja  $\tilde{f}(r, \varphi) = f(x, y)$  wenn  $(x, y)$  Polarkoordinaten  $(r, \varphi)$  hat. Satz 3.24 zeigt dann (gemeinsam mit Fubini), dass

$$\begin{aligned} \int_{\mathbb{R}^2} f d\lambda^2 &= \int_X f(x, y) d\lambda^2(x, y) = \int_Y \tilde{f}(r, \varphi) r d\lambda^2(r, \varphi) \\ &= \int_0^{2\pi} \int_0^\infty \tilde{f}(r, \varphi) r dr d\varphi = \int_0^\infty \int_0^{2\pi} \tilde{f}(r, \varphi) d\varphi r dr \end{aligned} \quad (3.39)$$

(sofern  $f \geq 0$  oder  $f$  integrierbar ist). Am einfachsten wird die Sache, wenn  $f$  etwa unter Rotationen invariant ist (sodass  $\tilde{f}(r, \varphi)$  nur von  $r$  abhängt) oder unter Streckungen (sodass  $\tilde{f}(r, \varphi)$  nur von  $\varphi$  abhängt).

**Beispiel 3.24 a)** Um die Fläche des Kreisringes  $A := \{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 2\}$  direkt auszurechnen, wollen wir  $\lambda^2(A) = \int_{\mathbb{R}^2} f d\lambda^2$  mit  $f := 1_A$  bestimmen. Der Integrand ist rotationsinvariant, daher ist  $\tilde{f}$  einfacher als  $f$ , konkret  $\tilde{f}(r, \varphi) = 1_{[1,2]}(r)$ . Somit laut (3.39),

$$\int_{\mathbb{R}^2} f d\lambda^2 = \int_0^\infty \int_0^{2\pi} \tilde{f}(r, \varphi) d\varphi r dr = \int_0^\infty 2\pi 1_{[1,2]}(r) r dr = 2\pi \int_1^2 r dr = 3\pi.$$

**b)** Der durchschnittliche Abstand  $\frac{1}{\lambda^2(A)} \int_A \|(x, y)\| d\lambda^2(x, y)$  von Punkten aus dem Ring  $A$  zum Ursprung sollte ein wenig größer sein als  $\frac{3}{2}$ , weil der äußere Teil des Ringes größere Fläche als der innere hat. Um diesen Mittelwert zu berechnen sei  $f(x, y) := 1_A(x, y) \|(x, y)\|$ , dann ist offenbar  $\tilde{f}(r, \varphi) = 1_{[1,2]}(r)r$ , und wir finden

$$\begin{aligned} \int_A \|(x, y)\| d\lambda^2(x, y) &= \int_{\mathbb{R}^2} f d\lambda^2 = \int_0^\infty \int_0^{2\pi} \tilde{f}(r, \varphi) d\varphi r dr = \\ &= \int_0^\infty 2\pi 1_{[1,2]}(r) r^2 dr = 2\pi \int_1^2 r^2 dr = \frac{14\pi}{3}. \end{aligned}$$

Wie beim Satz von Fubini erlaubt es auch kunstvolle Anwendung der mehrdimensionalen Substitutionsregel manchmal, schwierige eindimensionale Integrale in den Griff zu bekommen. Das nächste Beispiel stammt von C.F. Gauß.

**Beispiel 3.25** Wir zeigen erneut (vgl Bsp 3.22), dass

$$\int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}.$$

Es gilt nämlich (mittels Fubini und Übergang zu Polarkoordinaten)

$$\begin{aligned} \left( \int_{-\infty}^\infty e^{-x^2} dx \right)^2 &= \int_{-\infty}^\infty \left( \int_{-\infty}^\infty e^{-y^2} dy \right) e^{-x^2} dx = \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-(x^2+y^2)} dx dy \\ &= \int_0^\infty \int_0^{2\pi} e^{-r^2} r d\varphi dr = 2\pi \int_0^\infty e^{-r^2} r dr \\ &= 2\pi \left[ -\frac{e^{-r^2}}{2} \right] \Big|_{r=0}^\infty = \pi. \end{aligned}$$

**Mehr Beispiele.** Manchmal recht nützlich sind *Zylinderkoordinaten* in  $\mathbb{R}^3$ . Dabei drücken wir für Punkte  $(x, y, z) \in \mathbb{R}^3$  einfach die Projektion  $(x, y)$  wie oben durch Polarkoordinaten  $(r, \varphi)$  aus, und beschreiben  $(x, y, z)$  insgesamt durch  $(r, \varphi, z)$ . Dh wir setzen  $Y := (0, \infty) \times (0, 2\pi) \times \mathbb{R}$  und wenden  $\psi(r, \varphi, z) := (r \cos \varphi, r \sin \varphi, z)$  an um die zugehörigen Cartesischen Koordinaten zu bekommen. Damit finden wir

$$D\psi(r, \varphi) = \begin{pmatrix} \cos \varphi & -r \sin \varphi & 0 \\ \sin \varphi & r \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

und somit (wieder)  $|\det D\psi|(r, \varphi, z) = r$ , also (für integrierbares  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ )

$$\int_{\mathbb{R}^3} f(x, y, z) d\lambda^3(x, y, z) = \int_0^\infty \int_0^{2\pi} \int_{-\infty}^\infty f(r \cos \varphi, r \sin \varphi, z) r d\varphi dr dz.$$

### Beispiel 3.26 ...

*Kugelkoordinaten* (oder sphärische Polarkoordinaten) in  $\mathbb{R}^3$  erhalten wir indem wir  $(x, y, z) \in \mathbb{R}^3$  beschreiben durch  $(r, \varphi, \theta) \in Y := (0, \infty) \times (0, 2\pi) \times (-\frac{\pi}{2}, \frac{\pi}{2})$  mit

$$(x, y, z) = \psi(r, \varphi, \theta) := (r \cos \varphi \cos \theta, r \sin \varphi \cos \theta, r \sin \theta),$$

dh  $(r, \varphi, 0)$  beschreibt einen Punkt in der  $(x, y)$ -Ebene mit Polarkoordinaten  $(r, \varphi)$ . Wird der Strahl von  $(0, 0, 0)$  nach  $(r, \varphi, 0)$  dann innerhalb der gemeinsam mit der  $z$ -Achse aufgespannten senkrechten Ebene dann um den Winkel  $\theta$  aus der  $(x, y)$ -Ebene heraus, so bekommen wir  $(r, \varphi, \theta)$ . Eine unspektakuläre Rechnung zeigt, dass  $|\det D\psi|(r, \varphi, \theta) = r^2 \cos \theta$ . Daher ist, für integrierbares  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ ,

$$\begin{aligned} \int_{\mathbb{R}^3} f(x, y, z) d\lambda^3(x, y, z) \\ = \int_0^\infty \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} f(r \cos \varphi \cos \theta, r \sin \varphi \cos \theta, r \sin \theta) r^2 \cos \theta d\varphi dr d\theta. \end{aligned}$$

### Beispiel 3.27 ...

**C) Beweis der  $d$ -dimensionalen Transformationsformel.** Zum Ende stellen wir uns aber doch der Herausforderung, den zentralen Satz dieses Abschnittes zu beweisen. Das erfordert einen gewissen Aufwand, und einzelne technische Details werden unten weggelassen. Wenn Sie das Folgende aber durcharbeiten, werden Sie die Aussagen dieses Abschnittes sicher noch besser verstehen.

Zur Vorbereitung notieren wir einen praktischen Mittelwertsatz aus der mehrdimensionalen Differentialrechnung. Dabei bezeichnen wir für zwei Punkte  $y, z \in \mathbb{R}^d$  mit  $\gamma = \gamma_{[y,z]} : [0, 1] \rightarrow \mathbb{R}^d$  die naheliegende Parametrisierung  $\gamma(t) := y + t(z - y)$  des Segmentes  $[y, z] := \gamma([0, 1]) \subseteq \mathbb{R}^d$  welches diese verbindet. Offenbar ist  $\gamma$  eine differenzierbare Kurve, und die  $j$ te Komponente  $\gamma_j : [0, 1] \rightarrow \mathbb{R}$  von  $\gamma$  hat Ableitung  $\gamma'_j(t) = z_j - y_j$ .

### Proposition 3.13 (Mittelwertsatz für mehrdimensionale $\mathcal{C}^1$ -Abbildungen)

Sei  $Y \subseteq \mathbb{R}^d$  offen, und  $\varphi : Y \rightarrow \mathbb{R}^d$  eine  $\mathcal{C}^1$ -Abbildung. Sind  $y, z$  Punkte für die das ganze Segment  $[y, z]$  in  $Y$  liegt, so gilt

$$\varphi(z) - \varphi(y) = D\varphi[y, z] \cdot (z - y), \quad (3.40)$$

wobei die Matrix bzw lineare Abbildung  $D\varphi[y, z]$  die über  $[y, z]$  gemittelte Ableitung von  $\varphi$  ist,

$$D\varphi[y, z] := \int_0^1 D\varphi(y + t(z - y)) dt. \quad (3.41)$$

Hier wird einfach komponentenweise integriert, dh  $D\varphi[y, z] = (D\varphi[y, z]_{ij})_{1 \leq i, j \leq d}$  mit eindimensionalen Integralen

$$D\varphi[y, z]_{ij} = \int_0^1 \frac{\partial \varphi_i}{\partial x_j}(y + t(z - y)) dt.$$

**Beweis.** Sei  $\gamma = \gamma_{[y,z]}$  wie oben, und  $i \in \{1, \dots, d\}$  beliebig. Wir zeigen, dass (3.40) für die  $i$ te Komponente  $\varphi_i$  von  $\varphi$  gilt, dh

$$\varphi_i(z) - \varphi_i(y) = \sum_{j=1}^d D\varphi[y, z]_{ij}(z_j - y_j). \quad (3.42)$$

Betrachte dazu die  $i$ te Komponente der Bildkurve  $\varphi \circ \gamma$ , also  $\varphi_i \circ \gamma : [0, 1] \rightarrow \mathbb{R}$ . Dies ist eine stetig differenzierbare reelle Funktion, und die mehrdimensionale Kettenregel zeigt, dass

$$\begin{aligned} (\varphi_i \circ \gamma)'(t) &= D\varphi_i(\gamma(t)) D\gamma(t) \\ &= \sum_{j=1}^d \frac{\partial \varphi_i}{\partial y_j}(\gamma(t)) \gamma_j'(t) = \sum_{j=1}^d \frac{\partial \varphi_i}{\partial y_j}(y + t(z - y)) (z_j - y_j). \end{aligned}$$

Nach dem Hauptsatz der eindimensionalen Analysis gilt aber

$$\begin{aligned} \varphi_i(z) - \varphi_i(y) &= \varphi_i(\gamma(1)) - \varphi_i(\gamma(0)) = \int_0^1 (\varphi_i \circ \gamma)'(t) dt \\ &= \sum_{j=1}^d \left( \int_0^1 \frac{\partial \varphi_i}{\partial y_j}(y + t(z - y)) dt \right) (z_j - y_j), \end{aligned}$$

und das ist gerade (3.42). ■

Damit können wir nun leicht folgende Beobachtung beweisen: Ist die Ableitung einer  $\mathcal{C}^1$ -Abbildung auf einer Umgebung von  $y$  nahe an der Identität, dann ist das Bild ausreichend kleiner Würfel um  $y$  in vergleichbar kleinen Würfeln um den Bildpunkt enthalten. Um das zu präzisieren, schreiben wir  $W_r(y) := \prod_{j=1}^d (y_j - r, y_j + r]$  für den halboffenen  $d$ -dimensionalen Würfel mit Seitenlänge  $2r$  und Mittelpunkt  $y$ . Im weiteren verwenden wir auf  $\mathbb{R}^d$  der Einfachheit halber die Maximumsnorm  $\|y\|_\infty = \max_{1 \leq j \leq d} |y_j|$ . (Offene bzw abgeschlossene  $r$ -Umgebungen in dieser Norm sind dann nämlich offene bzw abgeschlossene Würfel.) Dazu passend nehmen wir für eine  $d \times d$ -Matrix  $M = (M_{ij})_{1 \leq i, j \leq d}$  als Norm  $\|M\| := \sum_{1 \leq i, j \leq d} |M_{ij}|$ , sodass  $\|M \cdot y\|_\infty \leq \|M\| \|y\|_\infty$ . Mit  $\text{Id}$  bezeichnen wir die Identität auf  $\mathbb{R}^d$  bzw die Einheitsmatrix.

**Lemma 3.7 ( $\mathcal{C}^1$ -Bilder von Würfeln bleiben in Würfeln)** Sei  $\varphi : Y \rightarrow \varphi(Y)$  ein  $\mathcal{C}^1$ -Diffeomorphismus offener Teilmengen von  $\mathbb{R}^d$  mit

$$\|D\varphi(y) - \text{Id}\| < \varepsilon \quad \text{für } y \in Y. \quad (3.43)$$

Dann gilt

$$\varphi(W_r(y)) \subseteq W_{(1+\varepsilon)r}(\varphi(y)) \quad \text{für } y \in Y \text{ und } r > 0 \text{ mit } W_r(y) \subseteq Y. \quad (3.44)$$

**Beweis.** Aufgrund obigen Mittelwertsatzes gilt für jeden Punkt  $z \in W_r(y)$  nämlich

$$\varphi(z) - \varphi(y) = D\varphi[y, z] \cdot (z - y).$$

Aber für  $w = z - y$  ist

$$\begin{aligned} \|D\varphi[y, z] \cdot w\|_\infty &= \left\| w + \left( \int_0^1 (D\varphi(y + tw) - \text{Id}) dt \right) \cdot w \right\| \\ &\leq \|w\|_\infty + \left\| \left( \int_0^1 (D\varphi(y + tw) - \text{Id}) dt \right) \cdot w \right\| \\ &\leq \|w\|_\infty + \left\| \int_0^1 (D\varphi(y + tw) - \text{Id}) dt \right\| \|w\|_\infty \\ &\leq \left( 1 + \int_0^1 \|D\varphi(y + tw) - \text{Id}\| dt \right) \|w\|_\infty \\ &< (1 + \varepsilon) \|w\|_\infty, \end{aligned}$$

also  $\|\varphi(z) - \varphi(y)\|_\infty < (1 + \varepsilon) \|z - y\|_\infty \leq (1 + \varepsilon)r$ , dh  $\varphi(z) \in W_{(1+\varepsilon)r}(\varphi(y))$ . ■

Eine kleine analytische Feststellung bereiten wir auch noch vor (siehe PS).



**Bemerkung 3.12** Für jede natürliche Zahl  $d \geq 1$  gibt es eine Konstante  $\kappa_d$  mit folgender Eigenschaft: Sind  $a, b > 0$  und  $\varepsilon \in (0, 1)$ , dann gilt

$$\frac{1}{(1 + \varepsilon)^d} \leq \frac{a}{b} \leq (1 + \varepsilon)^d \implies |a - b| \leq \kappa_d b \varepsilon.$$

Damit sind wir bereit für den

**Beweis von Satz 3.23.** (i) In (3.36) steht rechts und links jeweils ein Maß auf  $\mathcal{L}_Y$ . Das Argument unten zeigt, dass die beiden auf  $\mathcal{B}_Y$  übereinstimmen. Daraus sieht man, dass sie insbesondere genau dieselben Borel-messbaren Nullmengen wie  $\lambda^d$  besitzen. Satz 2.6 zeigt dann, dass sie auch auf  $\mathcal{L}_Y$  übereinstimmen.

Wir zeigen Gleichheit dieser Mengenfunktionen auf einer ausreichend reichhaltigen Familie  $\mathcal{H}$  von Mengen, und wenden dann die Eindeutigkeitsaussage aus Proposition 2.8 an. Konkret werden wir mit der Familie  $\mathcal{H}$  aller  $d$ -dimensionalen halboffenen Quader  $Q = \prod_{j=1}^d (a_j, b_j]$  mit  $a_j \leq b_j$  in  $\mathbb{Q}$  arbeiten, deren Abschluss  $\overline{Q}$  noch ganz in  $Y$  enthalten ist,  $\overline{Q} \subseteq Y$ . Weil wir längst wissen, dass  $\mathcal{I}_{\mathbb{R}^d}$  ein Halbring ist, erkennen wir leicht, dass dies auch für  $\mathcal{H}$  gilt. Es ist auch nicht schwer zu sehen, dass dieser  $\mathcal{B}_Y$  erzeugt. Weiters sind beide Maße  $\sigma$ -endlich auf  $\mathcal{H}$ . Wir bekommen also Gültigkeit von (3.36) auf  $\mathcal{B}_Y$  sobald wir nachweisen, dass

$$\lambda^d(\psi(Q)) = \int_Q |\det D\psi| d\lambda^d \quad \text{für } Q \in \mathcal{H}. \quad (3.45)$$

(ii) Es sei nun ein festes  $Q \in \mathcal{H}$  gegeben und  $\varepsilon \in (0, 1)$  beliebig. Wir nutzen die Stetigkeit der Ableitung  $D\psi$ , um  $Q$  in ausreichend kleine  $d$ -dimensionale Würfel zu zerlegen, auf denen  $D\psi$  und damit  $(D\psi)^{-1}$  und  $|\det D\psi(y)|$  jeweils wenig variieren.

Weil  $y \mapsto D\psi$  auf  $Y$  stetig ist, gilt dasselbe für  $y \mapsto \|D\psi\|$ , und auf der kompakten Menge  $\overline{Q}$  ist diese Funktion deshalb beschränkt, es gibt somit eine Konstante  $C_0$  mit

$$\|D\psi(y)\| \leq C_0 < \infty \quad \text{für } y \in Q. \quad (3.46)$$

Generell ist auch die Funktion  $M \mapsto |\det M|$  stetig, und so sehen wir, dass die Funktion  $J_\psi : Y \rightarrow (0, \infty)$  mit  $J_\psi(y) := |\det D\psi(y)|$  ebenso stetig ist. (Man beachte, dass  $J_\psi \neq 0$  auf  $Y$  weil  $\psi$  ein Diffeomorphismus ist.) Eingeschränkt auf die kompakte Menge  $\overline{Q}$  ist  $J_\psi$  daher sogar gleichmäßig stetig und beschränkt, dh es gibt eine Konstante  $C_1$  derart, dass

$$J_\psi(y) \leq C_1 < \infty \quad \text{für } y \in Q, \quad (3.47)$$

und ein  $\rho_1 > 0$  für welches

$$|J_\psi(y) - J_\psi(z)| < \varepsilon \quad \text{falls } z \in W_r(y) \subseteq Q \text{ mit } r \in (0, \rho_1]. \quad (3.48)$$

Als nächstes verwenden wir, dass auch das Invertieren  $M \mapsto M^{-1}$  von Matrizen eine stetige Abbildung ist, eben auf der Menge  $\text{GL}(d, \mathbb{R}) \subseteq \mathbb{R}^{d^2}$  der invertierbaren Matrizen. Die für uns interessanten Matrizen  $D\psi(y)$ ,  $y \in Q$ , liegen aber immer in  $\text{GL}(d, \mathbb{R})$ , also ist  $z \mapsto (D\psi(z))^{-1}$  stetig auf  $Y$ . Zuletzt ist die Matrixmultiplikation  $(M, M') \mapsto MM'$  ebenfalls stetig, und daher  $(y, z) \mapsto (D\psi(z))^{-1} D\psi(y)$  stetig auf  $Y^2$ , und gleichmäßig stetig auf der kompakten Teilmenge  $Q^2$ . Es gibt deshalb ein  $\rho \in (0, \rho_1)$  so, dass

$$\|(D\psi(z))^{-1} D\psi(y) - \text{Id}\| < \varepsilon \quad \text{falls } z \in W_r(y) \subseteq Q \text{ mit } r \in (0, \rho). \quad (3.49)$$

(Um sofort zu sehen, dass dies aus gleichmäßiger Stetigkeit folgt, beachte  $\text{Id} = (D\psi(y))^{-1} D\psi(y)$ , und verwende, dass aus  $z \in W_r(y) \subseteq Q$  gleich  $(y, z), (y, y) \in Q^2$  mit  $\|(y, z) - (y, y)\|_\infty < \rho$  folgt.)

(iii) Weil  $Q$  rationale Endpunkt besitzt, können wir diesen Quader in endlich viele Würfel  $W_l = W_r(y_l)$  mit  $r \in (0, \rho)$  und  $y_1, \dots, y_m \in Q$  zerlegen,  $Q = \bigsqcup_{l=1}^m W_l$ . Auf  $W_l$  vergleichen wir  $\psi$  mit  $\Psi_l(z) := \psi(y_l) + D\psi(y_l)(z - y_l) = x_l + D\psi(y_l)(z)$ ,

der affinen Taylorapproximation an  $\psi$  in  $y_l$  mit  $D\Psi_l(z) = D\psi(y_l)$  für alle  $z \in \mathbb{R}^d$ . Aus (3.33) in Proposition 3.12 wissen wir

$$\lambda^d(\Psi_l(W_l)) = \lambda^d(D\psi(y_l)(W_l)) = |\det D\psi(y_l)| \lambda^d(W_l) = J_\psi(y_l) \lambda^d(W_l). \quad (3.50)$$

Wir behaupten nun

$$\frac{1}{(1+\varepsilon)^d} \leq \frac{\lambda^d(\psi(W_l))}{\lambda^d(\Psi_l(W_l))} \leq (1+\varepsilon)^d \quad \text{für } l \in \{1, \dots, m\}, \quad (3.51)$$

woraus wegen Bemerkung 3.12 und (3.50) plus (3.47) folgt, dass

$$\left| \lambda^d(\psi(W_l)) - \lambda^d(\Psi_l(W_l)) \right| \leq \kappa_d C_1 \varepsilon \lambda^d(W_l) \quad \text{für } l \in \{1, \dots, m\}. \quad (3.52)$$

(iv) Wir beschränken uns auf den Nachweis der rechten Hälfte von (3.51). (Für die linke Hälfte kann man ein ähnliches Argument verwenden.) Dazu betrachten wir die  $\mathcal{C}^1$ -Abbildung  $\varphi := \Psi_l^{-1} \circ \psi$  auf  $W_l$ . Auch  $\Psi_l^{-1}$  ist eine affine Abbildung, und besitzt die konstante Ableitung  $D(\Psi_l^{-1}) = (D\psi(y_l))^{-1}$ , also ist

$$D\varphi(y) = D(\Psi_l^{-1} \circ \psi)(y) = (D\psi(y_l))^{-1} D\psi(y).$$

Da wir die  $W_l$  mit Seitenlänge kleiner als  $\rho$  gewählt haben, garantiert (3.49)

$$\|D\varphi(y) - \text{Id}\| = \|(D\psi(y_l))^{-1} D\psi(y) - \text{Id}\| < \varepsilon \quad \text{für } l \in \{1, \dots, m\},$$

und Lemma 3.7 impliziert daher  $\varphi(W_l) = \varphi(W_r(y_j)) \subseteq W_{(1+\varepsilon)r}(y_l)$  bzw. äquivalent dazu,

$$\psi(W_l) \subseteq \Psi_l(W_{(1+\varepsilon)r}(y_l)) = x_l + D\psi(y_l)(W_{(1+\varepsilon)r}(y_l)) = x_l + (1+\varepsilon)D\psi(y_l)(W_l).$$

Damit sehen wir aber, dass wirklich

$$\begin{aligned} \lambda^d(\psi(W_l)) &\leq \lambda^d(x_l + (1+\varepsilon)D\psi(y_l)(W_l)) \\ &= (1+\varepsilon)^d \lambda^d(D\psi(y_l)(W_l)) = (1+\varepsilon)^d J_\psi(y_l) \lambda^d(W_l). \end{aligned}$$

(v) Wir finden nun wegen  $\psi(Q) = \bigsqcup_{l=1}^m \psi(W_l)$  und (3.50),

$$\begin{aligned} \left| \lambda^d(\psi(Q)) - \int_Q |\det D\psi| d\lambda^d \right| &= \left| \sum_{l=1}^m \lambda^d(\psi(W_l)) - \sum_{l=1}^m \int_{W_l} J_\psi d\lambda^d \right| \\ &\leq \sum_{l=1}^m \left| \lambda^d(\psi(W_l)) - \lambda^d(\Psi_l(W_l)) \right| + \sum_{l=1}^m \left| J_\psi(y_l) \lambda^d(W_l) - \int_{W_l} J_\psi d\lambda^d \right|. \end{aligned}$$

Dabei ist für jedes  $l$  wegen (3.48)

$$\left| J_\psi(y_l) \lambda^d(W_l) - \int_{W_l} J_\psi d\lambda^d \right| \leq \int_{W_l} |J_\psi(y_l) - J_\psi| d\lambda^d < \varepsilon \lambda^d(W_l).$$

Gemeinsam mit (3.52) ergeben diese Abschätzungen

$$\left| \lambda^d(\psi(Q)) - \int_Q |\det D\psi| d\lambda^d \right| < (1 + \kappa_d C_1) \varepsilon \lambda^d(Q),$$

und weil  $\varepsilon \in (0, 1)$  beliebig war, folgt (3.45). ■

## 4. Ein Blick in die Stochastik

---

Stochastik ist die Mathematik des Zufalls, sie beinhaltet Wahrscheinlichkeitstheorie, Statistik und verwandte Gebiete. In diesem Kapitel besprechen wir, wie man Situationen in denen der Zufall eine Rolle spielt mathematisch beschreiben kann, und diskutieren (zwangsläufig knapp) zwei besonders prominente Ergebnisse aus der Wahrscheinlichkeitstheorie. Das passt deshalb gut zu einer Einführung in Maß- und Integrationstheorie, weil Wahrscheinlichkeiten - mathematisch betrachtet - einfach spezielle Maße sind, und man ohne Maßtheorie manche grundlegende Fragen der Stochastik kaum vernünftig formulieren (und erst recht nicht beweisen) kann.

### 4.1 Grundbegriffe - Der formale Rahmen

**A) Wahrscheinlichkeitsräume und Interpretation.** Wir beginnen mit der mathematischen Beschreibung eines *Zufallsexperimentes* (ZE), dh eines vom Zufall (mit)bestimmten Vorganges mit vorgegebenen möglichen Ergebnissen, den wir unter identischen Bedingungen real (zB Würfeln) oder wenigstens im Prinzip (Belastungstest für ein Atomkraftwerk) mehrfach wiederholen könn(t)en. Das mathematische Modell eines ZE ist durch einen normierten Maßraum gegeben:

**Definition 4.1** *Ein Wahrscheinlichkeitsmaß (W-Maß, W-Verteilung)  $P$  auf einem messbaren Raum  $(\Omega, \mathcal{F})$  ist ein normiertes Maß, dh  $P[\Omega] = 1$ . Eine Wahrscheinlichkeitsdichte (W-Dichte) auf einem Maßraum  $(\Omega, \mathcal{F}, \mu)$  (oder bezüglich  $\mu$ ) ist eine messbare Funktion  $h : \Omega \rightarrow [0, \infty]$  mit  $\int_{\Omega} h d\mu = 1$  (sodass  $h \odot \mu$  ein W-Maß ist).*

*Ein Wahrscheinlichkeitsraum (W-Raum) ist ein normierter Maßraum  $(\Omega, \mathcal{F}, P)$ . Die Menge  $\Omega \neq \emptyset$  ist der Grundraum, Mengen  $A \in \mathcal{F}$  werden als Ereignisse bezeichnet. Der Wert  $P[A]$  ist die Wahrscheinlichkeit (Wkeit) des Ereignisses  $A$ . Falls  $A = \{\omega\}$ , so schreiben wir einfach  $P[\omega]$  statt  $P[\{\omega\}]$ .*

**Die übliche Interpretation dieses Setups:** Die Elemente  $\omega$  der Menge  $\Omega$  repräsentieren die möglichen Ergebnisse des ZE. Das ZE selbst besteht dann in der Auswahl eines  $\omega \in \Omega$ , der *Realisierung* des ZE. Auf welche Art  $\omega$  ausgewählt wird, ist uns dabei im Detail nicht bekannt ("Zufall"). Man sagt, das *Ereignis*  $A$  tritt ein, falls die Realisierung  $\omega$  zu  $A$  gehört.

Mit  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$  wird die Familie jener Teilmengen festgelegt, über deren Wkeiten wir sprechen wollen bzw können. Durch die Maßtheorie sind wir gewarnt, und wissen, dass es gerade in überabzählbaren Mengen nicht immer günstig ist, gleich mit der Potenzmenge zu arbeiten. Offenbar tritt das "sichere Ereignis"  $\Omega$  immer ein. Sind  $A, B, A_1, A_2, \dots$  Ereignisse, so tritt  $A^c$  genau dann ein, wenn  $A$  nicht eintritt.  $A \cap B$  tritt genau dann ein, wenn  $A$  und  $B$  eintreten. Weiters tritt  $\bigcup_{n \geq 1} A_n$  genau dann ein, wenn wenigstens eines der  $A_n$  eintritt usw. Da interessante Ereignisse oft die Kombination abzählbar vieler einfacherer Ereignisse erfordern, sollen auch  $A^c$ ,  $A \cap B$  und  $\bigcup_{n \geq 1} A_n$  zu  $\mathcal{F}$  gehören. Das bedeutet aber gerade, dass  $\mathcal{F}$  eine  $\sigma$ -Algebra sein soll. Wir vereinbaren hier gleich, dass wir im Falle eines abzählbaren Grundraumes  $\Omega$  immer  $\mathcal{F} := \mathcal{P}(\Omega)$  wählen.

Die konkrete Wahl von  $\Omega$  und  $\mathcal{F}$  hängt insbesondere davon ab, welche Fragen man letztlich beantworten möchte. Interessieren wir uns beim Würfeln für die erzielte Augenzahl, oder auch für die Lage des Würfels auf dem Tisch?  $\Omega$  muss immer ausreichend detailliert sein, um die für uns relevante Information widerspiegeln zu können, sollte aber nicht unnötig kompliziert werden. Oft gibt es (wie

im folgenden Beispiel) eine naheliegende “kanonische” Wahl von  $\Omega$ , wir schauen uns aber auch bald Situationen an, wo es sich lohnt, nicht allzu minimalistisch vorzugehen (Beispiel 4.9 unten).

**Beispiel 4.1** Wie gesagt sei  $\mathcal{F} = \mathcal{P}(\Omega)$  wenn  $\Omega$  abzählbar ist. Ansonsten brauchen wir  $\mathcal{F}$  wenigstens so groß, dass die genannten Teilmengen Ereignisse werden. Wir sagen später noch mehr dazu.

**a) Würfel werfen.** Kanonische Wahl: die Menge der möglichen Augenzahlen,  $\Omega := \{1, \dots, 6\}$ , Interpretation  $\omega =$  “erzielte Augenzahl”. Ereignisse  $zB A := \{2, 4, 6\} \simeq$  “Augenzahl ist gerade” oder  $D := \{6\} \simeq$  “6er gewürfelt”.

**b) Münze werfen.** Kanonische Wahl: irgendein  $\Omega$  mit zwei Elementen,  $zB \Omega := \{0, 1\}$ , Interpretation  $\omega = 0$  für “Bild”,  $\omega = 1$  für “Zahl”. Ereignis  $A := \{0\} \simeq$  “Bild geworfen”

**a') Zwei(mal) Würfel werfen.** Kanonische Wahl: die Menge der möglichen Paare von Augenzahlen,  $\Omega := \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^2$ , Interpretation  $\omega_i =$  “Augenzahl beim  $i$ ten Wurf”. Ereignisse  $zB A := \{\omega : \omega_1 \in \{2, 4, 6\}\} = \{2, 4, 6\} \times \{1, \dots, 6\} \simeq$  “erste Augenzahl ist gerade”,  $B := \{\omega : \omega_2 \in \{2, 4, 6\}\} \simeq$  “zweite Augenzahl ist gerade” oder  $C := \{\omega : 2 \nmid \omega_1 + \omega_2\} \simeq$  “Summe der Augenzahlen ist ungerade”,  $D := \{\omega : 6 \in \{\omega_1, \omega_2\}\} \simeq$  “mindestens ein 6er”.

**b') Unendlich oft die Münze werfen.** Kanonische Wahl: die Menge der möglichen Folgen von Einzelergebnissen,  $\Omega := \{\omega = (\omega_i)_{i \geq 1} : \omega_i \in \{0, 1\}\} = \{0, 1\}^{\mathbb{N}^+}$ , Interpretation  $\omega_i =$  “Ergebnis  $i$ ter Münzwurf”. Interessante Ereignisse wären  $zB A_i := \{\omega : \omega_i = 0\} \simeq$  “im  $i$ ten Wurf Bild geworfen”,  $\bigcap_{i=1}^5 A_i \simeq$  “zu Beginn wird 5mal Bild geworfen”,  $B := \bigcup_{i \geq 1} A_i \simeq$  “es wird irgendwann Bild geworfen”,  $C := \{\omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \omega_i = \frac{1}{2}\} \simeq$  “langfristig in der Hälfte der Runden Zahl geworfen”. Beachten Sie bitte dass dies eine wesentlich kompliziertere Situation ist als zuvor, weil der Folgenraum  $\Omega$  überabzählbar ist. Um damit zurechtzukommen müssen wir wirklich Maßtheorie verwenden! Als geeignete  $\sigma$ -Algebra  $\mathcal{F}$  erweist sich hier übrigens die von den Mengen  $A_i$  erzeugte,  $\mathcal{F} := \sigma(A_i : i \geq 1)$ .

**c) Aus einer Menge  $M$  zwei unterschiedliche Elemente ziehen (Reihenfolge egal).** Kanonische Wahl: die Menge aller Teilmengen von  $M$  mit zwei Elementen,  $\Omega := \{\omega \subseteq M : |\omega| = 2\}$ , Interpretation  $\omega =$  “Menge der gezogenen Elemente”.  $A := \{\omega : m \in \omega\} \simeq$  “das Element  $m$  wurde gezogen”.

**d) Zufällig einen Punkt in  $(0, 1]$  wählen.** Kanonische Wahl: die Menge der möglichen Punkte,  $\Omega := (0, 1]$ , Interpretation  $\omega =$  “gewählter Punkt”,  $A := (0, \frac{1}{3}] \simeq$  “der Punkt liegt im linken Drittel”.

**e) Zufällig zwei Punkte in  $(0, 1]$  wählen.** Kanonische Wahl: die Menge der möglichen Paare von Punkten,  $\Omega := \{\omega = (\omega_1, \omega_2) : \omega_i \in (0, 1]\} = (0, 1]^2$ , Interpretation  $\omega_i =$  “ $i$ ter gewählter Punkt”,  $A := \{\omega : \omega_1 < \omega_2\} \simeq$  “erster Punkt liegt links vom zweiten”.

**f) Zufällige Bewegung eines Teilchens in  $\mathbb{R}^3$ .** Kanonische Wahl: die Menge der stetigen Kurven in  $\mathbb{R}^3$ ,  $\Omega := \mathcal{C}([0, \infty), \mathbb{R}^3) = \{\omega = (\omega_t)_{t \geq 0} : [0, \infty) \rightarrow \mathbb{R}^3 \mid \omega \text{ stetig}\}$ , Interpretation  $\omega_t =$  “Position des Teilchens zur Zeit  $t$ ”,  $A := \{\omega : \|\omega_t\| \leq 1 \text{ für } t \in [0, 1]\} \simeq$  “Teilchen entfernt sich bis Zeit  $t = 1$  höchstens 1 mm vom Startpunkt”.

Der messbare Raum  $(\Omega, \mathcal{F})$  bietet nur das Gerüst des Modells. Die entscheidende Zutat ist das W-Maß  $P$ . Durch  $P[A]$  wird quantifiziert, wie hoch wir die Chance des Eintretens von  $A$  einschätzen. Gerade bei einfachen realen ZE wie dem Würfeln oder dem Münzwurf legt man die Beobachtung zugrunde, dass ein gegebenes Ereignis  $A$  bei häufiger Wiederholung des ZE mit einer bestimmten relativen Häufigkeit aufzutreten scheint. Stellen wir uns etwa vor, die Münze wird  $n$ mal geworfen, und  $S_n(A)$  ist die Anzahl der Würfe, bei denen  $A =$  “Zahl” eintritt. Die (zufällige) relative Häufigkeit von “Zahl” unter diesen  $n$  Würfeln ist dann  $S_n(A)/n \in [0, 1]$ , und unsere Erfahrung suggeriert, dass es eine Zahl  $P[A]$

(die *asymptotische Häufigkeit*) gibt mit

$$P[A] = \lim_{n \rightarrow \infty} \frac{S_n(A)}{n} \in [0, 1]. \quad (4.1)$$

Für den Moment nehmen wir einfach mal optimistisch an, dass dies tatsächlich generell der Fall ist, und wir auch in anderen Situationen jedem Ereignis in  $(\Omega, \mathcal{F})$  eine Wkeit im Sinne von (4.1) zuordnen können. Dann ist  $P$  also eine Mengenfunktion  $P : \mathcal{F} \rightarrow [0, 1]$ , und wegen  $S_n(\emptyset) = 0$  und  $S_n(\Omega) = n$  ist jedenfalls  $P[\emptyset] = 0$  und  $P[\Omega] = 1$ . Sind  $A_1, \dots, A_m \in \mathcal{F}$  paarweise disjunkt, so kann bei jedem Versuch höchstens eines der  $A_j$  eintreten, und wir haben  $S_n(\biguplus_{j=1}^m A_j) = S_n(A_1) + \dots + S_n(A_m)$  für jedes  $m$ , und somit  $P[\biguplus_{j=1}^m A_j] = P[A_1] + \dots + P[A_m]$ . Wenn also an dem Phänomen der vermuteten Konvergenz (4.1) relativer Häufigkeiten etwas dran ist, dann definiert die asymptotische Häufigkeit  $P$  also eine endlich additive normierte Mengenfunktion auf  $(\Omega, \mathcal{F})$ . Mit stärkeren Hilfsmitteln aus der fortgeschrittenen Maßtheorie (und bei geeigneter Auslegung von (4.1)) kann man sogar sehen, dass  $P$   $\sigma$ -additiv, und damit ein W-Maß sein sollte. Die Interpretation über asymptotische Häufigkeiten versetzt uns also weitgehend zwangsläufig in den Rahmen der Maßtheorie! Beachten Sie, dass es vorerst aber wirklich nur eine Interpretation ist, in der Definition ist nicht von Häufigkeiten die Rede. Vielmehr werden wir später beweisen, dass  $P$  etwas mit Häufigkeiten zu tun hat. Dieser Satz, das Gesetz der Großen Zahl, bestätigt dann, dass der eben eingeführte formale Rahmen wirklich zu unserem intuitiven Verständnis passt, und ist entsprechend ein zentraler Pfeiler der Theorie.

Wie kommt man nun zu  $P$ ? Ein ganz einfacher Fall, anhand dessen man schon mal allerlei interessante Überlegungen anstellen kann, ist jener, dass  $\Omega$  endlich ist, und man aufgrund von Symmetrieeigenschaften davon ausgeht, dass alle  $\omega \in \Omega$  dieselbe Wahrscheinlichkeit besitzen (zB bei einer symmetrischen und daher *fairen* Münze). Das sehen wir uns im nächsten Unterabschnitt gleich genauer an. Ansonsten muss man auch in simplen Situationen eine plausible Annahme über  $P$  treffen, die sich zB auf Experimente (und die dort beobachteten relativen Häufigkeiten) stützt (zB bei einer verbogenen Münze oder einem gezinkten Würfel). Sehr oft kombiniert man solche Erwägungen für die Zutaten eines komplexeren ZE mit strukturellen Überlegungen, die Zusammenhänge zwischen diesen Komponenten betreffen. Das betrachten wir aber erst im nächsten Abschnitt.

**B) Gleichwahrscheinliche Ausgänge.** Eine grundlegende Klasse von Modellen bilden *Laplace-Experimente*. Dabei ist  $\Omega$  eine endliche Menge, und wir haben Grund zur Annahme, dass alle Einzelwahrscheinlichkeiten übereinstimmen,  $P[\omega'] = P[\omega'']$  für alle  $\omega', \omega'' \in \Omega$ . Dann folgt sofort  $P[\omega'] = 1/|\Omega|$  bzw  $P[A] = |A|/|\Omega|$  für alle  $A \in \mathcal{F} = P(\Omega)$  man spricht von der (*diskreten*) *uniformen Verteilung* auf  $\Omega$ , formal haben wir  $P = \frac{1}{|\Omega|} \#$  (das normierte Zählmaß). Um die Wkeit eines Ereignisses  $A$  zu berechnen, muss man also nur ein wenig Kombinatorik betreiben, und die Elemente von  $A$  zählen. Das kann allerdings durchaus herausfordernd sein. Wir sehen uns einige Standardsituationen und -fragen an.

**Beispiel 4.2 a) Fairer Würfel.** Ist der Würfel symmetrisch, so liegt es nahe, auf  $\Omega := \{1, \dots, 6\}$  die uniforme Verteilung zu wählen,  $P[\omega] = \frac{1}{6}$  für jedes  $\omega \in \Omega$  weil  $|\Omega| = 6$ . Für  $A \simeq$  "Augenzahl ist gerade" ist  $|A| = 3$  und daher  $P[A] = \frac{1}{2}$ .

**a') Zwei faire Würfel.** Wie im vorigen Beispiel sei  $\Omega := \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}\}$  mit  $|\Omega| = 36$ . Sind die Würfel symmetrisch, und werden sie unabhängig voneinander geworfen, so ist wieder die Laplace-Annahme gerechtfertigt. Für die Ereignisse aus Beispiel 4.1 a') finden wir  $|A| = |B| = |C| = 18$  und weiters zB  $|A \cap B| = |A \cap C| = 9$ ,  $|A \cap B \cap C| = 0$ , also  $P[A] = P[B] = P[C] = \frac{1}{2}$  und  $P[A \cap B] = P[A \cap C] = \frac{1}{4}$  während  $P[A \cap B \cap C] = 0$ . Schliesslich ist  $|D| = |\{(1, 6), \dots, (6, 6), (6, 5), \dots, (6, 1)\}| = 11$  und  $P[D] = \frac{11}{36}$ .

**c) Aus einer Menge  $M$  zwei unterschiedliche Elemente ziehen (Reihenfolge egal).** Man beschreibt dieses Szenario auch als "Ziehen ohne Zurücklegen". Mit  $\Omega := \{\omega \subseteq M : |\omega| = 2\}$  scheint die Laplace-Annahme plausibel, zB wenn

man sich vorstellt, dass aus einem Beutel mit unterschiedlich gefärbten oder nummerierten (aber ansonsten ununterscheidbaren) Kugeln blind zwei herausgenommen werden. Ist etwa  $M = \{R_1, \dots, R_4, B_1, \dots, B_3, G_1, \dots, G_3\}$ ,  $|M| = 10$ , mit roten, blauen und grünen Kugeln, dann ist  $|\Omega| = \binom{10}{2} = 45$ . Betrachte nun  $B := \{\omega : \omega \subseteq \{B_1, B_2, B_3\}\} \simeq$  "zwei blaue Kugeln gezogen" und  $C :=$  "zwei Kugeln unterschiedlicher Farbe gezogen". Im ersten Fall ist  $|B| = \binom{3}{2} = 3$  also  $P[B] = \frac{3}{45} = \frac{1}{15}$ . Im zweiten Fall ist es günstig,  $C^c$  zu beschreiben, und dann  $P[C] = 1 - P[C^c]$  zu nutzen. Es ist ja  $C^c = \{\omega : \omega \subseteq \{R_1, \dots, R_4\}\} \uplus \{\omega : \omega \subseteq \{B_1, B_2, B_3\}\} \uplus \{\omega : \omega \subseteq \{G_1, G_2, G_3\}\}$ , und somit  $|C^c| = \binom{4}{2} + \binom{3}{2} + \binom{3}{2} = 12$ . Insgesamt  $|C| = |\Omega| - |C^c| = 33$  und  $P[C] = \frac{33}{45} = \frac{11}{15}$ .

## 4.2 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Unsere intuitive Vorstellung von Wkeiten beruht auf dem Eindruck, dass ein gegebenes Ereignis  $A$  bei wiederholter Durchführung des ZE mit einer bestimmten asymptotischen Häufigkeit, eben  $P[A]$ , auftritt,

$$P[A] \approx \frac{S_n(A)}{n} \quad \text{wenn } n \text{ groß.}$$

Bisweilen sind wir nur an jenen Versuchen interessiert, deren Ergebnis einer weiteren Bedingung genügt. Wir beginnen mit einem sehr einfachen Beispiel.

**Beispiel 4.3** *Es wird wieder mehrmals gewürfelt, und ich gewinne jeweils einen "I ♥ Maths"-Sticker wenn dabei die Augenzahl 1, 3 oder 6 (insgesamt das Ereignis  $A$ ) auftritt. Bei einem fairen Würfel wäre dies laut unserer Interpretation langfristig bei der Hälfte aller Versuche der Fall,  $P[A] = \frac{1}{2}$ . Aus einem tiefsitzenden Aberglauben heraus fürchte ich jedoch ungerade Zahlen, und bestehe daher darauf, jene Versuche zu ignorieren, in denen eine ungerade Augenzahl erscheint. Wie häufig werde ich in den verbleibenden Runden gewinnen? Wir schreiben  $B$  für das Ereignis "gerade Augenzahl", und  $S_n(B)$  für die Anzahl jener Würfe unter den ersten  $n$  Versuchen, bei denen dies eintritt. Dann ist  $A \cap B$  das für mich günstige Ereignis, dass ich gewinne, weil eine der Glückszahlen 1, 3, 6 auftritt, und diese auch gewertet wird, weil sie gerade (also  $=6$ ) ist. Mit  $S_n(A \cap B)$  bezeichnen wir die Anzahl solcher Runden unter den ersten  $n$  Versuchen. Die gesuchte Häufigkeit ist dann*

$$\frac{S_n(A \cap B)}{S_n(B)} = \frac{S_n(A \cap B)}{n} \frac{n}{S_n(B)} \approx \frac{P[A \cap B]}{P[B]}.$$

Langfristig ist also unter der Bedingung, dass nur Runden gewertet werden, die  $B$  erfüllen, der Anteil der Runden in denen ich gewinne durch  $P[A \cap B]/P[B]$  gegeben. Wir werden das unten mit  $P[A | B]$  bezeichnen. Beim fairen Würfel ist das gerade  $\frac{1}{3}$ .

Wir lernen daraus nicht nur, dass es unklug ist, jedem Aberglauben zu folgen, sondern (fast so wichtig) auch, wie man asymptotische Häufigkeiten bestimmt, wenn man nur Versuche zählt, in denen eine vorgegebene Bedingung  $B$  erfüllt ist.

Man kann  $P[A | B]$  auch interpretieren als Wkeit des Eintretens von  $A$  wenn man schon erfahren hat, dass  $B$  eintritt: Wenn Sie mir verraten, dass eben eine gerade Zahl gewürfelt wurde, und ich darauf wetten soll, dass ich nun einen Sticker bekomme, dann wäre  $P[A | B]$  anstelle von  $P[A]$  ein fairer Wetteinsatz. Ich befinde mich dann ja nicht mehr in der Situation, dass ich das Ergebnis eines beliebigen Wurfes mit gleichwahrscheinlichen Ausgängen einschätzen soll. Vielmehr kann ich die fragliche Augenzahl als das Ergebnis eines modifizierten ZE ansehen, in dem nur gerade Zahlen gewertet werden.

Dem Beispiel folgend vereinbaren wir allgemein:

**Definition 4.2** Sind  $A, B$  Ereignisse im  $W$ -Raum  $(\Omega, \mathcal{F}, P)$  und  $P[B] > 0$ , so ist die bedingte Wahrscheinlichkeit von  $A$  gegeben  $B$  definiert als

$$P[A | B] := \frac{P[A \cap B]}{P[B]}.$$

Das auf  $B$  bedingte Maß  $A \mapsto P[A | B]$ ,  $A \in \mathcal{F}$  ist wieder ein  $W$ -Maß auf  $(\Omega, \mathcal{F})$ . (Dieses beschreibt das modifizierte ZE, in dem nur Ergebnisse aus  $B$  betrachtet werden.)

So wie  $P[A]$  die relative Größe von  $A$  in  $\Omega$  misst, erfasst  $P[A | B]$  analog eben die relative Größe von  $A \cap B$  in  $B$ . Die Interpretation aus dem Beispiel bietet sich weiterhin an. Einen besonders simplen Spezialfall notieren wir explizit:

**Beispiel 4.4** Ist  $(\Omega, \mathcal{F}, P)$  ein Laplacescher  $W$ -Raum, dann gilt für  $A, B \subseteq \Omega$  mit  $B \neq \emptyset$  immer

$$P[A | B] = \frac{P[A \cap B]}{P[B]} = \frac{|A \cap B| |\Omega|}{|\Omega| |B|} = \frac{|A \cap B|}{|B|},$$

dh das auf  $B$  bedingte Maß ist einfach die uniforme Verteilung auf  $B$ .

Wir halten nun ein paar (sehr) einfache Rechenregeln für bedingte Wahrscheinlichkeiten fest.

**Proposition 4.1 (Gesetz der vollständigen Wahrscheinlichkeit)** Bilden die Ereignisse  $B_1, \dots, B_m$  mit  $P[B_i] > 0$  eine Partition von  $\Omega$ , so gilt

$$P[A] = \sum_{i=1}^m P[B_i] P[A | B_i] \quad \text{für } A \in \mathcal{F}.$$

**Beweis.** Wegen  $\Omega = \bigsqcup_{i=1}^m B_i$  gilt  $A = \bigsqcup_{i=1}^m A \cap B_i$ , und mit  $P[A \cap B_i] = P[B_i] P[A | B_i]$  finden wir

$$P[A] = \sum_{i=1}^m P[A \cap B_i] = \sum_{i=1}^m P[B_i] P[A | B_i],$$

wie behauptet. ■

Auf diese Weise eine Partition  $\{B_1, \dots, B_m\}$  des Grundraumes  $\Omega$  zur Bestimmung von  $P[A]$  zu verwenden bedeutet einfach *Fallunterscheidungen* zu treffen und ganz elementar mit relativen Anteilen zu rechnen. Mehr steckt nicht dahinter.

**Beispiel 4.5** In der Bevölkerung  $\Omega$  Mesopotamiens setzt sich die Gruppe  $A \subseteq \Omega$  jener Bürger\*innen, die an das Auswahlaxiom glauben dafür ein, den regierenden Messdiener abzusetzen, weil dieser die Existenz nicht messbarer Mengen verleugnet. Die Menge  $B_i \subseteq \Omega$  jener, die in der  $i$ -ten der fünf Provinzen Mesopotamiens leben, hat einen bekannten Anteil von  $|B_i| / |\Omega|$  in der Gesamtbevölkerung. Für jede der Provinzen wissen wir zudem, wie groß dort der Anteil  $|A \cap B_i| / |B_i|$  der AC-Fans ist. Wählen wir nun (uniform verteilt) zufällig eine Bürger\*in aus, so beträgt die Wkeit, dass diese das Auswahlaxiom akzeptiert gerade

$$P[A] = \sum_{i=1}^5 P[B_i] P[A | B_i] = \sum_{i=1}^5 \frac{|B_i|}{|\Omega|} \frac{|A \cap B_i|}{|B_i|},$$

und wird so unmittelbar durch die gegebenen Daten ausgedrückt.

Äußerlich aufwendiger als Proposition 4.1, aber im Kern genauso simpel ist

**Proposition 4.2 (Formel von Bayes)** Bilden die Ereignisse  $B_1, \dots, B_m$  mit  $P[B_i] > 0$  eine Partition von  $\Omega$ , so gilt

$$P[B_j | A] = \frac{P[B_j] P[A | B_j]}{\sum_{i=1}^m P[B_i] P[A | B_i]} \quad \text{für } A \in \mathcal{F} \text{ mit } P[A] > 0.$$

**Beweis.** Wir schreiben einfach

$$P[B_j | A] = \frac{P[B_j \cap A]}{P[A]} = \frac{P[A \cap B_j]}{P[A]} = \frac{P[B_j] P[A | B_j]}{P[A]},$$

und setzen für  $P[A]$  den Ausdruck aus der vorigen Proposition ein. ■

Die Formel von Bayes ist sehr nützlich, weil sie uns erlaubt, die Rollen von  $A$  und  $B$  zu vertauschen. Offenbar geht es wieder nur um korrektes Rechnen mit Proportionen (und damit eigentlich um Schulwissen). Umso klarer ist, dass Sie das auch beherrschen müssen. Immerhin treten schon im Alltag oftmals Fragestellungen auf, die man damit beantworten kann.

**Beispiel 4.6** *Jüngste Erkenntnisse weisen darauf hin, dass die politischen Unruhen in Mesopotamien auf die Verbreitung des gefürchteten Morona-Virus zurückzuführen sind, welcher betroffene Personen sporadisch an der Existenz nicht-messbarer Mengen zweifeln lässt. Eine erfolgreiche Therapie mit Vermismeynnicht-Extrakt ist möglich, setzt aber zuverlässige Diagnosemöglichkeiten voraus. Zum Glück gibt es einen neuen Antikörper-Test, der eine Infektion in 95% der betroffenen Bevölkerung erkennt. Bei nicht betroffenen Personen zeigt er aber in 10% der Fälle fälschlich das tückische Virus an. Derzeit ist rund 1% der Bevölkerung betroffen. Mit welcher Wahrscheinlichkeit ist eine zufällig ausgewählte Person infiziert, falls sie ein positives Testergebnis bekommt?*

Betrachte die Ereignisse  $A := \text{"Person ist infiziert"}$  und  $B := \text{"Person bekommt positives Testergebnis"}$ . Laut Angabe ist  $P[B | A] = \frac{95}{100}$ ,  $P[B | A^c] = \frac{1}{10}$  und  $P[A] = \frac{1}{100}$ . Daher finden wir

$$\begin{aligned} P[A | B] &= \frac{P[A \cap B]}{P[B]} = \frac{P[A \cap B]}{P[A \cap B] + P[A^c \cap B]} \\ &= \frac{P[A]P[B | A]}{P[A]P[B | A] + P[A^c]P[B | A^c]} \\ &= \frac{\frac{1}{100} \cdot \frac{95}{100}}{\frac{1}{100} \cdot \frac{95}{100} + \frac{99}{100} \cdot \frac{1}{10}} = \frac{1 \cdot 95}{1 \cdot 95 + 99 \cdot 10} \approx 0.087. \end{aligned}$$

Beachten Sie bitte, dass es hier nicht nur auf die Qualität des Tests ankommt, sondern ganz wesentlich auch auf  $P[A]$ , die Verbreitung des Virus in der Bevölkerung! Das wird bei solchen Fragen manchmal übersehen. Hätten wir in diesem Beispiel etwa  $P[A] = \frac{1}{4}$ , bekämen wir  $P[A | B] = 0.76$ , ein dramatischer Unterschied!

**Bemerkung 4.1** Die beiden Proposition bleiben korrekt, wenn wir abzählbare Partitionen anstelle endlicher Partitionen verwenden. (Überzeugen Sie sich bitte davon.)

Wie oben skizziert können wir  $P[A | B]$  als aktualisierte Einschätzung von  $A$  betrachten, wenn wir erfahren, dass  $B$  eintritt. Wie stark sich  $P[A | B]$  vom ursprünglichen  $P[A]$  unterscheidet zeigt daher wie sehr unsere Einschätzung von  $A$  vom Eintreten der Bedingung  $B$  abhängt. In vielen natürlichen Situationen (etwa wenn  $A$  und  $B$  die Ergebnisse des ersten bzw zweiten Würfels beschreiben) beeinflusst Kenntnis von  $B$  die Situation für  $A$  aber überhaupt nicht.

Das eben skizzierte Szenario ist von größter Bedeutung. In dieser besonders wichtigen und übersichtlichen Grundsituation können wir nämlich beginnen, das *Zusammenspiel mehrerer ZE* zu verstehen, und das ist zu weiten Teilen das zentrale Anliegen der W-Theorie. Wir formulieren das fundamentale Konzept der *stochastischen Unabhängigkeit* zuerst ganz einfach für Paare von Ereignissen, und sehen uns danach noch verschiedene Erweiterungen an.

**Definition 4.3** Zwei Ereignisse  $A$  und  $B$  heißen (voneinander) unabhängig falls

$$P[A \cap B] = P[A] P[B].$$



Die Bedingung ist offenbar stets erfüllt sobald  $P[A]$  oder  $P[B]$  in  $\{0, 1\}$  liegt. Mit diesem Begriff verbunden ist anschaulich eben die Vorstellung, dass Kenntnis über das Eintreten oder Nichteintreten von  $B$  keine Information über das Eintreten von  $A$  beinhaltet. Kurz:  $A$  und  $B$  wissen nichts voneinander. Manchmal ist es nützlich, dieses Konzept auf Familien von Ereignissen auszudehnen.

**Definition 4.4** Sind  $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$  Familien von Ereignissen, so nennen wir  $\mathcal{A}$  und  $\mathcal{B}$  (voneinander) unabhängig falls für jede Wahl von  $A \in \mathcal{A}$  und  $B \in \mathcal{B}$  die Ereignisse  $A$  und  $B$  voneinander unabhängig sind.

In der üblichen Interpretation bedeutet dies: kein Ereignis aus  $\mathcal{A}$  enthält Information über irgendein Ereignis aus  $\mathcal{B}$ . Es gibt viele Situationen, in denen sich so etwas automatisch ergibt:

**Beispiel 4.7 a)** Sind die Ereignisse  $A$  und  $B$  voneinander unabhängig, dann sind auch  $\mathcal{A} := \{\emptyset, \Omega, A, A^c\}$  und  $\mathcal{B} := \{\emptyset, \Omega, B, B^c\}$  voneinander unabhängig.

**b)** In unserem Modell für zwei Würfel,  $\Omega := \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}\}$  mit uniformer Verteilung  $P$ , sind zB die Familien  $\mathcal{A}_1$  und  $\mathcal{A}_2$  mit  $\mathcal{A}_i := \{\{\omega : \omega_i = k\} : k \in \{1, \dots, 6\}\}$  voneinander unabhängig. Mit ausreichend Geduld können Sie überprüfen, dass sogar  $\sigma(\mathcal{A}_1)$  und  $\sigma(\mathcal{A}_2)$  voneinander unabhängig sind.

Wozu betrachten wir hier Familien  $\mathcal{A}$  und  $\mathcal{B}$  von Ereignissen, anstatt einfach bei  $A$  und  $B$  zu bleiben? Diese Idee bietet uns eine Möglichkeit, in einem stochastischen Modell das Konzept eines bestimmten *Kenntnisstandes* zu erfassen. Falls Sie in Beispiel b) oben das Ergebnis des ersten Würfels kennen, dann wissen Sie insbesondere für jedes Ereignis  $A$  aus  $\mathcal{A}_1$  (bzw  $\sigma(\mathcal{A}_1)$ ) ob es eingetreten ist, oder nicht. Gerade in komplizierteren Situationen erweist es sich als sehr nützlich, die gesamte verfügbare Information durch die Familie jener Ereignisse zu repräsentieren, über welche man bescheid weiß. (Am besten funktioniert diese Interpretation übrigens wenn  $\mathcal{A}$  und  $\mathcal{B}$   $\sigma$ -Algebren sind.)

**Unabhängige Kombination zweier Zufallsexperimente.** Häufig ist die Unabhängigkeit zwischen Komponenten eines größeren ZE ein wesentlicher Teil der Modellierungsannahmen. Stellen wir uns vor, wir hätten uns schon W-Räume  $(\Omega_1, \mathcal{F}_1, P_1)$  und  $(\Omega_2, \mathcal{F}_2, P_2)$  verschafft, die jeweils ein bestimmtes ZE angemessen beschreiben (zB einmal Würfeln und ein Münzwurf). Nun wollen wir ein Modell  $(\Omega, \mathcal{F}, P)$  für ein größeres ZE konstruieren, welches beide als voneinander unabhängige Telexperimente enthält, zB zwei gleichzeitig an verschiedenen Tischen im Casino ablaufende Spiele. Als Ergebnisraum bietet sich natürlich das Cartesische Produkt  $(\Omega, \mathcal{F}) := (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$  an mit  $\Omega = \{\omega = (\omega_1, \omega_2) : \omega_i \in \Omega_i\}$ . Damit beide Teile korrekt eingebettet sind, muss  $P[\{\omega : \omega_1 \in F_1\}] = P_1[F_1]$  für jedes  $F_1 \in \mathcal{F}_1$  gelten, und analog  $P[\{\omega : \omega_2 \in F_2\}] = P_2[F_2]$ . Wie muss  $P$  aber aussehen, damit die beiden Faktoren unabhängig sind? Damit meinen wir natürlich, dass jedes alleine durch das erste ZE bestimmte Ereignis  $A_1$  von jedem alleine durch das zweite ZE bestimmte Ereignis unabhängig ist. In anderen Worten: die Familie  $\mathcal{A}_1$  aller Ereignisse der Form  $\{\omega : \omega_1 \in F_1\}$  mit  $F_1 \in \mathcal{F}_1$  soll unabhängig von der Familie  $\mathcal{A}_2$  aller Ereignisse der Form  $\{\omega : \omega_2 \in F_2\}$  mit  $F_2 \in \mathcal{F}_2$  sein. Also:

$$P[\{\omega : \omega_1 \in F_1\} \cap \{\omega : \omega_2 \in F_2\}] = P[\{\omega : \omega_1 \in F_1\}] P[\{\omega : \omega_2 \in F_2\}].$$

Wir haben aber gerade überlegt, dass die rechte Seite gerade  $P_1[F_1] P_2[F_2]$  sein muss. Die linke Seite betreffend bemerken wir, dass  $\{\omega : \omega_1 \in F_1\} \cap \{\omega : \omega_2 \in F_2\} = F_1 \times F_2$ . Unsere Forderung lautet daher insgesamt

$$P[F_1 \times F_2] = P_1[F_1] P_2[F_2] \quad \text{für } F_i \in \mathcal{F}_i.$$

Das ist aber genau die definierende Bedingung für das Produktmaß  $P_1 \otimes P_2$ . Somit:

Die unabhängige Kombination zweier durch  $(\Omega_1, \mathcal{F}_1, P_1)$  und  $(\Omega_2, \mathcal{F}_2, P_2)$  (4.2) modellierter ZE wird durch  $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \otimes P_2)$  modelliert.

Was für ein günstiger Zufall, dass wir uns schon über Produkte allgemeiner Maßräume unterhalten haben!

**Viele Ereignisse, Ereignisfamilien und Teilerperimente.** Richtig spannend wird es oft erst dann, wenn wir es mit mehr als zwei, womöglich gar unendlich vielen Ereignissen oder Teilerperimenten zu tun haben.

**Definition 4.5** Sind  $A_i, i \in I$  Ereignisse im  $W$ -Raum  $(\Omega, \mathcal{F}, P)$ , wobei  $I$  eine beliebig große Indexmenge ist, so heißen diese (voneinander) unabhängig, falls für je endlich viele verschiedene Indices  $i_1, \dots, i_m \in I$  die Ereignisse  $A_{i_1}, \dots, A_{i_m}$  die Produktregel

$$P[A_{i_1} \cap \dots \cap A_{i_m}] = P[A_{i_1}] \cdots P[A_{i_m}]$$

erfüllen. Sind  $\mathcal{A}_i \subseteq \mathcal{F}, i \in I$  Familien von Ereignissen im  $W$ -Raum  $(\Omega, \mathcal{F}, P)$ , so heißen diese (voneinander) unabhängig, falls für jede Wahl von Ereignissen  $A_i \in \mathcal{A}_i, i \in I$  diese unabhängig sind.

Unabhängigkeit beliebig vieler Ereignisse (bzw Familien) wird also über die Produkt- oder Proportionalitätsregel für *sämtliche endlichen Teilkollektionen* definiert. Dabei reicht es nicht, immer nur Paare von Ereignissen zu betrachten. Das geht schon in sehr einfachen Situationen schief:

**Beispiel 4.8** In Beispiel 4.1 a') sind je zwei der Ereignisse  $A, B$  und  $C$  voneinander unabhängig, zusammen sind sie es aber nicht (z.B. weil  $A \cap B \cap C = \emptyset$ ).

Unsere früheren Überlegung zur unabhängigen Kombination zweier Teilerperimente übertragen sich in analoger Weise auf jede endliche Anzahl von Faktoren. Weil Unabhängigkeit aber ohnehin immer nur endlich viele Ereignisse betrachtet, können wir auch leicht erklären, was wir unter der unabhängigen Kombination einer unendlichen Folge  $(\Omega_i, \mathcal{F}_i, P_i), i \geq 1$  von Teilerperimenten verstehen wollen. Kanonisch nehmen wir den unendlichdimensionalen Produktraum  $\Omega := \Omega_1 \times \Omega_2 \times \dots$  und definieren  $\mathcal{F} = \bigotimes_{i \geq 1} \mathcal{F}_i$  als  $\sigma(\bigcup_{i \geq 1} X_i^{-1} \mathcal{F}_i)$  mit  $X_i : \Omega \rightarrow \Omega_i, \omega \mapsto \omega_i$ , die kleinste  $\sigma$ -Algebra bezüglich derer alle Projektionen  $X_i$  messbar sind. Für beliebiges  $m \geq 1$  möchten wir dann sicherstellen, dass

$$P[\bigcap_{i=1}^m \{\omega : \omega_i \in F_i\}] = \prod_{i=1}^m P[\{\omega : \omega_i \in F_i\}] \quad \text{für } F_i \in \mathcal{F}_i,$$

also

$$P[F_1 \times \dots \times F_m \times \Omega \times \Omega \times \dots] = \prod_{i=1}^m P_i[F_i] \quad \text{für } m \geq 1 \text{ und } F_i \in \mathcal{F}_i. \quad (4.3)$$

Ein Maß  $P$  mit dieser Eigenschaft nennen wir das (*unendliche*) *Produktmaß*  $\bigotimes_{i \geq 1} P_i$  der  $P_i$ . Das ist nicht nur ein Wunschtraum, die Maßtheorie macht's möglich:

**Satz 4.1 (Unendliche Produkte von  $W$ -Räumen)** Zu jeder Folge  $(\Omega_i, \mathcal{F}_i, P_i), i \geq 1$  von  $W$ -Räumen gibt es ein eindeutig bestimmtes Produktmaß  $\bigotimes_{i \geq 1} P_i$ .

Damit Sie den Rest des Studiums auch noch interessant finden, verrate ich den Beweis hier nicht. Dieser Satz ist aber für die Stochastik genauso fundamental wie jener über die Existenz des Lebesgue-Maßes für die Analysis. Die idealisierte Vorstellung einer unendlichen Folge unabhängiger Kopien eines beliebigen ZE besitzt wirklich eine mathematische Realisierung, und wir können logisch konsistent darüber sprechen.

## 4.3 Zufallsvariable, Erwartungswert und Momente

**A) Zufallsvariable.** Eine Zufallsvariable (ZV) beschreibt (in einem gegebenen Modell  $(\Omega, \mathcal{F}, P)$ ) eine vom Zufall, dh von  $\omega \in \Omega$  abhängige Größe oder, allgemeiner, ein anderes vom Zufall abhängiges Objekt. Wir beginnen mit zufälligen

Größen bzw reellen Zahlen. Da diese von  $\omega$  abhängen, werden sie einfach durch reellwertige Funktionen  $X : \Omega \rightarrow \mathbb{R}$  (oder gerne auch  $X : \Omega \rightarrow \overline{\mathbb{R}}$ ) beschrieben.

**Beispiel 4.9** Wir würfeln zweimal, und interessieren uns für das Maximum  $X$  der beiden erzielten Augenzahlen. Das könnten wir laut Standardrezept nun so modellieren, dass wir als Grundraum  $\Omega_X$  die Menge der möglichen Werte von  $X$ , also  $\{1, \dots, 6\}$  wählen, und dort ein  $W$ -Maß  $P_X$  fixieren, wo  $P_X[m]$  gerade die Wkeit dafür ist, als Maximum die Augenzahl  $m$  zu bekommen. Dieser Ansatz ist grundsätzlich OK, aber nicht nützlich: Woher sollen wir denn diese Wkeiten kennen? Müssen wir da zuerst wieder experimentieren?

Viel sinnvoller ist es, hier den Umstand zu nützen, dass wir schon wissen, wie wir zweifaches Würfeln passend beschreiben, und daraus die gesuchten Wkeiten korrekt zu bestimmen: Wir nehmen also, wie zuvor,  $\Omega := \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}\}$  mit uniformer Verteilung  $P$  als Modell fürs Würfeln. Das ist, wenn Sie den Raum betrachten zwar ein wenig komplizierter als  $\Omega_X$  oben, dafür kennen wir aber  $P$  (das zudem recht einfach ist). Nun beschreibt die Funktion  $X(\omega) := \max(\omega_1, \omega_2)$  das Maximum der Augenzahlen, und wir fragen nach den Wkeiten der Ereignisse  $A_m := \{\omega : X(\omega) = m\} = \{X = m\} = X^{-1}\{m\} \subseteq \Omega$ ,  $1 \leq m \leq 6$ , die durch unser  $P$  ja schon festgelegt sind. Dazu brauchen wir nur zu erkennen, dass  $A_m = \{(1, m), \dots, (m, m), \dots, (m, 1)\}$  gerade  $2m - 1$  Elemente hat, also  $P[X = m] = P[A_m] = |A_m| / |\Omega| = (2m - 1)/36$ . Damit wissen wir im nachhinein, dass oben  $P_X[m] = (2m - 1)/36$  die korrekte Wahl wäre.

So einfach das gerade war, so wichtig ist doch das Prinzip: Manchmal ist ein größerer, aber übersichtlicherer  $W$ -Raum die bessere Wahl, wenn wir dadurch hinter die Kulissen blicken und die gesuchte Größe als Funktion eines leichter zu verstehenden Experimentes interpretieren können.

**Reelle ZV.** Bei allgemeinem  $X : \Omega \rightarrow \mathbb{R}$  interessieren uns jetzt für den zufälligen Wert  $X(\omega)$  den wir bei Realisierung des ZE erhalten, und wollen insbesondere über die Wkeit sprechen, dass dieser Wert in einer vorgegebenen Menge  $B \subseteq \mathbb{R}$  liegt. Diese Forderung definiert ein Ereignis, nämlich<sup>1</sup>  $\{\omega \in \Omega : X(\omega) \in B\} = \{X \in B\} = X^{-1}B \subseteq \Omega$ , wir wollen also  $P[X \in B] = P[X^{-1}B]$  verstehen. Es ehrt Sie, wenn Sie jetzt schon unruhig werden. Warum sollte  $P[X^{-1}B]$  hier überhaupt definiert sein? Nun, damit das funktioniert, müssen wir tatsächlich noch eine Bedingung ergänzen. Sie wissen ja längst, dass es beliebig unappetitliche Mengen  $B \subseteq \mathbb{R}$  gibt, und wir lassen bei manchen lieber die Finger von der Frage ob  $X(\omega) \in B$  ist. Aber wenn wir die reiche Familie der Borel-Mengen behandeln können, dann können wir schon sehr zufrieden sein. Dies führt uns aber zu der Forderung, dass für jedes  $B \in \mathcal{B}_{\mathbb{R}}$  stets  $X^{-1}B \in \mathcal{F}$  sein sollte.

**Definition 4.6** Eine reelle Zufallsvariable (ZV) auf dem  $W$ -Raum  $(\Omega, \mathcal{F}, P)$  ist eine  $(\mathcal{F}, \mathcal{B}_{\mathbb{R}})$ -messbare Abbildung  $X : \Omega \rightarrow \mathbb{R}$ . Wir bezeichnen ZVen mit  $X, Y$  usw, in konkreten Situationen gerne unter Verwendung suggestiver Wahl der Buchstaben, wie  $S$  für Summen etc.

**Definition 4.7** Die Elemente von  $\{X \in B : B \in \mathcal{B}_{\mathbb{R}}\} = X^{-1}\mathcal{B}_{\mathbb{R}} =: \sigma(X)$  nennen wir die durch  $X$  bestimmten Ereignisse<sup>2</sup>.

**Bemerkung 4.2 a)** Per Definition ist auch jede konstante Funktion  $X$  auf  $\Omega$  eine ZV. (Solche ZV sind zwar langweilig, aber Sie wissen, dass es auch sonst nützlich ist, konstante Funktionen eben doch als Funktionen aufzufassen.)

**b)** Ist  $A \in \mathcal{F}$  ein Ereignis, so ist  $X := 1_A$  eine ZV, die das Eintreten von  $A$  erkennt.

**c)** Mit reellen ZV kann man, wie generell mit reellen Funktionen, (punktweise) rechnen,  $S := X_1 + X_2$  bedeutet  $S(\omega) := X_1(\omega) + X_2(\omega)$  etc.

**d)** Weiter oben haben wir vorgeschlagen, eine Familie von Ereignissen als einen

<sup>1</sup>Erinnern Sie sich bitte an die Schreibweise (2.8) für Urbilder, die wir auch beim Integrieren häufig verwendet haben.

<sup>2</sup>Das ist keine Standard-Terminologie. Sie scheint mir aber hilfreich zu sein.

bestimmten Kenntnisstand zu interpretieren. In diesem Sinne repräsentiert die  $\sigma$ -Algebra  $\sigma(X)$  jenen Kenntnisstand, den wir aus Beobachtung von  $X$  gewinnen können: Kennen wir  $X(\omega)$ , so wissen wir im Prinzip für jedes  $B \in \mathcal{B}_{\mathbb{R}}$  ob das Ereignis  $\{X \in B\}$  eintritt.

**Beispiel 4.10 a)** Beschreibt  $\Omega := \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}\}$  mit uniformer Verteilung  $P$  wie oben zwei Würfel, dann beschreiben die ZV  $X_j(\omega) := \omega_j$  das Ergebnis des  $j$ -ten Wurfes und,  $S := X_1 + X_2$  die Summe der Augenzahlen. Entsprechend ist  $zB \{S = 5\} = \{\omega : \omega_1 + \omega_2 = 5\}$  das Ereignis, dass die Summe der Augenzahlen genau 5 beträgt.

**b)** Beschreibt  $\Omega := [0, 1]^2 = \{\omega = (\omega_1, \omega_2) : \omega_i \in [0, 1]\}$  mit  $\mathcal{F} := \mathcal{B}_{[0,1]^2}$  und  $P := \lambda^2$  die zufällige Wahl eines Punktes im Einheitsquadrat  $[0, 1]^2$  mit der uniformen Verteilung, dann beschreibt  $X_j(\omega) := \omega_j$  die  $j$ -te Koordinate des Punktes, und  $D := |X_1 - X_2|/\sqrt{2}$  den Abstand des Punktes von der Diagonale. Weiters ist  $R := \sqrt{X_1^2 + X_2^2}$  der Abstand des Punktes von  $(0, 0)$ . Es ist dann  $\{R < \frac{1}{2}\}$  das Ereignis, dass der zufällig gewählte Punkt innerhalb eines Kreises vom Radius  $\frac{1}{2}$  um den Ursprung liegt.

Aufgrund der Messbarkeit einer reellen ZV können wir also für jede Borel-Menge  $B \in \mathcal{B}_{\mathbb{R}}$  über die Wkeit  $P[X \in B] = P[X^{-1}B]$  des durch  $X$  bestimmten Ereignisses  $\{X \in B\}$  sprechen. Gemeinsam repräsentieren diese Werte dann die gesamte Information, die wir in diesem Modell über  $X$  bekommen, wenn wir uns exklusiv auf diese ZV konzentrieren. Genau das war die Idee in Beispiel 4.9. Diese Wkeiten, den Wert von  $X$  in einem bestimmten Bereich  $B$  zu finden, ergeben zusammengefasst die *Verteilung von  $X$* . Das ist für Sie kein neues Objekt - Sie haben ja längst von Bildmaßen gehört.

**Definition 4.8** Die Verteilung einer reellen ZV  $X$  auf dem  $W$ -Raum  $(\Omega, \mathcal{F}, P)$  ist das Bildmaß  $P_X := P \circ X^{-1}$ .

Also ist  $P_X$  das  $W$ -Maß auf dem Bildraum  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  mit  $P_X[B] = P[X \in B]$  für  $B \in \mathcal{B}_{\mathbb{R}}$ . Der so erhaltene  $W$ -Raum  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P_X)$  beschreibt die zufällige Wahl des Wertes von  $X$ , und vergisst alles, was sonst noch so in  $(\Omega, \mathcal{F}, P)$  dargestellt ist. Als normiertes Borel-Maß auf  $\mathbb{R}$  kann  $P_X$  immer durch eine Verteilungsfunktion beschrieben werden (vgl. Abschnitt 2.3). Dabei wählt man diese immer so, dass ihre Werte in  $[0, 1]$  liegen, also ist die *Verteilungsfunktion (VF)* von  $X$  gegeben als

$$F_X : \mathbb{R} \rightarrow [0, 1], \quad F_X(t) := P_X((-\infty, t]) = P[X \leq t], \quad t \in \mathbb{R}.$$

Wir nennen  $X$  *diskret*, falls diese ZV nur abzählbar viele Werte annehmen kann, also ihr Bild abzählbar ist. In diesem Fall ist  $P_X$  ein diskretes Maß,  $P_X = \sum_x p_x \delta_x$  mit *Einzelwahrscheinlichkeiten*  $p_x = P_X[x] = P[X = x]$ . Als Gegenpol dazu können wir die Situation auffassen, in der  $P_X$  eine Dichte  $f_X$  bezüglich  $\lambda$  besitzt,  $P_X = f_X \odot \lambda$  also  $P_X[B] = \int_B f_X d\lambda$ . Wir sagen dann kurz,  $f_X$  sei die *Dichte von  $X$* . Häufig nennt man solches  $X$  eine *stetige ZV*<sup>3</sup>.

**Beispiel 4.11** Fortsetzung des vorigen Beispiels:

**a)** Die Verteilung des Maximums  $X$  ist gegeben durch die Einzelwahrscheinlichkeiten  $P_X[m] = P[X = m] = (2m - 1)/36$  für  $m \in \{1, \dots, 6\}$ .

**b)** Wir sehen, dass  $F_D(t) = P[D \leq t] = \lambda^2(\{(\omega_1, \omega_2) \in [0, 1]^2 : |\omega_1 - \omega_2| \leq t\sqrt{2}\})$ ,  $t \geq 0$  (Fläche innerhalb  $[0, 1]^2$  eines Streifens der Breite  $2t$  um die Diagonale), und  $F_D(t) = 0$  für  $t < 0$  weil  $D \geq 0$ . Man sieht leicht (PS Blatt 6)

$$F_D(t) = 2\sqrt{2}t - 2t^2 - 1, \quad t \in [0, 1/\sqrt{2}],$$

und wir finden entsprechend die Dichte

$$f_D(t) = F'_D(t) = 2\sqrt{2} - 4t, \quad t \in [0, 1/\sqrt{2}].$$

<sup>3</sup>Obwohl im Sinne der Maßtheorie die Bezeichnung *absolutstetig* korrekt wäre, siehe §3.5.D.

Die Verteilung von  $R$  ist gegeben durch die VF mit  $F_R(t) = P[R \leq t] = \lambda^2(\{(\omega_1, \omega_2) \in [0, 1]^2 : \sqrt{\omega_1^2 + \omega_2^2} \leq t\})$ ,  $t \geq 0$  (Fläche des Schnittes zwischen Einheitsquadrat und Kreis mit Radius  $t$ ). Wir finden (PS Blatt 6)

$$F_R(t) = \begin{cases} \frac{\pi}{4} t^2 & \text{für } t \in [0, 1], \\ t^2 \left( \arcsin\left(\frac{1}{t}\right) - \frac{\pi}{4} \right) + \sqrt{t^2 - 1} & \text{für } t > 1. \end{cases}$$

Differentiation liefert die Dichte

$$f_R(t) = F'_R(t) = \begin{cases} \frac{\pi}{2} t & \text{für } t \in [0, 1], \\ 2t \left( \arcsin\left(\frac{1}{t}\right) - \frac{1}{4}\pi \right) & \text{für } t > 1. \end{cases}$$

**Zufällige Tupel.** Bevor wir uns noch genauer mit reellen ZV und ihren Verteilungen befassen, besprechen wir noch kurz eine sehr nützliche Erweiterung des Begriffes. Man kann nämlich nicht nur zufällige numerische Größen, sondern überhaupt alle durch das zugrunde liegende ZE bestimmten zufälligen Objekte in analoger Weise beschreiben. Als ersten Schritt dazu bemerken wir, dass für reelle ZV  $X_1, \dots, X_d$  das Tupel  $(X_1, \dots, X_d) =: X$  eine messbare Abbildung  $X: \Omega \rightarrow \mathbb{R}^d$  definiert. Es erweist sich oft als sehr praktisch, die  $X_j$  auf diese Weise zu einem einzigen *zufälligen Vektor* (oder *zufälligen Punkt*) in  $\mathbb{R}^d$  zusammenzufassen. Die Verteilung dieses zufälligen Vektors ist dann einfach wieder das Bildmaß  $P_X = P_{(X_1, \dots, X_d)} := P \circ X^{-1}$ . In diesem Fall handelt es sich bei  $P_X$  um ein W-Maß auf  $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ , mit dem man insbesondere Cartesische Produktmengen  $B_1 \times \dots \times B_d$ ,  $B_j \in \mathcal{B}_{\mathbb{R}}$ , messen kann. Diese Verteilung dokumentiert also insbesondere die Wkeiten für das *gemeinsame Auftreten von Ereignissen*  $\{X_1 \in B_1\}, \dots, \{X_d \in B_d\}$  welche durch die einzelnen  $X_j$  bestimmt sind, da ja

$$P \left[ \bigcap_{j=1}^d \{X_j \in B_j\} \right] = P[(X_1, \dots, X_d) \in B_1 \times \dots \times B_d] = P_{(X_1, \dots, X_d)}[B_1 \times \dots \times B_d].$$

Man spricht daher von der *gemeinsamen Verteilung*  $P_{(X_1, \dots, X_d)}$  der ZV  $X_1, \dots, X_d$ . Diese ist für das Verständnis einer Situation in welcher mehrere ZV  $X_1, \dots, X_d$  zusammenwirken von zentraler Bedeutung, weil sie etwaige *Zusammenhänge zwischen den  $X_j$*  beschreibt, die man anhand der einzelnen Verteilungen  $P_{X_1}, \dots, P_{X_d}$  nicht erkennen kann.

**Beispiel 4.12 a)** Wird zweimal fair gewürfelt,  $\Omega := \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}\}$  mit uniformer Verteilung  $P$  und  $X_j(\omega) := \omega_j$ , dann sind  $P_{X_1}$  und  $P_{X_2}$ , die Verteilungen der einzelnen Würfe, natürlich jeweils durch die uniforme Verteilung auf  $\{1, \dots, 6\}$  gegeben. Da  $X := (X_1, X_2)$  einfach die Identität auf  $\Omega$  ist, haben wir also  $P_{(X_1, X_2)} = P$  mit  $P[(k, l)] = \frac{1}{36}$  für  $k, l \in \{1, \dots, 6\}$ .

**b)** Falls ich nun zu faul bin, zweimal zu würfeln, und stattdessen nur einmal würfle, und Ihnen dann zweimal dasselbe Ergebnis sage, befinden wir uns in einer anderen Situation, die wir zB durch  $\Omega := \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}\}$  mit einem W-Maß  $P$  beschreiben können, für welches  $P[(k, k)] = \frac{1}{6}$  und  $P[(k, l)] = 0$  falls  $k \neq l$  gilt. Schreiben wir wieder  $X_j(\omega) := \omega_j$ , und  $X := (X_1, X_2)$ , dann unterscheidet sich die gemeinsame Verteilung  $P_{(X_1, X_2)} = P$  ganz wesentlich von jener aus a). Sie können den Unterschied zwischen den beiden Varianten aber nicht erkennen, wenn Sie jeweils nur die Verteilungen  $P_{X_1}$  und  $P_{X_2}$  der einzelnen ZV betrachten, denn auch im zweiten Fall stimmen diese jeweils mit der uniformen Verteilung auf  $\{1, \dots, 6\}$  überein. (Das ist anschaulich klar, weil ja mit  $X_1$  auch  $X_2 = X_1$  das Ergebnis eines fairen Würfels ist. Bitte stellen Sie sicher, dass Sie das auch formal nachvollziehen können.)

**Allgemeine ZV\*.** Ganz allgemein (aber um nichts schwieriger) können wir andere Typen zufälliger Objekte analog beschreiben:

**Definition 4.9** Ist  $(\Omega', \mathcal{F}')$  ein messbarer Raum, dann ist eine Zufallsvariable (ZV) auf dem W-Raum  $(\Omega, \mathcal{F}, P)$  mit Werten in  $(\Omega', \mathcal{F}')$  (kurz: in  $\Omega'$ ) eine  $\mathcal{F}$ - $\mathcal{F}'$ -messbare Abbildung  $X: \Omega \rightarrow \Omega'$ . Man nennt  $X$  auch ein zufälliges Element von

$\Omega'$ , und die Elemente von  $\{\{X \in F'\} : F' \in \mathcal{F}'\} = X^{-1}\mathcal{F}' =: \sigma(X)$  die durch  $X$  bestimmten Ereignisse. Das Bildmaß  $P_X := P \circ X^{-1}$  auf  $(\Omega', \mathcal{F}')$  ist die Verteilung von  $X$ .

**Beispiel 4.13** Wir modellieren eine unendliche Folge von fairen 0-1-Münzwürfen indem wir  $\Omega := \{\omega = (\omega_i)_{i \geq 1} : \omega_i \in \{0, 1\}\} = \{0, 1\}^{\mathbb{N}^+}$  mit der unendlichen Produkt- $\sigma$ -Algebra  $\mathcal{F}$  und dem unendlichen Produktmaß  $P := \bigotimes_{i \geq 1} (\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)$  ausstatten (vgl Satz 4.1). Dabei repräsentiert  $X_j(\omega) := \omega_j$  das Ergebnis des  $j$ ten Wurfes. Wir können jeder 0-1-Folge  $\omega \in \Omega$  eine stetige Funktion  $f_\omega : [0, 1] \rightarrow \mathbb{R}$  zuordnen mittels

$$f_\omega(x) := \sum_{j \geq 1} \omega_j \frac{x^j}{j!}, \quad x \in [0, 1].$$

Bezeichnen wir diese Zuordnung mit  $X : \Omega \rightarrow \mathcal{C}[0, 1] =: \Omega'$ ,  $X(\omega) := f_\omega$ , so bekommen wir eine ZV mit Werten in einem Funktionenraum bzw eine zufällige stetige Funktion. Weil auf  $\mathcal{C}[0, 1]$  durch die Supremumsnorm  $\|\cdot\|$  eine schöne Metrik gegeben ist, nehmen wir auf diesem Raum am liebsten die zugehörige Borel- $\sigma$ -Algebra  $\mathcal{F}' := \mathcal{B}_{\mathcal{C}[0, 1]}$ . Diese enthält insbesondere die gleichmäßige  $\varepsilon$ -Umgebung um jede feste Funktion  $g$ . Messbarkeit von  $X$  folgt dann leicht aus der Stetigkeit dieser Abbildung bzgl einer geeigneten Metrik auf  $\Omega$ , und wir können damit zB legal von der Wkeit des Ereignisses sprechen, dass die zufällige Funktion  $X$  überall  $\varepsilon$ -nahe an  $g(x) = \sin(x)$ ,  $x \in [0, 1]$  liegt,

$$P[\|X - \sin\| < \varepsilon] = P_X[\{f \in \mathcal{C}[0, 1] : \|f - \sin\| < \varepsilon\}].$$

Die stochastische Regel, nach der  $X$  gewählt wird, ist also durch das W-Maß  $P_X$  auf dem Funktionenraum gegeben. (Wie bzw ob man solche Wkeiten konkret ausrechnen kann ist eine andere Geschichte...)

**Notation.** Weil es in vielen Situationen hauptsächlich auf die Verteilungen bestimmter ZV ankommt, erwähnen wir noch eine praktische Schreibweise. Sind  $X$  und  $Y$  ZV mit demselben Wertebereich  $(\Omega', \mathcal{F}')$ , dann bedeutet  $X \stackrel{d}{=} Y$  (mit  $d$  für "distribution"=Verteilung) einfach  $P_X = P_Y$ . Analog schreibt man für W-Maße  $Q$  auf  $(\Omega', \mathcal{F}')$  gelegentlich  $X \stackrel{d}{=} Q$  anstelle von  $P_X = Q$ .

**B) Erwartungswert und Momente.** Der letzte noch fehlende Grundbegriff ist jener des Erwartungswertes einer reellen ZV.

**Definition 4.10** Ist eine reelle ZV  $X : \Omega \rightarrow \mathbb{R}$  auf  $(\Omega, \mathcal{F}, P)$  quasi-integrierbar, so nennt man

$$\mathbb{E}[X] := \int_{\Omega} X dP = \int_{\Omega} X(\omega) dP(\omega) \in \overline{\mathbb{R}}$$

den Erwartungswert (EW) von  $X$ . Ist  $X$  integrierbar, also  $X \in \mathcal{L}^1(P)$  bzw  $\mathbb{E}[|X|] < \infty$ , so sagt man, der Erwartungswert von  $X$  existiert.

Der EW ist also einfach das Integral der Funktion  $X$  über den zugrunde liegenden Maßraum. Da letzterer per Definition normiert ist, ist  $\mathbb{E}[X]$  also immer der durch  $P$  gewichtete Mittelwert von  $X$ . Natürlich hat  $\mathbb{E}[\cdot]$  damit automatisch alle Eigenschaften des allgemeinen Integrals, ist also monoton und linear,

$$X \leq Y \text{ f.s.} \implies \mathbb{E}[X] \leq \mathbb{E}[Y] \quad \text{und} \quad \mathbb{E}[X + aY] = \mathbb{E}[X] + a\mathbb{E}[Y],$$

und erlaubt sogar die Anwendung der Konvergenzsätze aus Abschnitt 3.4. Stellt man sich die Verteilung von  $X$  als Masseverteilung auf  $\mathbb{R}$  vor, so gibt  $\mathbb{E}[X]$  gerade den Schwerpunkt derselben an. Offenbar ist

$$\mathbb{E}[1_A] = P[A] \quad \text{für } A \in \mathcal{F},$$

der EW verallgemeinert in diesem Sinne das Konzept der Wkeit. Er erlaubt auch eine intuitive Interpretation analog zu jener, mit der wir unsere Diskussion von

Wkeiten begonnen haben. Führt man das ZE wiederholt unabhängig durch, kann man jeweils unterschiedliche Werte  $X_1, \dots, X_n$  für  $X$  bekommen. Wie bei den Häufigkeiten des Auftretens von  $A$  scheinen sich aber die arithmetischen Mittel  $(X_1 + \dots + X_n)/n$  dieser Beobachtungen einem Grenzwert anzunähern, und wir stellen uns unter  $\mathbb{E}[X]$  diesen Grenzwert vor. (Gerechtfertigt wird dies erst im nächsten Abschnitt.)

**Beispiel 4.14** Ist  $(\Omega, \mathcal{F}, P)$  ein abzählbarer  $W$ -Raum, dann können wir den EW von  $X : \Omega \rightarrow \mathbb{R}$  ganz einfach explizit anschreiben als

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) dP(\omega) = \sum_{\omega \in \Omega} X(\omega) P[\omega].$$

Sehr oft ist es günstig, den EW mithilfe der Verteilung von  $X$  zu berechnen.

**Proposition 4.3** Ist  $X$  eine quasi-integrierbare reelle ZV auf  $(\Omega, \mathcal{F}, P)$ , dann gilt

$$\mathbb{E}[X] = \int_{\mathbb{R}} x dP_X(x), \quad (4.4)$$

und allgemeiner, wenn  $g$  messbar und  $g(X) = g \circ X$  quasi-integrierbar,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) dP_X(x). \quad (4.5)$$

**Beweis.** Wir brauchen nur (4.5) zu überprüfen. (Aussage (4.4) ist dann ja der spezielle Fall mit  $g(x) = x$ .) Das ist eine direkte Anwendung der allgemeinen Transformationsformel,

$$\mathbb{E}[g(X)] = \int_{\Omega} g \circ X dP = \int_{\mathbb{R}} g d(P \circ X^{-1}) = \int_{\mathbb{R}} g dP_X,$$

siehe Satz 3.7. ■

**Folgerung 4.1 a)** Bei ZV  $X$  mit abzählbarem Wertebereich  $J$  bekommen wir also

$$\mathbb{E}[X] = \sum_{j \in J} j P[X = j] \quad \text{und allgemein} \quad \mathbb{E}[g(X)] = \sum_{j \in J} g(j) P[X = j]. \quad (4.6)$$

**b)** Besitzt  $X$  (bzw  $P_X$ ) die Dichte  $f_X$  bezüglich des Lebesgue-Maßes, dann gilt

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx \quad \text{und allgemein} \quad \mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx. \quad (4.7)$$

(Laut Proposition 3.8 folgt aus  $P_X = f_X \odot \lambda$  gleich  $\int_{\mathbb{R}} g dP_X = \int_{\mathbb{R}} g(x) f_X(x) dx$ .)

**Beispiel 4.15 a)** Ist  $X$  das Maximum der Augenzahlen beim zweifachen Würfeln, dann finden wir gemäß Bsp 4.11, dass  $\mathbb{E}[X] = \sum_{m=1}^6 m(2m-1)/36 = \frac{161}{36} \approx 4.47$ .

**b<sub>i</sub>)** Für  $X$  mit uniformer Verteilung auf  $[0, 1]$ ,  $X \stackrel{d}{=} \mathcal{U}([0, 1])$  mit Dichte  $1_{[0, 1]}$ , finden wir  $\mathbb{E}[X] = \int_0^1 x dx = \frac{1}{2}$  und  $\mathbb{E}[X^2] = \int_0^1 x^2 dx = \frac{1}{3}$ .

**b<sub>ii</sub>)** Besitzt  $X$  die Standard-Normalverteilung  $\mathcal{N}(0, 1)$  mit Dichte  $\phi_{0,1}(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ ,  $x \in \mathbb{R}$ , so finden wir  $\mathbb{E}[X] = \int_0^1 x \phi_{0,1}(x) dx = 0$  und  $\mathbb{E}[X^2] = \int_0^1 x^2 \phi_{0,1}(x) dx = 1$  (siehe PS).

Dass wir hier gerne  $\mathbb{E}[X^2]$  ausrechnen ist ausnahmsweise kein Zufall. Die Integrale der Potenzen  $X^p$  mit  $p \in [1, \infty)$  liefern uns besonders nützliche Information über die Verteilung von  $X$ .

**Definition 4.11** Die reelle ZV  $X$  besitzt endliches  $p$ tes Moment (oder ihr  $p$ tes Moment existiert), falls  $\mathbb{E}[|X|^p] < \infty$  (d.h.  $X \in \mathcal{L}^p(P)$ ). In diesem Fall ist  $\mathbb{E}[X^p]$  das  $p$ te Moment von  $X$ .

Kenntnis von  $\mathbb{E}[|X|^p]$  bietet eine Möglichkeit, die Werten besonders großer Werte von  $X$  zu kontrollieren. Ganz einfach geht dies zB mithilfe simpler Ungleichungen.

**Proposition 4.4 (Markov-Ungleichung und Chebyshev-Ungleichung)** Ist  $X$  eine reelle ZV mit  $\mathbb{E}[|X|] < \infty$ , dann gilt

$$P[|X| \geq a] \leq \frac{\mathbb{E}[|X|]}{a} \quad \text{für } a > 0. \quad (4.8)$$

Falls sogar  $\mathbb{E}[X^2] < \infty$ , dann folgt

$$P[|X| \geq a] \leq \frac{\mathbb{E}[X^2]}{a^2} \quad \text{für } a > 0. \quad (4.9)$$

**Beweis.** Offenbar ist  $a1_{\{|X| \geq a\}} \leq |X|$ , und mittels Integration folgt  $aP[|X| \geq a] \leq \mathbb{E}[|X|]$ . Die zweite Ungleichung ergibt sich daraus, weil  $\{|X| \geq a\} = \{X^2 \geq a^2\}$ . ■

Falls  $\mathbb{E}[X]$  bekannt ist, können wir  $X$  auch von diesem Mittelwert aus betrachten, und als  $X = \mathbb{E}[X] + (X - \mathbb{E}[X])$  anschreiben, wobei wir die Konstante  $\mathbb{E}[X]$  als vorhersagbaren oder systematischen Anteil, und die Abweichung  $X - \mathbb{E}[X]$  vom EW als unsystematischen bzw zufälligen Anteil von  $X$  interpretieren. Wie unsystematisch oder zufällig  $X$  dann ist, versuchen wir mit der typischen Größe dieser Abweichung zu erfassen, zB mittels  $\mathbb{E}[|X - \mathbb{E}[X]|]$ . Als besonders schön und nützlich erweist sich allerdings die *mittlere quadratische Abweichung vom EW* (welche große Werte stärker betont), sie bekommt deshalb einen eigenen Namen.

**Definition 4.12** Die Varianz einer reellen ZV  $X$  ist  $\text{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] \in [0, \infty]$ . Sie wird oft auch mit  $\sigma^2$  bezeichnet. Die Standardabweichung von  $X$  ist  $\sqrt{\text{Var}[X]} = \sigma \geq 0$ .

Aus der Definition folgen unmittelbar die Rechenregeln

$$\text{Var}[X + c] = \text{Var}[X] \quad \text{und} \quad \text{Var}[cX] = c^2 \text{Var}[X] \quad \text{für } c \in \mathbb{R}, \quad (4.10)$$

und das Standard-Rezept um die Varianz auszurechnen (verwende  $\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$ ),

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \end{aligned} \quad (4.11)$$

**Beispiel 4.16 a)** Ist  $X$  das Maximum der Augenzahlen beim zweifachen Würfeln, dann ist  $\mathbb{E}[X^2] = \sum_{m=1}^6 m^2(2m-1)/36 = \frac{791}{36}$  und somit  $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{791}{36} - (\frac{161}{36})^2 = \frac{2555}{1296} \approx 1.97$ .

**b<sub>i</sub>)** Für  $X \stackrel{d}{=} \mathcal{U}([0, 1])$  ist (Bsp oben)  $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}$ . Die ZV  $Y$  besitzt die uniforme Verteilung auf  $[a, b]$ ,  $Y \stackrel{d}{=} \mathcal{U}([a, b])$ , wenn sie Dichte  $(b-a)^{-1}1_{[a,b]}$  hat. So eine ZV bekommen wir zB wenn  $Y := (b-a)X + a$  mit  $X \stackrel{d}{=} \mathcal{U}([0, 1])$ . Dann ist natürlich  $\mathbb{E}[Y] = (b-a)\mathbb{E}[X] + a = (a+b)/2$  und  $\text{Var}[Y] = (b-a)^2 \text{Var}[X] = (b-a)^2/12$ .

**b<sub>ii</sub>)** Besitzt  $X$  die Standard-Normalverteilung  $\mathcal{N}(0, 1)$ , dann sind  $\mathbb{E}[X] = 0$  und  $\text{Var}[X] = 1$  (Bsp oben). Für  $m \in \mathbb{R}$  und  $\sigma > 0$  hat dann die verschobene und skalierte ZV  $Y := \sigma X + m$  die Normalverteilung  $\mathcal{N}(m, \sigma^2)$  mit Dichte

$$\phi_{m,\sigma}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{\sigma^2}}, \quad x \in \mathbb{R}$$

(Substitution). Die Rechenregeln für EW und Varianz liefern  $\mathbb{E}[Y] = m$  und  $\text{Var}[Y] = \sigma^2$ .

Wichtig zu beachten ist, dass  $\text{Var}$  nicht additiv ist, dh im allgemeinen ist

$$\text{Var}[X + Y] \neq \text{Var}[X] + \text{Var}[Y],$$



(Weil zB  $\text{Var}[X+X] = \text{Var}[2X] = 4\text{Var}[X]$  nach obiger Regel.) Kehren wir aber zur Darstellung  $X = \mathbb{E}[X] + (X - \mathbb{E}[X])$  zurück. Indem wir in der letzten Proposition anstelle von  $X$  nun  $X - \mathbb{E}[X]$  einsetzen, sehen wir nun sofort, dass

$$P[|X - \mathbb{E}[X]| \geq a] \leq \frac{\mathbb{E}[|X - a|]}{a} \quad \text{für } a > 0,$$

bzw

$$P[|X - \mathbb{E}[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2} \quad \text{für } a > 0. \quad (4.12)$$

Die zweite Abschätzung (bitte einprägen!) illustriert, dass die Existenz höherer Momente viel bessere Kontrolle über große Werte der ZV ermöglicht. Wir sehen demnächst noch zwei sehr prominente Anwendungen. Besonders griffig ist die Formulierung (mit  $a = k\sigma$ ,  $\sigma^2 = \text{Var}[X]$ )

$$P[|X - \mathbb{E}[X]| \geq k\sigma] \leq \frac{1}{k^2} \quad \text{für } k \geq 1, \quad (4.13)$$

kurz: "Die Wkeit, dass  $X$  mehr als  $k$  Standardabweichungen vom EW entfernt ist, ist höchstens  $1/k^2$ ." Beachten Sie bitte, dass dies unabhängig von der konkreten Verteilung immer stimmt (sobald halt  $\mathbb{E}[X^2] < \infty$ )!

**C) Unabhängige Zufallsvariablen.** ZV sind voneinander unabhängig, wenn Ereignisse, die jeweils durch eine der ZV bestimmt sind, voneinander unabhängig sind. Wir beginnen mit dem Fall zweier reeller ZV.

**Definition 4.13** Die reellen Zufallsvariablen  $X$  und  $Y$  sind (voneinander) unabhängig, wenn die Ereignisse  $A := \{X \in F'\}$  und  $B := \{Y \in F''\}$  für jede Wahl von  $F', F'' \in \mathcal{B}_{\mathbb{R}}$  voneinander unabhängig sind. Äquivalent: die  $\sigma$ -Algebren  $\sigma(X)$  und  $\sigma(Y)$  sind voneinander unabhängig.

Intuitiv unmittelbar einleuchtend ist, dass Funktionen unabhängiger ZV wieder unabhängig sind:

**Proposition 4.5 (Funktionen unabhängiger ZV)** Die Zufallsvariablen  $X$  und  $Y$  seien voneinander unabhängig, und  $g_X, g_Y : \mathbb{R} \rightarrow \mathbb{R}$  Borel-messbare Abbildungen. Dann sind auch die ZV  $g_X(X) = g_X \circ X$  und  $g_Y(Y) = g_Y \circ Y$  voneinander unabhängig.

**Beweis.** Zu zeigen ist die Unabhängigkeit der  $\sigma$ -Algebren  $\sigma(g_X(X))$  und  $\sigma(g_Y(Y))$ . Das geht so einfach wie es sein sollte: Da die  $g_X, g_Y$  messbar sind, ist jeweils  $\sigma(g_X(X)) = (g_X \circ X)^{-1}\mathcal{B}_{\mathbb{R}} = X^{-1}(g_X^{-1}\mathcal{B}_{\mathbb{R}}) \subseteq X^{-1}(\mathcal{B}_{\mathbb{R}}) = \sigma(X)$  und genauso  $\sigma(g_Y(Y)) \subseteq \sigma(Y)$ . Nach Voraussetzung sind aber sogar diese größeren  $\sigma$ -Algebren voneinander unabhängig, also erst recht  $\sigma(g_X(X))$  und  $\sigma(g_Y(Y))$ . ■

Wie schon früher angesprochen, zeigen sich Zusammenhänge zwischen  $X$  und  $Y$  in der gemeinsamen Verteilung  $P_{(X,Y)}$ . Dies wird insbesondere deutlich in

**Proposition 4.6 (Unabhängigkeit und gemeinsame Verteilung)** Die ZV  $X$  und  $Y$  sind genau dann voneinander unabhängig, wenn

$$P_{(X,Y)} = P_X \otimes P_Y.$$

**Beweis.** Unabhängigkeit von  $X$  und  $Y$  bedeutet per Definition, dass für alle  $F', F'' \in \mathcal{B}_{\mathbb{R}}$  gilt

$$P[\{X \in F'\} \cap \{Y \in F''\}] = P[X \in F'] P[Y \in F''].$$

Die linke Seite ist aber  $P[\{X \in F'\} \cap \{Y \in F''\}] = P[(X,Y) \in F' \times F''] = P_{(X,Y)}[F' \times F'']$ , und die rechte Seite  $P[X \in F'] P[Y \in F''] = P_X[F'] P_Y[F'']$ . Unabhängigkeit ist also dasselbe wie

$$P_{(X,Y)}[F' \times F''] = P_X[F'] P_Y[F''] \quad \text{für } F', F'' \in \mathcal{B}_{\mathbb{R}},$$

dh eben  $P_{(X,Y)} = P_X \otimes P_Y$ . ■

Wie schon angedeutet, ist es oft sehr nützlich, ZV mit beliebigen messbaren Wertebereichen zuzulassen. Für  $X : \Omega \rightarrow \Omega'$  (mit  $\sigma$ -Algebra  $\mathcal{F}'$ ) verstehen wir unter Ereignissen, die durch  $X$  bestimmt sind eben solche der Form  $A := \{X \in F'\}$  mit  $F' \in \mathcal{F}'$ . Zusammengekommen ergeben diese gerade die  $\sigma$ -Algebra  $\sigma(X) = X^{-1}\mathcal{F}' = \{X^{-1}F' : F' \in \mathcal{F}'\}$ . (Mit diesem Typ von  $\sigma$ -Algebren sind Sie ja schon vertraut.)

**Definition 4.14** Die ZV  $X$  und  $Y$  (mit beliebigen Wertebereichen  $(\Omega', \mathcal{F}')$  und  $(\Omega'', \mathcal{F}'')$ ) sind (voneinander) unabhängig, wenn die Ereignisse  $A := \{X \in F'\}$  und  $B := \{Y \in F''\}$  für jede Wahl von  $F' \in \mathcal{F}'$  und  $F'' \in \mathcal{F}''$  voneinander unabhängig sind. Äquivalent: die  $\sigma$ -Algebren  $\sigma(X)$  und  $\sigma(Y)$  sind voneinander unabhängig.

In diesem Fall funktionieren die letzten beiden Beweise weiterhin (man muss nur die Bezeichnungen anpassen). Also bleiben die beiden Propositionen gültig (weshalb wir dort auch nicht explizit von reellen ZV sprechen). Eine andere Erweiterung ist auch noch wichtig: In interessanten Situationen haben wir es oft mit (unendlich) vielen ZV zu tun. Ganz analog zur Begriffsbildung bei Ereignissen vereinbart man allgemeiner für beliebig große Kollektionen von ZV:

**Definition 4.15** Die Zufallsvariablen  $X_i$ ,  $i \in I$  (mit beliebigen Wertebereichen  $(\Omega_i, \mathcal{F}_i)$ ) sind (voneinander) unabhängig, wenn die  $\sigma(X_i)$ ,  $i \in I$  voneinander unabhängig sind.

**Beispiel 4.17** Man sieht leicht: Sind die Ereignisse  $A_i$ ,  $i \in I$  unabhängig, dann auch die Zufallsvariablen  $X_i := 1_{A_i}$ ,  $i \in I$ .

Die beiden letzten Propositionen bleiben auch korrekt (mit ähnlichen Beweisen), wenn man nicht zwei, sondern beliebig viele unabhängige ZV betrachtet. Allerlei weitere Hilfsmittel, um Unabhängigkeit von ZV zu überprüfen und mit unabhängigen ZV gut umzugehen, lernen Sie in weiterführenden LVen noch kennen. Oft steht die Unabhängigkeit bestimmter ZV auch einfach als Modellierungsannahme schon am Beginn unserer Überlegungen. Ein sehr einleuchtendes Prinzip, das wir noch verwenden werden ist, dass unabhängige ZV zu disjunkten Tupel zusammengefasst wieder unabhängig sind. (Wir verzichten hier auf den Beweis.)

**Proposition 4.7 (Tupel unabhängiger ZV)** Die Zufallsvariablen  $X_k$ ,  $k \geq 1$  seien voneinander unabhängig. Dann sind auch die als Tupel gebildeten neuen ZV

$$(X_1, \dots, X_{m_1}), \dots, (X_{m_{j-1}+1}, \dots, X_{m_j}), (X_{m_j+1}, \dots)$$

voneinander unabhängig.

Eine Grundsituation, die wir auf jeden Fall besser verstehen möchten, ist jene von Summen unabhängiger reeller ZV. Wir machen uns zunächst klar, wie man die Verteilung so einer unabhängigen Summe beschreiben kann.

**Proposition 4.8 (Verteilung unabhängiger Summen)** Sind  $X, Y$  unabhängige reelle ZV, dann hat  $X + Y$  die Verteilungsfunktion

$$F_{X+Y}(t) := P[X + Y \leq t] = \int_{-\infty}^{\infty} F_Y(t - v) dP_X(v), \quad t \in \mathbb{R}.$$

Insbesondere, falls  $X, Y$  Dichten  $f_X, f_Y$  besitzen, so hat  $X + Y$  die Dichte

$$f_X * f_Y(w) := \int_{-\infty}^{\infty} f_X(v) f_Y(w - v) dv, \quad w \in \mathbb{R}.$$

Die Verteilung  $P_{X+Y}$  der unabhängigen Summe  $X+Y$  nennt man übrigens auch die *Faltung* der beiden Maße  $P_X$  und  $P_Y$ , geschrieben  $P_X * P_Y$ . Entsprechend heißt  $f_X * f_Y$  die *Faltung* der Dichten  $f_X$  und  $f_Y$ .

**Beweis.** (i) Für  $t \in \mathbb{R}$  sei  $A_t := \{(x, y) \in \mathbb{R}^2 : x + y \leq t\}$ , dann haben wir

$$P[X + Y \leq t] = P[(X, Y) \in A_t] = P_{(X, Y)}(A_t) = \int_{\mathbb{R}^2} 1_{A_t}(v, u) d(P_X \otimes P_Y)(v, u).$$

Wegen  $1_{A_t}(v, u) = 1_{(-\infty, t-v]}(u)$  daher

$$P[X + Y \leq t] = \int_{-\infty}^{\infty} \int_{-\infty}^{t-v} 1 dP_Y(u) dP_X(v) = \int_{-\infty}^{\infty} P[Y \leq t - v] dP_X(v).$$

(ii) Indem wir die letzte Identität mithilfe der Dichten anschreiben und  $w = u + v$  setzen erhalten wir

$$P[X + Y \leq t] = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{t-v} f_Y(u) du \right) f_X(v) dv = \int_{-\infty}^{\infty} \left( \int_{-\infty}^t f_Y(w - v) dw \right) f_X(v) dv.$$

Fubini erlaubt uns die Integrationsreihenfolge zu vertauschen (weil  $f_X, f_Y \geq 0$ ),

$$P[X + Y \leq t] = \int_{-\infty}^t \left( \int_{-\infty}^{\infty} f_X(v) f_Y(w - v) dv \right) dw \quad \text{für } t \in \mathbb{R},$$

und dies zeigt, dass der Ausdruck in der Klammer die Dichte von  $X + Y$  ist. ■

**Beispiel 4.18 a)** Sind  $X, Y \stackrel{d}{=} \mathcal{U}([0, 1])$  unabhängig und in  $[0, 1]$  uniform verteilt, dann hat  $X + Y$  also die Dichte

$$1_{[0,1]} * 1_{[0,1]}(w) = \int_{-\infty}^{\infty} 1_{[0,1]}(v) 1_{[0,1]}(w - v) dv = \lambda([0, 1] \cap [w - 1, w]).$$

**b)** Sind  $X \stackrel{d}{=} \mathcal{N}(m_X, \sigma_X^2)$  und  $Y \stackrel{d}{=} \mathcal{N}(m_Y, \sigma_Y^2)$  unabhängig, dann gilt  $X + Y \stackrel{d}{=} \mathcal{N}(m_X + m_Y, \sigma_X^2 + \sigma_Y^2)$ , siehe PS.

Eine weitere wichtige Beobachtung, die den Umgang mit unabhängigen ZV erheblich erleichtert und auch weitreichende theoretische Konsequenzen hat, ist

**Proposition 4.9 (Erwartungswert unabhängiger Produkte)** Sind  $X, Y$  unabhängige reelle ZV mit  $\mathbb{E}[|XY|] < \infty$ , dann gilt

$$\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]. \quad (4.14)$$

**Beweis.** Gemäß allgemeiner Transformationsformel und  $P_{(X, Y)} = P_X \otimes P_Y$  ist

$$\mathbb{E}[|XY|] = \int_{\Omega} |XY| dP = \int_{\mathbb{R}^2} |xy| dP_{(X, Y)}(x, y) = \int_{\mathbb{R}^2} |xy| d(P_X \otimes P_Y)(x, y) < \infty.$$

Also ist  $(x, y) \mapsto xy$  bezüglich  $P_X \otimes P_Y$  integrierbar, und wir können den Satz von Fubini anwenden. Dieser zeigt

$$\begin{aligned} \mathbb{E}[XY] &= \int_{\Omega} XY dP = \int_{\mathbb{R}^2} xy dP_{(X, Y)}(x, y) = \int_{\mathbb{R}^2} xy d(P_X \otimes P_Y)(x, y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} xy dP_X(x) dP_Y(y) = \int_{\mathbb{R}} x dP_X(x) \int_{\mathbb{R}} y dP_Y(y) = \mathbb{E}[X] \mathbb{E}[Y], \end{aligned}$$

wie behauptet. ■

Eine besonders schöne Konsequenz daraus finden wir in

**Proposition 4.10 (Satz des Pythagoras, zufällig)** Sind  $X, Y$  unabhängige reelle ZV, jeweils mit  $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$ , dann ist

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 0 \quad (4.15)$$

und

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]. \quad (4.16)$$

**Beweis.** Weil  $X - \mathbb{E}[X]$  und  $Y - \mathbb{E}[Y]$  voneinander unabhängig sind, ist nach obiger Proposition  $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[X - \mathbb{E}[X]] \mathbb{E}[Y - \mathbb{E}[Y]] = 0 \cdot 0 = 0$ . Weiters finden wir dann

$$\begin{aligned} \text{Var}[X + Y] &= \mathbb{E} \left[ ((X - \mathbb{E}[X]) + (Y - \mathbb{E}[Y]))^2 \right] \\ &= \mathbb{E} \left[ (X - \mathbb{E}[X])^2 \right] + \mathbb{E} \left[ (Y - \mathbb{E}[Y])^2 \right] + 2 \mathbb{E} [(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \text{Var}[X] + \text{Var}[Y] + 0, \end{aligned}$$

und das war's auch schon. ■

Der Ausdruck in (4.15) ist übrigens als *Covarianz* der ZV  $X$  und  $Y$  bekannt,

$$\text{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y].$$

Falls Sie sich an Abschnitt 3.6.D erinnern, erkennen Sie darin vielleicht das Innenprodukt im Raum  $\mathcal{L}^2$  bzw  $L^2$  wieder,

$$\text{Cov}[X, Y] = \langle X - \mathbb{E}[X], Y - \mathbb{E}[Y] \rangle.$$

Unabhängigkeit zweier ZV in  $\mathcal{L}^2$  bedeutet also insbesondere, dass deren unsystematische Anteile zueinander orthogonal sind! Die Umkehrung gilt allerdings nicht, aus  $\text{Cov}[X, Y] = 0$  (man nennt  $X, Y$  dann *unkorreliert*) folgt ja nicht die Unabhängigkeit von  $X$  und  $Y$ .

**Beispiel 4.19** Als ganz einfache Anwendung halten wir fest, dass für eine Folge  $(X_k)_{k \geq 1}$  unabhängiger identisch verteilter reeller ZVen die Partialsummen  $S_n := X_1 + \dots + X_n$  Erwartungswert  $n\mathbb{E}[X]$  und Varianz  $n\text{Var}[X_1]$  besitzen. Der Mittelwert  $S_n/n$  hat daher Erwartungswert  $\mathbb{E}[X_1]$  und Varianz  $\text{Var}[S_n/n] = \text{Var}[S_n]/n^2 = \text{Var}[X_1]/n$ . Der Durchschnitt mehrerer solcher ZV neigt also zu geringeren zufälligen Schwankungen als die einzelnen Zufallsgrößen. Das sehen wir uns im nächsten Abschnitt gleich genauer an.

## 4.4 Das Gesetz der Großen Zahl

Schon in der elementaren diskreten Stochastik und in einfachen stetigen Modellen gibt es allerlei interessante und nicht selten überraschende Einsichten zu gewinnen. Wir konnten hier nur ein paar wenige Aspekte ansprechen, eine große Vielfalt lohnender Themen lernen Sie in weiterführenden LVen kennen. Viele der wirklich tiefen Ergebnisse der W-Theorie sind aber asymptotischer Natur, und betreffen Aussagen über Situationen in welchen eine sehr hohe Zahl zufallsabhängiger Größen zusammenwirkt. Spektakulär ist dabei insbesondere die Erkenntnis, dass es dann manchmal gar nicht auf Details der einzelnen Komponenten (etwa die exakte Verteilung gewisser ZV) ankommt, sondern auf wenige Grundeigenschaften (zB Existenz von Erwartungswerten oder Endlichkeit der Varianzen) und die Struktur der Situation (etwa, dass eine Summe unabhängiger Größen vorliegt, die einzeln jeweils nur einen kleinen Beitrag leisten). Durch ausreichend viel Zufall entsteht wieder Ordnung. Wir sind nun endlich so weit, zwei grundlegende Typen solcher Grenzwertsätze (in einfacher Form) zu diskutieren, das Gesetz der Großen

Zahl und (im nächsten Abschnitt) den Zentralen Grenzwertsatz.

**A) Sanity Check.** Bislang haben wir uns damit beschäftigt, den offiziellen formalen Rahmen der W-Theorie einzuführen, und dort verschiedene plausible Begriffe zu besprechen. Ein paar Wkeiten haben wir auch ausgerechnet. Wie gut dies alles aber zu der intuitiven (und vagen) Vorstellung von Wkeiten als asymptotischen Häufigkeiten passt, ist momentan noch ungeklärt. Vielleicht hat unser Wahrscheinlichkeitsraum-Schiff längst die Umlaufbahn verlassen und driftet in mathematisch hübsche Gebiete ab, aus denen es leider kein Zurück zur realen Welt mehr gibt?

Obwohl allgemein die Frage nach der Beziehung zwischen der Mathematik und dem Rest der Welt durchaus interessant bleibt, können wir uns nun davon überzeugen, dass das hier vorgeschlagene Setup zur ursprünglichen Idee passt. Die als *Gesetz der Großen Zahl* (GGZ) bekannten mathematischen Ergebnisse sind der Anker, der unsere Theorie mit den Wahrscheinlichkeiten der Erfahrungswelt verbindet. Sie beweisen, dass für jedes Ereignis  $A'$  in einem beliebigen W-Raum  $(\Omega', \mathcal{F}', P')$  der Wert  $P'[A']$  stets der Grenzwert der relativen Häufigkeiten des Auftretens von  $A'$  bei fortgesetzten unabhängigen Wiederholungen des beschriebenen Experimentes ist.

Dazu betrachten wir wie oben eine Folge unabhängiger Kopien des ZE (also den unendlichen Produktraum  $(\Omega, \mathcal{F}, P) := (\prod_{i \geq 1} \Omega', \otimes_{i \geq 1} \mathcal{F}', \otimes_{i \geq 1} P')$ ) und schreiben  $A_i := \{\omega : \omega_i \in A'\} \in \mathcal{F}$  für das Ereignis, dass bei der  $i$ ten Wiederholung das Ereignis  $A'$  eintritt. Um zu zählen, betrachten wir am besten die unabhängigen ZV  $X_i := 1_{A_i}$ . Dann ist  $P'[A'] = P[A_i] = \mathbb{E}[X_i]$ , und die Summe  $S_n := X_1 + \dots + X_n$  die Anzahl der erfolgreichen unter den Versuchen  $1, \dots, n$  ( $n \geq 1$ ). Die relative Häufigkeit des Auftretens von  $A'$  in den ersten  $n$  Versuchen ist dann das Mittel  $S_n/n = (X_1 + \dots + X_n)/n$ , und die erhoffte Aussage kann in der Form

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \longrightarrow \mathbb{E}[X_1] \quad \text{für } n \rightarrow \infty \quad (4.17)$$

angeschrieben werden. Da Sie genau aufgepasst haben, stellen Sie natürlich gleich fest, dass es hier einigen Interpretationsspielraum gibt. Auf der linken Seite stehen ja lauter ZV, also Funktionen auf  $\Omega$ , und da gibt es bekanntlich einige unterschiedliche Konvergenzbegriffe. Das ist auch einer der Gründe dafür, dass oben von mehreren Ergebnissen die Rede war. Je nach Art der Konvergenz spricht man zB vom schwachen oder starken GGZ etc, siehe unten.

Die Formulierung in (4.17) ist so gewählt, dass sie bereits auf eine weitreichende Verallgemeinerung hindeutet. Neben dem grundlegenden Fall der 0-1-ZV  $X_i$  oben können wir auch für andere Folgen  $(X_k)_{k \geq 1}$  unabhängiger ZV die alle dieselbe Verteilung besitzen (*iid-Folgen*) den Durchschnitt  $S_n/n = (X_1 + \dots + X_n)/n$  der beobachteten zufälligen Werte studieren. Ist dann eine Aussage wie in (4.17) korrekt, so bekommen wir eine neue Interpretation für den Erwartungswert  $\mathbb{E}[X_1]$  als asymptotisches Mittel der realisierten Werte.

**B) Das Schwache Gesetz der Großen Zahl.** Die einfachste Fassung des GGZ ist tatsächlich ziemlich simpel. Sie ist deshalb nicht weniger grundlegend. Im Spezialfall unabhängiger (aber möglicherweise unfairer) Münzwürfe geht sie auf Jakob Bernoulli zurück (ca 1690, erst später veröffentlicht), dessen Argument allerdings ganz anders verläuft, schon weil viele der nun verwendeten Begriffe nicht zur Verfügung standen. Der folgende Satz wird als *schwaches GGZ* bezeichnet, weil die Art der Konvergenz vergleichsweise schwach ist, und das Ergebnis daher weniger weiterführende Schlussfolgerungen erlaubt als etwa das später beschriebene "starke" Gesetz.

**Definition 4.16** Sind  $V, V_1, V_2, \dots$  reelle ZV auf einem W-Raum  $(\Omega, \mathcal{F}, P)$ , so sagt man, die Folge  $(V_n)$  konvergiert in Wahrscheinlichkeit gegen  $V$ , geschrieben

$$V_n \xrightarrow{P} V \quad \text{für } n \rightarrow \infty,$$

falls für jedes  $\varepsilon > 0$  gilt, dass  $\lim_{n \rightarrow \infty} P[|V_n - V| > \varepsilon] = 0$ .

Damit formulieren wir nun endlich

**Satz 4.2 (Schwachtes Gesetz der Großen Zahl)** *Es sei  $(X_k)_{k \geq 1}$  eine iid-Folge reeller ZVen mit  $\mathbb{E}[X_k^2] < \infty$ , und  $S_n := X_1 + \dots + X_n$ . Dann gilt*

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mathbb{E}[X_1] \quad \text{für } n \rightarrow \infty. \quad (4.18)$$

*Ausführlicher geschrieben also*

$$\text{für jedes } \varepsilon > 0 \text{ gilt} \quad \lim_{n \rightarrow \infty} P \left[ \left| \frac{S_n}{n} - \mathbb{E}[X_1] \right| > \varepsilon \right] = 0. \quad (4.19)$$

Die Aussage dieses Satzes bleibt übrigens auch dann noch korrekt, wenn man anstelle von  $\mathbb{E}[X_k^2] < \infty$  (dh  $X_k \in \mathcal{L}^2(P)$ ) nur verlangt, dass  $\mathbb{E}[X_k]$  existiert (also  $\mathbb{E}[|X_k|] < \infty$  bzw  $X_k \in \mathcal{L}^1(P)$ ). Der  $\mathcal{L}^2$ -Fall erlaubt aber ein überraschend einfaches Argument, und man bekommt sogar noch eine bessere Aussage, nämlich

**Satz 4.3 ( $\mathcal{L}^2$ -Gesetz der Großen Zahl)** *Es sei  $(X_k)_{k \geq 1}$  eine iid-Folge reeller ZVen mit  $\mathbb{E}[X_k^2] < \infty$ , und  $S_n := X_1 + \dots + X_n$ . Dann gilt*

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \xrightarrow{\mathcal{L}^2} \mathbb{E}[X_1] \quad \text{für } n \rightarrow \infty. \quad (4.20)$$

*Ausführlicher geschrieben also*

$$\lim_{n \rightarrow \infty} \left\| \frac{S_n}{n} - \mathbb{E}[X_1] \right\|_2 = \lim_{n \rightarrow \infty} \mathbb{E} \left[ \left( \frac{S_n}{n} - \mathbb{E}[X_1] \right)^2 \right] = 0. \quad (4.21)$$

**Beweis.** Es ist, aufgrund der Unabhängigkeit (siehe Beispiel 4.19),

$$\mathbb{E} \left[ \left( \frac{S_n}{n} - \mathbb{E}[X_1] \right)^2 \right] = \text{Var} \left[ \frac{S_n}{n} \right] = \frac{\text{Var}[X_1]}{n},$$

und daher  $\lim_{n \rightarrow \infty} \mathbb{E} \left[ (S_n/n - \mathbb{E}[X_1])^2 \right] = 0$ . Entsprechend konvergieren auch die Quadratwurzeln  $\|S_n/n - \mathbb{E}[X_1]\|_2$  dieser Erwartungswerte gegen 0. ■

Damit ergibt sich auch ganz leicht der

**Beweis von Satz 4.2.** Mithilfe der Chebyshev-Ungleichung und (4.21) finden wir, dass für jedes  $\varepsilon > 0$ ,

$$P \left[ \left| \frac{S_n}{n} - \mathbb{E}[X_1] \right| > \varepsilon \right] \leq \frac{1}{\varepsilon^2} \mathbb{E} \left[ \left( \frac{S_n}{n} - \mathbb{E}[X_1] \right)^2 \right] \longrightarrow 0 \quad \text{für } n \rightarrow \infty,$$

wie behauptet. ■

**Bemerkung 4.3 a)** *Die beiden Sätze für identisch verteilte Folgen  $(X_k)$  in  $\mathcal{L}^2$  bleiben gültig, wenn man Unabhängigkeit ersetzt durch die schwächere Forderung, dass die  $X_k$  unkorreliert sind. (Die Beweise ändern sich nicht.)*

**b)** *Ersatzlos streichen kann man die Unabhängigkeit aber nicht, wie schon der einfache Fall  $X_1 = X_2 = X_3 = \dots$  zeigt.*

**C) Das Starke Gesetz der Großen Zahl.** Wir besprechen noch kurz die Aussage der "erwachsenen" Variante des GGZ. Wie angekündigt, unterscheidet es sich von Satz 4.2 durch die Art der Konvergenz. Das *starke GGZ* stellt nämlich sicher, dass die beobachteten Häufigkeiten *fast sicher* (*f.s.*), also f.ü. auf dem W-Raum, gegen den Erwartungswert konvergieren.

**Satz 4.4 (Starkes Gesetz der Großen Zahl)** *Es sei  $(X_k)_{k \geq 1}$  eine iid-Folge reeller ZVen mit  $\mathbb{E}[|X_k|] < \infty$ , und  $S_n := X_1 + \dots + X_n$ . Dann gilt*

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \longrightarrow \mathbb{E}[X_1] \quad \text{fast sicher} \quad \text{für } n \rightarrow \infty. \quad (4.22)$$

*Explizit geschrieben also*

$$\mathbb{P} \left[ \lim_{n \rightarrow \infty} \frac{S_n}{n} = \mathbb{E}[X_1] \right] = 1. \quad (4.23)$$

Nachdem f.s.-Konvergenz immer auch Konvergenz in Wkeit impliziert (aber nicht umgekehrt), ist dieses Ergebnis wirklich stärker als Satz 4.2. Tatsächlich ist es ein Resultat anderer Qualität: Im starken GGZ wird eine Aussage über das Ereignis  $\{\lim_{n \rightarrow \infty} S_n/n = \mathbb{E}[X_1]\}$  getroffen, das wirklich erst durch die unendlichen Folge  $(X_k)_{k \geq 1}$  definiert wird, und ein wesentlich komplizierteres Objekt ist als die im schwachen GGZ separat betrachteten Ereignisse  $\{|S_n/n - \mathbb{E}[X_1]| > \varepsilon\}$ .

Beweise des starken GGZ erfordern entsprechend auch deutlich mehr Anstrengung. Viele der erforderlichen Hilfsmittel haben Sie aber bereits kennengelernt. (Der wesentliche Punkt liegt übrigens in der Einsicht, dass die Mittel  $S_n/n$  fast überall auf  $\Omega$  gegen irgendetwas konvergieren. Aufgrund des schwachen Gesetzes wissen wir dann ja sofort, dass der Grenzwert gerade  $\mathbb{E}[X_1]$  sein muss.) Für genauere Erklärungen muss ich Sie aber leider auf die Fortsetzung dieser VO vertrösten.

Als Ausblick sei noch erwähnt, dass die Geschichte mit einem Beweis des starken GGZ keineswegs endet. Zwar klärt der wunderbare Satz 4.4 die eingangs formulierte Frage nach der Konsistenz zwischen der formalen W-Theorie und der intuitiven Vorstellung von asymptotischen Häufigkeiten in mathematisch höchst befriedigender Weise, er konzentriert sich aber auf die Idealsituation von iid-Folgen. Das ist auch gut so, denn so kann man überhaupt erst beginnen, etwas besser zu verstehen. Danach mag man sich aber vielleicht fragen, ob das betrachtete Phänomen der Konvergenz gegen einen Mittelwert wirklich darauf angewiesen ist, dass diese perfekten Bedingungen vorliegen (vgl Bemerkung 4.3). Das führt zu weiteren spannenden Fragen für die Mathematik! Tatsächlich gibt es noch wesentlich mächtigere Formen des starken GGZ, etwa den Ergodensatz aus der Theorie Dynamischer Systeme.

## 4.5 Der Zentrale Grenzwertsatz

In unserem zwangsläufig kurzen Spaziergang in die Stochastik haben wir bisher nur mit ein paar wenigen konkrete Verteilungen gearbeitet, oft um in möglichst einfachem Rahmen neue Konzepte zu illustrieren. Manche dieser W-Maße, insbesondere die uniforme Verteilung auf einer endlichen Menge oder einem Intervall sind offensichtlich wichtig, weil sie natürliche Grundsituationen beschreiben, andere (Binomialverteilung) gewinnt man aus diesen, wenn man naheliegende ZV (Summen) betrachtet. Grundsätzlich kommt aber jedes W-Maß in Frage, und es gibt bei der Modellierung realer Situationen zunächst keinen Grund zu erwarten, dass die Verteilung einer zufälligen Größe durch eine besonders hübsche Funktion gegeben sein sollte. Sie wissen aber sicher, dass trotzdem an vielen Stellen in der Stochastik die Normalverteilung auftaucht. Wir haben nur noch nicht erklärt warum.

Damit wenden wir uns dem zweiten besonders prominenten asymptotischen Ergebnis der W-Theorie zu, dem Zentralen Grenzwertsatz (o Grenzverteilungssatz). Auf diesem beruht die Bedeutung der Normalverteilung. Knapp formuliert garantiert dieses Resultat, dass man in wichtigen Situationen viele Details einer komplexen Situation garnicht zu kennen braucht, und dennoch die Verteilung einer zufälligen

Größe (bis auf einen kleinen Fehler) vorhersagen kann. Man muss nur wissen, dass diese ZV von einer bestimmten Bauart ist, nämlich eine Summe vieler kleiner voneinander unabhängiger Einzelkomponenten. Alleine aufgrund dieser Struktur weiss man dann, dass ihre Verteilung gut durch eine Normalverteilung approximiert wird! Das ist theoretisch höchst bemerkenswert und praktisch unheimlich nützlich! Im riesigen Zoo aller W-Maße ist die Normalverteilung das geflügelte Einhorn.

**A) Der Zentrale Grenzwertsatz.** Wir sehen uns dieses Ergebnis in der Grundsituation einer iid-Folge  $(X_k)_{k \geq 1}$  reeller ZVen an: Hier erweisen sich die Partialsummen  $S_n = X_1 + \dots + X_n$  für große  $n$  als beinahe normalverteilt - sofern die  $X_k$  (wie  $\mathcal{N}$  selbst) endliche positive Varianz besitzen. Wegen  $\mathbb{E}[S_n] = n\mathbb{E}[X_1]$  gibt es über die Erwartungswerte nichts interessantes zu sagen, und wir nehmen die  $X_k$  als zentriert an,  $\mathbb{E}[X_k] = 0$ , um uns Schreibarbeit zu ersparen. Analog setzen wir voraus, dass  $\text{Var}[X_k] = 1$  gilt. (Sollte dies nicht von vorneherein gelten, dann gilt es aber für  $X_k/\sqrt{\text{Var}[X_k]}$ .) Erinnern wir uns, dass die Varianz für unabhängige ZV additiv ist, so finden wir  $\text{Var}[S_n] = n\text{Var}[X_1] = n$ , dh durch die Kombination vieler zufälliger Einflüsse wird die Verteilung immer breiter, und  $S_n$  kann für große  $n$  immer stärker schwanken. Um diese ZV trotzdem mit einer bestimmten Verteilung vergleichen zu können, sehen wir sie deshalb mit dem richtigen Zoomfaktor an, und studieren  $S_n/\sqrt{n}$  (das hat wieder Varianz 1). Auf diese Art im passenden Maßstab betrachtet, sind unsere Summen dann beinahe normalverteilt. Genauer:

**Satz 4.5 (Der Zentrale Grenzwertsatz)** *Es sei  $(X_k)_{k \geq 1}$  eine iid-Folge reeller ZVen mit  $\mathbb{E}[X_k] = 0$  und  $\text{Var}[X_k] = 1$ . Bezeichnet  $Z$  eine  $\mathcal{N}(0, 1)$ -verteilte ZV, so gilt für  $a < b$  in  $\mathbb{R}$ ,*

$$\mathbb{P}\left[a < \frac{S_n}{\sqrt{n}} \leq b\right] \longrightarrow \mathbb{P}[a < Z \leq b] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt \quad \text{für } n \rightarrow \infty. \quad (4.24)$$

Die Aussage (4.24) ist ein Beispiel für eine weitere wichtige Konvergenzart, der *Verteilungskonvergenz von ZV*. Wir haben hier nicht genug Platz, um diese allgemein zu besprechen, Sie lernen mehr darüber in der Fortsetzung dieser VO.

**B. Stabilität.** Der Schlüssel zum Verständnis des Satzes und seines Beweises ist eine besondere Eigenschaft der Normalverteilung. Diese ist eine *stabile Verteilung* in dem Sinne, dass die Summe endlich vieler unabhängiger ZV mit dieser Verteilung wieder eine Verteilung desselben Typs besitzt (wenn man sie richtig skaliert).

**Lemma 4.1 (Stabilität der Normalverteilung)** *Es seien  $Y, Y_1, Y_2, \dots$  unabhängig und jeweils  $\mathcal{N}(0, 1)$ -verteilt. Dann ist auch*

$$\frac{Y_1 + \dots + Y_m}{\sqrt{m}} \stackrel{d}{=} Y, \quad \text{also } \mathcal{N}(0, 1)\text{-verteilt für } m \geq 1. \quad (4.25)$$

*Umgekehrt, falls  $W, W_1, W_2, \dots$  ua ZV mit derselben Verteilung und  $\mathbb{E}[W] = 0$ ,  $\text{Var}[W] = 1$  sind, und auch*

$$\frac{W_1 + \dots + W_m}{\sqrt{m}} \stackrel{d}{=} W \quad \text{für } m \geq 1 \quad (4.26)$$

*gilt, dann ist  $W$  zwangsläufig  $\mathcal{N}(0, 1)$ -verteilt.*

**Beweis.** Für (4.25) siehe Beispiel 4.18 bzw PS-Aufgabe 46:  $Y_1 + \dots + Y_m \stackrel{d}{=} \mathcal{N}(0, m)$ . Da der zweite Parameter hier die Varianz angibt, folgt nach den Rechenregeln für Var dass  $(Y_1 + \dots + Y_m)/\sqrt{m} \stackrel{d}{=} \mathcal{N}(0, 1)$ . Die zweite Aussage beweisen wir hier nicht separat, sie folgt aus Satz 4.5, dessen Beweis nur den ersten Teil unseres Lemmas verwenden wird. ■



Was hat diese Beobachtung mit dem Grenzwertungssatz zu tun? Nun, einerseits ist dessen Aussage offenbar in dem Spezialfall korrekt, in welchem die  $X_k$  bereits  $\mathcal{N}(0, 1)$ -verteilt sind. In dieser Situation haben wir ja laut (4.25) nicht nur Konvergenz, sondern überhaupt Gleichheit in (4.24).

Wir können zudem skizzieren, warum ähnliches auch für andere Folgen  $(X_k)$  gelten könnte. Die folgende Überlegung bleibt aus Platzgründen vage, mit etwas mehr Geduld kann man daraus aber ein wasserdichtes mathematisches Argument machen. Wir betrachten nun also allgemein eine Folge  $(X_k)_{k \geq 1}$  wie in Satz 4.5, und unterstellen zunächst mal recht optimistisch, dass es irgendeine Verteilung  $W$  gibt mit

$$\frac{S_n}{\sqrt{n}} \stackrel{d}{\approx} W \quad \text{falls } n \text{ groß.} \quad (4.27)$$

Hier schreiben wir ausnahmsweise recht vage  $V \stackrel{d}{\approx} W$  um anzudeuten, dass  $V$  und  $W$  fast dieselbe Verteilung haben. Wenn wir, bei festem  $m \geq 1$ , für einen großen Index der Form  $n = jm$  nun  $W_k := \frac{1}{\sqrt{j}}(X_{(k-1)j+1} + \dots + X_{kj})$ ,  $k \in \{1, \dots, m\}$  nennen, dann sehen wir, dass damit auch

$$\frac{W_1 + \dots + W_j}{\sqrt{j}} = \frac{S_n}{\sqrt{n}} \stackrel{d}{\approx} W.$$

Dabei sind aber die  $W_1, \dots, W_j$  i.i.d. und haben alle dieselbe Verteilung,  $W_i \stackrel{d}{=} W_1$ , und weil  $j$  groß ist, zeigt (4.27) mit  $j$  anstelle von  $n$ , dass  $W_1 = \frac{1}{\sqrt{j}}S_j \stackrel{d}{\approx} W$ . Wir sehen daher, dass die Verteilung von  $W$  annähernd die Eigenschaft (4.26) besitzen sollte. Nachdem letztere laut Lemma aber die Normalverteilung eindeutig charakterisiert, kann nur diese als Grenzwertung auftauchen.

**C. Beweis des Zentralen Grenzwertsatzes.** Beim Nachweis von Verteilungskonvergenz arbeitet man nur selten direkt mit den Verteilungen (oder Verteilungsfunktionen). Es ist nämlich oft viel einfacher anzuschauen, wie diese  $W$ -Maße gutmütige Funktionen integrieren. Zur Erinnerung: es ist ja  $\mathbb{E}[f(X)] = \int_{\mathbb{R}} f dP_X$  das Integral von  $f$  unter der Verteilung  $P_X$ . Natürlich kennt man  $P_X$  vollständig, wenn man das entsprechende Integral für jede Indikatorfunktion  $1_A$ ,  $A \in \mathcal{B}_{\mathbb{R}}$  kennt, weil ja  $P_X[A] = \int_{\mathbb{R}} 1_A dP_X$ . Letztere sind aber nicht immer sehr angenehm, die Mengen  $A$  dürfen ja doch recht wild sein. Günstiger ist es da oft, mit stetigen oder gar differenzierbaren Funktionen arbeiten zu können. Das geht tatsächlich sehr gut. Wir illustrieren dieses allgemeine Prinzip nur in einem speziellen Fall, den wir für den Beweis von Satz 4.5 benötigen. Dazu sei  $\mathcal{C}_b^3(\mathbb{R}) := \{f : \mathbb{R} \rightarrow \mathbb{R} : f \text{ ist } \mathcal{C}^3 \text{ und } f, f', f'', f''' \text{ sind beschränkt}\}$ .

**Lemma 4.2 (Verteilungskonvergenz via Integration von  $\mathcal{C}_b^3(\mathbb{R})$ -Funktionen)**

*Es seien  $Z, V, V_1, V_2, \dots$  reelle ZV, wobei  $Z$  eine stetige VF besitze. Dann gilt*

$$\mathbb{E}[f(V)] = \mathbb{E}[f(Z)] \quad \text{für } f \in \mathcal{C}_b^3(\mathbb{R}) \quad \implies \quad P_X = P_Z. \quad (4.28)$$

Weiters, falls

$$\mathbb{E}[f(V_n)] \longrightarrow \mathbb{E}[f(Z)] \quad \text{für } n \rightarrow \infty \quad \text{sobald } f \in \mathcal{C}_b^3(\mathbb{R}), \quad (4.29)$$

dann gilt für alle  $a < b$  in  $\mathbb{R}$ , dass

$$P[a < V_n \leq b] \longrightarrow P[a < Z \leq b] \quad \text{für } n \rightarrow \infty. \quad (4.30)$$

**Beweis.** Offenbar folgt (4.28) aus der zweiten Aussage. (Man wähle  $V_n = V$ .) Um (4.30) zu bekommen, müssen wir nur zeigen, dass für jedes  $b \in \mathbb{R}$ ,

$$P[V_n \leq b] \longrightarrow P[Z \leq b] \quad \text{für } n \rightarrow \infty. \quad (4.31)$$

Dazu sei jetzt  $b$  fest und  $\varepsilon > 0$  beliebig. Weil  $Z$  eine stetige VF besitzt, gibt es ein  $\eta > 0$  mit

$$P[Z \leq b - \eta] > P[Z \leq b] - \varepsilon \quad \text{und} \quad P[Z \leq b + \eta] < P[Z \leq b] + \varepsilon.$$

Wir wählen nun Funktionen  $f, g \in \mathcal{C}_b^3(\mathbb{R})$  mit  $1_{(-\infty, b-\eta]} \leq f \leq 1_{(-\infty, b]}$  und  $1_{(-\infty, b]} \leq g \leq 1_{(-\infty, b+\eta]}$ . Dann sind

$$\mathbb{E}[f(V_n)] \leq \mathbb{P}[V_n \leq b] \leq \mathbb{E}[g(V_n)],$$

sowie

$$\mathbb{E}[f(Z)] \geq \mathbb{P}[Z \leq b - \eta] \quad \text{und} \quad \mathbb{E}[g(Z)] \leq \mathbb{P}[Z \leq b + \eta].$$

Wegen  $\mathbb{E}[f(V_n)] \rightarrow \mathbb{E}[f(Z)]$  und  $\mathbb{E}[g(V_n)] \rightarrow \mathbb{E}[g(Z)]$  gibt es dann ein  $N$  mit

$$\mathbb{P}[Z \leq b] - \varepsilon < \mathbb{E}[f(V_n)] \leq \mathbb{P}[V_n \leq b] \leq \mathbb{E}[g(V_n)] < \mathbb{P}[Z \leq b] + \varepsilon \quad \text{für } n \geq N,$$

Und weil  $\varepsilon > 0$  beliebig war, folgt (4.31). ■

Als Vorbereitung für den Beweis des Satzes wiederholen wir noch:

**Bemerkung 4.4 (Satz von Taylor)** Nach diesem grundlegenden Satz der Analysis gibt es für  $f \in \mathcal{C}_b^3(\mathbb{R})$  und  $s, t \in \mathbb{R}$  einen Punkt  $\xi$  zwischen  $s$  und  $t$  derart, dass  $f(t) = f(s) + f'(s)(t-s) + f''(\xi)(t-s)^2/2$ , bzw. (indem wir  $f''(s)(t-s)^2/2$  einfügen)

$$f(t) - f(s) - f'(s)(t-s) - f''(s)\frac{(t-s)^2}{2} = (f''(\xi) - f''(s))\frac{(t-s)^2}{2}. \quad (4.32)$$

Damit sind wir bereit für den

**Beweis von Satz 4.5.** (i) Es seien  $Z$  und  $(X_k)_{k \geq 1}$  wie in der Aussage des Satzes, insbesondere also  $\mathbb{E}[X_k] = 0$  und  $\mathbb{E}[X_k^2] = 1$ . Wir nehmen eine beliebige Funktion  $f \in \mathcal{C}_b^3(\mathbb{R})$ . Laut Lemma reicht es zu zeigen, dass

$$\mathbb{E}[f(S_n/\sqrt{n})] \longrightarrow \mathbb{E}[f(Z)] \quad \text{für } n \rightarrow \infty. \quad (4.33)$$

Dazu vergleichen wir  $(X_k)_{k \geq 1}$  mit einer dazu unabhängigen Folge  $(Y_k)_{k \geq 1}$  unabhängiger  $\mathcal{N}(0, 1)$ -verteilter ZV. Wir wissen ja schon, dass  $Z_n := \frac{1}{\sqrt{n}}(Y_1 + \dots + Y_n)$  dann ebenfalls  $\mathcal{N}(0, 1)$ -verteilt ist, also  $\mathbb{E}[f(Z_n)] = \mathbb{E}[f(Z)]$  gilt. Daher folgt (4.33) falls

$$\mathbb{E}[f(S_n/\sqrt{n})] - \mathbb{E}[f(Z_n)] \longrightarrow 0 \quad \text{für } n \rightarrow \infty. \quad (4.34)$$

(ii) Eine schöne Idee die es erlaubt, dies nachzuweisen stammt von J.W.Lindeberg. Sie ist ziemlich genau 100 Jahre alt. Um diese Differenz gut zu kontrollieren, tauschen wir schrittweise in der Summe  $X_1 + \dots + X_n$  einen Term  $X_k$  nach dem anderen gegen die entsprechende Variable  $Y_k$  aus, und schätzen ab, wie weit dies jeweils den Erwartungswert verändern kann. Seien dazu

$$Z_{n,i} := \frac{1}{\sqrt{n}}(X_1 + \dots + X_{i-1} + X_i + Y_{i+1} + \dots + Y_n), \quad i \in \{0, \dots, n\}, \text{ und}$$

$$S_{n,i} := \frac{1}{\sqrt{n}}(X_1 + \dots + X_{i-1} + Y_{i+1} + \dots + Y_n), \quad i \in \{1, \dots, n\}.$$

Dann haben wir

$$Z_{n,0} = Z_n, \quad Z_{n,n} = S_n/\sqrt{n}, \quad \text{und} \quad Z_{n,i} - S_{n,i} = X_i/\sqrt{n},$$

sodass

$$\mathbb{E}[f(S_n/\sqrt{n})] - \mathbb{E}[f(Z_n)] = \sum_{i=1}^n \mathbb{E}[f(Z_{n,i})] - \mathbb{E}[f(Z_{n,i-1})]. \quad (4.35)$$

(iii) Sei nun  $\varepsilon > 0$ . Wir betrachten die Summanden in der letzten Zeile. Um sie in den Griff zu bekommen, wählen wir eine Konstante  $M$  mit  $|f|, |f'|, |f''|, |f'''| \leq M < \infty$  und setzen  $\delta := \varepsilon/M > 0$ . Nach dem Mittelwertsatz der Differentialrechnung (auf  $f''$  angewandt) finden wir dann

$$|x - y| < \delta \implies |f''(x) - f''(y)| \leq \delta \sup |f'''| \leq \varepsilon. \quad (4.36)$$

Indem wir im Satz von Taylor in der Form von (4.32)  $t := Z_{n,i}$  und  $s := S_{n,i}$  setzen, bekommen wir

$$\begin{aligned} f(Z_{n,i}) - f(S_{n,i}) - f'(S_{n,i})\frac{X_i}{\sqrt{n}} - f''(S_{n,i})\frac{X_i^2}{2n} \\ = (f''(C_{n,i}) - f''(S_{n,i}))\frac{X_i^2}{2n} =: R_{n,i}, \end{aligned} \quad (4.37)$$

wobei  $C_{n,i}$  zwischen  $Z_{n,i}$  und  $S_{n,i}$  liegt. Dabei ist wegen (4.36)

$$|R_{n,i}| \leq Q_{n,i} := \begin{cases} \frac{\varepsilon}{2n} X_i^2 & \text{falls } |X_i| \leq \delta\sqrt{n}, \\ \frac{M}{n} X_i^2 & \text{sonst.} \end{cases}$$

Weil  $X_i^2 \geq 1_{\{|X_i| > \delta\sqrt{n}\}} X_i^2 \searrow 0$  f.ü. mit  $X_i^2$  integrierbar, haben wir  $\mathbb{E}[1_{\{|X_i| > \delta\sqrt{n}\}} X_i^2] \searrow 0$  (dominierte Konvergenz). Also gibt es ein  $n_0$  mit  $\mathbb{E}[1_{\{|X_i| > \delta\sqrt{n}\}} X_i^2] < \varepsilon/2M$  für  $n \geq n_0$ . Für die obere Schranke  $Q_{n,i}$  erhalten wir durch Integration über die genannten Teilmengen damit

$$\mathbb{E}[1_{\{|X_i| \leq \delta\sqrt{n}\}} Q_{n,i}] \leq \frac{\varepsilon}{2n} \mathbb{E}[X_i^2] = \frac{\varepsilon}{2n},$$

und

$$\mathbb{E}[1_{\{|X_i| > \delta\sqrt{n}\}} Q_{n,i}] \leq \frac{M}{n} \mathbb{E}[1_{\{|X_i| > \delta\sqrt{n}\}} X_i^2] < \frac{\varepsilon}{2n} \quad \text{für } n \geq n_0.$$

Insgesamt also

$$\mathbb{E}[|R_{n,i}|] \leq \mathbb{E}[Q_{n,i}] < \frac{\varepsilon}{n} \quad \text{für } n \geq n_0. \quad (4.38)$$

(iv) Da die  $X_i$  per Annahme zentriert sind, ist  $\mathbb{E}[X_i] = 0$ . Weil die Summe in der Definition von  $S_{n,i}$  die ZV  $X_i$  nicht enthält, sind  $S_{n,i}$  und  $X_i$  voneinander unabhängig, und wir haben  $\mathbb{E}[f''(S_{n,i})X_i^2] = \mathbb{E}[f''(S_{n,i})] \mathbb{E}[X_i^2] = \mathbb{E}[f''(S_{n,i})]$  (siehe Proposition 4.9). Indem wir in der Gleichung (4.37) auf beiden Seiten integrieren, erhalten wir deshalb

$$\mathbb{E}[f(Z_{n,i})] - \mathbb{E}[f(S_{n,i})] - \frac{\mathbb{E}[f''(S_{n,i})]}{2n} = \mathbb{E}[R_{n,i}],$$

und in Anbetracht von (4.38),

$$\left| \mathbb{E}[f(Z_{n,i})] - \mathbb{E}[f(S_{n,i})] - \frac{\mathbb{E}[f''(S_{n,i})]}{2n} \right| \leq \mathbb{E}[|R_{n,i}|] < \frac{\varepsilon}{n} \quad \text{für } n \geq n_0.$$

Ganz analog zeigt man

$$\left| \mathbb{E}[f(Z_{n,i-1})] - \mathbb{E}[f(S_{n,i})] - \frac{\mathbb{E}[f''(S_{n,i})]}{2n} \right| \leq \mathbb{E}[|R_{n,i}|] < \frac{\varepsilon}{n} \quad \text{für } n \geq n_0,$$

(wobei halt  $Y_i$  die Rolle von  $X_i$  übernimmt). Gemeinsam liefern diese beiden Abschätzungen mittels Dreiecksungleichung

$$|\mathbb{E}[f(Z_{n,i})] - \mathbb{E}[f(Z_{n,i-1})]| < \frac{2\varepsilon}{n} \quad \text{für } n \geq n_0,$$

und dann (vergleiche (4.35))

$$|\mathbb{E}[f(S_n/\sqrt{n})] - \mathbb{E}[f(Z_n)]| \leq \sum_{i=1}^n |\mathbb{E}[f(Z_{n,i})] - \mathbb{E}[f(Z_{n,i-1})]| < 2\varepsilon \quad \text{für } n \geq n_0.$$

Nachdem  $\varepsilon > 0$  beliebig war, folgt nun (4.34). ■