

Skript zur Vorlesung  
Maß- und Wahrscheinlichkeitstheorie

# **Teil I: Maß- und Integrationstheorie**

Robert Hable  
Christina Schneider \*

Wintersemester 2013/14

LMU München  
Institut für Statistik

\* Originalfassung Kapitel 1-10: Robert Hable im Wintersemester  
2007/08

# Inhaltsverzeichnis

<b>1</b>	<b>Maßtheorie <math>\leftrightarrow</math> Wahrscheinlichkeitstheorie</b>	<b>1</b>
<b>2</b>	<b><math>\sigma</math>-Algebren</b>	<b>3</b>
<b>3</b>	<b>Maße</b>	<b>8</b>
3.1	Definitionen . . . . .	8
3.2	Der Begriff „Wahrscheinlichkeitstheorie“ . . . . .	9
3.3	Eigenschaften von Maßen . . . . .	10
<b>4</b>	<b>Wichtige Maße</b>	<b>13</b>
4.1	Das Dirac-Maß . . . . .	13
4.2	Das Zählmaß . . . . .	13
4.2.1	Endlicher Fall . . . . .	13
4.2.2	Abzählbar unendlicher Fall . . . . .	13
4.2.3	Überabzählbar unendlicher Fall . . . . .	13
4.3	Das Lebesgue-Maß . . . . .	14
4.3.1	Ziel . . . . .	14
4.3.2	Das eindimensionale Lebesgue-Maß $\lambda$ auf $(\mathbb{R}, \mathfrak{B})$ . . . . .	14
4.3.3	Das n-dimensionale Lebesgue-Maß $\lambda^n$ auf $(\mathbb{R}^n, \mathfrak{B}^{\otimes n})$ . . . . .	17
4.3.4	$\lambda$ kann nicht auf ganz $\mathcal{P}(\mathbb{R})$ sinnvoll definiert werden! . . . . .	18
<b>5</b>	<b>Messbare Abbildungen und Bildmaße</b>	<b>20</b>
5.1	Einführende Beispiele . . . . .	20
5.2	Definitionen und Eigenschaften . . . . .	21
5.3	Reellwertige Funktionen . . . . .	23
<b>6</b>	<b>Integration</b>	<b>27</b>
6.1	Idee . . . . .	27
6.2	Definition des allgemeinen Integrals . . . . .	28
6.3	Eigenschaften des Integrals . . . . .	33
6.4	Die Transformationsformel – Integration bzgl. dem Bildmaß . . . . .	37
<b>7</b>	<b>Dichtefunktionen</b>	<b>41</b>
7.1	Der Satz von Radon–Nikodym . . . . .	41
7.2	Integralberechnungen . . . . .	43
7.2.1	Dirac-Maß . . . . .	43
7.2.2	Zählmaß . . . . .	43
7.2.3	Lebesgue-Maß und Lebesgue-Dichten . . . . .	44
7.3	Einschub: Zufallsvariablen . . . . .	44

7.4	Transformationssatz für Dichten . . . . .	47
7.5	Ausblick: Gemischte Verteilungen . . . . .	50
7.6	Ausblick: Stochastische Prozesse . . . . .	52
<b>8</b>	<b>Unabhängigkeit</b>	<b>54</b>
8.1	Unabhängigkeit bei Mengensystemen . . . . .	54
8.2	Unabhängigkeit bei Zufallsvariablen . . . . .	55
<b>9</b>	<b>Produktmaße</b>	<b>58</b>
9.1	2-faches Produkt $\Omega_1 \times \Omega_2$ . . . . .	58
9.1.1	Produkte von $\sigma$ -Algebren . . . . .	58
9.1.2	Produkte von Maßen . . . . .	60
9.2	$n$ -faches Produkt $\Omega_1 \times \cdots \times \Omega_n$ . . . . .	62
9.3	Faltung von Wahrscheinlichkeitsmaßen . . . . .	65
<b>10</b>	<b>Bedingter Erwartungswert</b>	<b>69</b>
10.1	Einleitung . . . . .	69
10.1.1	Elementare bedingte Wahrscheinlichkeiten . . . . .	69
10.1.2	Der Informationsgehalt einer $\sigma$ -Algebra . . . . .	70
10.1.3	Der Informationsgehalt einer Zufallsvariable . . . . .	73
10.2	Definition und grundlegende Eigenschaften . . . . .	74
10.3	Faktorierte bedingte Erwartung . . . . .	82
10.4	Bedingte Dichten . . . . .	84
10.5	Bedingte Wahrscheinlichkeiten . . . . .	89
10.5.1	Idee: Wahrscheinlichkeiten als spezielle Erwartungswerte . . . . .	89
10.5.2	Definition: Bedingte Wahrscheinlichkeiten . . . . .	90
10.5.3	Einschub: Markov-Kerne . . . . .	92
10.5.4	Verallgemeinerter Satz von Fubini . . . . .	94
10.5.5	Berechnung der bedingten Verteilung . . . . .	97
<b>11</b>	<b>Einschub: Zufällige Maße auf <math>(\mathbb{R}^p, \mathfrak{B}^p)</math></b>	<b>99</b>
<b>12</b>	<b>Das Riemann-Stieltjes-Integral</b>	<b>101</b>
12.1	Treppenfunktionen . . . . .	101
12.2	Das bestimmte Integral . . . . .	102
12.3	Eine Alternative Vorgehensweise . . . . .	103
12.4	Das Riemann-Stieltjes-Integral . . . . .	104
<b>13</b>	<b>Gesetze der Großen Zahlen und Grenzwertsätze</b>	<b>109</b>
13.1	Konvergenzbegriffe . . . . .	109
13.2	Gesetze der Großen Zahlen . . . . .	111
13.3	Die Kolmogorov'schen Sätze . . . . .	112
13.4	Der Zentrale Grenzwertsatz . . . . .	112
	Literatur . . . . .	114

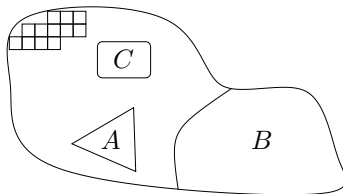
# 1 Maßtheorie ↔ Wahrscheinlichkeitstheorie

Was sind Wahrscheinlichkeiten überhaupt?

→ Wir wissen es nicht!

**Aber:** Wir können Wahrscheinlichkeiten mathematisch modellieren – und zwar mit Hilfe der Maßtheorie.

Die Maßtheorie beschäftigt sich eigentlich mit „messen“; und zwar zuallererst mit dem Messen von Flächen.



Ein Maß  $\mu$  ordnet jeder Teilfläche einen „gemessenen“ Wert zu:

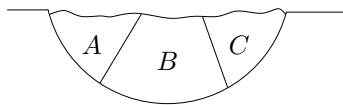
$$\mu(B) = 52 \text{ (m}^2\text{)}$$

d.h.: Teilfläche  $B$  besitzt den Flächeninhalt 52 bzgl. dem Maß  $\mu$  (also in  $\text{m}^2$ ).

Wir hätten aber vielleicht auch ein anderes Maß  $\nu$  wählen können:

$$\nu(B) = 1,668089 \text{ (sächsische Quadratellen)}$$

**Anderes Beispiel:** ein Teich



Wir messen die Wassermenge in Litern – das entsprechende Maß nennen wir  $\mu$ :

$$\mu(A) = 8, \quad \mu(B) = 15, \quad \mu(C) = 6$$

(Klar:  $\mu(A \cup B) = 23$ )

Möglicherweise interessieren wir uns gar nicht für die Wassermengen, sondern für die darin enthaltenen Algenmengen.

→ Wir brauchen ein anderes Maß  $\nu$ , das die Algenmenge in kg angibt:

$$\nu(A) = 3, \quad \nu(B) = 1, \quad \nu(C) = 7$$

(Klar:  $\mu(A \cup B) = 4$ )

Was hat das jetzt mit Wahrscheinlichkeitstheorie zu tun? Zunächst mal nichts.

Um 1930 waren die mathematischen Grundlagen der Maßtheorie weitgehend entwickelt (insb. durch Arbeiten von Cauchy, Riemann, Borel, Lebesgue). Im Gegensatz dazu war die Stochastik kein richtiges Teilgebiet der Mathematik, denn: Es fehlte eine exakte mathematische Formalisierung von Wahrscheinlichkeiten.

„Probability is the most important concept in modern science, especially as nobody has the slightest notion what it means.“

(Bertrand Russell in einer Vorlesung 1929)

1933 wendet Kolmogorov die Maßtheorie auf Wahrscheinlichkeiten an: (Kolmogorov (1933), siehe auch <http://www.mathematik.com/Kolmogorov/>): Wahrscheinlichkeiten werden mit Hilfe der Maßtheorie mathematisch formalisiert:

	Maßtheorie	Wahrscheinlichkeitstheorie
$A$	Menge	Ereignis
$\emptyset$	leere Menge	unmögliches Ereignis
$\mu(A) = \frac{1}{3}$	Menge $A$ hat Masse $\frac{1}{3}$	Wahrscheinlichkeit, dass Ereignis $A$ eintritt ist $\frac{1}{3}$
...	...	...

→ Die Objekte werden nur uminterpretiert. Vom mathematischen Standpunkt aus betrachtet spielt die Interpretation natürlich keine Rolle; d.h. seit Kolmogorov gilt: „Wahrscheinlichkeitstheorie ist Maßtheorie!“

Wir wissen immer noch nicht, was Wahrscheinlichkeiten eigentlich sind, aber wir können sie nun mathematisch modellieren. Dazu heißt es in Hoffmann-Jørgensen (1994a), Introduction

„The history demonstrates that the notions of randomness and probabilities are difficult – very difficult – to apprehend, and that the notions are fairly new and alien to humans. (I don’t think that mankind ever was meant to apprehend randomness and probabilities.) Second, the history and a wealth of examples (...) demonstrate that our intuition about probability is poor and often takes the wrong track. Thus, when you evaluate probabilities trust your computations and doubt your intuition. For this reason the interpretation of probabilities presents a difficult and fundamental problem, and the past – from the emerge of probabilities in 1550 until today – contains numerous examples of misinterpretations and miscalculations.“

**Tipp:** Wer in dieser Vorlesung Probleme damit hat, sich etwas unter den Begriffen vorzustellen, der soll sich keine Ereignisse und Wahrscheinlichkeiten vorstellen, sondern stattdessen ans Messen von Flächen denken – das ist viel einfacher!

## 2 $\sigma$ -Algebren

Sei  $\Omega$  eine Menge.

$$\mathcal{P}(\Omega) = \{A \mid A \subset \Omega\}$$

ist die Menge aller möglichen Teilmengen von  $\Omega$ .

D.h.: Die Elemente der Menge  $\mathcal{P}(\Omega)$  sind selbst wieder Mengen – und zwar Teilmengen von  $\Omega$ .

**Beispiel 2.1**  $\Omega = \mathbb{R}$

$$\begin{aligned} \{1; 2; 5\} \subset \mathbb{R} &\Rightarrow \{1; 2; 5\} \in \mathcal{P}(\mathbb{R}) \\ [-2; 7] \subset \mathbb{R} &\Rightarrow [-2; 7] \in \mathcal{P}(\mathbb{R}) \end{aligned}$$

**Definition 2.2**

$$\mathcal{P}(\Omega) = \{A \mid A \subset \Omega\}$$

heißt Potenzmenge von  $\Omega$ . Statt  $\mathcal{P}(\Omega)$  wird häufig auch  $2^\Omega$  geschrieben.

**Bemerkung 2.3**

a)  $\emptyset \in \mathcal{P}(\Omega)$

b) Für  $A \in \mathcal{P}(\Omega)$  ist

$$A^C = \{\omega \in \Omega \mid \omega \notin A\}$$

das Komplement von  $A$ . Es wird auch mit  $\bar{A}$  oder  $\neg A$  notiert.

c) Sei  $I$  eine Indexmenge (z.B.  $I = \{1, \dots, n\}$ ,  $I = \mathbb{N}$ ,  $I = \mathbb{R}$  oder sonst irgendeine Menge) und sei  $\mathcal{A} \subset \mathcal{P}(\Omega)$ . Dann ist der Ausdruck

$$\text{„Sei } (A_i)_{i \in I} \subset \mathcal{A} \text{“}$$

gleichbedeutend mit

„Für jedes  $i \in I$  sei  $A_i$  eine Teilmenge von  $\Omega$  so dass  $A_i \in \mathcal{A}$ .“

d) Sei  $I$  eine Indexmenge und  $(A_i)_{i \in I} \subset \mathcal{P}(\Omega)$ . Dann ist

$$\bigcup_{i \in I} A_i = \{\omega \in \Omega \mid \exists i_0 \in I, \text{ so dass } \omega \in A_{i_0}\}$$

und

$$\bigcap_{i \in I} A_i = \{\omega \in \Omega \mid \omega \in A_i \quad \forall i \in I\}$$

**Satz 2.4 (Morgan'sche Regeln)**

Sei  $I$  eine Indexmenge und  $(A_i)_{i \in I} \subset \mathcal{P}(\Omega)$ . Dann ist

$$\left( \bigcup_{i \in I} A_i \right)^C = \bigcap_{i \in I} A_i^C \quad \text{und} \quad \left( \bigcap_{i \in I} A_i \right)^C = \bigcup_{i \in I} A_i^C$$

$\mathcal{P}(\Omega)$  enthält die Objekte, die wir messen wollen (vgl. Beispiele in Kapitel 1).

**Aber:** Wir verlangen nicht, dass wir alle Teilmengen  $A \subset \Omega$  messen können. Statt ganz  $\mathcal{P}(\Omega)$  betrachten wir oft nur bestimmte Teilmengen von  $\Omega$ , also eine  $\sigma$ -Algebra

$$\mathcal{A} \subset \mathcal{P}(\Omega)$$

**Definition 2.5** Sei  $\Omega$  eine Menge. Eine  $\sigma$ -Algebra auf einer Menge  $\Omega$  ist ein  $\mathcal{A} \subset \mathcal{P}(\Omega)$  mit:

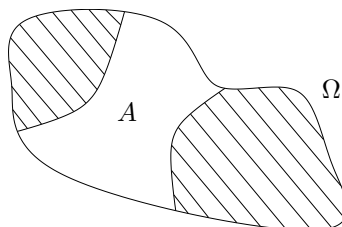
- (i)  $\Omega \in \mathcal{A}$
- (ii) Falls  $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ .
- (iii) Für  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$  ist auch

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$$

**Bemerkung 2.6**

a) Als Beispiel dient wieder das Messen von Flächen:

- Wenn wir die Teilfläche  $A \subset \Omega$  messen können, dann ist es plausibel, dass wir auch die Teilfläche  $A^c \subset \Omega$  messen können.  
 → (ii) in Def. 2.5



- Wenn wir die Teilflächen  $A_1, A_2 \in \mathcal{P}(\Omega)$  messen können, dann können wir auch  $A_1 \cup A_2$  messen.  
 → (iii) in Def. 2.5 ist noch ein kleines bisschen mehr.
- Natürlich können wir auch  $\Omega$  selbst messen.  
 → (i) in Def. 2.5

b) Warum verlangen wir eigentlich nicht, dass wir alle  $A \in \mathcal{P}(\Omega)$  messen können?

Das geht nicht immer gut (v.a. in Fällen, die für die Statistik extrem wichtig sind); vgl. Kapitel 4.3.4 über das Lebesgue-Maß.

c) Es gibt in der Maßtheorie noch weitere Objekte, die der  $\sigma$ -Algebra ähnlich sind: Algebra, Ring, Dynkin-System.

### Beispiel 2.7

- Die Menge  $\mathcal{F} = \{\emptyset, \{a\}, \{b\}, \{c, d\}, \{a, c, d\}, \{b, c, d\}, \Omega\}$  ist keine  $\sigma$ -Algebra auf  $\Omega = \{a, b, c, d\}$ , da  $\{a, b\} \notin \mathcal{F}$
- $\mathcal{P}(\Omega)$  ist  $\sigma$ -Algebra auf  $\Omega$
- $\{\emptyset, \Omega\}$  ist die trivialste und kleinste  $\sigma$ -Algebra auf  $\Omega$
- $\{A \subset \mathbb{N} : A \text{ oder } A^c \text{ abzählbar}\}$  ist  $\sigma$ -Algebra auf  $\Omega = \mathbb{N}$

### Notation 2.8 Statt

„Sei  $\Omega$  eine Menge und  $\mathcal{A}$  eine  $\sigma$ -Algebra auf  $\Omega$ .“

sagen wir auch

„Sei  $\Omega$  eine Menge mit  $\sigma$ -Algebra  $\mathcal{A}$ .“

oder

„Sei  $(\Omega, \mathcal{A})$  ein Messraum.“

Und falls  $A \in \mathcal{A}$  sagen wir auch

„ $A$  ist  $\mathcal{A}$ -messbar.“

### Satz 2.9 (Eigenschaften einer $\sigma$ -Algebra)

Sei  $(\Omega, \mathcal{A})$  ein Messraum. Dann gilt:

a)  $\emptyset \in \mathcal{A}$

b)  $A_1 \in \mathcal{A}, \dots, A_n \in \mathcal{A} \Rightarrow A_1 \cup \dots \cup A_n \in \mathcal{A}$

c) Für  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$  ist auch

$$\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$$

d)  $A_1 \in \mathcal{A}, \dots, A_n \in \mathcal{A} \Rightarrow A_1 \cap \dots \cap A_n \in \mathcal{A}$

e)  $A_1 \in \mathcal{A}, A_2 \in \mathcal{A} \Rightarrow A_1 \setminus A_2 \in \mathcal{A}$

### Satz 2.10 (Durchschnitt von $\sigma$ -Algebren)

Sei  $\Omega$  eine Menge, sei  $I$  eine Indexmenge und für jedes  $i \in I$  sei  $\mathcal{A}_i$  eine  $\sigma$ -Algebra auf  $\Omega$ . Dann ist auch

$$\bigcap_{i \in I} \mathcal{A}_i = \{A \subset \Omega \mid A \in \mathcal{A}_i \forall i \in I\}$$

eine  $\sigma$ -Algebra auf  $\Omega$ .

**Beweis:**  $\rightsquigarrow$  Übung

□



**Satz 2.11 (Kleinste  $\sigma$ -Algebra)**

Sei  $\Omega$  eine Menge und  $\mathcal{E} \subset \mathcal{P}(\Omega)$ . Dann existiert eine kleinste  $\sigma$ -Algebra  $\sigma(\mathcal{E})$  auf  $\Omega$ , die  $\mathcal{E}$  enthält; d.h.:

Es existiert ein  $\sigma(\mathcal{E}) \subset \mathcal{P}(\Omega)$ , so dass gilt:

- a)  $\mathcal{E} \subset \sigma(\mathcal{E})$
- b)  $\sigma(\mathcal{E})$  ist eine  $\sigma$ -Algebra auf  $\Omega$ .
- c) Falls  $\mathcal{C}$  eine weitere  $\sigma$ -Algebra auf  $\Omega$  ist mit  $\mathcal{E} \subset \mathcal{C}$ , so ist  $\sigma(\mathcal{E}) \subset \mathcal{C}$

**Definition 2.12**

(i)  $\sigma(\mathcal{E})$  aus Satz 2.11 heißt von  $\mathcal{E}$  erzeugte  $\sigma$ -Algebra.

(ii)  $\mathcal{E}$  heißt Erzeuger der  $\sigma$ -Algebra  $\sigma(\mathcal{E})$ .

**Beweis von Satz 2.11:** Sei

$$\Psi = \{ \mathcal{A} \subset \mathcal{P}(\Omega) \mid \mathcal{E} \subset \mathcal{A}, \mathcal{A} \text{ ist eine } \sigma\text{-Algebra} \}$$

Wegen  $\mathcal{P}(\Omega) \in \Psi$  ist  $\Psi$  nicht leer. Also existiert

$$\mathcal{D} := \bigcap_{\mathcal{A} \in \Psi} \mathcal{A}$$

und es gilt:

- $\mathcal{E} \subset \mathcal{D} \subset \mathcal{P}(\Omega)$
- $\mathcal{D}$  ist eine  $\sigma$ -Algebra auf  $\Omega$  (folgt aus Satz 2.10).
- Falls  $\mathcal{C}$  eine weitere  $\sigma$ -Algebra auf  $\Omega$  mit  $\mathcal{E} \subset \mathcal{C}$  ist, so ist  $\mathcal{C} \in \Psi$  und daher

$$\mathcal{D} = \bigcap_{\mathcal{A} \in \Psi} \mathcal{A} \subset \mathcal{C}$$

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

Der Beweis liefert eine Anschauung für „kleinste  $\sigma$ -Algebra“:  $\sigma(\mathcal{E})$  ist der Durchschnitt aller  $\sigma$ -Algebren, die  $\mathcal{E}$  enthalten.

**Beispiel 2.13**

- Sei  $\Omega = \{1; 2; \dots; 7\}$  und  $\mathcal{E} = \{\{1; 2\}; \{6\}\}$ . Dann ist

$$\sigma(\mathcal{E}) = \left\{ \emptyset; \{1; 2\}; \{3; 4; 5; 6; 7\}; \{6\}; \{1; 2; 3; 4; 5; 7\}; \{1; 2; 6\}; \{3; 4; 5; 7\}; \{1; 2; 3; 4; 5; 6; 7\} \right\}$$

- Sei  $\Omega$  eine Menge und  $A \subset \Omega$ ; sei  $\mathcal{E} = \{A\}$ . Dann ist

$$\sigma(\mathcal{E}) = \{ \emptyset; A; A^C; \Omega \}$$

- Sei  $\Omega = \mathbb{N}$  und  $\mathcal{E} = \{\{n\} \mid n \in \mathbb{N}\}$ . Dann ist

$$\sigma(\mathcal{E}) = \mathcal{P}(\mathbb{N})$$

- Sei  $\Omega = \mathbb{N}$  und  $\mathcal{F} = \{\{1, \dots, n\} \mid n \in \mathbb{N}\}$ . Dann ist

$$\sigma(\mathcal{F}) = \mathcal{P}(\mathbb{N})$$

- Sei  $\Omega$  eine Menge und  $\mathcal{A}$  eine  $\sigma$ -Algebra auf  $\Omega$ . Dann ist natürlich

$$\sigma(\mathcal{A}) = \mathcal{A}$$

# 3 Maße

## 3.1 Definitionen

Nachdem im vorherigen Kapitel die Objekte betrachtet wurden, die wir messen wollen, kommen wir nun zu den Objekten, mit denen wir messen.

**Definition 3.1** Sei  $\Omega$  eine Menge und  $\mathcal{A}$  eine  $\sigma$ -Algebra auf  $\Omega$ . Ein Maß  $\mu$  auf der  $\sigma$ -Algebra  $\mathcal{A}$  ist eine Abbildung

$$\mu : \mathcal{A} \mapsto [0; \infty]$$

so dass gilt:

(i)  $\mu(\emptyset) = 0$

(ii) Für  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$  mit  $A_n \cap A_m = \emptyset$  (falls  $n \neq m$ ) ist

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

**Notation 3.2** Statt

„Sei  $\Omega$  eine Menge,  $\mathcal{A}$  eine  $\sigma$ -Algebra auf  $\Omega$  und  $\mu$  ein Maß auf  $\mathcal{A}$ .“

sagen wir auch

„Sei  $\mu$  ein Maß auf dem Messraum  $(\Omega, \mathcal{A})$ .“

oder

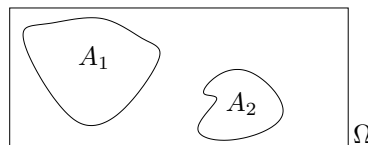
„Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum.“

**Bemerkung 3.3**

a) Als Beispiel dient wieder das Messen von Flächen:

- Es ist unmittelbar einleuchtend, dass der Flächeninhalt jeder Fläche  $A$  nicht negativ ist:  $\mu(A) \geq 0$ .
- Natürlich ist der Flächeninhalt der leeren Menge gleich null:  $\mu(\emptyset) = 0$ .  
→ (i) in Def. 3.1
- Plausibel ist auch, dass der Gesamtflächeninhalt einer Fläche die Summe der Flächeninhalte seiner Teilflächen ist:

$$A_1 \cap A_2 = \emptyset \quad \Rightarrow \quad \mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$$



→ (ii) in Def. 3.1 ist noch ein kleines bisschen mehr ( $\sigma$ -Additivität).

- b) Gemäß der Definition 3.1 ist es auch möglich, dass  $\mu(A) = \infty$  für manche  $A \in \mathcal{A}$  (z.B. Lebesgue-Maß; vgl. Abschnitt 4.3).
- c) Es gibt in der Maßtheorie noch weitere Objekte, die dem Begriff Maß ähnlich sind: Inhalt, Prämaß, signiertes Maß.

**Definition 3.4** Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum.

(i)  $\mu$  heißt endliches Maß, falls

$$\mu(\Omega) < \infty$$

(In diesem Fall gilt dann automatisch:  $\mu(A) < \infty \forall A \in \mathcal{A}$ )

(ii)  $\mu$  heißt Wahrscheinlichkeitsmaß, falls

$$\mu(\Omega) = 1$$

(iii)  $\mu$  heißt  $\sigma$ -endliches Maß, falls  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$  existiert, so dass

$$\bigcup_{n=1}^{\infty} A_n = \Omega \quad \text{und} \quad \mu(A_n) < \infty \quad \forall n \in \mathbb{N}$$

**Bemerkung 3.5**  $\sigma$ -endliche Maße verhalten sich im wesentlichen wie endliche Maße. Im Zusammenhang mit „Dichten“ spielen sie eine wichtige Rolle in der Statistik (vgl. Lebesgue-Maß in Abschnitt 4.3, Satz von Radon-Nikodym in Abschnitt 7.1).

Achtung:  $\sigma$ -endliche Maße müssen aber nicht endlich sein; d.h.  $\mu(\Omega) = \infty$  ist möglich.

Und speziell für Wahrscheinlichkeitsmaße  $\mu = P$  definieren wir:

**Definition 3.6** Sei  $(\Omega, \mathcal{A}, P)$  ein Maßraum, wobei  $P$  ein Wahrscheinlichkeitsmaß sei. Dann heißt  $(\Omega, \mathcal{A}, P)$  Wahrscheinlichkeitsraum.

## 3.2 Der Begriff „Wahrscheinlichkeitstheorie“

Falls das Maß  $\mu$  auf dem Messraum  $(\Omega, \mathcal{A})$  die Bedingung

$$\mu(\Omega) = 1$$

erfüllt, können wir z.B.

$$\mu(A) = \frac{1}{5} \tag{3.1}$$

interpretieren als

„Die Wahrscheinlichkeit, dass Ereignis  $A$  eintritt ist  $\frac{1}{5}$ .“

Nach wie vor können wir aber (3.1) genauso gut auch interpretieren als

„Die Fläche  $A$  hat Flächeninhalt  $\frac{1}{5}$ .“

Für die eigene Vorstellung sind Flächeninhalte einfacher als Wahrscheinlichkeiten; mathematisch gesehen spielt die Interpretation keine Rolle.

Eine etwas neutralere Bezeichnung für *Wahrscheinlichkeitsmaß* wäre *normiertes Maß* (wegen  $\mu(\Omega) = 1$ ). Das mathematische Gebiet *Wahrscheinlichkeitstheorie* könnte genauso gut auch *Theorie normierter Maße* heißen.

### 3.3 Eigenschaften von Maßen

#### Satz 3.7 (Elementare Eigenschaften von Maßen)

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

a) Endliche Additivität:

Für ein  $n \in \mathbb{N}$  seien  $A_1 \in \mathcal{A}, \dots, A_n \in \mathcal{A}$  so dass  $A_i \cap A_j = \emptyset$  (für  $i \neq j$ ).

Dann ist

$$\mu(A_1 \cup \dots \cup A_n) = \mu(A_1) + \dots + \mu(A_n)$$

b) Isotonie:

$$A \in \mathcal{A}, B \in \mathcal{A}, \quad A \subset B \quad \Rightarrow \quad \mu(A) \leq \mu(B)$$

c) Subtraktivität:

$$A \in \mathcal{A}, B \in \mathcal{A}, \quad A \subset B, \quad \mu(A) < \infty \quad \Rightarrow \quad \mu(B \setminus A) = \mu(B) - \mu(A)$$

d) Sub-Additivität:

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{A} \quad \Rightarrow \quad \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

e) Stetigkeit von unten:

Sei

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}; \quad A_1 \subset A_2 \subset A_3 \subset \dots \subset \bigcup_{n=1}^{\infty} A_n = A \in \mathcal{A}$$

eine isotone Folge messbarer Mengen mit messbarem Grenzwert. Dann gilt

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

f) Stetigkeit von oben:

Sei

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}; \quad A_1 \supset A_2 \supset A_3 \supset \dots \supset \bigcap_{n=1}^{\infty} A_n = A \in \mathcal{A}$$

eine antitone Folge messbarer Mengen mit messbarem Grenzwert, wobei  $\mu(A_1) < \infty$ . Dann gilt

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

**Beweis:** a) – d) und f)  $\rightsquigarrow$  Übungen

Beweis von e):

Setze  $B_1 := A_1$  und  $B_n := A_n \setminus A_{n-1} \quad \forall n \geq 2$ . Dann ist

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n, \quad \bigcup_{n=1}^m B_n = \bigcup_{n=1}^m A_n = A_m \quad (3.2)$$

und

$$B_n \cap B_m = \emptyset \quad \forall n \neq m \quad (3.3)$$

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \stackrel{(3.2)}{=} \mu\left(\bigcup_{n=1}^{\infty} B_n\right) \stackrel{(3.3)}{=} \sum_{n=1}^{\infty} \mu(B_n) = \\ &= \lim_{m \rightarrow \infty} \sum_{n=1}^m \mu(B_n) = \lim_{m \rightarrow \infty} \mu\left(\bigcup_{n=1}^m B_n\right) \stackrel{(3.2)}{=} \lim_{m \rightarrow \infty} \mu(A_m) = \\ &= \lim_{n \rightarrow \infty} \mu(A_n) \end{aligned}$$

□

**Beobachtung:**

Sei  $(\Omega, \mathcal{A})$  ein Messraum und  $\mu$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$ . Spätestens aus Satz 3.7 wird folgendes deutlich:

Durch die Angabe des Wertes  $\mu(A) = 0,7$  für ein  $A \in \mathcal{A}$ , weiß man schon viel mehr über  $\mu$  als nur diesen einen Wert  $\mu(A) = 0,7$ . Zum Beispiel folgt daraus schon

$$\mu(B) \leq 0,7 \quad \forall B \subset A$$

und

$$\mu(C) \leq 0,3 \quad \forall C \in \mathcal{A} \text{ mit } A \cap C = \emptyset$$

und auch

$$\mu(A^C) = 0,3$$

Aus jeden weiteren angegebenen Wert von  $\mu$ , könnten wir eine noch viel größere Anzahl von zusätzlichen Werten ableiten. Somit leuchtet auch sofort ein, dass Folgendes möglich ist:

Seien  $\mu$  und  $\nu$  zwei Maße auf  $(\Omega, \mathcal{A})$  und  $\mathcal{F} \subset \mathcal{A}$ . Unter Umständen kann dann aus

$$\mu(F) = \nu(F) \quad \forall F \in \mathcal{F}$$

schon folgen, dass

$$\mu(A) = \nu(A) \quad \forall A \in \mathcal{A}$$

Der folgende sog. *Eindeutigkeitssatz* beschreibt Situationen, in denen dies tatsächlich der Fall ist. Es sollte auch nicht weiter überraschen, dass die Rolle von  $\mathcal{F}$  ein Erzeuger von  $\mathcal{A}$  einnimmt. Allerdings sind nicht alle Erzeuger von  $\mathcal{A}$  hierzu ausreichend – aber sog.  $\cap$ -stabile Erzeuger sind es.

**Definition 3.8** Sei  $(\Omega, \mathcal{A})$  ein Messraum. Ein  $\mathcal{E} \subset \mathcal{A}$  heißt  $\cap$ -stabiler Erzeuger von  $\mathcal{A}$ , falls gilt

(i)  $\mathcal{E}$  ist ein Erzeuger von  $\mathcal{A}$ ; d.h.  $\sigma(\mathcal{E}) = \mathcal{A}$  (vgl. Def. 2.12).

(ii)  $\mathcal{E}$  ist  $\cap$ -stabil, d.h.

$$E_1 \in \mathcal{E}, E_2 \in \mathcal{E} \quad \Rightarrow \quad E_1 \cap E_2 \in \mathcal{E}$$

**Satz 3.9 (Eindeutigkeitssatz)**

Sei  $(\Omega, \mathcal{A})$  ein Messraum und  $\mathcal{E}$  ein  $\cap$ -stabiler Erzeuger von  $\mathcal{A}$ .

a) Seien  $\mu$  und  $\nu$  endliche Maße auf  $(\Omega, \mathcal{A})$ , so dass

$$\mu(\Omega) = \nu(\Omega) \quad \text{und} \quad \mu(E) = \nu(E) \quad \forall E \in \mathcal{E}$$

Dann ist  $\mu = \nu$

b) Seien  $\mu$  und  $\nu$  Maße auf  $(\Omega, \mathcal{A})$  und sei

$$(E_n)_{n \in \mathbb{N}} \subset \mathcal{E} \quad \text{mit} \quad \bigcup_{n=1}^{\infty} E_n = \Omega$$

Falls außerdem

$$\mu(E_n) = \nu(E_n) < \infty \quad \forall n \in \mathbb{N} \quad \text{und} \quad \mu(E) = \nu(E) \quad \forall E \in \mathcal{E}$$

dann ist  $\mu = \nu$

## 4 Wichtige Maße

### 4.1 Das Dirac-Maß

Sei  $(\Omega, \mathcal{A})$  ein Messraum und sei  $x \in \Omega$ . Setze

$$\delta_x(A) = 1 \quad \text{für jedes } A \in \mathcal{A} \text{ mit } x \in A$$

und

$$\delta_x(A) = 0 \quad \text{für jedes } A \in \mathcal{A} \text{ mit } x \notin A$$

Dann ist

$$\delta_x : \mathcal{A} \rightarrow [0; \infty], \quad A \mapsto \delta_x(A)$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$ .  $\delta_x$  heißt Dirac-Maß an der Stelle  $x$ . Das Dirac-Maß  $\delta_x$  mißt also, ob  $x$  jeweils in  $A$  liegt.

### 4.2 Das Zählmaß

#### 4.2.1 Endlicher Fall

Sei  $\Omega = \{1; 2; 3; \dots; n\}$  und  $\mathcal{A} = \mathcal{P}(\Omega)$ .

Für  $A \subset \{1; 2; 3; \dots; n\}$  bezeichne  $\sharp(A)$  die Anzahl der Elemente in  $A$ . Zum Beispiel

$$\sharp(\{1; 2; 7; 13\}) = 4; \quad \sharp(\Omega) = n$$

Einfache Rechnungen zeigen, dass

$$\sharp : A \mapsto \sharp(A)$$

ein Maß auf  $(\{1; \dots; n\}, \mathcal{P}(\{1; \dots; n\}))$  ist.  $\sharp$  heißt Zählmaß.

#### 4.2.2 Abzählbar unendlicher Fall

Sei  $\Omega = \mathbb{N}$  und  $\mathcal{A} = \mathcal{P}(\mathbb{N})$ .

Für  $A \subset \mathbb{N}$  bezeichne  $\sharp(A)$  wieder die Anzahl der Elemente in  $A$ . Zum Beispiel

$$\sharp(\{1; 2; 7; 13\}) = 4; \quad \sharp(\Omega) = \sharp(\mathbb{N}) = \infty$$

Sei  $C$  die Menge der Primzahlen in  $\mathbb{N}$ . Dann ist

$$\sharp(C) = \infty \quad (\text{Euklid; ca. 300 v.Chr.})$$

$\sharp$  ist ein  $\sigma$ -endliches Maß auf  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ .

#### 4.2.3 Überabzählbar unendlicher Fall

Sei  $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{P}(\mathbb{R}))$  und  $\sharp(A)$  bezeichne wieder die Menge der Elemente von  $A \subset \mathbb{R}$ .  $\sharp(A)$  ist ein Maß auf  $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ . Ist  $\sharp$  hier ein  $\sigma$ -endliches Maß?

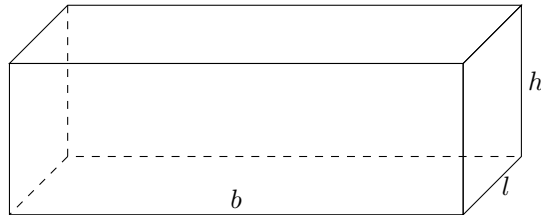


## 4.3 Das Lebesgue-Maß

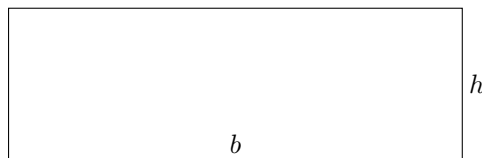
### 4.3.1 Ziel

Wir wollen jetzt ein Maß definieren, dass uns wie gewohnt

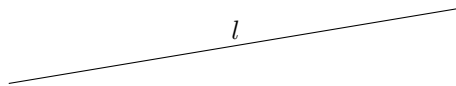
- im  $\mathbb{R}^3$  Volumen mißt: Der Quader soll den Messwert  $b l h$  haben.



- im  $\mathbb{R}^2$  Flächeninhalte mißt: Das Rechteck soll den Messwert  $b h$  haben.



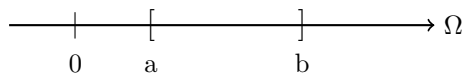
- im  $\mathbb{R}^1$  Streckenlängen mißt: Die Strecke soll den Messwert  $l$  haben.



Zunächst betrachten wir den eindimensionalen Fall, also  $\Omega = \mathbb{R}$ .

### 4.3.2 Das eindimensionale Lebesgue-Maß $\lambda$ auf $(\mathbb{R}, \mathfrak{B})$

Betrachte ein Intervall  $[a; b] \subset \mathbb{R} = \Omega$



Das gesuchte Maß  $\lambda$  soll dann erfüllen

$$\lambda([a; b]) = b - a$$

Natürlich können wir das Maß  $\lambda$  nicht nur auf Intervalle definieren – wir brauchen jetzt eine geeignete  $\sigma$ -Algebra auf  $\Omega = \mathbb{R}$ .

**Einfachste Idee:** Wir nehmen die kleinste  $\sigma$ -Algebra auf  $\mathbb{R}$ , die alle Intervalle enthält. Der Erzeuger unserer  $\sigma$ -Algebra ist also

$$\mathcal{E} = \left\{ [a; b] \mid a \in \mathbb{R}, b \in \mathbb{R}, a < b \right\}$$

(vgl. Def. 2.12). Die daraus resultierende  $\sigma$ -Algebra

$$\mathfrak{B} := \sigma(\mathcal{E})$$

heißt Borel- $\sigma$ -Algebra.

**Satz 4.1 (Eigenschaften von  $\mathfrak{B}$ )**

a)  $\emptyset \in \mathfrak{B}, \mathbb{R} \in \mathfrak{B}$

b)  $\{c\} \in \mathfrak{B} \quad \forall c \in \mathbb{R}$

c) Für alle  $a \in \mathbb{R}, b \in \mathbb{R}$  mit  $a < b$  ist

$$[a, b] \in \mathfrak{B} \quad [a, b) \in \mathfrak{B} \quad (a, b] \in \mathfrak{B} \quad (a, b) \in \mathfrak{B}$$

d)  $\mathbb{N} \in \mathfrak{B}, \mathbb{Q} \in \mathfrak{B}, \mathbb{Q}^C \in \mathfrak{B}$

Natürlich sind auch jeweils die Komplemente, die (abzählbaren) Vereinigungen und die (abzählbaren) Durchschnitte all dieser Mengen in  $\mathfrak{B}$ .

**Beweis:**  $\rightsquigarrow$  Übung □

**Bemerkung 4.2** Andere Erzeuger der Borelschen sigma-Algebra sind zum Beispiel die Menge aller offenen Intervalle  $(a, b)$ , die Menge aller halboffenen Intervalle  $(a, b]$  (bzw.  $[a, b)$ ) sowie die Menge aller einseitig unbeschränkten Intervalle  $(a, \infty)$  (bzw.  $[a, \infty), (-\infty, b), (-\infty, b]$ ).

**Satz 4.3** Alle abgeschlossenen Teilmengen von  $\mathbb{R}$  und auch alle offenen Teilmengen von  $\mathbb{R}$  sind in  $\mathfrak{B}$ .

**Definition 4.4** Eine Menge  $B \subset \mathbb{R}$  heißt Borel-Menge, falls

$$B \in \mathfrak{B}$$

**Bemerkung 4.5** Nicht jede Teilmenge von  $\mathbb{R}$  ist eine Borel-Menge!

**Aber:** Jede Teilmenge von  $\mathbb{R}$ , die Ihr bisher in Eurem Leben gesehen habt, ist eine Borel-Menge.

Es ist aufwendig und mathematisch anspruchsvoll, Mengen zu konstruieren, die nicht in  $\mathfrak{B}$  liegen.

**Dennoch** muss man schon immer erst zeigen, dass die gerade betrachtete Menge tatsächlich eine Borel-Menge ist. (Das gilt natürlich nicht für die praktische Arbeit als Statistiker.)

**Satz 4.6 (Das Lebesgue-Maß)**

Es gibt genau ein Maß  $\lambda$  auf dem Messraum  $(\mathbb{R}, \mathfrak{B})$ , so dass

$$\lambda([a; b]) = b - a \quad \forall a \in \mathbb{R}, b \in \mathbb{R}, a < b \quad (4.1)$$

Dieses  $\lambda$  heißt (eindimensionales) Lebesgue-Maß.

**Zum Beweis:** Der Beweis, dass  $\lambda$  tatsächlich existiert ist sehr aufwendig. Dabei verwendete Begriffe sind: Ring, Prämaß, Fortsetzungssatz (Carathéodory), äußeres Maß.

Wir zeigen jetzt aber noch die Eindeutigkeit. Das heißt, wir zeigen, dass es auf  $(\mathbb{R}, \mathfrak{B})$  kein weiteres Maß geben kann, das (4.1) erfüllt.

Nehmen wir dazu an, dass neben  $\lambda$  auch  $\mu$  ein Maß auf  $(\mathbb{R}, \mathfrak{B})$  sei, dass (4.1) erfüllt. Dann gilt

$$\lambda([a; b]) = b - a = \mu([a; b]) \quad \forall a \in \mathbb{R}, b \in \mathbb{R}, a < b$$

Sei

$$\bar{\mathcal{E}} = \left\{ [a; b] \mid a \in \mathbb{R}, b \in \mathbb{R}, a < b \right\} \cup \{ \emptyset \} \cup \left\{ \{c\} \mid c \in \mathbb{R} \right\}$$

Für

$$\lambda(E) = \mu(E) \quad \forall E \in \bar{\mathcal{E}} \quad (4.2)$$

ist nur noch zu zeigen, dass  $\lambda(\{c\}) = \mu(\{c\}) \quad \forall c \in \mathbb{R}$ :

Für ein  $c \in \mathbb{R}$  sei  $\lambda(\{c\}) = p > 0$ . Dann ist

$$p = \lambda(\{c\}) \leq \lambda\left([c; c + \frac{1}{2}p]\right) = \frac{1}{2}p$$

und daraus folgt  $p \leq 0$ ; ein WIDERSPRUCH.

Somit folgt  $\lambda(\{c\}) = 0$ . Analog folgt auch  $\mu(\{c\}) = 0$ .

Insbesondere gilt für  $E_n := [-n; n] \quad n \in \mathbb{N}$

$$\bigcup_{n=1}^{\infty} E_n = \mathbb{R} \quad \text{und} \quad \lambda(E_n) = \mu(E_n) = 2n < \infty \quad \forall n \in \mathbb{N} \quad (4.3)$$

Gemäß der Definition von  $\mathfrak{B}$  ist  $\bar{\mathcal{E}}$  ein Erzeuger von  $\mathfrak{B}$ . Weil der Durchschnitt zweier Intervalle wieder ein Intervall, ein Punkt oder die leere Menge ist, ist  $\bar{\mathcal{E}}$  sogar ein  $\cap$ -stabiler Erzeuger von  $\mathfrak{B}$  (vgl. Def. 3.8).

Aus (4.2), (4.3) und dem Eindeutigkeitssatz 3.9 folgt schließlich  $\lambda = \mu$ .  $\square$

#### Satz 4.7 (Translationsinvarianz)

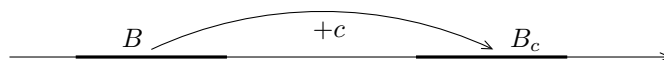
Sei  $B \in \mathfrak{B}$ ,  $c \in \mathbb{R}$  und

$$B_c := c + B = \{c + b \mid b \in B\}$$

Dann ist

$$\lambda(B_c) = \lambda(B)$$

D.h.: Durch Verschieben verändert sich die Streckenlänge nicht.



**Beweis:**  $\rightsquigarrow$  Übung  $\square$

### 4.3.3 Das n-dimensionale Lebesgue-Maß $\lambda^n$ auf $(\mathbb{R}^n, \mathfrak{B}^{\otimes n})$

Sei  $n \in \mathbb{N}$  fest. Sei

$$\mathcal{E}_n = \left\{ [a_1; b_1] \times \cdots \times [a_n; b_n] \mid a_i \in \mathbb{R}, b_i \in \mathbb{R}, a_i < b_i \forall i \in \{1; \dots; n\} \right\}$$

Dabei ist

$$[a_1; b_1] \times \cdots \times [a_n; b_n] = \{(c_1, \dots, c_n) \mid c_i \in [a_i; b_i] \forall i \in \{1; \dots; n\}\}$$

Sei

$$\mathfrak{B}^{\otimes n} := \sigma(\mathcal{E}_n)$$

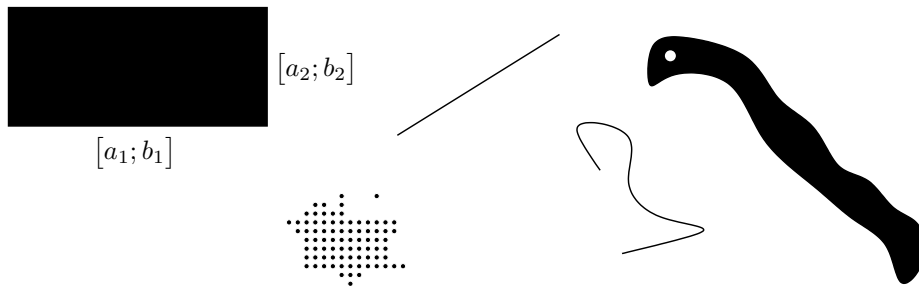
die von  $\mathcal{E}_n$  erzeugte  $\sigma$ -Algebra (vgl. Def. 2.12).  $\mathfrak{B}^{\otimes n}$  heißt (*n*-dimensionale) **Borel- $\sigma$ -Algebra**.

**Satz 4.8** *Alle abgeschlossenen Teilmengen von  $\mathbb{R}^n$  und auch alle offenen Teilmengen von  $\mathbb{R}^n$  sind in  $\mathfrak{B}^{\otimes n}$ .*

**Definition 4.9** *Eine Menge  $B \subset \mathbb{R}^n$  heißt Borel-Menge, falls*

$$B \in \mathfrak{B}^{\otimes n}$$

Die Aussagen aus Bemerkung 4.5 gelten auch für  $(\mathbb{R}^n, \mathfrak{B}^{\otimes n})$ .  
Beispiele für Borel-Mengen in  $\mathfrak{B}^{\otimes 2}$  sind:



#### **Satz 4.10 (Das n-dimensionale Lebesgue-Maß)**

*Es gibt genau ein Maß  $\lambda^n$  auf dem Messraum  $(\mathbb{R}^n, \mathfrak{B}^{\otimes n})$ , so dass für alle*

$$a_i \in \mathbb{R}, b_i \in \mathbb{R}, a_i < b_i, i \in \{1; \dots; n\}$$

*gilt*

$$\lambda^n([a_1; b_1] \times \cdots \times [a_n; b_n]) = (b_1 - a_1) \cdot (b_2 - a_2) \cdot \dots \cdot (b_n - a_n) \quad (4.4)$$

*Dieses  $\lambda^n$  heißt (*n*-dimensionales) Lebesgue-Maß.*

**Bemerkung 4.11** *Das Lebesguemaß  $\lambda_n$  auf  $(\mathbb{R}^n, \mathfrak{B}^{\otimes n})$  ist  $\sigma$ -endlich (und nicht endlich).*

**Satz 4.12 (Translationsinvarianz)**

Sei  $B \in \mathfrak{B}^{\otimes n}$ ,  $c \in \mathbb{R}^n$  und

$$B_c := c + B = \{c + b \mid b \in B\}$$

Dann ist

$$\lambda(B_c) = \lambda(B)$$

D.h.: Durch Verschieben verändert sich nicht der Rauminhalt.

**Beweis:** Geht genau wie im eindimensionalen Fall.  $\square$

**Bemerkung 4.13** Für das Lebesgue-Maß  $\lambda$  aus Abschnitt 4.3.2 und für  $n = 1$  gilt natürlich

$$\lambda^1 = \lambda \quad \text{und} \quad (\mathbb{R}^1, \mathfrak{B}^{\otimes 1}) = (\mathbb{R}, \mathfrak{B})$$

**4.3.4  $\lambda$  kann nicht auf ganz  $\mathcal{P}(\mathbb{R})$  sinnvoll definiert werden!**

Dieser Abschnitt soll die Frage aus Kapitel 2 (Bemerkung 2.6 (b)) beantworten. Wieso brauchen wir eigentlich  $\sigma$ -Algebren und betrachten nicht immer gleich die ganze Potenzmenge  $\mathcal{P}(\Omega)$ ?

Das Lebesgue-Maß soll (vgl. Abschnitt 4.3.1) für eine Teilmenge  $B \subset \mathbb{R}$  die Länge angeben. Zum Beispiel

$$\lambda([2; 5] \cup [7; 20]) = (5 - 2) + (20 - 7) = 16$$

Durch verschieben einer Menge  $B$  sollte sich natürlich nicht die Länge von  $B$  verändern (Translationsinvariants, vgl. Satz 4.7).

Bisher haben wir  $\lambda$  nicht für alle Teilmengen  $A \subset \mathbb{R}$  von  $\mathbb{R}$  definiert, sondern nur für Borel-Mengen – d.h. wir haben  $\lambda$  nur auf der  $\sigma$ -Algebra  $\mathfrak{B}$  definiert. Im Folgenden wird gezeigt: Egal wie wir  $\lambda(A)$  für Mengen  $A \subset \mathbb{R}$ ,  $A \notin \mathfrak{B}$  definieren würden,  $\lambda$  wäre dann nicht mehr translationsinvariant. Das heißt: Durch Verschieben von manchen Mengen  $A$  würde sich ihre „Länge“ verändern!

Wir können also  $\lambda$  nicht sinnvoll für alle  $A \subset \mathbb{R}$  definieren! Daher brauchen wir eine geeignete  $\sigma$ -Algebra (nämlich  $\mathfrak{B}$ ).

**Satz 4.14** Es gibt kein Maß  $\tilde{\lambda}$  auf  $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ , so dass

$$a) \quad \tilde{\lambda}([a; b]) = b - a \quad \forall a \in \mathbb{R}, b \in \mathbb{R}, a < b$$

$$b) \quad \tilde{\lambda}(A_c) = \tilde{\lambda}(A) \quad \forall A \subset \mathbb{R}, \forall c \in \mathbb{R}$$

$$\text{wobei jeweils } A_c = c + A = \{c + a \mid a \in A\}$$

**Beweis:** Angenommen es gäbe so ein Maß  $\tilde{\lambda}$ .

Sei  $A \subset [0; 1]$ , so dass gilt:

$$\forall x \in [0, 1] \quad \exists! y \in A, \quad \text{so dass } x - y \in \mathbb{Q}$$

(So eine Menge  $A$  existiert nach dem Auswahlaxiom.)

Setze

$$B := \bigcup_{r \in [-1; 1] \cap \mathbb{Q}} A_r$$

wobei  $A_r = r + A = \{r + a \mid a \in A\}$ . Dann gilt:

1)  $B$  ist die Vereinigung abzählbar vieler Mengen  $A_r$ , wobei

$$A_r \cap A_{r'} = \emptyset \quad \text{für } r \neq r'$$

2)  $[0; 1] \subset B \subset [-1; 2]$  also

$$1 = \tilde{\lambda}([0; 1]) \leq \tilde{\lambda}(B) \leq \tilde{\lambda}([-1; 2]) = 3 \quad (4.5)$$

Außerdem gilt

$$\tilde{\lambda}(B) \stackrel{1)}{=} \sum_{r \in [-1; 1] \cap \mathbb{Q}} \tilde{\lambda}(A_r) = \sum_{r \in [-1; 1] \cap \mathbb{Q}} \tilde{\lambda}(A) \quad (4.6)$$

Aus (4.6) folgt

$$\tilde{\lambda}(B) = \begin{cases} \infty & \text{falls } \tilde{\lambda}(A) > 0 \\ 0 & \text{falls } \tilde{\lambda}(A) = 0 \end{cases}$$

Die ist jedoch ein Widerspruch zu (4.5).

Also: Es kann kein solches  $\tilde{\lambda}$  geben! □

**Bemerkung 4.15** *Die Situation ist in der Tat sogar noch viel schlimmer:*

„[Der erste Gödelsche Unvollständigkeitssatz] besagt, dass in einem widerspruchsfreien Axiomensystem, das genügend reichhaltig ist, um den üblichen Aufbau der Arithmetik (natürliche Zahlen) sicherzustellen und überdies 'hinreichend einfach' ist, es immer Aussagen gibt, die aus diesem weder bewiesen noch widerlegt werden können.“

*Aus dem Wikipedia-Artikel zu Kurt Gödel*

*Man weiß, dass folgende Aussage weder bewiesen noch widerlegt werden kann:*

Man kann das Lebesgue-Maß  $\lambda$  zu einem (nicht notwendigerweise translationsinvarianten) Maß auf  $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$  fortsetzen.

Vgl. Hoffmann-Jørgensen (1994b), S. 513

# 5 Messbare Abbildungen und Bildmaße

## 5.1 Einführende Beispiele

### Ein Beispiel aus der Stochastik:

Zweifacher Würfelwurf:

$$\Omega = \{1; \dots; 6\} \times \{1; \dots; 6\}$$

Schreibe  $(\omega_1, \omega_2) = \omega \in \Omega$ .

$\omega_1$ : 1. Wurf

$\omega_2$ : 2. Wurf

Die Angabe der Wahrscheinlichkeiten modellieren wir mit folgendem (Wahrscheinlichkeits-) Maß  $P$ :

$$P(\{(\omega_1, \omega_2)\}) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

(D.h.:  $P = \frac{1}{36} \cdot \#$ , vgl. Abschnitt 4.2.1.)

Uns interessiert aber nicht  $(\omega_1, \omega_2)$ , sondern nur die Summe aus beiden Würfeln:

$$S((\omega_1, \omega_2)) = \omega_1 + \omega_2$$

Das modellieren wir durch die Abbildung

$$S : \Omega \rightarrow \Omega', \quad (\omega_1, \omega_2) \mapsto \omega_1 + \omega_2$$

wobei  $\Omega' = \{2; \dots; 12\}$ .

Wir haben also neben  $(\Omega, \mathcal{P}(\Omega))$  nun noch einen zweiten Messraum  $(\Omega', \mathcal{P}(\Omega'))$ . Die Angabe der Wahrscheinlichkeiten für Elemente  $\omega'$  aus  $\Omega'$  modellieren wir mit dem (Wahrscheinlichkeits-) Maß  $P'$ , dass gegeben ist durch

$$P'(\{\omega'\}) = P(\{\omega \in \Omega \mid S(\omega) = \omega'\})$$

Zum Beispiel

$$P'(\{4\}) = P(\{(1; 3); (3; 1); (2; 2)\}) = \frac{1}{36} \cdot \#(\{(1; 3); (3; 1); (2; 2)\}) = \frac{3}{36}$$

Allgemein ist für  $A' \subset \Omega'$  dann

$$P'(A') = P(\{\omega \in \Omega \mid S(\omega) \in A'\}) = P(S^{-1}(A'))$$

wobei  $S^{-1}(A') = \{\omega \in \Omega \mid S(\omega) \in A'\}$  das Urbild von  $A'$  unter  $S$  bezeichnet. Auf diese Weise erzeugt  $S$  aus dem Maß  $P$  ein neues Maß

$$P' : \mathcal{P}(\Omega') \rightarrow [0; \infty], \quad A' \mapsto P(S^{-1}(A'))$$

Man schreibt auch

$$P' = S(P) \quad \text{bzw.} \quad P'(A') = [S(P)](A')$$

und nennt  $P' = S(P)$  auch Bildmaß von  $P$  unter  $S$ .

**Ein weiteres Beispiel:**

Sei jetzt  $\Omega = \mathbb{R}$ ,  $\Omega' = \{0; 1\}$  und  $T : \mathbb{R} \rightarrow \{0; 1\}$  irgendeine Abbildung. Betrachten wir nun das Lebesgue-Maß  $\lambda$  auf  $(\mathbb{R}, \mathfrak{B})$ . Wie vorher soll auf  $(\Omega', \mathcal{P}(\Omega'))$  das Bildmaß von  $\lambda$  unter  $T$  definiert werden – und zwar durch

$$\mathcal{P}(\Omega') \rightarrow [0; \infty], \quad A' \mapsto \lambda(T^{-1}(A'))$$

**Problem:** Es kann passieren, dass

$$T^{-1}(A') \notin \mathfrak{B}$$

In diesem Fall ist  $\lambda(T^{-1}(A'))$  aber gar nicht definiert.

**Daher** können nur Abbildungen  $T$  verwendet werden, bei denen dieses Problem nicht auftritt. Dies führt uns zur Definition der sog. *messbaren* Abbildungen.

## 5.2 Definitionen und Eigenschaften

**Definition 5.1** Seien  $(\Omega, \mathcal{A})$  und  $(\Omega', \mathcal{A}')$  Messräume. Eine Abbildung

$$T : \Omega \rightarrow \Omega'$$

heißt  $\mathcal{A}/\mathcal{A}'$ -messbar, falls

$$T^{-1}(A') \in \mathcal{A} \quad \forall A' \in \mathcal{A}'$$

**Notation 5.2** Die Kurzschreibweise

$$T : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$$

bedeutet:

$T$  ist eine Abbildung

$$T : \Omega \rightarrow \Omega',$$

die  $\mathcal{A}/\mathcal{A}'$ -messbar ist.

**Definition 5.3** Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum,  $(\Omega', \mathcal{A}')$  ein Messraum und

$$T : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$$

Das durch

$$\mu'(A') = \mu(T^{-1}(A')) \quad \forall A' \in \mathcal{A}' \tag{5.1}$$

definierte Maß  $\mu'$  auf  $(\Omega', \mathcal{A}')$  heißt Bildmaß von  $\mu$  unter  $T$ .



**Notation 5.4** Für das Bildmaß  $\mu'$  von  $\mu$  unter  $T$  schreiben wir auch

$$\mu' = T(\mu)$$

Das heißt:  $[T(\mu)](A') = \mu(T^{-1}(A'))$

**Satz 5.5** Durch Gleichung (5.1) wird tatsächlich ein Maß  $\mu'$  auf  $(\Omega', \mathcal{A}')$  definiert. Außerdem gilt:

a)  $\mu$  ist endliches Maß  $\Rightarrow T(\mu)$  ist endliches Maß

b)  $\mu$  ist Wahrscheinlichkeitsmaß  $\Rightarrow T(\mu)$  ist Wahrscheinlichkeitsmaß

**Beweis:**  $\rightsquigarrow$  Übung! □

**Bemerkung 5.6** Die Aussage

$$\mu \text{ ist } \sigma\text{-endliches Maß} \Rightarrow T(\mu) \text{ ist } \sigma\text{-endliches Maß}$$

ist im Allgemeinen falsch.

Gegenbeispiel zu dieser Aussage:

Sei  $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathfrak{B})$  und  $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathfrak{B})$ . Betrachte  $T : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$T(x) = 1 \quad \forall x \in \mathbb{R}$$

$T$  ist  $\mathfrak{B}/\mathfrak{B}$ -messbar! Das Lebesgue-Maß  $\lambda$  auf  $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathfrak{B})$  ist  $\sigma$ -endlich, aber das Bildmaß  $T(\lambda)$  ist nicht  $\sigma$ -endlich:

Für alle  $B' \in \mathcal{A}' = \mathfrak{B}$  ist

$$[T(\lambda)](B') = \begin{cases} 0 & \text{falls } 1 \notin B' \\ \infty & \text{falls } 1 \in B' \end{cases}$$

und so ein Maß kann nicht  $\sigma$ -endlich sein, wie man leicht sieht.

**Satz 5.7 (Kriterium für Messbarkeit)**

Seien  $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$  Messräume und

$$T : \Omega \rightarrow \Omega'$$

eine Abbildung. Sei  $\mathcal{E}'$  ein Erzeuger von  $\mathcal{A}'$ , also  $\sigma(\mathcal{E}') = \mathcal{A}'$ . Es gilt:

$T$  ist genau dann  $\mathcal{A}/\mathcal{A}'$ -messbar, wenn

$$T^{-1}(E') \in \mathcal{A} \quad \forall E' \in \mathcal{E}'$$

d.h. wenn  $T$   $\mathcal{A}/\mathcal{E}'$ -messbar ist.

**Satz 5.8 (Verkettung von messbaren Abbildungen)**

Seien  $(\Omega_1, \mathcal{A}_1)$ ,  $(\Omega_2, \mathcal{A}_2)$  und  $(\Omega_3, \mathcal{A}_3)$  drei Messräume. Für

$$T_1 : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$$

und

$$T_2 : (\Omega_2, \mathcal{A}_2) \rightarrow (\Omega_3, \mathcal{A}_3)$$

ist auch

$$T_2 \circ T_1 : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_3, \mathcal{A}_3)$$

Ist außerdem  $\mu_1$  ein Maß auf  $(\Omega_1, \mathcal{A}_1)$ , so gilt

$$(T_2 \circ T_1)(\mu_1) = T_2(T_1(\mu_1))$$

**Beweis:** Folgt unmittelbar aus den Definitionen! □

**Satz 5.9** Sei  $(\Omega, \mathcal{A})$  ein Messraum und  $(\Omega', \mathcal{A}') = (\mathbb{R}^k, \mathfrak{B}^{\otimes k})$ .

a) Seien  $f : \Omega \rightarrow \mathbb{R}^k$  und  $g : \Omega \rightarrow \mathbb{R}^k$   $\mathcal{A}/\mathfrak{B}^{\otimes k}$ -messbare Funktionen. Sei  $c \in \mathbb{R}$ . Dann sind auch

$$f + g, \quad f - g, \quad f \cdot g, \quad c \cdot f$$

$\mathcal{A}/\mathfrak{B}^{\otimes k}$ -messbare Funktionen. Ist außerdem  $g > 0$ , so ist auch

$$\frac{f}{g}$$

eine  $\mathcal{A}/\mathfrak{B}^{\otimes k}$ -messbare Funktion.

b) Sei nun  $(\Omega, \mathcal{A}) = (\mathbb{R}^j, \mathfrak{B}^{\otimes j})$ . Es gilt: Jede stetige Funktion

$$f : \mathbb{R}^j \rightarrow \mathbb{R}^k$$

ist  $\mathfrak{B}^{\otimes j}/\mathfrak{B}^{\otimes k}$ -messbar

## 5.3 Reellwertige Funktionen

In diesem Abschnitt ist  $(\Omega, \mathcal{A})$  ein Messraum und  $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathfrak{B})$ . Wir betrachten Funktionen

$$f : \Omega \rightarrow \mathbb{R}$$

**Notation 5.10** Anstelle von

$f$  ist  $\mathcal{A}/\mathfrak{B}$ -messbar

schreiben wir auch einfach nur

$f$  ist  $\mathcal{A}$ -messbar

Diese Funktionen spielen eine besondere Rolle; im nächsten Kapitel werden wir ein Integral auf solche Funktionen definieren. (Als Anwendung werden wir schließlich den Erwartungswert von Zufallsvariablen in der Stochastik erhalten.)

Zunächst notieren wir noch ein paar Aussagen zur Messbarkeit:

**Satz 5.11** Sei  $(\Omega, \mathcal{A})$  ein Messraum und  $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathfrak{B})$ .

a) Sei  $f : \Omega \rightarrow \mathbb{R}$ , so dass

$$f^{-1}((a, \infty)) \in \mathcal{A} \quad \forall a \in \mathbb{R}$$

Dann ist  $f$   $\mathcal{A}$ -messbar.

Die analoge Aussage gilt auch für  $[a; \infty)$  anstelle von  $(a; \infty)$ .

b) Sei  $(f_n)_{n \in \mathbb{N}}$  eine Folge  $\mathcal{A}$ -messbarer Funktionen

$$f_n : \Omega \rightarrow \mathbb{R}$$

so dass

$$-\infty < \inf_{n \in \mathbb{N}} f_n(\omega) \leq \sup_{n \in \mathbb{N}} f_n(\omega) < \infty \quad \forall \omega \in \Omega$$

Dann sind die Funktionen

$$\omega \mapsto \inf_{n \in \mathbb{N}} f_n(\omega) \quad \text{und} \quad \omega \mapsto \sup_{n \in \mathbb{N}} f_n(\omega)$$

$\mathcal{A}$ -messbar.

Falls  $(f_n)_{n \in \mathbb{N}}$  außerdem punktweise gegen ein

$$f : \Omega \rightarrow \mathbb{R}$$

konvergiert (d.h.  $\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega) \quad \forall \omega \in \Omega$ ), so ist auch  $f$  eine  $\mathcal{A}$ -messbare Funktion.

c) Seien  $f : \Omega \rightarrow \mathbb{R}$  und  $g : \Omega \rightarrow \mathbb{R}$   $\mathcal{A}$ -messbare Funktionen. Dann sind auch

$$\omega \mapsto \min\{f(\omega); g(\omega)\} \quad \text{und} \quad \omega \mapsto \max\{f(\omega); g(\omega)\}$$

$\mathcal{A}$ -messbare Funktionen.

**Beweisskizze:**

a) Folgt aus Satz 5.7, weil

$$\mathcal{F}' = \left\{ (a; \infty) \mid a \in \mathbb{R} \right\}$$

ein Erzeuger von  $\mathfrak{B}$  ist.

Beweis für  $[a; \infty)$  anstelle von  $(a; \infty)$  geht analog.

b) Messbarkeit von  $\inf_{n \in \mathbb{N}} f_n$  folgt aus

$$\begin{aligned} \left( \inf_{n \in \mathbb{N}} f_n \right)^{-1}([a; \infty)) &= \{ \omega \in \Omega \mid \inf_{n \in \mathbb{N}} f_n(\omega) \in [a; \infty) \} \\ &= \{ \omega \in \Omega \mid \forall n \in \mathbb{N} : f_n(\omega) \geq a \} \\ &= \bigcap_{n \in \mathbb{N}} f_n^{-1}([a; \infty)) \end{aligned}$$

und aus Teil a).

Messbarkeit von  $\sup_{n \in \mathbb{N}} f_n$  folgt analog.

Messbarkeit von  $f = \lim_{n \rightarrow \infty} f_n$  folgt aus der Messbarkeit von  $\liminf_n f_n$ , wobei die Messbarkeit von  $\liminf_n$  ähnlich bewiesen wird wie die Messbarkeit von  $\inf$ .

c) Folgt aus Teil b), z.B. mit  $f_1 := f, f_2 := g, f_3 := f, f_4 := g, \dots$

□

**Definition 5.12** Sei  $(\Omega, \mathcal{A})$  ein Messraum.

(i) Für  $A \subset \Omega$  sei

$$I_A : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto I_A(\omega)$$

die Funktion mit

$$I_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \notin A \end{cases}$$

$I_A$  heißt Indikatorfunktion von  $A$ .

(ii) Sei  $n \in \mathbb{N}$ ,  $A_1 \in \mathcal{A}, \dots, A_n \in \mathcal{A}$  und  $\alpha_1 \in \mathbb{R}, \dots, \alpha_n \in \mathbb{R}$ . Eine Funktion  $s : \Omega \rightarrow \mathbb{R}$  der Gestalt

$$s(\omega) = \sum_{i=1}^n \alpha_i I_{A_i}(\omega) \quad \forall \omega \in \Omega$$

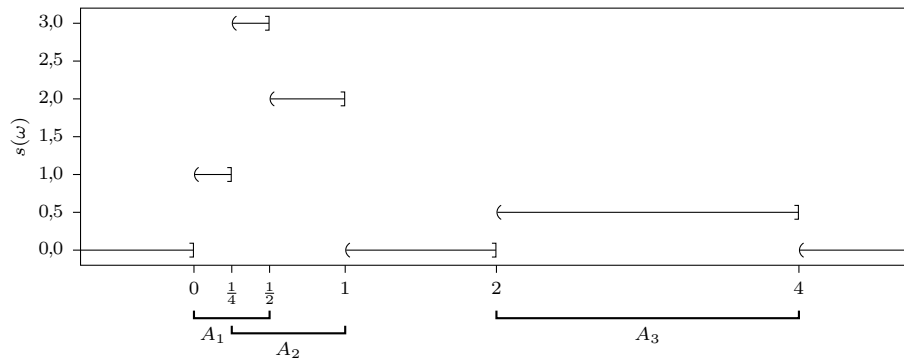
heißt ( $\mathcal{A}$ -) einfache Funktion.

**Beispiel 5.13** Das Bild zeigt die einfache Funktion mit  $\Omega = \mathbb{R}, \mathcal{A} = \mathfrak{B}$  sowie

$i$	$1$	$2$	$3$
$A_i$	$(0; \frac{1}{2}]$	$(\frac{1}{4}; 1]$	$(2; 4]$
$\alpha_i$	$1$	$2$	$\frac{1}{2}$

also

$$s(\omega) = 1 \cdot I_{(0; \frac{1}{2}]}(\omega) + 2 \cdot I_{(\frac{1}{4}; 1]}(\omega) + \frac{1}{2} \cdot I_{(2; 4]}(\omega)$$



**Satz 5.14** Sei  $(\Omega, \mathcal{A})$  ein Messraum. Dann gilt

- Die Indikatorfunktion  $I_A$  von  $A$  ist  $\mathcal{A}$ -messbar genau dann, wenn  $A \in \mathcal{A}$ .
- Jede  $\mathcal{A}$ -einfache Funktion ist  $\mathcal{A}$ -messbar.

**Beweis:**

- Sei  $B \in \mathfrak{B}$ .

$$I_A^{-1}(B) = \{\omega \in \Omega \mid I_A(\omega) \in B\} = \begin{cases} A & \text{falls } 1 \in B \wedge 0 \notin B \\ A^C & \text{falls } 1 \notin B \wedge 0 \in B \\ \emptyset & \text{falls } 1 \notin B \wedge 0 \notin B \\ \Omega & \text{falls } 1 \in B \wedge 0 \in B \end{cases}$$

Demnach gilt für alle  $B \in \mathfrak{B}$ :  $I_A^{-1}(B) \in \mathcal{A} \iff A \in \mathcal{A}$ .

- Folgt aus a) und Satz 5.9 a).

□

Der nachfolgende Satz ist *das* entscheidende Hilfsmittel, um im nächsten Kapitel Integrale über Funktionen

$$f : \Omega \rightarrow \mathbb{R}$$

zu definieren.

**Satz 5.15** Sei  $f : \Omega \rightarrow \mathbb{R}$  eine  $\mathcal{A}$ -messbare Funktion mit  $f(\omega) \geq 0 \quad \forall \omega \in \Omega$ . Dann gilt:

Es existiert eine Folge  $(s_n)_{n \in \mathbb{N}}$  von Funktionen

$$s_n : \Omega \rightarrow \mathbb{R}$$

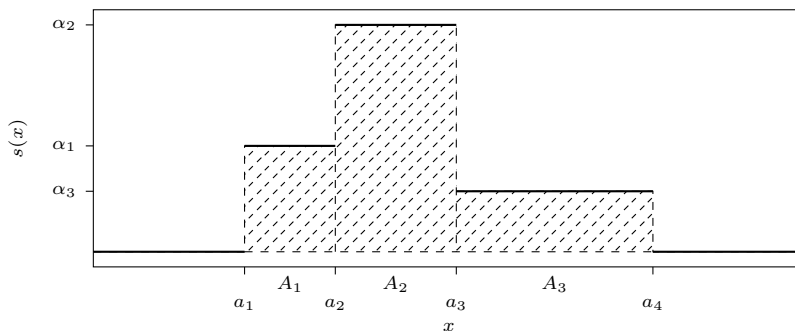
mit folgenden Eigenschaften:

- $s_n$  ist eine  $\mathcal{A}$ -einfache Funktion,  $n \in \mathbb{N}$ .
- $s_n(\omega) \geq 0 \quad \forall \omega \in \Omega, n \in \mathbb{N}$ .
- $s_n \nearrow f$ , d.h.:

- $s_1(\omega) \leq s_2(\omega) \leq s_3(\omega) \leq s_4(\omega) \leq \dots \leq f(\omega) \quad \forall \omega \in \Omega$
- $\lim_{n \rightarrow \infty} s_n(\omega) = f(\omega) \quad \forall \omega \in \Omega$

# 6 Integration

## 6.1 Idee



$$s(x) = \sum_{i=1}^3 \alpha_i I_{A_i}(x)$$

Integration:

$$\begin{aligned} \int_{\mathbb{R}} s(x) dx &= \int_{a_1}^{a_2} s(x) dx + \int_{a_2}^{a_3} s(x) dx + \int_{a_3}^{a_4} s(x) dx \\ &= \alpha_1 \cdot (a_2 - a_1) + \alpha_2 \cdot (a_3 - a_2) + \alpha_3 \cdot (a_4 - a_3) \\ &= \alpha_1 \cdot \lambda(A_1) + \alpha_2 \cdot \lambda(A_2) + \alpha_3 \cdot \lambda(A_3) \end{aligned} \quad (6.1)$$

↪ Auf diese Weise können wir die einfache Funktion  $s$  bzgl. des Lebesguemaßes  $\lambda$  integrieren.

Statt

$$\int_{\mathbb{R}} s(x) dx$$

schreiben wir ab jetzt aber

$$\int s(x) \lambda(dx) \quad \text{oder} \quad \int_{\mathbb{R}} s d\lambda$$

Mit Hilfe von Gleichung (6.1) können wir die einfache Funktionen  $s$  aber auch ganz analog bzgl. jedem anderen Maß  $\mu$  integrieren:

$$\int s d\mu = \alpha_1 \cdot \mu(A_1) + \alpha_2 \cdot \mu(A_2) + \alpha_3 \cdot \mu(A_3)$$

Mit Hilfe von Satz 5.15 werden wir dann (im nächsten Abschnitt) das Integral über messbare Funktionen  $f$  definieren – und zwar als

$$\int f d\mu = \lim_{n \rightarrow \infty} \int s_n d\mu \quad \text{wobei} \quad s_n \nearrow f$$

## 6.2 Definition des allgemeinen Integrals

**Vorgehen:**

- [1.] Integration einfacher Funktionen
- [2.] Integration positiver messbarer Funktionen
- [3.] Integration messbarer Funktionen

Sei im Folgenden  $(\Omega, \mathcal{A})$  ein Messraum und  $\mu$  ein Maß auf  $(\Omega, \mathcal{A})$ . Wir wollen Funktionen

$$f : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathfrak{B})$$

bzgl. dem Maß  $\mu$  integrieren.

### [1.] Integration einfacher Funktionen

Sei  $s : \Omega \rightarrow \mathbb{R}$  eine einfache Funktion; d.h.

$$s(\omega) = \sum_{i=1}^n \alpha_i I_{A_i}(\omega) \quad \forall \omega \in \Omega$$

wobei  $n \in \mathbb{N}$ ,  $A_1 \in \mathcal{A}, \dots, A_n \in \mathcal{A}$  und  $\alpha_1 \in \mathbb{R}, \dots, \alpha_n \in \mathbb{R}$ .

Wir definieren

$$\boxed{\int s \, d\mu := \sum_{i=1}^n \alpha_i \mu(A_i)} \quad (6.2)$$

**Bemerkung 6.1** Die einfache Funktion  $s = \sum_{i=1}^n \alpha_i I_{A_i}$  könnte auch eine andere Darstellung haben:

$$s(\omega) = \sum_{j=1}^m \beta_j I_{B_j}(\omega) \quad \forall \omega \in \Omega$$

mit  $m \in \mathbb{N}$ ,  $B_1 \in \mathcal{A}, \dots, B_m \in \mathcal{A}$  und  $\beta_1 \in \mathbb{R}, \dots, \beta_m \in \mathbb{R}$ . Also:

$$\sum_{i=1}^n \alpha_i I_{A_i} = \sum_{j=1}^m \beta_j I_{B_j}$$

obwohl  $m \neq n$ ,  $B_1 \neq A_1, \dots, \beta_1 \neq \alpha_1, \dots$

Eine leichte Rechnung zeigt, dass in diesem Fall aber auch

$$\sum_{i=1}^n \alpha_i \mu(A_i) = \sum_{j=1}^m \beta_j \mu(B_j)$$

Die Definition in (6.2) hängt also nicht von der Darstellung von  $s$  ab. Das Integral für einfache Funktionen ist daher „wohldefiniert“.

## [2.] Integration positiver messbarer Funktionen

Sei  $f : \Omega \rightarrow \mathbb{R}$  eine  $\mathcal{A}$ -messbare Funktion, so dass

$$f(\omega) \geq 0 \quad \forall \omega \in \Omega$$

Nach Satz 5.15 existiert eine Folge  $(s_n)_{n \in \mathbb{N}}$  von Funktionen

$$s_n : \Omega \rightarrow \mathbb{R}$$

mit folgenden Eigenschaften:

- $s_n$  ist eine  $\mathcal{A}$ -einfache Funktion,  $n \in \mathbb{N}$ .
- $s_n(\omega) \geq 0 \quad \forall \omega \in \Omega, n \in \mathbb{N}$ .
- $s_n \nearrow f$

Wir definieren das Integral von  $f$  bzgl.  $\mu$  (wie bereits angekündigt) durch

$$\int f d\mu := \lim_{n \rightarrow \infty} \int s_n d\mu \quad (6.3)$$

### Bemerkung 6.2

a) Der Limes der Folge

$$\left( \int s_n d\mu \right)_{n \in \mathbb{N}}$$

existiert immer. Es kann aber durchaus sein, dass

$$\lim_{n \rightarrow \infty} \int s_n d\mu = \infty$$

b) Es kann sein, dass es noch weitere Folgen  $(t_n)_{n \in \mathbb{N}}$  gibt, so dass gilt

- $t_n$  ist eine  $\mathcal{A}$ -einfache Funktion,  $n \in \mathbb{N}$ .
- $t_n(\omega) \geq 0 \quad \forall \omega \in \Omega, n \in \mathbb{N}$ .
- $t_n \nearrow f$

In diesem Fall gilt aber

$$\lim_{n \rightarrow \infty} \int t_n d\mu = \lim_{n \rightarrow \infty} \int s_n d\mu$$

Die Definition in (6.3) hängt also nicht von der gewählten Folge  $(s_n)_{n \in \mathbb{N}}$  ab. Das Integral für positive messbare Funktionen ist daher „wohldefiniert“.

## [3.] Integration messbarer Funktionen

Sei  $f : \Omega \rightarrow \mathbb{R}$  eine  $\mathcal{A}$ -messbare Funktion.

Setze

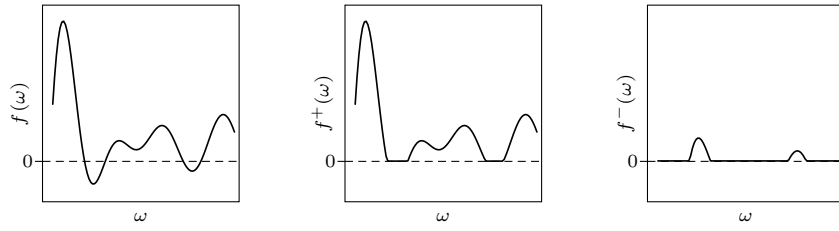
$$f^+ : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto \max\{f(\omega); 0\}$$

als *Positivteil* von  $f$  und

$$f^- : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto -\min\{f(\omega); 0\}$$

als *Negativteil* von  $f$ .





Es gilt

- $f = f^+ - f^-$
- $f^+(\omega) \geq 0 \quad \forall \omega \in \Omega$  und  $f^-(\omega) \geq 0 \quad \forall \omega \in \Omega$
- $f^+$  und  $f^-$  sind  $\mathcal{A}$ -messbar (Satz 5.11)

Durch Gleichung (6.3) sind damit

$$\int f^+ d\mu \quad \text{und} \quad \int f^- d\mu$$

schon definiert.

Falls

$$\int f^+ d\mu < \infty \quad \text{und} \quad \int f^- d\mu < \infty$$

können wir daher für das Integral von  $f$  definieren

$$\boxed{\int f d\mu := \int f^+ d\mu - \int f^- d\mu} \quad (6.4)$$

Probleme ergeben sich, falls

$$\int f^+ d\mu = \infty \quad \text{oder} \quad \int f^- d\mu = \infty$$

Daher folgende Definition:

**Definition 6.3** Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum. Eine Funktion

$$f : \Omega \rightarrow \mathbb{R}$$

heißt integrierbar bzgl.  $\mu$ , falls gilt:

(i)  $f$  ist  $\mathcal{A}$ -messbar.

(ii)  $\int f^+ d\mu < \infty$  und  $\int f^- d\mu < \infty$

Die Menge der bzgl.  $\mu$  integrierbaren Funktionen wird mit

$$\mathcal{L}_1(\Omega, \mathcal{A}, \mu) := \left\{ f : \Omega \rightarrow \mathbb{R} \mid f \text{ ist integrierbar bzgl. } \mu \right\}$$

bezeichnet.

**Also:** Nicht jede messbare Funktion ist integrierbar!

### Der Integrationsbereich

Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum und  $f : \Omega \rightarrow \mathbb{R}$  eine bzgl.  $\mu$  integrierbare Funktion.

$$\int f d\mu$$

bedeutet: Es wird über den ganzen Definitionsbereich integriert.

$\int f d\mu$  ist also eine abkürzende Schreibweise für

$$\int_{\Omega} f d\mu$$

Falls  $f$  nur über eine Teilmenge  $A \in \mathcal{A}$  von  $\Omega$  integriert werden soll, so ist

$$\int_A f d\mu = \int_{\Omega} f \cdot I_A d\mu = \int \underbrace{f \cdot I_A}_{=: f'} d\mu$$

Falls  $f$  bzgl.  $\mu$  integrierbar ist, so ist auch  $f' = f \cdot I_A$  für jedes  $A \in \mathcal{A}$  bzgl.  $\mu$  integrierbar.

### Schreibweisen

$$\int f d\mu = \int f(\omega) \mu(d\omega) = \int f(\omega) d\mu(\omega)$$

bzw.

$$\int_A f d\mu = \int_A f(\omega) \mu(d\omega) = \int_A f(\omega) d\mu(\omega)$$

### Fazit

Wir haben jetzt in großer Allgemeinheit ein Integral bzgl. beliebiger Maße über eine große Klasse von Funktionen definiert.

Zusammenhang mit dem Riemann-Integral aus der Schule/Uni:

Sei  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  eine Riemann-integrierbare Funktion. Dann ist  $f$  auch integrierbar bzgl. des Lebesguemaßes  $\lambda^k$  im Sinne von Definition 6.3 und es gilt:

$$\int_{\mathbb{R}^k} f(x) dx = \int f d\lambda^k$$

Die Umkehrung gilt aber nicht! Sei  $\Omega = \mathbb{R}$  und  $f : \mathbb{R} \rightarrow \mathbb{R}$  gleich der Indikatorfunktion von  $\mathbb{Q}$  auf  $\mathbb{R}$  („Dirichletsche Sprungfunktion“)

$$f = I_{\mathbb{Q}}$$

Dann ist  $f$  nicht Riemann-Integrierbar. Aber wegen  $\mathbb{Q} \in \mathfrak{B}$  ist  $f$  messbar (vgl. Satz 5.14 a)) und wegen

$$\int f d\lambda = \lambda(\mathbb{Q}) = \sum_{q \in \mathbb{Q}} \lambda(\{q\}) = 0$$

( $\mathbb{Q}$  ist abzählbar!) ist  $f$  integrierbar im Sinne von Definition 6.3.

D.h.: Es können jetzt viel mehr Funktionen  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  integriert werden!

Während das Riemann-Integral außerdem nur für (bestimmte)  $\Omega \subset \mathbb{R}^k$  definiert ist, können wir nun auch auf beliebige Messräume  $(\Omega, \mathcal{A})$  und bzgl. beliebiger Maße integrieren.

Zum Unterschied zwischen der Definition des Riemann-Integrals und der Definition des Lebesgue-Integrals noch ein Auszug aus dem Wikipedia-Artikel Lebesgue-Integral:

Der wesentliche Unterschied im Vorgehen bei der Integration nach Riemann bzw. Lebesgue besteht darin, dass beim Riemann-Integral der Definitionsbereich (Abzisse) beim Lebesgue-Integral jedoch die Bildmenge (Ordinate) der Funktion unterteilt wird. An obigen Beispielen lässt sich bereits erkennen, dass sich dieser Unterschied durchaus als entscheidend herausstellen kann.

### Henri Lebesgue über den Vergleich zwischen Riemann- und Lebesgue-Integral

„Man kann sagen, dass man sich bei dem Vorgehen von Riemann verhält wie ein Kaufmann ohne System, der Geldstücke und Banknoten zählt in der Reihenfolge, wie er sie in die Hand bekommt; während wir vorgehen wie ein umsichtiger Kaufmann, der sagt:

Ich habe  $m(E_1)$  Münzen zu einer Krone, macht  $1 \cdot m(E_1)$   
 ich habe  $m(E_2)$  Münzen zu zwei Kronen, macht  $2 \cdot m(E_2)$   
 ich habe  $m(E_3)$  Münzen zu fünf Kronen, macht  $5 \cdot m(E_3)$ .

usw., ich habe also insgesamt

$$S = 1 \cdot m(E_1) + 2 \cdot m(E_2) + 5 \cdot m(E_3) + \dots \text{ [Kronen]}$$

Die beiden Verfahren führen sicher den Kaufmann zum gleichen Resultat, weil er - wie reich er auch sei - nur eine endliche Zahl von Banknoten zu zählen hat; aber für uns, die wir unendlich viele Indivisiblen zu addieren haben, ist der Unterschied zwischen beiden Vorgehensweisen wesentlich.“

*(H. Lebesgue, 1926; zitiert nach J.Elstrodt)*

Und wie rechnet man jetzt ein Integral bzgl. einem Maß  $\mu$  aus? Später! Zunächst müssen noch einige sehr wichtige Eigenschaften des Integrals festgehalten werden.

### 6.3 Eigenschaften des Integrals

**Satz 6.4** Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum.

a) Falls  $f_1 \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$  und  $f_2 \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$ , so ist auch  $f_1 + f_2 \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$  und es gilt

$$\int f_1 + f_2 d\mu = \int f_1 d\mu + \int f_2 d\mu$$

b) Falls  $f \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$  und  $c \in \mathbb{R}$ , so ist auch  $cf \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$  und es gilt

$$\int cf d\mu = c \int f d\mu$$

c) Sei  $c \in \mathbb{R}$  eine Konstante und sei  $\mu$  ein endliches Maß. Dann ist  $c \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$  und

$$\int c d\mu = c \cdot \mu(\Omega)$$

d) Für  $A \in \mathcal{A}$  ist

$$\int I_A d\mu = \mu(A)$$

und mit  $B \in \mathcal{A}$  ist

$$\int_B I_A d\mu = \mu(A \cap B)$$

e) Für  $f_1 \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$  und  $f_2 \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$  gilt

$$f_1(\omega) \leq f_2(\omega) \quad \forall \omega \in \Omega \quad \Rightarrow \quad \int f_1 d\mu \leq \int f_2 d\mu$$

Insbesondere folgt aus  $f \geq 0$  auch  $\int f d\mu \geq 0$ .

f) Für  $\mathcal{A}$ -messbare Funktionen  $f : \Omega \rightarrow \mathbb{R}$  gilt

$$|f| \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu) \quad \Leftrightarrow \quad f \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$$

In diesem Fall ist

$$\left| \int f d\mu \right| \leq \int |f| d\mu$$

g) Sei  $f : \Omega \rightarrow \mathbb{R}$  eine beschränkte  $\mathcal{A}$ -messbare Funktion und sei  $\mu$  ein endliches Maß. Dann ist  $f \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$ .

**Bemerkung 6.5** Die Aussagen gelten in analoger Weise auch, falls der Integrationsbereich nicht ganz  $\Omega$  sondern ein  $A \in \mathcal{A}$  ist, d.h.  $\int_A$  anstelle von  $\int$ .

**Satz 6.6 (Monotone Konvergenz)**

Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum. Sei  $(f_n)_{n \in \mathbb{N}}$  eine Folge von  $\mathcal{A}$ -messbaren Funktionen  $f_n : \Omega \rightarrow \mathbb{R}$ , so dass

$$0 \leq f_1(\omega) \leq f_2(\omega) \leq f_3(\omega) \leq \dots \leq \lim_{n \rightarrow \infty} f_n(\omega) < \infty \quad \forall \omega \in \Omega$$

Dann ist

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int \lim_{n \rightarrow \infty} f_n(\omega) \mu(d\omega)$$

Und als einfache Folgerung erhält man:

**Korollar 6.7 ( $\sigma$ -Additivität des Integrals)**

Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum und sei  $(f_k)_{k \in \mathbb{N}}$  eine Folge von  $\mathcal{A}$ -messbaren Funktionen  $f_k: \Omega \rightarrow \mathbb{R}$ , so dass

$$f_k(\omega) \geq 0 \quad \forall \omega \in \Omega$$

für jedes  $k \in \mathbb{N}$ . Sei außerdem

$$\sum_{k=1}^{\infty} f_k(\omega) < \infty \quad \forall \omega \in \Omega$$

Dann ist

$$\int \sum_{k=1}^{\infty} f_k(\omega) \mu(d\omega) = \sum_{k=1}^{\infty} \int f_k(\omega) \mu(d\omega)$$

(und die beiden Integrale existieren gemäß (6.3).)

**Beweis:**  $\rightsquigarrow$  Übung! □

**Lemma 6.8 (Lemma von Fatou)**

Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum. Sei  $(f_n)_{n \in \mathbb{N}}$  eine Folge von  $\mathcal{A}$ -messbaren Funktionen  $f_n: \Omega \rightarrow \mathbb{R}$  mit

$$f_n(\omega) \geq 0 \quad \forall \omega \in \Omega$$

Sei außerdem

$$\liminf_{n \rightarrow \infty} f_n(\omega) < \infty \quad \forall \omega \in \Omega$$

Dann ist

$$\Omega \rightarrow \mathbb{R}, \quad \omega \mapsto \liminf_{n \rightarrow \infty} f_n(\omega)$$

eine  $\mathcal{A}$ -messbare Funktion und

$$0 \leq \int \liminf_{n \rightarrow \infty} f_n(\omega) \mu(d\omega) \leq \liminf_{n \rightarrow \infty} \int f_n(\omega) \mu(d\omega)$$

**Fast überall bestehende Eigenschaften:**

**Definition 6.9** Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum. Eine Menge  $N \in \mathcal{A}$  heißt  $\mu$ -Nullmenge, falls

$$\mu(N) = 0$$

**Beispiel 6.10**

- $\emptyset$  ist  $\mu$ -Nullmenge für jedes Maß  $\mu$ .
- Alle abzählbaren Teilmengen von  $\mathbb{R}^n$  sind  $\lambda^n$ -Nullmengen, z.B. auch  $\mathbb{Q}$ .
- Alle Hyperebenen im  $\mathbb{R}^n$  sind  $\lambda^n$ -Nullmengen.
- Falls  $A$  eine  $\mu$ -Nullmenge ist und  $B \subset A$ , dann ist auch  $B$  eine  $\mu$ -Nullmenge.

Fast überall bestehende Eigenschaften sind Eigenschaften, die auf ganz  $\Omega$  mit Ausnahme von einer  $\mu$ -Nullmenge bestehen. Etwas formaler: Eine Eigenschaft  $E(\omega)$  mit  $\{\omega \in \Omega \mid E(\omega) \text{ gilt}\} \in \mathcal{A}$  und  $\mu(\{\omega \in \Omega \mid E(\omega) \text{ gilt nicht}\}) = 0$  besteht  $\mu$ -fast überall. Man sagt auch: Solche Eigenschaften bestehen  $\mu$ -fast sicher bzw. in abgekürzter Schreibweise  $\mu$ -f.s.

Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum; die wichtigsten Beispiele für fast überall bestehende Eigenschaften sind:

- Seien  $f : \Omega \rightarrow \mathbb{R}$  und  $g : \Omega \rightarrow \mathbb{R}$  zwei  $\mathcal{A}$ -messbare Funktionen, so dass

$$\mu(\{\omega \mid f(\omega) \neq g(\omega)\}) = 0 \quad (6.5)$$

Hierfür schreiben wir auch

$$f = g \quad \mu\text{-f.s.}$$

- Seien  $f : \Omega \rightarrow \mathbb{R}$  und  $g : \Omega \rightarrow \mathbb{R}$  zwei  $\mathcal{A}$ -messbare Funktionen, so dass

$$\mu(\{\omega \mid f(\omega) > g(\omega)\}) = 0 \quad (6.6)$$

Hierfür schreiben wir auch

$$f \leq g \quad \mu\text{-f.s.}$$

- Sei  $(f_n)_{n \in \mathbb{N}}$  eine Folge  $\mathcal{A}$ -messbarer Funktionen  $f_n : \Omega \rightarrow \mathbb{R}$  und  $f : \Omega \rightarrow \mathbb{R}$  eine  $\mathcal{A}$ -messbare Funktion, so dass

$$\mu(\{\omega \mid (f_n(\omega))_{n \in \mathbb{N}} \text{ konvergiert nicht gegen } f(\omega)\}) = 0 \quad (6.7)$$

Hierfür schreiben wir auch

$$f_n \xrightarrow[n \rightarrow \infty]{} f \quad \mu\text{-f.s.}$$

oder

$$\lim_{n \rightarrow \infty} f_n = f \quad \mu\text{-f.s.}$$

**Übung 6.11** Man zeige, dass die Teilmengen von  $\Omega$ , die in den Gleichungen (6.5) und (6.6) auftreten, tatsächlich in  $\mathcal{A}$  liegen.

**Satz 6.12** Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum und seien  $f : \Omega \rightarrow \mathbb{R}$  und  $g : \Omega \rightarrow \mathbb{R}$  zwei  $\mathcal{A}$ -messbare Funktionen. Es gilt:

$$a) \quad f \geq 0 \quad \text{und} \quad \int f \, d\mu = 0 \quad \Rightarrow \quad f = 0 \quad \mu\text{-f.s.}$$

$$b) \quad f = g \quad \mu\text{-f.s.} \quad \Leftrightarrow \quad \int_A f \, d\mu = \int_A g \, d\mu \quad \forall A \in \mathcal{A}$$

**Beweis:**

a) Setzte

$$A_n := \{\omega \in \Omega \mid f(\omega) \geq \frac{1}{n}\} \quad \forall n \in \mathbb{N}$$

Dann ist  $I_{A_n} \leq n \cdot f \quad \forall n \in \mathbb{N}$  und somit

$$\mu(A_n) = \int I_{A_n} d\mu \leq n \cdot \int f d\mu = 0$$

Also

$$\begin{aligned} \mu(\{\omega \mid f(\omega) \neq 0\}) &= \mu(\{\omega \mid f(\omega) > 0\}) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \\ &\leq \sum_{n=1}^{\infty} \mu(A_n) = 0 \end{aligned}$$

b)  $\rightsquigarrow$  Übung!

□

**Satz 6.13 (Majorisierter Konvergenz; Satz von Lebesgue)**

Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum und sei  $(f_n)_{n \in \mathbb{N}}$  eine Folge  $\mathcal{A}$ -messbarer Funktionen  $f_n : \Omega \rightarrow \mathbb{R}$ , so dass

$$f_n \xrightarrow[n \rightarrow \infty]{} f \quad \mu\text{-f.s.}$$

für eine Funktion  $f : \Omega \rightarrow \mathbb{R}$ . Sei außerdem  $g \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$ , so dass

$$|f_n| \leq g \quad \mu\text{-f.s.}$$

Dann ist

$$f_n \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu) \quad \forall n \in \mathbb{N} \quad \text{und} \quad f \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$$

Außerdem gilt

$$\lim_{n \rightarrow \infty} \int f_n(\omega) \mu(d\omega) = \int \underbrace{\lim_{n \rightarrow \infty} f_n(\omega)}_{f(\omega)} \mu(d\omega)$$

und

$$\lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0$$

Zum Abschluss dieses Abschnitts noch ein Satz aus der reellen Analysis, der in der Statistik nützlich sein kann, um Regularitätsvoraussetzungen zu überprüfen:

**Satz 6.14 (Vertauschen von Integration und Differentiation)**

Es seien  $a \in \mathbb{R}$  und  $b \in \mathbb{R}$  mit  $a < b$ . Sei außerdem

$$f : (a; b) \times \mathbb{R}^n \rightarrow \mathbb{R}$$

eine Funktion mit folgenden Eigenschaften:

a)  $f(t, \cdot) \in \mathcal{L}_1(\mathbb{R}^n, \mathfrak{B}^{\otimes n}, \lambda^n)$  für alle  $t \in (a; b)$ .

b)  $f(\cdot, x)$  ist differenzierbar für alle  $x \in \mathbb{R}^n$ .

c) Es gibt ein  $g \in \mathcal{L}_1(\mathbb{R}^n, \mathfrak{B}^{\otimes n}, \lambda^n)$ , so dass

$$\left| \frac{\partial}{\partial t} f(t, x) \right| \leq g(x) \quad \forall t \in (a; b), \quad \forall x \in \mathbb{R}^n$$

Dann ist  $\frac{\partial}{\partial t} f(t, \cdot) \in \mathcal{L}_1(\mathbb{R}^n, \mathfrak{B}^{\otimes n}, \lambda^n)$  für alle  $t \in (a; b)$  und

$$\int_{\mathbb{R}^n} \frac{\partial}{\partial t} f(t, x) \lambda^n(dx) = \frac{d}{dt} \int_{\mathbb{R}^n} f(t, x) \lambda^n(dx)$$

## 6.4 Die Transformationsformel – Integration bzgl. dem Bildmaß

Seien  $(\Omega, \mathcal{A})$  und  $(\Omega', \mathcal{A}')$  zwei Messräume und  $\mu$  ein Maß auf  $(\Omega, \mathcal{A})$ . Sei

$$T : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$$

d.h.  $T$  ist eine  $\mathcal{A}/\mathcal{A}'$ -messbare Abbildung  $T : \Omega \rightarrow \Omega'$  (vgl. Notation 5.2).

Das Bildmaß von  $\mu$  unter  $T$  ist definiert durch

$$T(\mu) : \mathcal{A}' \mapsto \mu(T^{-1}(A'))$$

(vgl. Definition 5.3).

Der nachfolgende Satz beschreibt den Zusammenhang zwischen den Integralen bzgl.  $\mu$  und  $T(\mu)$ .

### Satz 6.15 (Transformationsformel)

Seien  $(\Omega, \mathcal{A})$  und  $(\Omega', \mathcal{A}')$  zwei Messräume,  $\mu$  ein Maß auf  $(\Omega, \mathcal{A})$  und

$$T : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$$

Für  $\mathcal{A}'$ -messbare Abbildungen  $f' : \Omega' \rightarrow \mathbb{R}$  gilt dann:

$$f' \in \mathcal{L}_1(\Omega', \mathcal{A}', T(\mu)) \quad \Leftrightarrow \quad f' \circ T \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu) \quad (6.8)$$

In diesem Fall ist außerdem

$$\int f' d[T(\mu)] = \int f' \circ T d\mu \quad (6.9)$$

bzw.

$$\int_{A'} f' d[T(\mu)] = \int_{T^{-1}(A')} f' \circ T d\mu \quad \forall A' \in \mathcal{A}' \quad (6.10)$$

**Beweis:**

[0.] Angenommen  $f' = I_{A'}$  für ein  $A' \in \mathcal{A}'$ . Dann ist

$$\begin{aligned} \int I_{A'} d[T(\mu)] &= [T(\mu)](A') = \mu(T^{-1}(A')) = \int I_{T^{-1}(A')} d\mu \\ &= \int I_{A'} \circ T d\mu \end{aligned}$$

d.h. (6.9) für diesen Spezialfall.



[1.] Angenommen  $f' = s' = \sum_{i=1}^n \alpha'_i I_{A'_i}$  ist eine  $\mathcal{A}'$ -einfache Funktion. Dann ist

$$\begin{aligned} \int f' d[T(\mu)] &\stackrel{[0.]}{=} \sum_{i=1}^n \alpha'_i \int I_{A'_i} \circ T d\mu = \int \left( \sum_{i=1}^n \alpha'_i I_{A'_i} \right) \circ T d\mu = \\ &= \int f' \circ T d\mu \end{aligned}$$

d.h. (6.9) für diesen Spezialfall.

[2.] Angenommen  $f' \geq 0$ . Dann existiert nach Satz 5.15 eine Folge  $(s'_n)_{n \in \mathbb{N}}$  von Funktionen

$$s'_n : \Omega' \rightarrow \mathbb{R}$$

mit folgenden Eigenschaften:

- a)  $s'_n$  ist eine  $\mathcal{A}'$ -einfache Funktion,  $n \in \mathbb{N}$ .
- b)  $s'_n(\omega') \geq 0 \quad \forall \omega' \in \Omega', n \in \mathbb{N}$ .
- c)  $s'_n \nearrow f'$

Dann ist aber auch  $(s'_n \circ T)_{n \in \mathbb{N}}$  eine Folge von Funktionen

$$s'_n \circ T : \Omega \rightarrow \mathbb{R}$$

mit folgenden Eigenschaften:

- a)  $s'_n \circ T$  ist eine  $\mathcal{A}$ -einfache Funktion,  $n \in \mathbb{N}$ .
- b)  $s'_n \circ T(\omega) \geq 0 \quad \forall \omega \in \Omega, n \in \mathbb{N}$ .
- c)  $s'_n \circ T \nearrow f' \circ T$

Aus dem Satz von der monotonen Konvergenz (Satz 6.6) folgt schließlich

$$\begin{aligned} \int f' d[T(\mu)] &\stackrel{\text{S. 6.6}}{=} \lim_{n \rightarrow \infty} \int s'_n d[T(\mu)] \stackrel{[1.]}{=} \lim_{n \rightarrow \infty} \int s'_n \circ T d\mu = \\ &\stackrel{\text{S. 6.6}}{=} \int f' \circ T d\mu \end{aligned}$$

d.h. (6.9) für diesen Spezialfall.

[3.] Sei nun  $f' : \Omega' \rightarrow \mathbb{R}$  eine beliebige  $\mathcal{A}'$ -messbare Funktion. Es gilt

$$f' = (f')^+ - (f')^- \quad \text{vgl. Abschnitt 6.2 [3.]}$$

Um (6.8) zu zeigen, wird nun die Äquivalenz folgender drei Aussagen gezeigt:

- (i)  $f' \in \mathcal{L}_1(\Omega', \mathcal{A}', T(\mu))$
- (ii)  $(f')^+ \in \mathcal{L}_1(\Omega', \mathcal{A}', T(\mu))$  und  $(f')^- \in \mathcal{L}_1(\Omega', \mathcal{A}', T(\mu))$
- (iii)  $(f' \circ T)^+ \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$  und  $(f' \circ T)^- \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$
- (iv)  $f' \circ T \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$

Die Äquivalenzen (i)  $\Leftrightarrow$  (ii) und (iii)  $\Leftrightarrow$  (iv) folgen unmittelbar aus der Definition 6.3.

Es bleibt noch zu zeigen (ii)  $\Leftrightarrow$  (iii):

Wegen  $(f')^+ \circ T = (f' \circ T)^+$  und  $(f')^- \circ T = (f' \circ T)^-$  gilt

$$\int (f')^+ d[T(\mu)] \stackrel{[2.]}{=} \int (f')^+ \circ T d\mu = \int (f' \circ T)^+ d\mu \quad (6.11)$$

und

$$\int (f')^- d[T(\mu)] \stackrel{[2.]}{=} \int (f')^- \circ T d\mu = \int (f' \circ T)^- d\mu \quad (6.12)$$

Dies zeigt (ii)  $\Leftrightarrow$  (iii).

Somit folgt schließlich (i)  $\Leftrightarrow$  (iv), d.h. (6.8).

Aus (6.11), (6.12) und der Integraldefinition (6.4) folgt außerdem (6.9).

Sei nun  $A' \in \mathcal{A}'$ . Dann folgt aus

$$\begin{aligned} \int_{A'} f' d[T(\mu)] &= \int f' \cdot I_{A'} d[T(\mu)] = \int (f' \cdot I_{A'}) \circ T d\mu \\ &= \int (f' \circ T) \cdot (I_{A'} \circ T) d\mu = \int (f' \circ T) \cdot I_{T^{-1}(A')} d\mu \\ &= \int_{T^{-1}(A')} f' \circ T d\mu \end{aligned}$$

auch (6.10). □

**Bemerkung 6.16** *Der Beweis von Satz 6.15 wurde in die Schritte [0.], [1.], [2.] und [3.] zerlegt, wobei diese Schritte denen bei der Definition des Integrals (vgl. Abschnitt 6.2) entsprachen. Dies ist eine Beweistechnik, die in Zusammenhang mit Integralen häufig verwendet wird.*

Aus der reellen Analysis kennen wir folgende Transformationsformel:

**Satz 6.17 (Transformationsformel aus der reellen Analysis)**

*Für zwei offene Teilmengen  $G \subset \mathbb{R}^n$  und  $G' \subset \mathbb{R}^n$  sei  $\phi : G \rightarrow G'$   $\mathcal{C}_1$ -invertierbar. Sei  $f' : \mathbb{R}^n \rightarrow \mathbb{R}^n$   $\mathfrak{B}^{\otimes n}/\mathfrak{B}^{\otimes n}$ -messbar.*

*Falls  $f'$  über  $G$  integrierbar bzgl.  $\lambda^n$  ist, so ist*

$$\int_G f' \circ \phi(x) \lambda^n(dx) = \int_{G'} f'(x') \cdot \frac{1}{|(\det J_\phi)(x')|} \lambda^n(dx') \quad .$$

**Notation 6.18**  $J_\phi$  bezeichnet die Jacobi-Matrix von  $\phi$ , d.h.

$$J_\phi = \left( \frac{\partial \phi_i}{\partial x_j} \right)_{\substack{i=1,\dots,n \\ j=1,\dots,n}} = \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1} & \frac{\partial \phi_1}{\partial x_2} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \frac{\partial \phi_n}{\partial x_2} & \cdots & \frac{\partial \phi_n}{\partial x_n} \end{pmatrix}$$

**Bemerkung 6.19** *In der Situation von Satz 6.17 folgt aus der Transformationsformel (Satz 6.15) sofort*

$$\int_G f' \circ \phi(x) \lambda^n(dx) = \int_{G'} f'(x') [\phi(\lambda^n)](dx')$$

*Um die Transformationsformel aus der reellen Analysis (Satz 6.17) zu beweisen, müßte nun nur noch gezeigt werden, dass*

$$[\phi(\lambda^n)](B') = \int_{B'} \frac{1}{|(\det J_\phi)(x')|} \lambda^n(dx') \quad \forall B' \in \mathfrak{B}^{\otimes n}, B' \subset G'$$

*Das heißt, es müßte nur noch das Bildmaß von  $\lambda^n$  unter  $\phi$  bestimmt werden. Der Beweis hiervon ist allerdings sehr lang!*

## 7 Dichtefunktionen

### 7.1 Der Satz von Radon–Nikodym

In diesem Abschnitt ist  $(\Omega, \mathcal{A})$  ein Messraum und  $\mu$  ein Maß auf  $(\Omega, \mathcal{A})$ .

**Satz 7.1 (Dichtefunktion eines Maßes)**

Sei  $(\Omega, \mathcal{A}, \mu)$  ein Maßraum. Sei  $f : \Omega \rightarrow \mathbb{R}$  eine  $\mathcal{A}$ -messbare Funktion mit

$$f(\omega) \geq 0 \quad \forall \omega \in \Omega$$

Dann wird durch

$$\nu : \mathcal{A} \rightarrow [0; \infty], \quad A \mapsto \int_A f \, d\mu$$

ein Maß  $\nu$  auf  $(\Omega, \mathcal{A})$  definiert. Für

$$\nu(A) = \int_A f \, d\mu \quad \forall A \in \mathcal{A}$$

schreiben wir auch

$$d\nu = f \, d\mu$$

Falls  $d\nu = f \, d\mu$ , so gilt auch

$$\int g \, d\nu = \int g f \, d\mu \quad \forall g \in \mathcal{L}_1(\Omega, \mathcal{A}, \nu) \quad (7.1)$$

**Beweis:**  $\rightsquigarrow$  Übung!

*Hinweis zum Beweis von (7.1): Beweis zuerst für  $g = I_A$ , dann für einfache Funktionen  $g = s$ , dann für  $g \geq 0$ , dann für allgemeines  $g$ .  $\square$*

**Definition 7.2** Sei  $(\Omega, \mathcal{A})$  ein Messraum, seien  $\mu$  und  $\nu$  zwei Maße auf  $(\Omega, \mathcal{A})$  so dass

$$d\nu = f \, d\mu$$

für eine  $\mathcal{A}$ -messbare Funktion

$$f : \Omega \rightarrow \mathbb{R} \quad \text{mit} \quad f(\omega) \geq 0 \quad \forall \omega \in \Omega$$

Dann heißt  $f$  Dichte oder Dichtefunktion von  $\nu$  bzgl.  $\mu$ .

Sei  $(\Omega, \mathcal{A})$  ein Messraum und seien  $\mu$  und  $\nu$  zwei Maße auf  $(\Omega, \mathcal{A})$ .

**Frage:** Existiert dann immer eine Dichte  $f$  von  $\nu$  bzgl.  $\mu$ ?

**Dazu:** Angenommen  $f$  ist eine Dichte von  $\nu$  bzgl.  $\mu$ . Und sei  $A \in \mathcal{A}$ , so dass  $\mu(A) = 0$ .

Dann ist

$$I_A = 0 \quad \mu\text{-f.s.}$$

und somit auch

$$I_A f = 0 \quad \mu\text{-f.s.}$$

und aus Satz 6.12 b) folgt

$$\nu(A) = \int I_A f d\mu = 0$$

Das heißt: Für  $A \in \mathcal{A}$  gilt:

$$\mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0 \quad (7.2)$$

Sei zum Beispiel  $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathfrak{B})$ ,  $\mu = \lambda$  und  $\nu = \delta_1$  (Dirac-Maß, vgl. Abschnitt 4.1). Dann ist

$$\lambda(\{1\}) = 0 \quad \text{aber} \quad \delta_1(\{1\}) = 1$$

$\delta_1$  kann also keine Dichte bzgl.  $\lambda$  besitzen.

**Antwort:** Nein! (7.2) ist eine notwendige Bedingung.

(7.2) ist nicht nur notwendig für die Existenz einer Dichte sondern auch hinreichend. Dies ist die Aussage des folgenden Satzes von Radon–Nikodym. Zunächst aber halten wir Bedingung (7.2) in einer Definition fest:

**Definition 7.3** Seien  $\mu$  und  $\nu$  Maße auf dem Maßraum  $(\Omega, \mathcal{A})$ , so dass für jedes  $A \in \mathcal{A}$  gilt

$$\mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0$$

Dann sagt man

$\nu$  wird dominiert von  $\mu$ .

oder

$\nu$  ist absolut stetig bzgl.  $\mu$ .

Notation:  $\nu \ll \mu$

**Satz 7.4 (Radon–Nikodym)**

Seien  $\mu$  und  $\nu$   $\sigma$ -endliche Maße auf dem Messraum  $(\Omega, \mathcal{A})$ . Dann sind folgende Aussagen äquivalent:

a)  $\nu$  besitzt eine Dichte bzgl.  $\mu$ .

b)  $\nu \ll \mu$

**Bemerkung 7.5** Beim Satz 7.4 in dieser Form ist es eine wichtige Voraussetzung, dass sowohl  $\mu$  als auch  $\nu$   $\sigma$ -endlich sind.

Es gibt noch eine allgemeinere Form dieses Satzes, bei der  $\nu$  auch nicht- $\sigma$ -endlich sein kann. Dann müssten wir allerdings auch Dichtefunktionen  $f$  betrachten, die als Funktionswert  $f(\omega) = \infty$  haben können.

Für die Stochastik spielt das aber kaum eine Rolle. Üblicherweise ist hier  $\nu$  ein Wahrscheinlichkeitsmaß (also ein endliches und damit natürlich auch  $\sigma$ -endliches Maß) und  $\mu$  ist ein  $\sigma$ -endliches Maß – meistens das Zählmaß  $\sharp$  oder das Lebesguemaß  $\lambda$ . Diese beiden für die Stochastik besonders wichtigen Fälle betrachten wir im nächsten Abschnitt.

## 7.2 Integralberechnungen

### 7.2.1 Dirac-Maß

Sei  $(\Omega, \mathcal{A})$  ein Messraum. Sei  $\omega_0 \in \Omega$  und  $\delta_{\omega_0}$  das Dirac-Maß an der Stelle  $\omega_0$ . Dann ist

$$\int_A f d\delta_{\omega_0} = f(\omega_0) \cdot \delta_{\omega_0}(A) \quad (7.3)$$

**Formeller Beweis von (7.3):** Setze  $c_0 := f(\omega_0)$ . Es gilt

$$f = c_0 \cdot \delta_{\omega_0}\text{-f.s.}$$

Aus Satz 6.12 b) und Satz 6.4 b) folgt dann

$$\int_A f d\delta_{\omega_0} \stackrel{\text{S. 6.12}}{=} \int_A c_0 d\delta_{\omega_0} \stackrel{\text{S. 6.4}}{=} c_0 \int_A 1 d\delta_{\omega_0} = f(\omega_0) \cdot \delta_{\omega_0}(A)$$

□

### 7.2.2 Zählmaß

#### a) Endlicher Fall

Sei  $\Omega = \{1, \dots, n\}$ ,  $\mathcal{A} = \mathcal{P}(\Omega)$  und  $\sharp$  das Zählmaß hierauf (vgl. Abschnitt 4.2.1). Für jede Funktion  $f : \{1, \dots, n\} \rightarrow \mathbb{R}$  gilt

$$\int f d\sharp = \sum_{i=1}^n f(i) \quad (7.4)$$

**Formeller Beweis von (7.4):** Setze  $\alpha_i = f(i) \quad \forall i \in \{1, \dots, n\}$ . Wegen

$$f(\omega) = \sum_{i=1}^n \alpha_i I_{\{i\}}(\omega)$$

ist  $f$  eine einfache Funktion und aus der Integraldefinition (6.2) für einfache Funktionen folgt sofort die Behauptung. □

**Hinweis:** Sei  $\nu$  ein weiteres Maß auf  $(\{1, \dots, n\}, \mathcal{P}(\{1, \dots, n\}))$ . Aus  $\nu(A) = 0$  für  $A \subset \{1, \dots, n\}$  folgt  $A = \emptyset$  und daher auch  $\nu(A) = 0$ .

Folglich:  $\nu \ll \sharp$  und  $\nu$  besitzt eine Dichte bzgl.  $\sharp$ .

#### b) Abzählbar unendlicher Fall

Sei  $\Omega = \mathbb{N}$ ,  $\mathcal{A} = \mathcal{P}(\mathbb{N})$  und  $\sharp$  das Zählmaß hierauf (vgl. Abschnitt 4.2.2).

Dann gilt

- Für jede positive Funktion  $f : \mathbb{N} \rightarrow \mathbb{R}$  ist

$$\int f d\sharp = \sum_{i=1}^{\infty} f(i)$$

- Für jedes Maß  $\nu$  auf  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  gilt

$$\nu \ll \sharp$$

- Für jedes  $\sigma$ -endliche Maß  $\nu$  auf  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  ist die Dichte von  $\nu$  bzgl.  $\sharp$  gegeben durch

$$i \mapsto \nu(\{i\})$$

- Für jedes  $\sigma$ -endliche Maß  $\nu$  auf  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  und jede positive Funktion  $g: \mathbb{N} \rightarrow \mathbb{R}$  ist

$$\int g d\nu = \sum_{i=1}^{\infty} g(i) \cdot \nu(\{i\})$$

**Beweis:**  $\rightsquigarrow$  Übung!

□

### 7.2.3 Lebesgue-Maß und Lebesgue-Dichten

Sei  $(\Omega, \mathcal{A}) = (\mathbb{R}^n, \mathfrak{B}^{\otimes n})$ .

Sei  $\nu$  ein  $\sigma$ -endliches Maß auf  $(\mathbb{R}^n, \mathfrak{B}^{\otimes n})$ .  $\nu$  heißt absolut stetig, falls

$$\nu \ll \lambda^n$$

Das heißt:  $\nu$  ist absolut stetig genau dann, wenn  $\nu$  eine Dichte bzgl. dem Lebesgue-Maß  $\lambda^n$  besitzt:

$$d\nu = f d\lambda^n, \quad f \geq 0 \quad \mathfrak{B}^{\otimes n} \text{-messbar}$$

Eine solche Dichte  $f$  heißt Lebesgue-Dichte.

#### Beispiel 7.6 (1-dim. Normalverteilung)

Sei  $n = 1$  und  $\nu = \mathcal{N}(a, \sigma^2)$ . Es gilt

$$d\mathcal{N}(a, \sigma^2) = f_{a, \sigma^2} d\lambda$$

wobei

$$f_{a, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) \quad \forall x \in \mathbb{R}$$

Aus Satz 7.1 folgt für  $g \in \mathcal{L}_1(\mathbb{R}, \mathfrak{B}, \mathcal{N}(a, \sigma^2))$

$$\int g d\mathcal{N}(a, \sigma^2) \stackrel{(7.1)}{=} \int g f_{a, \sigma^2} d\lambda = \int g(x) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) \lambda(dx) \quad (7.5)$$

### 7.3 Einschub: Zufallsvariablen

Seien  $(\Omega, \mathcal{A})$  und  $(\Omega', \mathcal{A}')$  Messräume. In der Stochastik heißen die  $\mathcal{A}/\mathcal{A}'$ -messbaren Abbildungen

$$X: \Omega \rightarrow \Omega'$$

auch Zufallsvariablen.

Meistens ist in der Statistik  $\Omega' = \mathbb{R}^n$  oder  $\Omega' = \{1, \dots, n\}$ .

- Was ist eigentlich die exakte mathematische Formalisierung für den Ausdruck

„Sei  $X$  eine standardnormalverteilte Zufallsvariable.“  
bzw.  $X \sim \mathcal{N}(0; 1)$ .

Das bedeutet:

Wir haben einen Messraum  $(\Omega, \mathcal{A})$  und ein Wahrscheinlichkeitsmaß  $P$  auf  $(\Omega, \mathcal{A})$ .  $X$  ist dann eine  $\mathcal{A}$ -messbare Funktion

$$X : \Omega \rightarrow \mathbb{R}$$

und das Bildmaß von  $P$  unter  $X$  ist („Die Verteilung von  $X$ “)

$$P_X := X(P) = \mathcal{N}(0, 1)$$

$$\text{D.h.: } P(X^{-1}(B)) = (\mathcal{N}(0, 1))(B) \quad \forall B \in \mathfrak{B}$$

Normalerweise wird  $(\Omega, \mathcal{A})$  und  $P$  nicht erwähnt oder gar genauer angegeben.

Der Raum  $\Omega$  enthält die Ereignisse  $\omega \in \Omega$ , die irgendwie zufällig geschehen. Von diesem „irgendwie“ weiß man nicht viel. Im Experiment beobachten wir nicht das zugrundeliegende Ereignis  $\omega$  sondern nur die Auswirkung dieses Ereignisses  $\omega$ ; die Beobachtung ist ein Datenpunkt

$$x = X(\omega)$$

Wir berechnen nun den Erwartungswert von  $X$ . Sei dazu

$$g : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto g(x) \quad \text{mit} \quad g(x) = x \quad \forall x \in \mathbb{R}$$

Der Erwartungswert von  $X$  ist

$$\begin{aligned} \mathbb{E}[X] &= \int X dP = \int_{\Omega} X(\omega) P(d\omega) = \int_{\Omega} g \circ X(\omega) P(d\omega) \\ &\stackrel{(6.9)}{=} \int_{\mathbb{R}} g(x) [X(P)](dx) = \int_{\mathbb{R}} x (\mathcal{N}(0, 1))(dx) \\ &\stackrel{(7.5)}{=} \int_{\mathbb{R}} x e^{-\frac{1}{2}x^2} \lambda(dx) = \int_{-\infty}^{\infty} x e^{-\frac{1}{2}x^2} dx = \dots \end{aligned}$$

Normalerweise wird nicht nur ein Wert beobachtet, d.h. man hat nicht nur einen Datenpunkt  $x$  sondern mehrere Daten

$$x_1, \dots, x_n$$

Dies wird dann durch entsprechend viele Zuvallsvariablen

$$X_i : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathfrak{B}), \quad i \in \{1; \dots; n\}$$

modelliert.



Man beachte hierbei besonders, dass die verschiedenen  $X_i$  alle auf demselben Messraum  $(\Omega, \mathcal{A})$  definiert sind. Und ein und dasselbe Ereignis  $\omega_0 \in \Omega$  hat dann zu den Beobachtungen

$$x_1 = X_1(\omega_0), \quad x_2 = X_2(\omega_0), \quad x_3 = X_3(\omega_0), \quad \dots, \quad x_n = X_n(\omega_0)$$

geführt. Im *i.i.d.* - Fall sind dann  $X_1, \dots, X_n$  stochastisch unabhängig und

$$X_1(P) = \dots = X_n(P)$$

- Sei nun  $X$  eine binäre Zufallsvariable, die nach  $\text{Bin}(1; \frac{1}{5})$  verteilt ist. Dabei bedeute

$$X = 1 \quad \text{dass es regnet}$$

und

$$X = 0 \quad \text{dass es nicht regnet}$$

Die Wahrscheinlichkeit, dass es regnet ist  $\frac{1}{5}$ .

Wir haben also einen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  und das Bildmaß von  $X$  bzgl.  $P$  ist

$$X(P) = \text{Bin}(1; \frac{1}{5})$$

Dabei enthält  $\Omega$  alle möglichen Gesamtwetterlagen (inkl. physikalischer und chemischer Vorgänge z.B. auch auf molekularer Ebene). Jedes  $\omega_0 \in \Omega$  beschreibt eine spezifische Wettersituation. Naturgesetze bestimmen dann, ob es in der Situation  $\omega_0$  regnet oder nicht:

$$X(\omega_0) = 1 \quad \text{oder} \quad X(\omega_0) = 0$$

Wie wahrscheinlich welche Wettersituation  $\omega$  ist, wird von dem Wahrscheinlichkeitsmaß  $P$  auf  $(\Omega, \mathcal{A})$  angegeben. Die Wahrscheinlichkeit, dass eine Wettersituation vorliegt, die zu Regen führt, ist dann

$$P(\{\omega \in \Omega \mid X(\omega) = 1\}) = P(X^{-1}(\{1\})) = [X(P)](\{1\})$$

Normalerweise wird  $\Omega$  nicht näher spezifiziert und  $P$  kennen wir gar nicht; aber aufgrund statistischer Untersuchungen „wissen“ wir, dass

$$X(P) = \text{Bin}(1; \frac{1}{5})$$

also

$$[X(P)](\{1\}) = \frac{1}{5}$$

und das reicht uns.

## 7.4 Transformationsatz für Dichten

Ist man an einer Transformation  $T$  einer Zufallsvariablen  $X$  interessiert, so gibt die Transformationsformel aus Satz 6.15 keine praktische Hilfestellung im für die Statistik relevanten Fall, dass  $P_X := X(P)$  eine bekannte Lebesgue-Dichte besitzt. Der Satz trifft eine Aussage auf dem Niveau der beteiligten Wahrscheinlichkeitsmaße:

$$\int_B f(y)[T(P_X)](dy) = \int_{T^{-1}(B)} (f \circ T)(x)P_X(dx) \quad \forall B \in \mathfrak{B}$$

Der folgende Satz zeigt, wie sich die Dichte des transformierten Wahrscheinlichkeitsmaßes  $T(P_X)$  aus der Dichte von  $P_X$  berechnen lässt.

**Satz 7.7** Sei  $f : \mathbb{R}^p \rightarrow \mathbb{R}, (x_1, \dots, x_p) \mapsto f(x_1, \dots, x_p)$  die  $\lambda^p$ -Dichte eines Wahrscheinlichkeitsmaßes  $P_X$ . Seien  $G, G' \in \mathfrak{B}^{\otimes p}$  offen und die Abbildung

$$T : G \rightarrow G' \\ (x_1, \dots, x_p) \mapsto \left( \underbrace{T_1(x_1, \dots, x_p)}_{=:y_1}, \dots, \underbrace{T_p(x_1, \dots, x_p)}_{=:y_p} \right)$$

bijektiv und samt  $T^{-1}$  messbar und differenzierbar.

Dann gilt für die  $\lambda^p$ -Dichte  $g$  von  $T(P_X)$ :

$$g(y_1, \dots, y_p) = \left| \det J_{T^{-1}}(y_1, \dots, y_p) \right| \cdot f\left(T^{-1}(y_1, \dots, y_p)\right) \\ = \left| \det J_T(T^{-1}(y_1, \dots, y_p)) \right|^{-1} \cdot f\left(T^{-1}(y_1, \dots, y_p)\right)$$

**Bemerkung 7.8** Im eindimensionalen Fall ( $p = 1$ ) vereinfacht sich die Dichtetransformationsformel zu

$$g(y) = \left| (T^{-1})'(y) \right| \cdot f\left(T^{-1}(y)\right)$$

**Beispiel 7.9** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Sei  $X$  eine  $\mathcal{A}$ -messbare Zufallsvariable, so dass das Wahrscheinlichkeitsmaß  $X(P)$  die Lebesgue-Dichte  $f(x) = \theta e^{-\theta x} I_{(0; \infty)}(x)$  besitzt; oder kurz:  $X \sim \text{Exp}(\theta)$ .

Die Abbildung

$$T : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \\ x \mapsto x^2$$

ist bijektiv mit Umkehrfunktion

$$T^{-1} : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \\ y \mapsto \sqrt{y}$$

mit Ableitung

$$\frac{d}{dy} T^{-1}(y) = \frac{1}{2} y^{-\frac{1}{2}} \quad .$$

Mit dem Satz 7.7 können wir nun einfach die zur Zufallsvariable  $Y := T(X)$  gehörige Lebesgue-Dichte angeben:

$$g(y) = \left| \frac{1}{2} y^{-\frac{1}{2}} \right| \cdot f(\sqrt{y}) = \frac{1}{2} y^{-\frac{1}{2}} \theta e^{-\theta \sqrt{y}}$$

für  $y > 0$ . Für  $y \leq 0$  setze  $g(y) = 0$ .

Eine Verallgemeinerung der Transformationsformel auf nicht-bijektive Abbildungen  $T$  ist möglich, indem man  $T$  in bijektive Abbildungen aufspaltet. Dies wird anhand von zwei Beispielen dargestellt.

**Beispiel 7.10** Betrachte die  $\lambda$ -Dichte der Laplace-Verteilung (auch Doppelpotentialverteilung genannt):

$$f(x) = \frac{1}{2} \theta e^{-\theta |x|} \quad x \in \mathbb{R}$$

Die Abbildung

$$\begin{aligned} T : \mathbb{R} &\rightarrow \mathbb{R}_0^+ \\ x &\mapsto x^2 \end{aligned}$$

ist nicht bijektiv, aber eine Aufspaltung in zwei bijektive Abbildungen ist möglich:

$$\begin{array}{ll} T_- : \mathbb{R}^- \rightarrow \mathbb{R}^+ & T_+ : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \\ x \mapsto x^2 & x \mapsto x^2 \\ T_-^{-1} : \mathbb{R}^+ \rightarrow \mathbb{R}^- & T_+^{-1} : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \\ y \mapsto -\sqrt{y} & y \mapsto +\sqrt{y} \end{array}$$

Somit ist die transformierte Dichte

$$\begin{aligned} g(y) &= \left| \frac{d}{dy} T_-^{-1}(y) \right| \cdot f(T_-^{-1}(y)) + \left| \frac{d}{dy} T_+^{-1}(y) \right| \cdot f(T_+^{-1}(y)) \\ &= \left| -\frac{1}{2} y^{-\frac{1}{2}} \right| \cdot \frac{1}{2} \theta e^{-\theta |-\sqrt{y}|} + \left| \frac{1}{2} y^{-\frac{1}{2}} \right| \cdot \frac{1}{2} \theta e^{-\theta |\sqrt{y}|} \\ &= \frac{1}{2} y^{-\frac{1}{2}} \theta e^{-\theta \sqrt{y}} \end{aligned}$$

für  $y \in \mathbb{R}^+$ , sonst setze  $g(y) = 0$ . Es ergibt sich also die gleiche Dichte wie in Beispiel 7.9.

**Beispiel 7.11** Uns interessiert im Folgenden die Dichte der Orderstatistik, d.h. der Abbildung

$$(X_1, \dots, X_n) \xrightarrow{\circ} (X_{(1)}, \dots, X_{(n)}),$$

wobei  $(X_1, \dots, X_n)$  die gemeinsame Dichte  $f$  haben.

Wir betrachten zunächst den Spezialfall  $n = 2$ . Definiere dazu die Mengen

$$G_1 := \{(x_1, x_2) \in \mathbb{R} \times \mathbb{R} : x_1 \leq x_2\}$$

und

$$G_2 := \{(x_1, x_2) \in \mathbb{R} \times \mathbb{R} : x_1 > x_2\},$$

wobei also  $G_1$  die Vektoren in richtiger Reihenfolge enthält. Dann sind

$$\begin{array}{ll} o_1 : G_1 \rightarrow G_1 & o_2 : G_2 \rightarrow G_1 \\ (x_1, x_2) \mapsto (x_1, x_2) & (x_2, x_1) \mapsto (x_1, x_2) \\ o_1^{-1} : G_1 \rightarrow G_1 & o_2^{-1} : G_1 \rightarrow G_2 \\ (x_{(1)}, x_{(2)}) \mapsto (x_{(1)}, x_{(2)}) & (x_{(1)}, x_{(2)}) \mapsto (x_{(2)}, x_{(1)}) \end{array}$$

zwei bijektive, stetige Abbildungen mit differenzierbaren Umkehrabbildungen. Die Anwendung von Satz 7.7 auf jede einzelne Abbildung ergibt für  $o_k, k \in \{1, 2\}$ :

$$\underbrace{\left| \det J_{o_k^{-1}}(x_{(1)}, x_{(2)}) \right|}_{=1} \cdot f\left(o_k^{-1}(x_{(1)}, x_{(2)})\right)$$

Die Dichte der Orderstatistik ist dann

$$g(x_{(1)}, x_{(2)}) = \sum_{k=1}^2 f\left(o_k^{-1}(x_{(1)}, x_{(2)})\right) = f(x_{(1)}, x_{(2)}) + f(x_{(2)}, x_{(1)})$$

für  $x_{(1)} \leq x_{(2)}$ .

In Kapitel 8 wird die Unabhängigkeit (von Zufallsvariablen) behandelt. Ist diese gegeben, so folgt

$$g(x_{(1)}, x_{(2)}) = 2 \cdot f_{X_{(1)}}(x_{(1)})f_{X_{(2)}}(x_{(2)}) \quad .$$

Sind die beiden Zufallsvariablen unabhängig und identisch mit Dichte  $f_X$  verteilt, so vereinfacht sich die Dichte der Orderstatistik zu

$$g(x_{(1)}, x_{(2)}) = 2 \cdot f_X(x_{(1)})f_X(x_{(2)}) \quad .$$

Im allgemeinen Fall von  $n$  Zufallsvariablen gibt es  $n!$  Permutationen für die Reihenfolge und somit eine Aufspaltung in  $n!$  einzelne bijektive Abbildungen  $G_1, \dots, G_{n!}$ . Die Dichte der Orderstatistik für unabhängige und identisch verteilte Zufallsvariablen ist dann

$$g(x_{(1)}, \dots, x_{(n)}) = n! \cdot f_X(x_{(1)}) \cdots f_X(x_{(n)}) \quad x_{(1)} \leq \dots \leq x_{(n)} \quad .$$

Sind beispielsweise  $X_1, \dots, X_n \stackrel{iid}{\sim} U(0; 1)$ , so gilt

$$g(x_{(1)}, \dots, x_{(n)}) = \begin{cases} n! & \text{falls } x_{(1)} \leq \dots \leq x_{(n)} \\ 0 & \text{sonst} \end{cases}$$

## 7.5 Ausblick: Gemischte Verteilungen

Sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathfrak{B})$ . Seien  $x_1 \in \mathbb{R}, x_2 \in \mathbb{R}, \dots, x_k \in \mathbb{R}$ , so dass gilt:

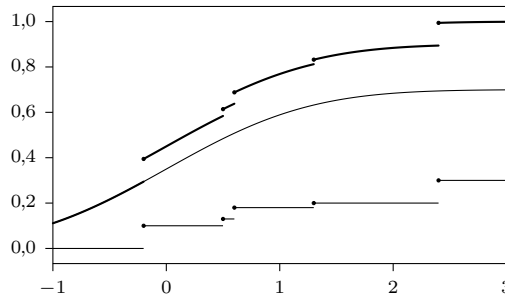
a)  $P$  ist auf  $\mathbb{R} \setminus \{x_1, \dots, x_k\}$  absolut stetig, also

$$P(B) = \int_B f(x) \lambda(dx) \quad \forall B \in \mathfrak{B} \quad \text{mit } x_j \notin B \quad \forall j \in \{1, \dots, k\}$$

für eine Lebesgue-Dichte  $f$ .

b)  $P(\{x_j\}) > 0 \quad \forall j \in \{1, \dots, k\}$

$P$  besitzt also einen absolut stetigen und einen diskreten Teil;  $P$  besitzt somit keine Lebesgue-Dichte. Vergleiche hierzu auch den Satz von Radon-Nikodym (Satz 7.4).



Gibt es ein anderes Maß  $\mu$  auf  $(\mathbb{R}, \mathfrak{B})$  bzgl. welchem  $P$  eine Dichte  $\hat{f}$  besitzt? Betrachte

$$\mu := \lambda + \delta_{x_1} + \dots + \delta_{x_k}$$

Klar:  $\mu$  ist ein  $\sigma$ -endliches Maß!

Wir zeigen nun

$$P \ll \mu \tag{7.6}$$

Sei dazu  $B \in \mathfrak{B}$ , so dass  $\mu(B) = 0$ . Also

$$\underbrace{\lambda(B)}_{\geq 0} + \underbrace{\delta_{x_1}(B)}_{\geq 0} + \dots + \underbrace{\delta_{x_k}(B)}_{\geq 0} = 0$$

und somit

$$\lambda(B) = 0, \quad \underbrace{\delta_{x_1}(B) = 0, \delta_{x_2}(B) = 0, \dots, \delta_{x_k}(B) = 0}_{\Rightarrow x_j \notin B \quad \forall j \in \{1, \dots, k\}}$$

Schließlich folgt daher

$$P(B) \stackrel{x_j \notin B}{=} \int_B f(x) \lambda(dx) = \int f(x) I_B(x) \lambda(dx) \stackrel{\text{S. 6.12b)}}{=} 0$$

Also ist

$$P \ll \lambda + \delta_{x_1} + \dots + \delta_{x_k} = \mu$$

und damit existiert auch eine Dichte von  $P$  bzgl.  $\mu$ .

Es gilt

$$P(B) = \int_B f(x) \lambda(dx) \quad \forall B \in \mathfrak{B} \quad \text{mit } x_j \notin B \quad \forall j \in \{1, \dots, k\}$$

und

$$\alpha_j := P(\{x_j\}) > 0 \quad j \in \{1, \dots, k\}$$

Setze nun

$$\hat{f}(x) := f(x) \quad \forall x \in \mathbb{R} \setminus \{x_1, \dots, x_k\}$$

und

$$\hat{f}(x) := \alpha_j \quad \text{falls } x = x_j, \quad i \in \{1, \dots, k\}$$

Dabei ist  $f$  bereits passend normiert, so dass

$$\int_{\mathbb{R} \setminus \{x_1, \dots, x_k\}} f d\lambda = \int_{\mathbb{R}} f d\lambda = 1 - \sum_{j=1}^k \alpha_j$$

Dann ist

$$\hat{f} : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto \hat{f}(x)$$

die Dichte von  $P$  bzgl.  $\mu = \lambda + \delta_{x_1} + \dots + \delta_{x_k}$ , das heißt

$$P(B) = \int \hat{f}(x) \mu(dx) \quad \forall B \in \mathfrak{B}$$

**Beweis:**

$$\begin{aligned} P(B) &= \int_B \hat{f} d(\lambda + \delta_{x_1} + \dots + \delta_{x_k}) \\ &= \int_B \hat{f} d\lambda + \sum_{j=1}^k \int_B \hat{f} d\delta_{x_j} \\ &\stackrel{\text{Satz 6.12}}{=} \int_{B \setminus \{x_1, \dots, x_k\}} f d\lambda + \sum_{j=1}^k \alpha_j \cdot \delta_{x_j}(B) \\ &= \int_{B \setminus \{x_1, \dots, x_k\}} f d\lambda + \sum_{j: x_j \in B} P(\{x_j\}) \end{aligned}$$

□

**Fazit:** Wir haben nun eine mathematische Theorie, mit der wir einheitlich alle erdenklichen Verteilungen behandeln können. Lästige (und verwirrende) Fallunterscheidungen (stetig, diskret, ...) sind damit nicht mehr nötig!

## 7.6 Ausblick: Stochastische Prozesse

In diesem Abschnitt soll am Beispiel „Stochastische Prozesse“ kurz angedeutet werden, wie allgemein und weitreichend der maßtheoretische Aufbau tatsächlich ist.

Sei  $T = [a; b] \subset \mathbb{R}$  mit  $a \in \mathbb{R}$ ,  $b \in \mathbb{R}$  und  $a < b$ .  $T$  heißt *Parameterraum*. Man kann sich  $T$  als einen Zeitraum vorstellen,  $t = a$  ist der Anfangszeitpunkt,  $t = b$  ist der Endzeitpunkt. Sei  $(\Omega, \mathcal{A})$  ein Messraum und sei für jedes  $t \in T$

$$\begin{aligned} X_t &: \Omega \rightarrow \mathbb{R} \\ \omega &\mapsto X_t(\omega) \end{aligned}$$

eine  $\mathcal{A}$ -messbare Zufallsvariable. Die Gesamtheit aller Zufallsvariablen

$$X_t, \quad t \in T$$

heißt *Stochastischer Prozess*.

Man bezeichnet  $(X_t)_{t \in T}$  auch als Familie von Zufallsvariablen.

$X_t : \Omega \rightarrow \mathbb{R}$  kann zum Beispiel jeweils eine Zufallsvariable sein, die den Wert einer Aktie zum Zeitpunkt  $t$  beschreibt: Ist das zufällige Ereignis  $\omega_0$  eingetreten, dann beträgt der Wert der Aktie z.B.

$$X(\omega_0) = 164$$

Hält man ein  $\omega_0 \in \Omega$  fest und betrachtet dann die Wertentwicklung im Zeitraum  $T$  bei vorliegendem  $\omega_0$ , so erhält man eine Funktion

$$X_\bullet(\omega_0) : [a; b] \rightarrow \mathbb{R}, \quad t \mapsto X_t(\omega_0)$$

Eine solche Funktion  $t \mapsto X_t(\omega_0)$  bei festgehaltenem  $\omega_0$  nennt man *Pfad*. Im Folgenden nehmen wir nun an, dass für jedes  $\omega \in \Omega$  die Funktion

$$X_\bullet(\omega) : [a; b] \rightarrow \mathbb{R}, \quad t \mapsto X_t(\omega)$$

stetig ist. Das heißt, wir betrachten der Einfachheit halber jetzt mal nur stetige Pfade (z.B. bei Aktienkurse ist das durchaus sinnvoll und üblich). Sei

$$\mathcal{C}([a; b])$$

die Menge aller stetigen Funktionen auf  $T = [a; b]$ . Also

$$\mathcal{C}([a; b]) = \{f : [a; b] \rightarrow \mathbb{R} \mid f \text{ ist stetig}\}$$

und für jedes  $\omega \in \Omega$  ist

$$X_\bullet(\omega) \in \mathcal{C}([a; b])$$

Natürlich ist es recht umständlich und schwierig bei einem Stochastischen Prozess immer an „Familien von Zufallsvariablen“

$$(X_t)_{t \in T}$$

zu denken. Das sollte man deswegen auch nicht tun. Stattdessen definieren wir eine zusätzliche Zufallsvariable

$$X : \Omega \rightarrow \mathcal{C}([a; b]), \quad \omega \mapsto X_{\bullet}(\omega)$$

Das heißt,  $X$  ist eine Zufallsvariable, die einem zufälligen Ereignis  $\omega$  gleich seinen ganzen Pfad zuordnet. Damit beschreibt  $X$  schon unseren ganzen Stochastischen Prozess. Statt bei einem Stochastischen Prozess an eine Familie von Zufallsvariablen  $(X_t)_{t \in T}$  zu denken, ist es einfacher, nur an dieses  $X$  zu denken.

Setzen wir nun

$$\Omega' := \mathcal{C}([a; b])$$

Jetzt brauchen wir noch eine geeignete  $\sigma$ -Algebra  $\mathcal{A}'$  auf  $\Omega'$ . Die genaue Definition einer geeigneten  $\sigma$ -Algebra  $\mathcal{A}'$  würde jetzt zu weit führen; wichtig ist im Moment nur: es gibt so etwas.

Damit ist unser gesamter Stochastischer Prozess  $(X_t)_{t \in T}$  auch ganz einfach darstellbar als eine  $\mathcal{A}/\mathcal{A}'$ -messbare Funktion

$$X : \Omega \rightarrow \Omega'$$

Ein Stochastischer Prozess ist also gar nichts neues. Mit solchen  $\mathcal{A}/\mathcal{A}'$ -messbaren Funktion beschäftigen wir uns schon die gesamte Vorlesung über.

Wenn zum Beispiel  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  ist, so können wir nun auch wie gewohnt das Bildmaß  $P'$  von  $X$  unter  $P$  definieren:

$$P'(A') = P(X^{-1}(A')) \quad \forall A' \in \mathcal{A}'$$

Dass  $P'$  eigentlich was ganz kompliziertes ist – nämlich ein Wahrscheinlichkeitsmaß auf einem Raum von Funktionen, muss uns (auf dieser Abstraktionsebene) nicht stören. Mit den maßtheoretischen Grundlagen aus dieser Vorlesung ist so etwas immer noch gut handhabbar.

Mehr zu Stochastische Prozesse findet sich z.B. in van der Vaart (1997) ab Kapitel 18.



## 8 Unabhängigkeit

In der Statistik ist die Unabhängigkeit von Ereignissen oder Zufallsvariablen ein zentraler Begriff und schon in der Schule lernt man, dass Mengen  $A_1, \dots, A_n$  unabhängig sind, wenn

$$\begin{aligned}
 P(A_1 \cap A_2) &= P(A_1) \cdot P(A_2) \\
 P(A_1 \cap A_3) &= P(A_1) \cdot P(A_3) \\
 P(A_1 \cap A_4) &= P(A_1) \cdot P(A_4) \\
 &\dots \quad \dots \\
 P(A_1 \cap A_2 \cap A_3) &= P(A_1) \cdot P(A_2) \cdot P(A_3) \\
 P(A_1 \cap A_2 \cap A_4) &= P(A_1) \cdot P(A_2) \cdot P(A_4) \\
 P(A_1 \cap A_2 \cap A_5) &= P(A_1) \cdot P(A_2) \cdot P(A_5) \\
 &\dots \quad \dots \\
 P(A_1 \cap A_3 \cap A_4) &= P(A_1) \cdot P(A_3) \cdot P(A_4) \\
 P(A_1 \cap A_3 \cap A_5) &= P(A_1) \cdot P(A_3) \cdot P(A_5) \\
 &\dots \quad \dots \\
 &\dots \quad \dots \\
 P(A_2 \cap A_3) &= P(A_2) \cdot P(A_3) \\
 P(A_2 \cap A_4) &= P(A_2) \cdot P(A_4) \\
 &\dots \quad \dots \\
 &\dots \quad \dots \\
 &\dots \quad \dots
 \end{aligned}$$

oder in formaler Schreibweise

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i) \quad \text{für alle Teilmengen } I \subset \{1, 2, \dots, n\} \quad (8.1)$$

### 8.1 Unabhängigkeit bei Mengensystemen

In diesem Abschnitt ist  $(\Omega, \mathcal{A})$  ein Messraum und  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$ .

Seien  $\mathcal{D}_1 \subset \mathcal{A}$ ,  $\mathcal{D}_2 \subset \mathcal{A}$ ,  $\dots$ ,  $\mathcal{D}_n \subset \mathcal{A}$ . So ein  $\mathcal{D}_i$  bezeichnet man auch als *Mengensystem*.

In naheliegender Weise wird nun die Unabhängigkeit von ganzen Mengensystemen definiert:  $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$  sind *stochastisch unabhängig bzgl.  $P$* , falls gilt

$$D_1, D_2, \dots, D_n \text{ sind stochastisch unabhängig bzgl. } P$$

für alle  $D_1 \in \mathcal{D}_1$ ,  $D_2 \in \mathcal{D}_2$ ,  $\dots$ ,  $D_n \in \mathcal{D}_n$ .

Entsprechend Gleichung (8.1) lautet die formale Definition

**Definition 8.1** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und seien

$$\mathcal{D}_1 \subset \mathcal{A}, \quad \mathcal{D}_2 \subset \mathcal{A}, \quad \dots, \quad \mathcal{D}_n \subset \mathcal{A}$$

$\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$  heißen stochastisch unabhängig bzgl.  $P$ , falls

$$P\left(\bigcap_{i \in I} D_i\right) = \prod_{i \in I} P(D_i) \quad \forall D_i \in \mathcal{D}_i \quad \forall i \in I \quad (8.2)$$

für alle Teilmengen  $I \subset \{1, 2, \dots, n\}$ .

**Satz 8.2 ( $\cap$ -stabile Erzeuger und Unabhängigkeit von  $\sigma$ -Algebren)**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und seien

$$\mathcal{A}_1 \subset \mathcal{A}, \quad \mathcal{A}_2 \subset \mathcal{A}, \quad \dots, \quad \mathcal{A}_n \subset \mathcal{A}$$

weitere  $\sigma$ -Algebren auf  $\Omega$  und sei  $\mathcal{E}_i$  jeweils ein  $\cap$ -stabiler Erzeuger von  $\mathcal{A}_i$ .

Seien  $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_n$  stochastisch unabhängig bzgl.  $P$ .

Dann sind auch  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$  stochastisch unabhängig bzgl.  $P$ .

## 8.2 Unabhängigkeit bei Zufallsvariablen

Sei  $(\Omega, \mathcal{A})$  ein Messraum und sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$ .

Seien  $(\Omega'_1, \mathcal{A}'_1)$  und  $(\Omega'_2, \mathcal{A}'_2)$  zwei weitere Messräume. Sei

$$X_1 : \Omega \rightarrow \Omega'_1$$

eine  $\mathcal{A}/\mathcal{A}'_1$ -messbare Abbildung (Zufallsvariable) und sei

$$X_2 : \Omega \rightarrow \Omega'_2$$

eine  $\mathcal{A}/\mathcal{A}'_2$ -messbare Abbildung (Zufallsvariable)

Unabhängigkeit zwischen zwei Zufallsvariablen  $X_1$  und  $X_2$  soll in etwa bedeuten, dass

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2)$$

Maßtheoretisch sauber aufgeschrieben heißt das

$$\begin{aligned} P(\{\omega \in \Omega \mid X_1(\omega) = x_1\} \cap \{\omega \in \Omega \mid X_2(\omega) = x_2\}) \\ = P(\{\omega \in \Omega \mid X_1(\omega) = x_1\}) \cdot P(\{\omega \in \Omega \mid X_2(\omega) = x_2\}) \end{aligned}$$

bzw.

$$P\left(X_1^{-1}(\{x_1\}) \cap X_2^{-1}(\{x_2\})\right) = P\left(X_1^{-1}(\{x_1\})\right) \cdot P\left(X_2^{-1}(\{x_2\})\right) \quad (8.3)$$

bzw.

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) \quad (8.4)$$

für  $A_1 = X_1^{-1}(\{x_1\})$  und  $A_2 = X_2^{-1}(\{x_2\})$ .

Damit sind wir also wieder bei der Unabhängigkeit von Ereignismengen angelangt. Auf diese Weise können wir die stochastische Unabhängigkeit von Zufallsvariablen auf die stochastische Unabhängigkeit von Mengen (bzw. Mengensystemen) zurückführen.

**Definition 8.3 (Stochastische Unabhängigkeit)**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Seien

$$(\Omega'_1, \mathcal{A}'_1), \quad (\Omega'_2, \mathcal{A}'_2), \quad \dots, \quad (\Omega'_n, \mathcal{A}'_n)$$

Messräume und sei für  $i \in \{1, 2, \dots, n\}$

$$X_i : \Omega \rightarrow \Omega'_i$$

jeweils eine  $\mathcal{A}/\mathcal{A}'_i$ -messbare Zufallsvariable. Setze für  $i \in \{1, 2, \dots, n\}$  jeweils

$$\mathcal{A}_i := X_i^{-1}(\mathcal{A}'_i) = \{X_i^{-1}(A'_i) \mid A'_i \in \mathcal{A}'_i\} \subset \mathcal{A}$$

$X_1, X_2, \dots, X_n$  heißen stochastisch unabhängig bzgl.  $P$ , falls  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$  stochastisch unabhängig bzgl.  $P$  sind.

Mit anderen Worten:

$X_1, X_2, \dots, X_n$  sind stochastisch unabhängig bzgl.  $P$  genau dann, wenn

$$P\left(\bigcap_{i=1}^n X_i^{-1}(A'_i)\right) = \prod_{i=1}^n P(X_i^{-1}(A'_i)) \quad \forall A'_i \in \mathcal{A}'_i \quad \forall i \in \{1, 2, \dots, n\}$$

**Bemerkung 8.4**

- a) Das Mengensystem  $\mathcal{A}_i$  ist eine  $\sigma$ -Algebra auf  $\Omega$ .
- b) Hier tauchen jetzt keine Teilindexmengen  $I \subset \{1, 2, \dots, n\}$  mehr auf wie in Gleichung (8.2). Wenn obige Gleichung erfüllt ist, so gilt sie automatisch für alle Teilmengen  $I$  mit

$$A'_i = \Omega'_i \quad \forall i \in \{1, 2, \dots, n\} \setminus I$$

da

$$X_i^{-1}(\Omega'_i) = \{\omega \in \Omega \mid X_i(\omega) \in \Omega'_i\} = \Omega \in \mathcal{A}_i$$

Und wieder genügt ein  $\cap$ -stabiler Erzeuger  $\mathcal{E}'_i$  statt der ganzen  $\sigma$ -Algebra  $\mathcal{A}'_i$ :

**Satz 8.5 ( $\cap$ -stabile Erzeuger und Unabhängigkeit)**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Seien

$$(\Omega'_1, \mathcal{A}'_1), \quad (\Omega'_2, \mathcal{A}'_2), \quad \dots, \quad (\Omega'_n, \mathcal{A}'_n)$$

Messräume und sei für  $i \in \{1, 2, \dots, n\}$

$$X_i : \Omega \rightarrow \Omega'_i$$

jeweils eine  $\mathcal{A}/\mathcal{A}'_i$ -messbare Zufallsvariable. Sei außerdem für  $i \in \{1, 2, \dots, n\}$  jeweils  $\mathcal{E}'_i$  ein  $\cap$ -stabiler Erzeuger von  $\mathcal{A}'_i$  mit  $\Omega'_i \in \mathcal{E}'_i$ . Dann gilt:

$X_1, X_2, \dots, X_n$  sind stochastisch unabhängig bzgl.  $P$ , falls

$$P\left(\bigcap_{i=1}^n X_i^{-1}(E'_i)\right) = \prod_{i=1}^n P(X_i^{-1}(E'_i)) \quad \forall E'_i \in \mathcal{E}'_i \quad \forall i \in \{1, 2, \dots, n\}$$

Der Vollständigkeit halber halten wir noch fest:

**Satz 8.6** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Seien

$$X_1 \in \mathcal{L}_1(\Omega, \mathcal{A}, P), \quad X_2 \in \mathcal{L}_1(\Omega, \mathcal{A}, P), \quad \dots, \quad X_n \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$$

stochastisch unabhängige Zufallsvariablen. Dann ist

$$\mathbb{E}_P \left[ \prod_{i=1}^n X_i \right] = \prod_{i=1}^n \mathbb{E}_P [X_i] \quad (8.5)$$

und

$$\text{Var} \left( \sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) \quad (8.6)$$

(Hierbei ist (8.6) ein Spezialfall des Satzes von Bienaymé, bei dem nur paarweise Unkorreliertheit anstelle von Unabhängigkeit verlangt wird.)

# 9 Produktmaße

## 9.1 2-faches Produkt $\Omega_1 \times \Omega_2$

### 9.1.1 Produkte von $\sigma$ -Algebren

Seien  $(\Omega_1, \mathcal{A}_1)$  und  $(\Omega_2, \mathcal{A}_2)$  Messräume.

Zum Beispiel

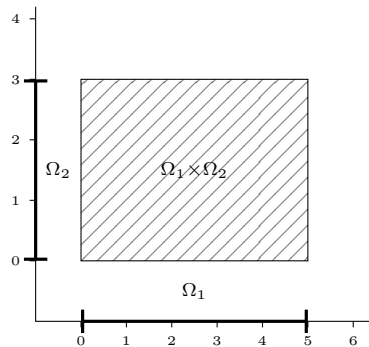
$$\Omega_1 = [0; 5], \quad \Omega_2 = [0; 3]$$

Wir betrachten nun den Produktraum (kartesisches Produkt) von  $\Omega_1$  und  $\Omega_2$

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) \mid \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

Zum Beispiel

$$\Omega_1 \times \Omega_2 = [0; 5] \times [0; 3] = \{(\omega_1, \omega_2) \in \mathbb{R}^2 \mid 0 \leq \omega_1 \leq 5, 0 \leq \omega_2 \leq 3\}$$

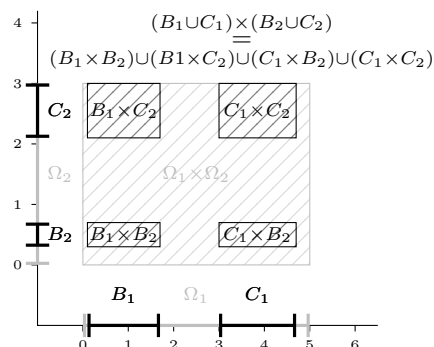
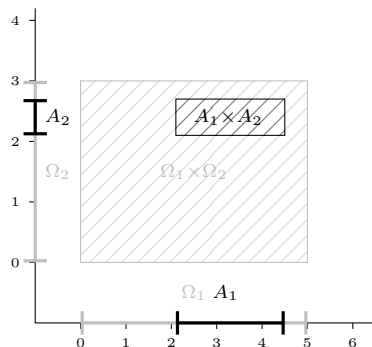


Was sind nun die messbaren Mengen in  $\Omega_1 \times \Omega_2$ ? Das heißt: Was ist eine geeignete  $\sigma$ -Algebra  $\mathcal{A}$  auf  $\Omega = \Omega_1 \times \Omega_2$ ?

Wenn  $A_1 \subset \Omega_1$  messbar ist (also  $A_1 \in \mathcal{A}_1$ ) und  $A_2 \subset \Omega_2$  messbar ist (also  $A_2 \in \mathcal{A}_2$ ), dann soll auch

$$A_1 \times A_2 = \{(a_1, a_2) \mid a_1 \in A_1, a_2 \in A_2\} \subset \Omega$$

messbar sein (also  $A_1 \times A_2 \in \mathcal{A}$ ).



**Einfachste Idee:**

Nehme kleinste  $\sigma$ -Algebra auf  $\Omega = \Omega_1 \times \Omega_2$ , die

$$\mathcal{E} = \mathcal{A}_1 \times \mathcal{A}_2 = \{A_1 \times A_2 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$$

enthält.

**Definition 9.1 (Produkt- $\sigma$ -Algebra)**

Seien  $(\Omega_1, \mathcal{A}_1)$  und  $(\Omega_2, \mathcal{A}_2)$  zwei Messräume. Sei  $\mathcal{A}$  die von

$$\mathcal{E} = \mathcal{A}_1 \times \mathcal{A}_2 = \{A_1 \times A_2 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$$

erzeugte  $\sigma$ -Algebra auf  $\Omega = \Omega_1 \times \Omega_2$ , d.h.

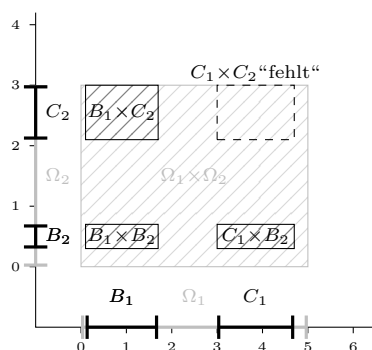
$$\mathcal{A} = \sigma(\mathcal{A}_1 \times \mathcal{A}_2) \quad .$$

Dann heißt  $\mathcal{A}$  Produkt- $\sigma$ -Algebra von  $\mathcal{A}_1$  und  $\mathcal{A}_2$  auf  $\Omega = \Omega_1 \times \Omega_2$ . Für die Produkt- $\sigma$ -Algebra  $\mathcal{A}$  schreibt man auch

$$\mathcal{A} := \mathcal{A}_1 \otimes \mathcal{A}_2$$

**Achtung:** Nicht jede Menge  $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$  ist darstellbar als „Rechtecksmenge“

$$A = A_1 \times A_2; \quad A_1 \in \mathcal{A}_1, \quad A_2 \in \mathcal{A}_2$$



**Beispiel 9.2 (2-dim. Borel-Mengen)**

$$\mathfrak{B}^{\otimes 2} = \mathfrak{B} \otimes \mathfrak{B}$$

Der Beweis von  $\mathfrak{B}^{\otimes 2} = \mathfrak{B} \otimes \mathfrak{B}$  ist nicht ganz trivial und benötigt ein bisschen Topologie.

### 9.1.2 Produkte von Maßen

Seien  $(\Omega_1, \mathcal{A}_1)$  und  $(\Omega_2, \mathcal{A}_2)$  zwei Messräume. Sei  $\mu_1$  ein Maß auf  $(\Omega_1, \mathcal{A}_1)$  und  $\mu_2$  ein Maß auf  $(\Omega_2, \mathcal{A}_2)$ .

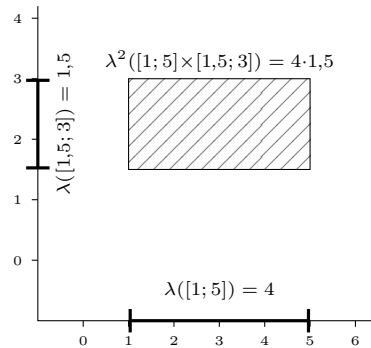
Wir wollen nun ein Maß  $\mu$  auf  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$  definieren, so dass

$$\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2) \quad \forall A_1 \in \mathcal{A}_1, \quad \forall A_2 \in \mathcal{A}_2 \quad (9.1)$$

#### Beispiel 9.3 (2-dim. Lebesguemaß $\lambda^2$ )

Sei  $\Omega_1 = \mathbb{R}$  und  $\Omega_2 = \mathbb{R}$ .

$$\lambda^2([a_1; b_1] \times [a_2; b_2]) = (b_1 - a_1) \cdot (b_2 - a_2) = \lambda^1([a_1; b_1]) \cdot \lambda^1([a_2; b_2])$$



An dieser Stelle ist noch nicht klar, ob durch (9.1) immer ein Maß  $\mu$  auf  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$  definiert ist.

Als Hilfsmittel brauchen wir zunächst folgendes Lemma:

**Lemma 9.4** Seien  $(\Omega_1, \mathcal{A}_1)$ ,  $(\Omega_2, \mathcal{A}_2)$  und  $(\Omega', \mathcal{A}')$  Messräume. Sei

$$f : \Omega_1 \times \Omega_2 \rightarrow \Omega'$$

eine  $\mathcal{A}_1 \otimes \mathcal{A}_2 / \mathcal{A}'$ -messbare Abbildung. Dann ist für jedes  $\omega_1 \in \Omega_1$

$$f_{\omega_1} : \Omega_2 \rightarrow \Omega', \quad \omega_2 \mapsto f(\omega_1, \omega_2)$$

eine  $\mathcal{A}_2 / \mathcal{A}'$ -messbare Abbildung und für jedes  $\omega_2 \in \Omega_2$

$$f_{\omega_2} : \Omega_1 \rightarrow \Omega', \quad \omega_1 \mapsto f(\omega_1, \omega_2)$$

eine  $\mathcal{A}_1 / \mathcal{A}'$ -messbare Abbildung.

Außerdem benötigen wir folgenden berühmten Satz:

#### Satz 9.5 (Existenz des Produktmaßes)

Seien  $(\Omega_1, \mathcal{A}_1)$  und  $(\Omega_2, \mathcal{A}_2)$  zwei Messräume. Sei  $\mu_1$  ein  $\sigma$ -endliches Maß auf  $(\Omega_1, \mathcal{A}_1)$  und  $\mu_2$  ein  $\sigma$ -endliches Maß auf  $(\Omega_2, \mathcal{A}_2)$ .

Dann gibt es genau ein Maß  $\mu$  auf  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$  mit

$$\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2) \quad \forall A_1 \in \mathcal{A}_1, \quad \forall A_2 \in \mathcal{A}_2 \quad (9.2)$$

Dieses Maß  $\mu$  ist ebenfalls  $\sigma$ -endlich.

**Beweisskizze:** Sei  $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ . Nach Lemma 9.4 ist für jedes  $\omega_1 \in \Omega_1$

$$f_{\omega_1} : \Omega_2 \rightarrow \mathbb{R}, \quad \omega_2 \mapsto I_A(\omega_1, \omega_2)$$

$\mathcal{A}_2$ -messbar. Also ist

$$\omega_1 \mapsto \int I_A(\omega_1, \omega_2) \mu_2(d\omega_2)$$

wohldefiniert und nach dem Satz von Tonelli (vgl. Bauer (1992); S. 155, oder Satz 9.3) sogar  $\mathcal{A}_1$ -messbar. Setze

$$\mu(A) = \int_{\Omega_1} \int_{\Omega_2} I_A(\omega_1, \omega_2) \mu_2(d\omega_2) \mu_1(d\omega_1)$$

Dies definiert ein Maß auf  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ , woraus die Existenz folgt.

Die Eindeutigkeit folgt aus dem Eindeutigkeitssatz 3.9.

Die  $\sigma$ -Endlichkeit folgt aus der  $\sigma$ -Endlichkeit von  $\mu_1$  und  $\mu_2$ . □

**Definition 9.6 (Produktmaß)**

Das Maß  $\mu$  aus Satz 9.5 heißt Produktmaß von  $\mu_1$  und  $\mu_2$ . Für das Produktmaß  $\mu$  schreiben wir auch

$$\mu_1 \otimes \mu_2$$

Der nachfolgende Satz beschreibt, wie man Integrale bzgl.  $\mu_1 \otimes \mu_2$  berechnen kann:

**Satz 9.7 (Satz von Fubini)**

Seien  $(\Omega_1, \mathcal{A}_1)$  und  $(\Omega_2, \mathcal{A}_2)$  zwei Messräume. Sei  $\mu_1$  ein  $\sigma$ -endliches Maß auf  $(\Omega_1, \mathcal{A}_1)$  und  $\mu_2$  ein  $\sigma$ -endliches Maß auf  $(\Omega_2, \mathcal{A}_2)$ .

Sei  $f \in \mathcal{L}_1(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ . Dann ist

$$f_{\omega_1} : \Omega_2 \rightarrow \mathbb{R}, \quad \omega_2 \mapsto f(\omega_1, \omega_2)$$

für  $\mu_1$ -fast alle  $\omega_1 \in \Omega_1$  integrierbar bzgl.  $\mu_2$  und

$$\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}(\omega_2) \mu_2(d\omega_2)$$

ist integrierbar bzgl.  $\mu_1$ . Die analoge Aussage mit vertauschten Indizes 1 und 2 gilt ebenso.

Außerdem gilt

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) \mu_1 \otimes \mu_2(d(\omega_1, \omega_2)) &= \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \mu_1(d\omega_1) \\ &= \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1) \mu_2(d\omega_2) \end{aligned}$$

Der Satz von Fubini besagt insbesondere, dass beim Integrieren die Reihenfolge der Integration vertauscht werden darf.



## 9.2 $n$ -faches Produkt $\Omega_1 \times \dots \times \Omega_n$

### Definition 9.8 (Produkt- $\sigma$ -Algebra)

Seien  $(\Omega_1, \mathcal{A}_1), \dots, (\Omega_n, \mathcal{A}_n)$  Messräume. Sei  $\mathcal{A}$  die von

$$\mathcal{E} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n = \{A_1 \times \dots \times A_n \mid A_i \in \mathcal{A}_i \ \forall i \in \{1, \dots, n\}\}$$

erzeugte  $\sigma$ -Algebra auf  $\Omega = \Omega_1 \times \dots \times \Omega_n$ .

Dann heißt  $\mathcal{A}$  Produkt- $\sigma$ -Algebra von  $\mathcal{A}_1, \dots, \mathcal{A}_n$  auf  $\Omega = \Omega_1 \times \dots \times \Omega_n$ . Für die Produkt- $\sigma$ -Algebra  $\mathcal{A}$  schreibt man auch

$$\mathcal{A} = \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$$

### Satz 9.9 (Existenz des Produktmaßes)

Seien  $(\Omega_1, \mathcal{A}_1), \dots, (\Omega_n, \mathcal{A}_n)$  Messräume. Für jedes  $i \in \{1, \dots, n\}$  sei  $\mu_i$  ein  $\sigma$ -endliches Maß auf  $(\Omega_i, \mathcal{A}_i)$ .

Dann gibt es genau ein Maß  $\mu$  auf  $(\Omega_1 \times \dots \times \Omega_n, \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n)$  mit

$$\mu(A_1 \times \dots \times A_n) = \mu_1(A_1) \cdot \dots \cdot \mu_n(A_n) \quad \forall A_i \in \mathcal{A}_i \ i \in \{1, \dots, n\} \quad (9.3)$$

Dieses Maß  $\mu$  ist ebenfalls  $\sigma$ -endlich.

**Beweisidee:** Wie beim 2-fachen Produkt. □

### Definition 9.10 (Produktmaß)

Das Maß  $\mu$  aus Satz 9.9 heißt Produktmaß von  $\mu_1, \dots, \mu_n$ . Für das Produktmaß  $\mu$  schreiben wir auch

$$\mu_1 \otimes \dots \otimes \mu_n$$

### Notation 9.11

- Falls  $\mu = \mu_1 = \dots = \mu_n$ , schreiben wir auch

$$\mu^{\otimes n} = \underbrace{\mu \otimes \dots \otimes \mu}_{n\text{-mal}} = \mu_1 \otimes \dots \otimes \mu_n$$

bzw. falls  $\mathcal{A} = \mathcal{A}_1 = \dots = \mathcal{A}_n$ , schreiben wir auch

$$\mathcal{A}^{\otimes n} = \underbrace{\mathcal{A} \otimes \dots \otimes \mathcal{A}}_{n\text{-mal}} = \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$$

- Abkürzend schreiben wir auch

$$\times_{i=1}^n \Omega_i = \Omega_1 \times \dots \times \Omega_n$$

und

$$\otimes_{i=1}^n \mathcal{A}_i = \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$$

und

$$\otimes_{i=1}^n \mu_i = \mu_1 \otimes \dots \otimes \mu_n$$

Also zum Beispiel auch

$$\begin{aligned} \mathcal{L}_1(\Omega_1 \times \dots \times \Omega_n, \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n, \mu_1 \otimes \dots \otimes \mu_n) \\ = \mathcal{L}_1(\times_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i, \otimes_{i=1}^n \mu_i) \end{aligned}$$

**Satz 9.12 (Satz von Fubini)**

Seien  $(\Omega_1, \mathcal{A}_1), \dots, (\Omega_n, \mathcal{A}_n)$  Messräume. Für jedes  $i \in \{1, \dots, n\}$  sei  $\mu_i$  ein  $\sigma$ -endliches Maß auf  $(\Omega_i, \mathcal{A}_i)$ .

Sei

$$f \in \mathcal{L}_1(\times_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i, \otimes_{i=1}^n \mu_i)$$

Dann ist

$$\begin{aligned} & \int_{\times_{i=1}^n \Omega_i} f(\omega_1, \dots, \omega_n) \otimes_{i=1}^n \mu_i (d(\omega_1, \dots, \omega_n)) \\ &= \int_{\Omega_1} \int_{\Omega_2} \dots \int_{\Omega_n} f(\omega_1, \omega_2, \dots, \omega_n) \mu_n(d\omega_n) \dots \mu_2(d\omega_2) \mu_1(d\omega_1) \end{aligned} \quad (9.4)$$

und die Reihenfolge der Integration in (9.4) darf beliebig vertauscht werden.

**Satz 9.13 (Satz von Tonelli)**

Seien

$$(\Omega_1, \mathcal{A}_1, \mu_1), \dots, (\Omega_n, \mathcal{A}_n, \mu_n)$$

Messräume und sei

$$f : \Omega_1 \times \dots \times \Omega_n \rightarrow \mathbb{R}$$

eine  $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$ -messbare Funktion mit

$$f(\omega) \geq 0 \quad \forall \omega \in \Omega$$

Dann gilt

$$\begin{aligned} & \int_{\times_{i=1}^n \Omega_i} f(\omega_1, \dots, \omega_n) \otimes_{i=1}^n \mu_i (d(\omega_1, \dots, \omega_n)) \\ &= \int_{\Omega_1} \int_{\Omega_2} \dots \int_{\Omega_n} f(\omega_1, \omega_2, \dots, \omega_n) \mu_n(d\omega_n) \dots \mu_2(d\omega_2) \mu_1(d\omega_1) \end{aligned} \quad (9.5)$$

und die Reihenfolge der Integration in (9.5) darf beliebig vertauscht werden.

Mit Hilfe des Satzes von Tonelli kann man also die Integrierbarkeit einer Funktion bzgl. des Produktmaßes zeigen. Der Satz von Fubini setzt diese Integrierbarkeit ja voraus.

**Beispiel 9.14** Es gilt

$$\mathfrak{B} \otimes \dots \otimes \mathfrak{B} = \mathfrak{B}^{\otimes n} \quad (9.6)$$

und

$$\lambda \otimes \dots \otimes \lambda = \lambda^n \quad (9.7)$$

Der Beweis von (9.6) benötigt ein bisschen Topologie.

Der Beweis von (9.7) ist eine einfache Übung.

Zum Abschluss noch ein Messbarkeitskriterium für Funktionen

$$f : \Omega \rightarrow \times_{i=1}^n \Omega'_i$$

**Satz 9.15** Seien  $(\Omega, \mathcal{A}), (\Omega'_1, \mathcal{A}'_1), \dots, (\Omega'_n, \mathcal{A}'_n)$  Messräume.  
Für  $i \in \{1, \dots, n\}$  sei  $f_i$  eine Funktion

$$f_i : \Omega \rightarrow \Omega'_i$$

Setze

$$f : \Omega \rightarrow \times_{i=1}^n \Omega'_i, \quad \omega \mapsto (f_1(\omega), \dots, f_n(\omega))$$

Dann gilt:

$f$  ist  $\mathcal{A} / \otimes_{i=1}^n \mathcal{A}'_i$ -messbar genau dann, wenn  $f_i$  für jedes  $i \in \{1, \dots, n\}$  eine  $\mathcal{A} / \mathcal{A}'_i$ -messbare Funktion ist.

**Beweis:**

- Sei zunächst  $f$  eine  $\mathcal{A} / \otimes_{i=1}^n \mathcal{A}'_i$ -messbare Funktion.

Sei  $i \in \{1, \dots, n\}$  und  $A'_i \in \mathcal{A}'_i$ .

Dann ist

$$A' := \Omega'_1 \times \dots \times \Omega'_{i-1} \times A'_i \times \Omega'_{i+1} \times \dots \times \Omega'_n \in \times_{i=1}^n \mathcal{A}'_i \subset \otimes_{i=1}^n \mathcal{A}'_i \quad (9.8)$$

und es gilt

$$\begin{aligned} f_i^{-1}(A'_i) &= \underbrace{f_1^{-1}(\Omega'_1)}_{=\Omega} \cap \dots \cap \underbrace{f_{i-1}^{-1}(\Omega'_{i-1})}_{=\Omega} \cap f_i^{-1}(A'_i) \cap \underbrace{f_{i+1}^{-1}(\Omega'_{i+1})}_{=\Omega} \cap \dots \cap \underbrace{f_n^{-1}(\Omega'_n)}_{=\Omega} \\ &= f^{-1}(A') \stackrel{(9.8)}{\in} \mathcal{A} \end{aligned}$$

aufgrund der  $\mathcal{A} / \otimes_{i=1}^n \mathcal{A}'_i$ -Messbarkeit von  $f$ .

Da dies für jedes  $A'_i \in \mathcal{A}'_i$  gilt, folgt die  $\mathcal{A} / \mathcal{A}'_i$ -Messbarkeit von  $f_i$ .

- Sei nun  $f_i$  für jedes  $i \in \{1, \dots, n\}$  eine  $\mathcal{A} / \mathcal{A}'_i$ -messbare Funktion. Sei

$$\mathcal{E}' = \mathcal{A}'_1 \times \dots \times \mathcal{A}'_n = \{A'_1 \times \dots \times A'_n \mid A'_i \in \mathcal{A}'_i \forall i \in \{1, \dots, n\}\}$$

Dies ist ein Erzeuger von  $\otimes_{i=1}^n \mathcal{A}'_i$ . Aufgrund von Satz 5.7 reicht es daher zu zeigen, dass

$$f^{-1}(A'_1 \times \dots \times A'_n) \in \mathcal{A} \quad \forall A'_1 \times \dots \times A'_n \in \mathcal{E}'$$

Sei also  $A'_1 \times \dots \times A'_n \in \mathcal{E}'$ , d.h.

$$A'_i \in \mathcal{A}'_i \quad \forall i \in \{1, \dots, n\}$$

Dann ist

$$f^{-1}(A'_1 \times \dots \times A'_n) = \bigcap_{i=1}^n f_i^{-1}(A'_i) \in \mathcal{A}$$

aufgrund der  $\mathcal{A} / \mathcal{A}'_i$ -Messbarkeit der Funktionen  $f_i$ . □

**Übung 9.16 (Unabhängigkeit und Produktmaß)**

Seien  $(\Omega, \mathcal{A}), (\Omega'_1, \mathcal{A}'_1), (\Omega'_2, \mathcal{A}'_2)$  Messräume. Sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$ . Sei

$$X_1 : \Omega \rightarrow \Omega'_1$$

eine  $\mathcal{A}/\mathcal{A}'_1$ -messbare Zufallsvariable und sei

$$X_2 : \Omega \rightarrow \Omega'_2$$

eine  $\mathcal{A}/\mathcal{A}'_2$ -messbare Zufallsvariable. Setze

$$X : \Omega \rightarrow \Omega_1 \times \Omega_2, \quad \omega \mapsto (X_1(\omega), X_2(\omega))$$

Dann gilt:

$X_1$  und  $X_2$  sind stochastisch unabhängig, genau dann wenn

$$X(P) = [X_1(P)] \otimes [X_2(P)]$$

**Übung 9.17** Für  $i \in \{1, \dots, n\}$  sei  $P_i$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathfrak{B})$  mit Lebesgue-Dichte  $f_i$ . Dann ist

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad (x_1, \dots, x_n) \mapsto f_1(x_1) \cdot \dots \cdot f_n(x_n)$$

eine Lebesgue-Dichte von  $P_1 \otimes \dots \otimes P_n$  auf  $\mathbb{R}^n$ , das heißt

$$d(P_1 \otimes \dots \otimes P_n) = f d\lambda^n$$

### 9.3 Faltung von Wahrscheinlichkeitsmaßen

**Ziel:** Berechnung der Verteilung der Summe von Zufallsvariablen, am besten dargestellt durch ihre Dichte.

**Voraussetzungen:**

- Wahrscheinlichkeitsmaße  $\mu_1, \dots, \mu_n$  auf  $(\mathbb{R}, \mathfrak{B}^{\otimes m})$
- Produktmaß  $\mu_1 \otimes \dots \otimes \mu_n$  auf  $\underbrace{\mathbb{R}^m \times \dots \times \mathbb{R}^m}_{n\text{-mal}}$  versehen mit der  $\sigma$ -Algebra

$$\underbrace{\mathfrak{B}^{\otimes m} \otimes \dots \otimes \mathfrak{B}^{\otimes m}}_{n\text{-mal}}$$

- Darstellung der Summe durch die Abbildung

$$A_n : \mathbb{R}^m \times \dots \times \mathbb{R}^m \longrightarrow \mathbb{R}^m$$

$$(x_1, \dots, x_n) \longmapsto \sum_{i=1}^n x_i$$

$A_n$  ist messbar (vgl. Satz 5.9).

**Definition 9.18 (Faltungsprodukt)**

Unter obigen Voraussetzungen heißt

$$\mu_1 * \dots * \mu_n := A_n(\mu_1 \otimes \dots \otimes \mu_n)$$

Faltungsprodukt von  $\mu_1, \dots, \mu_n$  (Verteilung der Summe).

Dies ist ein Maß auf  $(\mathbb{R}, \mathfrak{B}^{\otimes m})$  und nach Definition 5.3 gilt für  $B \in \mathfrak{B}^{\otimes m}$

$$[\mu_1 * \cdots * \mu_n](B) = [\mu_1 \otimes \cdots \otimes \mu_n](A_n^{-1}(B)) \quad .$$

**Bemerkung 9.19** Sei

$$(X_1, \dots, X_{n+1}) \sim \mu_1 \otimes \cdots \otimes \mu_{n+1}$$

(d.h. insbesondere, dass  $X_1, \dots, X_{n+1}$  stochastisch unabhängig sind (vgl. Übung 9.16)) und

$$\begin{aligned} B_{n+1} : \mathbb{R}^m \times \cdots \times \mathbb{R}^m &\longrightarrow \mathbb{R}^m \times \mathbb{R}^m \\ (x_1, \dots, x_{n+1}) &\longmapsto \left( \sum_{i=1}^n x_i, x_{n+1} \right) = (A_n(x_1, \dots, x_n), x_{n+1}) \end{aligned}$$

Definiere das Maß

$$\tilde{P} := B_{n+1}(\mu_1 \otimes \cdots \otimes \mu_{n+1})$$

und die (Hilfs-)Abbildung  $\hat{A}_n$  so, dass

$$\hat{A}_n^{-1}(C_1 \times \mathbb{R}^m) = A_n^{-1}(C_1) \times \mathbb{R}^m$$

mit  $C_1 \in \mathbb{R}^m$ .

Außerdem sei  $pr_i : \mathbb{R}^m \times \cdots \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ , mit  $pr_i(x_1, \dots, x_i, \dots, x_n) = x_i$  die  $i$ -te Projektionsabbildung.

Mit  $A \in \mathfrak{B}^{\otimes m}$  und  $B \in \mathfrak{B}^{\otimes m}$  ist dann

$$\begin{aligned} B_{n+1}^{-1}(A \times B) &= \hat{A}_n^{-1}(A \times \mathbb{R}^m) \cap pr_{n+1}^{-1}(B) \\ &= \underbrace{A_n^{-1}(A)}_{\in (\mathbb{R}^m)^{\times n}} \times \mathbb{R}^m \cap \mathbb{R}^m \times \cdots \times \mathbb{R}^m \times B \\ &= A_n^{-1}(A) \times B \end{aligned}$$

und somit gilt

$$\begin{aligned} \tilde{P}(A \times B) &= [\mu_1 \otimes \cdots \otimes \mu_{n+1}](B_{n+1}^{-1}(A \times B)) \\ &= [\mu_1 \otimes \cdots \otimes \mu_{n+1}](A_n^{-1}(A) \times B) \\ &= [\mu_1 \otimes \cdots \otimes \mu_n](A_n^{-1}(A)) \cdot \mu_{n+1}(B) \\ &= [\mu_1 * \cdots * \mu_n](A) \cdot \mu_{n+1}(B) \end{aligned}$$

weshalb  $\sum_{i=1}^n X_i$  und  $X_{n+1}$  unabhängig sind.

Insbesondere gilt

$$A_{n+1} = A_2 \circ B_{n+1} \quad ,$$

d.h. man kann sich auf zweifache Faltungen beschränken, z.B.

$$(x_1, x_2, x_3) \mapsto (x_1 + x_2, x_3) \mapsto (x_1 + x_2) + x_3$$

**Lemma 9.20** Sei  $f \geq 0$  eine messbare reellwertige Funktion. Dann ist

$$\begin{aligned} \int f d(\mu_1 * \mu_2) &= \int f \circ A_2 d(\mu_1 \otimes \mu_2) \\ &\stackrel{\text{S. 9.13}}{=} \iint f(x+y) \mu_1(dy) \mu_2(dx) \quad . \end{aligned}$$

Mit  $f = I_B, B \in \mathfrak{B}^{\otimes m}$ , gilt somit

$$\begin{aligned} [\mu_1 * \mu_2](B) &= \int I_B d(\mu_1 * \mu_2) \\ &= \iint I_B(x+y) \mu_1(dx) \mu_2(dy) \\ &= \int \left( \int_{B-y} \mu_1(dx) \right) \mu_2(dy) \\ &= \int \mu_1(B-y) \mu_2(dy) \quad . \end{aligned}$$

Analog auch mit vertauschten Rollen von  $\mu_1$  und  $\mu_2$ .

**Beispiel 9.21**

$$[\delta_a * \mu](B) = \int \mu(B-x) \delta_a(dx) = \mu(B-a) = \mu(T_{-a}(B))$$

Für  $\mu = \delta_b$ :

$$[\delta_a * \delta_b](B) = \delta_b(T_{-a}(B)) = \begin{cases} 1 & b \in T_{-a}(B) \Leftrightarrow a+b \in B \\ 0 & \text{sonst} \end{cases} = \delta_{a+b}(B)$$

**Satz 9.22** Seien  $\mu_1$  und  $\mu_2$  Wahrscheinlichkeitsmaße auf  $(\mathbb{R}^m, \mathfrak{B}^{\otimes m})$ . Das Maß  $\mu_1$  besitze die Lebesgue-Dichte  $f$ .  
Dann gilt für jede Borelmenge  $B \in \mathfrak{B}^{\otimes m}$

$$[\mu_1 * \mu_2](B) = \int_B \left( \int f(x-y) \mu_2(dy) \right) \lambda^m(dx)$$

Somit besitzt  $\mu_1 * \mu_2$  (mittels  $f$  von  $\mu_1$ ) ebenfalls eine Lebesgue-Dichte.

**Beweis:** Definiere die Translationsabbildung durch  $T_y(x) := x + y$ . Mit der Translationsinvarianz des Lebesgue-Maßes, der Transformationsformel, sowie dem Satz von Fubini folgert man die Behauptung:

$$\begin{aligned} [\mu_1 * \mu_2](B) &\stackrel{\text{L. 9.20}}{=} \iint I_B(x+y) \mu_1(dx) \mu_2(dy) \\ &= \int \left( \int I_B(x+y) f(x) \lambda^m(dx) \right) \mu_2(dy) \\ &\stackrel{\text{S. 4.12}}{=} \int \left( \int I_B(x+y) f(x) T_{-y}(\lambda^m)(dx) \right) \mu_2(dy) \\ &\stackrel{\text{S. 6.15}}{=} \int \left( \int I_B(T_{-y}(x+y)) f(T_{-y}(x)) \lambda^m(dx) \right) \mu_2(dy) \\ &\stackrel{\text{S. 9.7}}{=} \iint I_B(x) f(x-y) \mu_2(dy) \lambda^m(dx) \\ &= \int I_B(x) \left( \int f(x-y) \mu_2(dy) \right) \lambda^m(dx) \end{aligned}$$

□

**Notation 9.23** Die Lebesgue-Dichte von  $\mu_1 * \mu_2$  wird mit  $f * \mu_2$  bezeichnet:

$$[f * \mu_2](z) := \int f(z - y) \mu_2(dy)$$

Um Verwechslungen vorzubeugen, wird hier der Buchstabe  $z$  für die Summenvariable (Ergebnis der Faltung) verwendet.

**Korollar 9.24** Besitzen nun sowohl  $\mu_1$  als auch  $\mu_2$  Lebesgue-Dichten  $d\mu_1 = f d\lambda^m$  und  $d\mu_2 = g d\lambda^m$ , dann gilt

$$[f * \mu_2](z) = \int f(z - y)g(y) \lambda^m(dy) =: [f * g](z)$$

und die Dichte des Faltungsprodukts wird mit  $f * g$  bezeichnet.

**Bemerkung 9.25** Sind  $X_1$  und  $X_2$  unabhängige und identisch mit Lebesgue-Dichte  $f$  verteilte Zufallsvariablen, dann besitzt  $Z := X_1 + X_2$  die Lebesgue-Dichte

$$[f * f](z) = \int f(z - x)f(x) \lambda(dx) .$$

**Beispiel 9.26** Sei  $f$  die  $\lambda$ -Dichte der stetigen Gleichverteilung auf dem Intervall  $[0; 1]$ , d.h.  $f = I_{[0;1]}$ . Seien  $X_1, X_2 \stackrel{i.i.d.}{\sim} U(0; 1)$  und  $Z := X_1 + X_2$ . Dann ist die  $\lambda$ -Dichte der Faltung  $Z$  gleich

$$[f * f](z) = \int I_{[0;1]}(z - x)I_{[0;1]}(x) \lambda(dx) .$$

Der Integrand nimmt nur die Werte 0 oder 1 an. Er hat den Wert 1 genau dann, wenn

$$\begin{aligned} & 0 \leq z - x \leq 1 \quad \text{und} \quad 0 \leq x \leq 1 \\ \iff & x \leq z \quad \text{und} \quad x \geq z - 1 \quad \text{und} \quad 0 \leq x \leq 1 \\ \iff & \max(0, z - 1) \leq x \leq \min(1, z) \\ \iff & \begin{cases} 0 \leq x \leq z & \text{falls } 0 \leq z \leq 1 \\ z - 1 \leq x \leq 1 & \text{falls } 1 < z \leq 2 \end{cases} \end{aligned}$$

Somit ergibt sich für die Dichte

$$[f * f](z) = \int_{\max(0, z-1)}^{\min(1, z)} \lambda(dx) = \begin{cases} z & \text{falls } 0 \leq z \leq 1 \\ 1 - (z - 1) = 2 - z & \text{falls } 1 < z \leq 2 \end{cases}$$

# 10 Bedingter Erwartungswert

## 10.1 Einleitung

### 10.1.1 Elementare bedingte Wahrscheinlichkeiten

Die bedingte Wahrscheinlichkeit

$$P(A|B)$$

hat die intuitive Bedeutung:

$P(A|B)$  gibt die Wahrscheinlichkeit an, dass Ereignis  $A$  eintritt, wenn man schon weiß, dass Ereignis  $B$  eingetreten ist.

$P(A|B)$  errechnet sich mit Hilfe der Formel

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Entsprechend hat der (elementare) bedingte Erwartungswert

$$\mathbb{E}_P[Y|B]$$

die intuitive Bedeutung:

$\mathbb{E}_P[Y|B]$  ist der Erwartungswert von  $Y$ , wenn man schon weiß, dass Ereignis  $B$  eingetreten ist.

$\mathbb{E}_P[Y|B]$  errechnet sich mit Hilfe der Formel

$$\mathbb{E}_P[Y|B] = \frac{1}{P(B)} \int_B Y dP$$

Für unseren allgemeinen maßtheoretischen Aufbau sind das aber keine tragfähigen Konzepte, mit denen man arbeiten könnte.

**Problem:** Diese Definitionen sind nicht möglich, falls

$$P(B) = 0$$

Aber dieser Fall ist auch wichtig, wie sich im folgendem Beispiel zeigt.

**Beispiel:** Seien  $Y$  und  $T$  Zufallsvariablen. Sei  $T$  normalverteilt. Die Beobachtung für  $T$  in einem Experiment sei zum Beispiel

$$T = 0.2641$$

Uns interessiert jetzt

$$P(Y \in A' | T = 0.2641)$$



also

$$P(Y^{-1}(A')|T^{-1}(\{0.2641\})) = P(A|B)$$

wobei  $A := Y^{-1}(A')$  und  $B := T^{-1}(\{0.2641\})$ .

Allerdings ist

$$P(B) = P(T^{-1}(\{0.2641\})) = \mathcal{N}(0,1)(0.2641) = 0$$

Diese Art von bedingter Wahrscheinlichkeit und bedingtem Erwartungswert betrachten wir daher im Folgenden nicht mehr. Stattdessen betrachten wir ein anderes (maßtheoretisches) Konzept, nämlich den *allgemeinen bedingten Erwartungswert*.

Viele Bücher über Wahrscheinlichkeitstheorie stellen dieses Konzept des allgemeinen bedingten Erwartungswertes als eine einfache, direkte Verallgemeinerung des elementaren bedingten Erwartungswertes hin.

Das ist aus zwei Gründen problematisch:

- Der allgemeine bedingte Erwartungswert ist nicht einfach!
- Der allgemeine bedingte Erwartungswert ist keine direkte Verallgemeinerung des elementaren Erwartungswertes! Das heißt: Der elementare bedingte Erwartungswert ist kein Spezialfall des *allgemeinen bedingten Erwartungswertes*.

Natürlich bestehen trotzdem gewisse Beziehungen zwischen dem elementaren bedingten Erwartungswert und dem *allgemeinen bedingten Erwartungswert*.

Trotzdem sollte man sich das Ganze nicht schwerer als nötig machen. Am Besten vergißt man vorübergehend den elementaren bedingten Erwartungswert (bedingte Wahrscheinlichkeiten) und denkt beim *allgemeinen bedingten Erwartungswert* NICHT an das „Bedingen nach irgendwas“, sondern an den „Informationsgehalt einer  $\sigma$ -Algebra“ (vgl. Abschnitt 10.1.2) – auch wenn der Name *allgemeiner bedingter Erwartungswert* etwas anderes nahelegt.

### 10.1.2 Der Informationsgehalt einer $\sigma$ -Algebra

Sei  $\Omega$  eine Menge und  $\mathcal{A}$  eine  $\sigma$ -Algebra auf  $\Omega$ .

*Information* lässt sich formalisieren als eine Menge  $\mathcal{C}$  von Ereignissen  $C \in \mathcal{A}$

$$\mathcal{C} \subset \mathcal{A}$$

von denen wir entscheiden können, ob sie eingetreten sind oder nicht.

$\mathcal{C}$  enthält also genau die Teilmengen  $C \subset \Omega$  (= Ereignisse), von denen wir wissen, ob sie eingetreten sind oder nicht.

Wenn wir entscheiden können, ob  $C \subset \Omega$  eingetreten ist, dann ist klar, dass wir auch entscheiden können, ob das Gegenereignis  $C^c$  eingetreten ist. Das heißt:

$$C \in \mathcal{C} \quad \Rightarrow \quad C^c \in \mathcal{C} \quad (10.1)$$

Wenn wir für jedes  $n \in \mathbb{N}$  entscheiden können, ob  $C_n \subset \Omega$  eingetreten ist, dann ist es plausibel, dass wir entscheiden können, ob  $\bigcup_{n=1}^{\infty} C_n$  eingetreten ist. Das heißt

$$(C_n)_{n \in \mathbb{N}} \subset \mathcal{C} \quad \Rightarrow \quad \bigcup_{n=1}^{\infty} C_n \in \mathcal{C} \quad (10.2)$$

Wir wissen auch, ob das unmögliche Ereignis  $\emptyset$  eingetreten ist. (Es ist natürlich nicht eingetreten.) Also

$$\emptyset \in \mathcal{C} \tag{10.3}$$

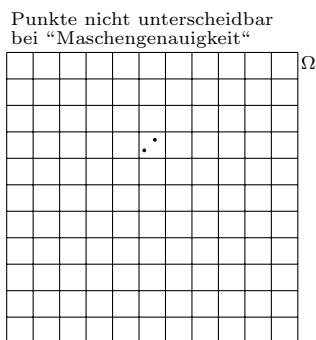
Aus (10.1) – (10.3) folgt, dass  $\mathcal{C}$  eine  $\sigma$ -Algebra ist. Somit läßt sich Information also durch eine  $\sigma$ -Algebra

$$\mathcal{C} \subset \mathcal{A}$$

formalisieren.

Eine  $\sigma$ -Algebra  $\mathcal{C}$  auf  $\Omega$  mit  $\mathcal{C} \subset \mathcal{A}$  heißt Unter- $\sigma$ -Algebra von  $\mathcal{A}$ .

Information kann man sich auch als ein Netz auf  $\Omega$  vorstellen, so dass zwei Ergebnisse eines Experiments nicht voneinander unterschieden werden können, wenn sie in der selben Masche des Netzes auf  $\Omega$  liegen. Auf einer Weltkarte kann man nicht zwischen dem EZB-Gebäude in Frankfurt und der Paulskirche in Frankfurt unterscheiden. Aber ein Stadtplan von Frankfurt enthält genügend Informationen, um die beiden Gebäude voneinander zu unterscheiden.



**Beispiel 10.1** Sei  $\Omega = \mathbb{R}$  und  $\mathcal{A} = \mathfrak{B}$ : Sei  $\omega \in \mathbb{R}$  das Ergebnis eines normalverteilten Experimentes, z.B.

$$\omega = 0,2417269253751908503217094378\dots$$

Allerdings kann man das Ergebnis normalerweise gar nicht so genau beobachten (Messgenauigkeit!). Zum Beispiel zeigt das Messgerät das Ergebnis nur auf zwei Nachkommastellen an. Das angezeigte Ergebnis ist dann 0,24 und man weiß nur, dass  $\omega \in [0,235; 0,245)$ . Das heißt also

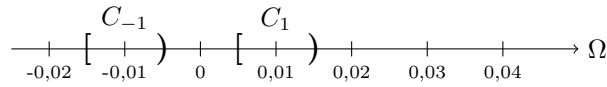
$$\mathcal{C} = \sigma \left( \left\{ \left[ \frac{k}{100} - 0,005; \frac{k}{100} + 0,005 \right) \mid k \in \mathbb{Z} \right\} \right)$$

wobei  $\mathcal{C} \subsetneq \mathfrak{B}$ . Setze

$$C_k = \left[ \frac{k}{100} - 0,005; \frac{k}{100} + 0,005 \right) \quad \forall k \in \mathbb{Z}$$

Dann ist

$$\mathcal{C} = \left\{ \bigcup_{k \in K} C_k \mid K \subset \mathbb{Z} \right\}$$



Sei  $(\Omega, \mathcal{A})$  wieder ein Messraum und

$$\mathcal{C} \subset \mathcal{A}$$

eine Unter- $\sigma$ -Algebra von  $\mathcal{A}$ , welche die zur Verfügung stehende Information repräsentiert.

Sei

$$Y : \Omega \rightarrow \mathbb{R}$$

eine  $\mathcal{C}$ -messbare Zufallsvariable.

Die  $\sigma$ -Algebra  $\mathcal{C}$  repräsentiert die verfügbare Information. Aufgrund der  $\mathcal{C}$ -Messbarkeit von  $Y$  kennen wir dann das Ergebnis von  $Y$ , denn:

Wir wissen für alle  $a \in \mathbb{R}$ , ob

$$C = Y^{-1}((a; \infty))$$

eingetreten ist, weil  $Y^{-1}((a; \infty)) \in \mathcal{C}$ .

Also:  $Y$  ist durch die verfügbare Information vollständig bestimmt.

Sei nun  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  und sei

$$Y : \Omega \rightarrow \mathbb{R}$$

eine  $\mathcal{A}$ -messbare Zufallsvariable.

Wir wollen jetzt eine „beste Schätzung“ von  $Y$  basierend auf der verfügbaren Information  $\mathcal{C}$  suchen.

Die  $\mathcal{C}$ -messbaren Zufallsvariablen sind die Zufallsvariablen, die durch unsere verfügbare Information (repräsentiert durch  $\mathcal{C}$ ) vollständig bestimmt sind.

Eine „beste Schätzung“ von  $Y$  basierend auf der verfügbaren Information  $\mathcal{C}$  ist daher eine  $\mathcal{C}$ -messbare Zufallsvariable  $Y_0$ , die „so nahe wie möglich“ an  $Y$  liegt.

Eine  $\mathcal{C}$ -messbare Zufallsvariable  $Y_0$  liegt „so nahe wie möglich“ an  $Y$  falls

$$\int_C Y - Y_0 dP = 0 \quad \forall C \in \mathcal{C} \quad (10.4)$$

oder anders geschrieben

$$\int_C Y dP = \int_C Y_0 dP \quad \forall C \in \mathcal{C} \quad (10.5)$$

Für  $Y_0$  schreiben wir dann auch

$$Y_0 = \mathbb{E}_P[Y | \mathcal{C}]$$

**Beispiel 10.2 (Fortsetzung von Beispiel 10.1)**

Wie vorher ist  $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathfrak{B})$  und  $P = \mathcal{N}(0, 1)$ . Sei  $Y$  eine  $\mathfrak{B}$ -messbare Zufallsvariable

$$Y : \mathbb{R} \rightarrow \mathbb{R}$$

Da wir die Ereignisse  $\omega$  nicht genau beobachten können, sondern nur wissen, ob ein

$$\omega \in C_k$$

eingetreten ist, kennen wir  $Y$  nicht vollständig. Wir wissen nur, ob

$$Y \in \left\{ Y(\omega) \mid \omega \in C_k \right\}$$

Eine naheliegende Schätzung für  $Y$  auf  $C_k$  wäre der relative Mittelwert

$$\frac{1}{\mathcal{N}(0, 1)(C_k)} \cdot \int_{C_k} Y d\mathcal{N}(0, 1)$$

Insgesamt wäre dann also eine  $\mathcal{C}$ -messbare Schätzung von  $Y$

$$Y_0 : \omega \mapsto \sum_{k \in \mathbb{Z}} c_k \cdot I_{C_k}(\omega)$$

wobei

$$c_k = \frac{1}{\mathcal{N}(0, 1)(C_k)} \cdot \int_{C_k} Y d\mathcal{N}(0, 1) \quad \forall k \in \mathbb{Z}$$

Und dieses  $Y_0$  erfüllt tatsächlich Gleichung (10.5): Für  $C = \bigcup_{k \in K} C_k \in \mathcal{C}$  ist

$$\begin{aligned} \int_C Y_0 d\mathcal{N}(0, 1) &= \int_C \sum_{k \in K} c_k I_{C_k}(\omega) \mathcal{N}(0, 1)(d\omega) \\ &= \sum_{k \in K} c_k \cdot \mathcal{N}(0, 1)(C_k) \\ &\stackrel{\text{einsetzen}}{=} \sum_{k \in K} \int_{C_k} Y d\mathcal{N}(0, 1) \\ &= \int_C Y d\mathcal{N}(0, 1) \end{aligned}$$

**10.1.3 Der Informationsgehalt einer Zufallsvariable**

Sei  $(\Omega, \mathcal{A})$  ein Messraum und

$$Y : \Omega \rightarrow \mathbb{R}$$

eine  $\mathcal{A}$ -messbare Zufallsvariable.

Sei  $(\Omega', \mathcal{A}')$  ein weiterer Messraum und

$$T : \Omega \rightarrow \Omega'$$

eine  $\mathcal{A}/\mathcal{A}'$ -messbare Zufallsvariable, die wir beobachten können.

Das heißt: Für jedes  $A' \in \mathcal{A}'$  wissen wir, ob

$$T \in A' \quad \text{oder} \quad T \notin A'$$

Oder mit anderen Worten: Wir können für jedes  $A' \in \mathcal{A}'$  entscheiden, ob das Ereignis

$$T^{-1}(A') \in \mathcal{A}$$

eingetreten ist.

Also wird unsere Information über die beobachtete Zufallsvariable  $T$  durch

$$\mathcal{C} = \{T^{-1}(A') \mid A' \in \mathcal{A}'\}$$

repräsentiert.

**Beachte:** Dieses  $\mathcal{C} \subset \mathcal{A}$  ist eine Unter- $\sigma$ -Algebra von  $\mathcal{A}$ .

Der Informationsgehalt einer Zufallsvariable läßt sich also durch eine Unter- $\sigma$ -Algebra von  $\mathcal{A}$  ausdrücken.

## 10.2 Definition und grundlegende Eigenschaften

Ein beste Schätzung  $Y_0$  von  $Y$  ergibt sich aus Gleichung (10.5) unter der Annahme, dass  $\mathcal{C}$  unsere zur Verfügung stehende Information darstellt. Eine solche „beste Schätzung“ nennt man *Bedingte Erwartung von  $Y$  gegeben  $\mathcal{C}$* .

**Definition 10.3** Sei  $(\Omega, \mathcal{A})$  ein Messraum,  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  und  $Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ .

Sei  $\mathcal{C} \subset \mathcal{A}$  eine Unter- $\sigma$ -Algebra von  $\mathcal{A}$ . Sei

$$Y_0 : \Omega \rightarrow \mathbb{R}$$

mit folgenden Eigenschaften

(i)  $Y_0 \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$

(ii)  $Y_0$  ist  $\mathcal{C}$ -messbar.

(iii) Für alle  $C \in \mathcal{C}$  ist

$$\int_C Y \, dP = \int_C Y_0 \, dP \quad (10.6)$$

Dann heißt  $Y_0$  bedingte Erwartung von  $Y$  gegeben  $\mathcal{C}$  (unter  $P$ ) und man schreibt

$$Y_0 = \mathbb{E}_P[Y | \mathcal{C}]$$

Diese Definition wirft sofort zwei Fragen auf:

1. Existiert zu jedem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , zu jedem  $Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$  und zu jeder Unter- $\sigma$ -Algebra  $\mathcal{C} \subset \mathcal{A}$  eine bedingte Erwartung von  $Y$  gegeben  $\mathcal{C}$  unter  $P$ ?
2. Ist die bedingte Erwartung von  $Y$  gegeben  $\mathcal{C}$  unter  $P$  eindeutig oder gibt es zu einem  $Y$  mehrere bedingte Erwartungen?

Die Antwort auf beide Fragen ist (im Wesentlichen) „ja“, wie der nachfolgende Satz zeigt:

**Satz 10.4 (Existenz und Eindeutigkeit der bedingten Erwartung)**

a) Zu jedem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , zu jedem  $Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$  und zu jeder Unter- $\sigma$ -Algebra  $\mathcal{C} \subset \mathcal{A}$  existiert eine bedingte Erwartung von  $Y$  gegeben  $\mathcal{C}$  unter  $P$ , also ein

$$Y_0 = \mathbb{E}_P[Y | \mathcal{C}]$$

b) Seien  $Y_0$  und  $X_0$  zwei bedingte Erwartungen von  $Y$  gegeben  $\mathcal{C}$  unter  $P$ , dann ist

$$Y_0 = X_0 \quad P\text{-fast sicher}$$

**Beweis:**

a) Sei  $Y^+$  der Positivteil von  $Y$  und  $Y^-$  der Negativteil von  $Y$ . Also

$$Y^+ \geq 0, \quad Y^- \geq 0 \quad \text{und} \quad Y = Y^+ - Y^-$$

Setze

$$\nu^+(A) = \int_A Y^+ dP \quad \forall A \in \mathcal{A}$$

und

$$\nu^-(A) = \int_A Y^- dP \quad \forall A \in \mathcal{A}$$

Gemäß Satz 7.1 und Satz 7.4 sind  $\nu^+$  und  $\nu^-$  zwei Maße auf  $(\Omega, \mathcal{A})$  so dass

$$\nu^+ \ll P \quad \text{und} \quad \nu^- \ll P \quad (10.7)$$

Wegen  $\mathcal{C} \subset \mathcal{A}$  werden durch

$$\hat{P}(C) = P(C) \quad \forall C \in \mathcal{C}$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{C})$  und durch

$$\hat{\nu}^+(C) = \nu^+(C), \quad \hat{\nu}^-(C) = \nu^-(C) \quad \forall C \in \mathcal{C}$$

zwei Maße auf  $(\Omega, \mathcal{C})$  definiert.

Wir betrachten also jetzt den Messraum  $(\Omega, \mathcal{C})$ . Offensichtlich folgt aus (10.7) auch

$$\hat{\nu}^+ \ll \hat{P} \quad \text{und} \quad \hat{\nu}^- \ll \hat{P} \quad (10.8)$$

Nach Satz 7.4 existieren dann

$$Y_0^+ \in \mathcal{L}_1(\Omega, \mathcal{C}, \hat{P}) \quad \text{und} \quad Y_0^- \in \mathcal{L}_1(\Omega, \mathcal{C}, \hat{P})$$

so dass

$$\hat{\nu}^+(C) = \int_C Y_0^+ d\hat{P}, \quad \hat{\nu}^-(C) = \int_C Y_0^- d\hat{P} \quad \forall C \in \mathcal{C}$$

Setze  $Y_0 = Y_0^+ - Y_0^-$ . Dann gilt:

- $Y_0 \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$
- $Y_0$  ist  $\mathcal{C}$ -messbar.
- Für alle  $C \in \mathcal{C}$  ist

$$\begin{aligned} \int_C Y_0 dP &= \int_C Y_0^+ d\hat{P} - \int_C Y_0^- d\hat{P} = \hat{\nu}^+(C) - \hat{\nu}^-(C) \\ &= \nu^+(C) - \nu^-(C) = \int_C Y^+ dP - \int_C Y^- dP \\ &= \int_C Y dP \end{aligned}$$

Also ist

$$Y_0 = \mathbb{E}_P[Y|\mathcal{C}]$$

und die bedingte Erwartung von  $Y$  gegeben  $\mathcal{C}$  unter  $P$  existiert.

- b) Sei  $X_0$  neben  $Y_0$  eine weitere bedingte Erwartung von  $Y$  gegeben  $\mathcal{C}$  unter  $P$ . Dann folgt aus

$$\int_C Y_0 d\hat{P} = \int_C Y dP = \int_C X_0 d\hat{P} \quad \forall C \in \mathcal{C}$$

und Satz 6.12 b)

$$Y_0 = X_0 \quad \hat{P}\text{-fast sicher}$$

Und hieraus folgt schließlich

$$Y_0 = X_0 \quad P\text{-fast sicher}$$

□

$\mathcal{C}$  stellt unsere zur Verfügung stehende Information dar. Falls diese Information von einer beobachteten Zufallsvariable

$$T : \Omega \rightarrow \Omega'$$

stammt (vgl. Abschnitt 10.1.3), so schreiben wir auch

$$\mathbb{E}_P[Y|\mathcal{C}] = \mathbb{E}_P[Y|T]$$

Dies ist der Inhalt folgender Definition:

**Definition 10.5** Sei  $(\Omega, \mathcal{A})$  ein Messraum,  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  und  $Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ .

Sei  $(\Omega', \mathcal{A}')$  ein weiterer Messraum und

$$T : \Omega \rightarrow \Omega'$$

eine  $\mathcal{A}/\mathcal{A}'$ -messbare Zufallsvariable. Setze

$$\mathcal{C} = \{T^{-1}(A') \mid A' \in \mathcal{A}'\}$$

Dann heißt die bedingte Erwartung von  $Y$  gegeben  $\mathcal{C}$  unter  $P$  ( $\mathbb{E}_P[Y|\mathcal{C}]$ ) auch bedingte Erwartung von  $Y$  gegeben  $T$  (unter  $P$ ) und man schreibt

$$\mathbb{E}_P[Y|T]$$

anstelle von  $\mathbb{E}_P[Y|\mathcal{C}]$

Eine entsprechende Definition ist auch bei mehreren Zufallsvariablen

$$T_i : \Omega \rightarrow \Omega'$$

möglich und sinnvoll. Diese Situation liegt vor allem bei Stochastischen Prozessen häufig vor (vgl. Markov-Ketten).

**Definition 10.6** Sei  $(\Omega, \mathcal{A})$  ein Messraum,  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  und  $Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ .

Sei  $I$  eine Indexmenge. Für jedes  $i \in I$  sei  $(\Omega'_i, \mathcal{A}'_i)$  ein Messraum und

$$T_i : \Omega \rightarrow \Omega'_i$$

eine  $\mathcal{A}/\mathcal{A}'_i$ -messbare Zufallsvariable. Setze

$$\mathcal{C}_i = \{T_i^{-1}(A'_i) \mid A'_i \in \mathcal{A}'_i\}$$

und

$$\mathcal{C} = \sigma\left(\bigcup_{i \in I} \mathcal{C}_i\right)$$

Dann heißt die bedingte Erwartung von  $Y$  gegeben  $\mathcal{C}$  unter  $P$  ( $\mathbb{E}_P[Y|\mathcal{C}]$ ) auch bedingte Erwartung von  $Y$  gegeben  $(T_i)_{i \in I}$  (unter  $P$ ) und man schreibt

$$\mathbb{E}_P[Y | T_i, i \in I]$$

anstelle von  $\mathbb{E}_P[Y|\mathcal{C}]$ .

Bevor wir einige Eigenschaften der bedingten Erwartung festhalten, betrachten wir zwei Beispiele

**Beispiel 10.7 (Völlige Unwissenheit)**

Sei  $(\Omega, \mathcal{A})$  ein Messraum,  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  und  $Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ .

Völlige Unwissenheit über das eingetretene Ereignis wird durch die Unter- $\sigma$ -Algebra

$$\mathcal{C} = \{\emptyset, \Omega\}$$

repräsentiert. Wir wissen nämlich nur, dass das unmögliche Ereignis  $\emptyset$  nicht und dass das sichere Ereignis  $\Omega$  eingetreten ist; von den anderen Ereignissen  $A \in \mathcal{A}$  wissen wir nicht, ob sie eingetreten sind.

In diesem Fall ist

$$\mathbb{E}_P[Y|\mathcal{C}] = \mathbb{E}_P[Y]$$

Der gewöhnliche Erwartungswert ist in diesem Fall die „beste Schätzung“ für  $Y$ .



**Beispiel 10.8 (Konkretes statistisches Beispiel)** Seien  $y^{(1)}, \dots, y^{(n)}$  unabhängig identisch nach  $P$  verteilte Beobachtungen auf  $\mathbb{R}$

$$y^{(i)} \in \mathbb{R}, \quad y^{(i)} \sim P, \quad i = 1, \dots, n$$

Das heißt also: Der zugrundeliegende Messraum ist  $(\mathbb{R}^n, \mathfrak{B}^{\otimes n})$  und das entsprechende Wahrscheinlichkeitsmaß auf  $(\mathbb{R}^n, \mathfrak{B}^{\otimes n})$  ist

$$P^{\otimes n} = \underbrace{P \otimes \dots \otimes P}_{n\text{-mal}}$$

Der zugrundeliegende Wahrscheinlichkeitsraum für dieses Experiment ist somit

$$(\mathbb{R}^n, \mathfrak{B}^{\otimes n}, P^{\otimes n})$$

Beim Sammeln der Daten wurde nicht auf die Reihenfolge geachtet. Sei

$$(y_1, \dots, y_n)$$

der Beobachtungsvektor in der richtigen Reihenfolge. Wir wissen also nicht, ob

$$(y_1, y_2, y_3, \dots, y_n) = (y^{(1)}, y^{(2)}, y^{(3)}, \dots, y^{(n)})$$

oder ob

$$(y_1, y_2, y_3, \dots, y_n) = (y^{(7)}, y^{(5)}, y^{(n)}, \dots, y^{(3)})$$

Formal läßt sich das jetzt so aufschreiben:

Sei

$$S_n = \{ \tau : \{1, \dots, n\} \rightarrow \{1, \dots, n\} \mid \tau \text{ ist bijektiv} \}$$

die Menge der Permutationen von  $\{1, \dots, n\}$ . Jedes  $\tau \in S_n$  permutiert also die Koordinaten und wir schreiben

$$\tau((y_1, y_2, \dots, y_n)) = (y_{\tau(1)}, y_{\tau(2)}, \dots, y_{\tau(n)})$$

Unser Wissensstand wird nicht durch  $\mathfrak{B}^{\otimes n}$  ausgedrückt sondern durch

$$\mathcal{C} = \{ C \in \mathfrak{B}^{\otimes n} \mid \tau(C) = C \quad \forall \tau \in S_n \}$$

Denn wir wissen nicht, welches  $(y_1, y_2, y_3, \dots, y_n)$  eingetreten ist. Wir wissen nur, welche Ereignismenge

$$\bigcup_{\tau \in S_n} \{ \tau((y_1, y_2, \dots, y_n)) \}$$

eingetreten ist.

Entsprechend wissen wir nicht, ob  $B \in \mathfrak{B}^{\otimes n}$  eingetreten ist. Wir wissen nur, ob

$$\bigcup_{\tau \in S_n} \tau(B)$$

eingetreten ist.

Sei

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

eine  $\mathfrak{B}^{\otimes n}/\mathfrak{B}$ -messbare Funktion. Wir wollen jetzt

$$\mathbb{E}_{P^{\otimes n}}[f|\mathcal{C}]$$

berechnen.

Ergebnis:

$$\mathbb{E}_{P^{\otimes n}}[f|\mathcal{C}](y_1, y_2, \dots, y_n) = \frac{1}{n!} \sum_{\tau \in S_n} f(y_{\tau(1)}, y_{\tau(2)}, \dots, y_{\tau(n)})$$

$$\forall (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$$

Sei nun

$$f(y_1, y_2, \dots, y_n) = y_1 + y_2 + \dots + y_n$$

Dann ist

$$\mathbb{E}_{P^{\otimes n}}[f|\mathcal{C}] = f$$

Der tieferliegende Grund für diese Gleichheit ist:

*f* ist schon  $\mathcal{C}$ -messbar! Und daraus folgt immer  $\mathbb{E}_{P^{\otimes n}}[f|\mathcal{C}] = f$

Dies ist eine grundlegende Eigenschaft der bedingten Erwartung; weitere Eigenschaften sind in nachfolgendem Satz festgehalten.

### Grundlegende Eigenschaften:

Wie in der Stochastik üblich schreiben wir das Integral bzgl. einem Wahrscheinlichkeitsmaß  $P$  auch häufig als Erwartungswert

$$\mathbb{E}_P[f] = \int f dP$$

**Satz 10.9** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und seien

$$f : \Omega \rightarrow \mathbb{R}, \quad f_1 : \Omega \rightarrow \mathbb{R}, \quad f_2 : \Omega \rightarrow \mathbb{R}$$

bzgl.  $P$  integrierbare Funktionen. Sei  $\mathcal{C}$  eine Unter- $\sigma$ -Algebra von  $\mathcal{A}$ .

Dann gilt:

a)  $\mathbb{E}_P[f|\mathcal{C}] \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$

b)  $\mathbb{E}_P[\mathbb{E}_P[f|\mathcal{C}]] = \mathbb{E}_P[f]$

c)  $f$  ist  $\mathcal{C}$ -messbar  $\Rightarrow \mathbb{E}_P[f|\mathcal{C}] = f$   $P$ -f.s.

d)  $f = g$   $P$ -f.s.  $\Rightarrow \mathbb{E}_P[f|\mathcal{C}] = \mathbb{E}_P[g|\mathcal{C}]$   $P$ -f.s.

e)  $f = \text{konst.} = \alpha$   $\Rightarrow \mathbb{E}_P[f|\mathcal{C}] = \alpha$   $P$ -f.s.

f) Für  $\alpha_1, \alpha_2 \in \mathbb{R}$  ist

$$\mathbb{E}_P[\alpha_1 f_1 + \alpha_2 f_2|\mathcal{C}] = \alpha_1 \mathbb{E}_P[f_1|\mathcal{C}] + \alpha_2 \mathbb{E}_P[f_2|\mathcal{C}] \quad P\text{-f.s.}$$

$$g) f_1 \leq f_2 \quad P\text{-f.s.} \quad \Rightarrow \quad \mathbb{E}_P[f_1|\mathcal{C}] \leq \mathbb{E}_P[f_2|\mathcal{C}] \quad P\text{-f.s.}$$

$$h) |\mathbb{E}_P[f|\mathcal{C}]| \leq \mathbb{E}_P[|f||\mathcal{C}] \quad P\text{-f.s.}$$

Für Folgen von integrierbaren Funktionen gelten analoge Aussagen zum Satz 6.6 von der monotonen Konvergenz und zum Satz 6.13 von der majorisierten Konvergenz:

**Satz 10.10** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Für jedes  $n \in \mathbb{N}$  sei  $f_n \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ , so dass

$$0 \leq f_1(\omega) \leq f_2(\omega) \leq \dots \leq \lim_{n \rightarrow \infty} f_n(\omega) < \infty \quad \forall \omega \in \Omega$$

Dann ist

$$\lim_{n \rightarrow \infty} \mathbb{E}_P[f_n|\mathcal{C}] = \mathbb{E}_P\left[\lim_{n \rightarrow \infty} f_n|\mathcal{C}\right] \quad P\text{-f.s.}$$

**Satz 10.11** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Sei  $g \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$  und für jedes  $n \in \mathbb{N}$  sei  $f_n \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ , so dass

$$|f_k| \leq g \quad \forall n \in \mathbb{N}$$

und

$$\lim_{n \rightarrow \infty} f_n = f \quad P\text{-f.s.}$$

Dann ist

$$\lim_{n \rightarrow \infty} \mathbb{E}_P[f_n|\mathcal{C}] = \mathbb{E}_P[f|\mathcal{C}] \quad P\text{-f.s.}$$

**Glättungseigenschaften:**

**Satz 10.12 (Glättungssatz)**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $\mathcal{C}$  eine Unter- $\sigma$ -Algebra von  $\mathcal{A}$ . Sei

$$f \in \mathcal{L}_1(\Omega, \mathcal{A}, P) \quad \text{und} \quad g \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$$

so dass auch

$$f \cdot g \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$$

Sei  $g$  außerdem  $\mathcal{C}$ -messbar.

Dann gilt

$$a) \mathbb{E}_P[g \cdot f|\mathcal{C}] = g \cdot \mathbb{E}_P[f|\mathcal{C}]$$

$$b) \mathbb{E}_P[g \cdot f] = \mathbb{E}_P[g \cdot \mathbb{E}_P[f|\mathcal{C}]]$$

**Satz 10.13 (Iteriertes Bedingen)**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum, seien  $\mathcal{C}$  und  $\mathcal{D}$  Unter- $\sigma$ -Algebren von  $\mathcal{A}$ , so dass

$$\mathcal{C} \subset \mathcal{D} \subset \mathcal{A}$$

Sei

$$f \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$$

Dann ist

$$\mathbb{E}_P\left[\mathbb{E}_P[f|\mathcal{D}]\middle|\mathcal{C}\right] = \mathbb{E}_P[f|\mathcal{C}]$$

**Beweis:** Für alle  $C \in \mathcal{C}$  ist auch  $C \in \mathcal{D}$  und somit

$$\int_C \mathbb{E}_P[f|\mathcal{D}] dP = \int_C f dP = \int_C \mathbb{E}_P[f|\mathcal{C}] dP$$

□

### Bedingte Erwartung und Unabhängigkeit

**Satz 10.14** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und

$$Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$$

Sei  $(\Omega', \mathcal{A}')$  ein weiterer Messraum und

$$T : \Omega \rightarrow \Omega'$$

eine  $\mathcal{A}/\mathcal{A}'$ -messbare Zufallsvariable.

Falls  $Y$  und  $T$  stochastisch unabhängig sind, dann ist

$$\mathbb{E}_P[Y|T] = \mathbb{E}_P[Y] \quad P\text{-f.s.}$$

**Beweis:** Zu zeigen ist, dass mit

$$\mathcal{C} := \{T^{-1}(A') | A' \in \mathcal{A}'\} \subset \mathcal{A}$$

für alle  $C \in \mathcal{C}$  gilt:

$$\int_C Y dP = \int_C \mathbb{E}_P[Y] dP$$

Sei  $A' \in \mathcal{A}'$ . Mit Transformationsformel (Satz 6.15) und Fubini (Satz 9.7) gilt:

$$\begin{aligned} \int_{T^{-1}(A')} \mathbb{E}_P[Y] dP &\stackrel{\mathbb{E}_P[Y] \text{ konstant}}{=} \mathbb{E}_P[Y] \int_{T^{-1}(A')} dP \\ &\stackrel{\text{Trafo}}{=} \mathbb{E}_P[Y] \int_{A'} dT(P) \\ &= \int_{A'} \mathbb{E}_P[Y] dT(P) \\ &= \int_{A'} \int_{\Omega} Y dP dT(P) \\ &\stackrel{\text{Trafo}}{=} \int_{\Omega'} I_{A'}(\omega') \int_{\mathbb{R}} y[Y(P)](dy)[T(P)](d\omega') \\ &\stackrel{\text{Fubini}}{=} \int_{\Omega' \times \mathbb{R}} I_{A'}(\omega') y[T(P) \times Y(P)](d(\omega', y)) \\ &\stackrel{\text{Übung 9.16}}{=} \int_{\Omega' \times \mathbb{R}} I_{A'}(\omega') y\left[\left(\frac{T}{Y}\right)(P)\right](d(\omega', y)) \\ &\stackrel{\text{Trafo}}{=} \int_{\Omega} (I_{A'} \circ T) \cdot Y dP \\ &= \int_{T^{-1}(A')} Y dP \end{aligned}$$

□

**Satz 10.15 (Alternative Formulierung von Satz 10.14)**

Es sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und

$$Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$$

Sei  $\mathcal{C}$  eine Unter- $\sigma$ -Algebra von  $\mathcal{A}$ , die von  $Y$  stochastisch unabhängig ist; das heißt, die  $\sigma$ -Algebren

$$\{Y^{-1}(B) \mid B \in \mathfrak{B}\} \quad \text{und} \quad \mathcal{C}$$

sind stochastisch unabhängig. Dann ist

$$\mathbb{E}_P[Y \mid \mathcal{C}] = \mathbb{E}_P[Y] \quad P\text{-f.s.}$$

**10.3 Faktorisierte bedingte Erwartung**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $(\Omega', \mathcal{A}')$  ein weiterer Messraum und

$$T : \Omega \rightarrow \Omega'$$

eine  $\mathcal{A}/\mathcal{A}'$ -messbare Zufallsvariable.

Wenn wir  $T$  beobachten können, spiegelt

$$\mathcal{C} = \{T^{-1}(A') \mid A' \in \mathcal{A}'\}$$

die zur Verfügung stehende Information wider.  
Jede Funktion der Form

$$\tilde{g} \circ T \quad \text{mit} \quad \tilde{g} : (\Omega', \mathcal{A}') \rightarrow (\mathbb{R}, \mathfrak{B}) \quad (10.9)$$

ist natürlich  $\mathcal{C}$ -messbar. Nachfolgendes Lemma zeigt, dass aber auch jede  $\mathcal{C}$ -messbare Funktion eine Darstellung der Form (10.9) hat.

**Lemma 10.16 (Faktorisierungslemma)**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und

$$g : \Omega \rightarrow \mathbb{R}$$

eine  $\mathcal{A}$ -messbare Funktion.

Sei  $(\Omega', \mathcal{A}')$  ein weiterer Messraum und

$$T : \Omega \rightarrow \Omega'$$

eine  $\mathcal{A}/\mathcal{A}'$ -messbare Zufallsvariable. Setze

$$\mathcal{C} = \{T^{-1}(A') \mid A' \in \mathcal{A}'\}$$

Dann gilt:  $g$  ist genau dann  $\mathcal{C}$ -messbar, wenn  $g$  darstellbar ist als

$$g = \tilde{g} \circ T \quad \text{mit} \quad \tilde{g} : (\Omega', \mathcal{A}') \rightarrow (\mathbb{R}, \mathfrak{B}) \quad (10.10)$$

Dies ermöglicht folgende Definition:

**Definition 10.17** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und

$$f : \Omega \rightarrow \mathbb{R}$$

eine bzgl.  $P$  integrierbare Funktion.

Sei  $(\Omega', \mathcal{A}')$  ein weiterer Messraum und

$$T : \Omega \rightarrow \Omega'$$

eine  $\mathcal{A}/\mathcal{A}'$ -messbare Zufallsvariable.

Nach Lemma 10.16 gibt es eine  $\mathcal{A}'$ -messbare Funktion

$$\tilde{g} : \Omega' \rightarrow \mathbb{R}$$

so dass

$$\mathbb{E}_P[f|T] = \tilde{g} \circ T$$

Dieses  $\tilde{g}$  heißt faktorisierte bedingte Erwartung von  $f$  gegeben  $T$  (unter  $P$ ) und man schreibt

$$\tilde{g}(t) = \mathbb{E}_P[f|T=t] \quad \forall t \in \Omega'$$

**Satz 10.18** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und

$$f : \Omega \rightarrow \mathbb{R}$$

eine bzgl.  $P$  integrierbare Funktion. Sei  $(\Omega', \mathcal{A}')$  ein weiterer Messraum und

$$T : \Omega \rightarrow \Omega'$$

eine  $\mathcal{A}/\mathcal{A}'$ -messbare Zufallsvariable. Dann gilt:

Die faktorisierte bedingte Erwartung von  $f$  gegeben  $T$

$$\tilde{g} : \Omega' \rightarrow \mathbb{R}, \quad \tilde{g}(t) = \mathbb{E}_P[f|T=t] \quad \forall t \in \Omega'$$

ist integrierbar bzgl. dem Bildmaß  $T(P)$  und

$$\begin{aligned} \int_{A'} \mathbb{E}_P[f|T=t] [T(P)](dt) &= \int_{T^{-1}(A')} \mathbb{E}_P[f|T](\omega) P(d\omega) = \\ &= \int_{T^{-1}(A')} f(\omega) P(d\omega) \end{aligned} \quad (10.11)$$

$$\forall A' \in \mathcal{A}'.$$

Die faktorisierte bedingte Erwartung von  $f$  gegeben  $T$  ist nur  $T(P)$ -fast sicher eindeutig bestimmt.

**Beweis:** Die Integrierbarkeit von  $\tilde{g}$  bzgl.  $T(P)$  und Gleichung (10.11) folgen direkt aus der Transformationsformel (Satz 6.15) und der Definition der bedingten Erwartung (Definition 10.3)

Die  $T(P)$ -fast sichere Eindeutigkeit folgt aus (10.11) und Satz 6.12 b).  $\square$

**Bemerkung 10.19** An dieser Stelle ergibt sich (zum ersten Mal) tatsächlich ein direkter Zusammenhang mit dem elementaren bedingten Erwartungswert aus Abschnitt 10.1.1:

Aus (10.11) folgt für  $t \in \Omega'$

$$\begin{aligned} \int_{\{T=t\}} f dP &= \int_{\{\omega \in \Omega \mid T(\omega)=t\}} f(\omega) P(d\omega) \\ &= \int_{\{t\}} \mathbb{E}_P[f \mid T = s] [T(P)](ds) \\ &= [T(P)](\{t\}) \cdot \mathbb{E}_P[f \mid T = t] \\ &= P(T = t) \cdot \mathbb{E}_P[f \mid T = t] \end{aligned}$$

Also

$$\mathbb{E}_P[f \mid T = t] = \frac{1}{P(T = t)} \int_{\{T=t\}} f dP \quad (10.12)$$

Das heißt,  $\mathbb{E}_P[f \mid T = t]$  ist genau der elementare Bedingte Erwartungswert von  $f$  gegeben  $T = t$ . Allerdings ergibt Gleichung (10.12) keinen Sinn, falls

$$P(T = t) = 0$$

Im Allgemeinen wird genau dieser Fall  $P(T = t) = 0$  für die meisten Punkte  $t \in \Omega'$  eintreten (z.B. falls  $T$  normalverteilt). Der elementare bedingte Erwartungswert kann dann nicht definiert werden, aber die Definition der faktorisierten bedingten Erwartung ist auch hier problemlos möglich.

**Bemerkung 10.20 (Ergänzung zum Beweis von Satz 10.18)**

Sei  $(\Omega, \mathcal{A})$  ein Messraum und  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$   
Sei  $(\Omega', \mathcal{A}')$  ein weiterer Messraum und

$$T : \Omega \rightarrow \Omega'$$

eine  $\mathcal{A}/\mathcal{A}'$ -messbare Zufallsvariable. Setze

$$\mathcal{C} = \{T^{-1}(A') \mid A' \in \mathcal{A}'\}$$

Seien  $\tilde{h}_1 \in \mathcal{L}_1(\Omega', \mathcal{A}', T(P))$  und  $\tilde{h}_2 \in \mathcal{L}_1(\Omega', \mathcal{A}', T(P))$ .

Dann folgt aus der Transformationsformel (Satz 6.15) und aus Satz 6.12b) folgende Äquivalenz:

$$\tilde{h}_1 \circ T = \tilde{h}_2 \circ T \quad P\text{-f.s.} \quad \Leftrightarrow \quad \tilde{h}_1 = \tilde{h}_2 \quad T(P)\text{-f.s.}$$

## 10.4 Bedingte Dichten

Natürlich stellt sich nun die Frage, wie man (faktorisierte) bedingte Erwartungen ausrechnet. Dies ist in der Praxis häufig wesentlich einfacher, als es auf den ersten Blick scheint.

In diesem Abschnitt ist  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum.

$(\Omega_1, \mathcal{A}_1)$  und  $(\Omega_2, \mathcal{A}_2)$  sind zwei weitere Messräume.

$$S : \Omega \rightarrow \Omega_1$$

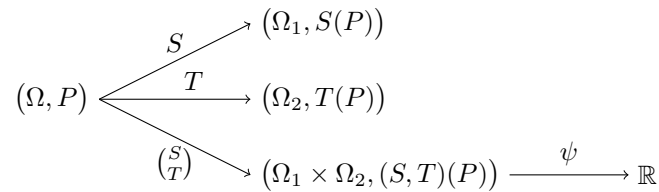
ist eine  $\mathcal{A}/\mathcal{A}_1$ -messbare Zufallsvariable und

$$T : \Omega \rightarrow \Omega_2$$

ist eine  $\mathcal{A}/\mathcal{A}_2$ -messbare Zufallsvariable.

Dann ist das Bildmaß  $(S, T)(P)$  ein Wahrscheinlichkeitsmaß auf

$$(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$$



### Annahme:

Sei  $\mu_1$  ein  $\sigma$ -endliches Maß auf  $(\Omega_1, \mathcal{A}_1)$  und  $\mu_2$  ein  $\sigma$ -endliches Maß auf  $(\Omega_2, \mathcal{A}_2)$ , so dass

$$(S, T)(P) \ll \mu_1 \otimes \mu_2$$

Das heißt also nach dem Satz von Radon-Nikodym (Satz 7.4):

Es existiert eine Dichte

$$f_{(S, T)} : \Omega_1 \times \Omega_2 \rightarrow [0; \infty), \quad (s, t) \mapsto f_{(S, T)}(s, t)$$

von  $(S, T)(P)$  bzgl.  $\mu_1 \otimes \mu_2$ :

$$d[(S, T)(P)] = f_{(S, T)} d(\mu_1 \otimes \mu_2)$$

### Berechnung der bedingten Erwartung über die bedingte Dichte

Sei

$$\psi : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$$

eine  $\mathcal{A}_1 \otimes \mathcal{A}_2$ -messbare Funktion, so dass

$$\psi(S, T) : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto \psi(S(\omega), T(\omega))$$

integrierbar bzgl.  $P$  ist, d.h.  $\psi(S, T) \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ .

Wir wollen nun

$$\mathbb{E}_P[\psi(S, T) | T] \quad \text{und} \quad \mathbb{E}_P[\psi(S, T) | T = t]$$

berechnen.



Aufgrund der Voraussetzungen ist automatisch auch

$$T(P) \ll \mu_2$$

Sei dann  $f_T : \Omega_2 \rightarrow [0; \infty)$  die Dichte von  $T(P)$  bzgl.  $\mu_2$ :

$$d[T(P)] = f_T d\mu_2$$

( $f_T$  lässt sich übrigens ganz einfach aus  $f_{(S,T)}$  mit Hilfe des Satzes von Fubini (Satz 9.7) berechnen. Wie?)

Setze

$$f_{S|T}(s|t) := \frac{f_{(S,T)}(s,t)}{f_T(t)} \quad \text{falls } f_T(t) > 0$$

und

$$f_{S|T}(s|t) := 0 \quad \text{falls } f_T(t) = 0$$

Dann gilt

$$\mathbb{E}_P[\psi(S,T)|T=t] = \int_{\Omega_1} \psi(s,t) f_{S|T}(s|t) \mu_1(ds) \quad (10.13)$$

und

$$\mathbb{E}_P[\psi(S,T)|T](\omega) = \int_{\Omega_1} \psi(s,T(\omega)) f_{S|T}(s|T(\omega)) \mu_1(ds) \quad (10.14)$$

**Beweis von (10.13) und (10.14):**

Sei zunächst

$$\psi : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}, \quad \omega \mapsto \psi(s,t)$$

eine *nicht-negative* bzgl.  $(S,T)(P)$  integrierbare Funktion. Sei  $A_2 \in \mathcal{A}_2$  fest. Trivialerweise ist

$$f_T > 0 \quad T(P)\text{-fast sicher}$$

und somit existiert eine  $T(P)$ -Nullmenge  $N_2 \in \mathcal{A}_2$  (Def. 6.9), so dass

$$f_T > 0 \quad \forall t \in \Omega_2 \setminus N_2$$

und

$$f_{(S,T)}(s,t) = f_{S|T}(s|t) f_T(t) \quad \forall s \in \Omega_1, \quad \forall t \in \Omega_2 \setminus N_2$$

Setze

$$\Psi(t) = \int_{\Omega_1} \psi(s,t) f_{S|T}(s|t) \mu_1(ds) \quad \forall t \in \Omega_2$$

Zeige nun  $\Psi \in \mathcal{L}_1(\Omega_2, \mathcal{A}_2, T(P))$ :

Die Funktion

$$\psi f_{(S,T)} : (s,t) \mapsto \psi(s,t) f_{(S,T)}(s,t)$$

ist  $\mathcal{A}_1 \otimes \mathcal{A}_2$ -messbar und nicht-negativ. Wegen der Integrierbarkeit von  $\psi(S,T)$  bzgl.  $P$  und

$$\begin{aligned} \int_{T^{-1}(A_2)} \psi(S,T) dP &= \int_{T^{-1}(A_2)} \psi \circ (S,T) dP \\ &\stackrel{(*)}{=} \int_{\Omega_1 \times A_2} \psi(s,t) [(S,T)(P)](d(s,t)) \\ &= \int_{\Omega_1 \times A_2} \psi(s,t) f_{(S,T)}(s,t) [\mu_1 \otimes \mu_2](d(s,t)) \end{aligned} \quad (10.15)$$

ist  $\psi f_{(S,T)}$  auch integrierbar bzgl.  $\mu_1 \otimes \mu_2$ . Hierbei folgt (\*) aus der Transformationsformel (Satz 6.15).

Nach dem Satz von Fubini (Satz 9.7) ist dann

$$t \mapsto \int_{\Omega_1} \psi(s, t) f_{(S,T)}(s, t) \mu_1(ds)$$

eine  $\mathcal{A}_2$ -messbare (reellwertige) Funktion. Somit ist auch

$$\Psi : t \mapsto I_{N_2^C}(t) \frac{1}{f_T(t)} \int_{\Omega_1} \psi(s, t) f_{(S,T)}(s, t) \mu_1(ds)$$

eine  $\mathcal{A}_2$ -messbare Funktion und

$$\begin{aligned} \int_{\Omega_2} \Psi(t) [T(P)](dt) &= \int_{\Omega_2} \frac{I_{N_2^C}(t)}{f_T(t)} \int_{\Omega_1} \psi(s, t) f_{(S,T)}(s, t) \mu_1(ds) [T(P)](dt) \\ &= \int_{\Omega_2} \frac{I_{N_2^C}(t)}{f_T(t)} \int_{\Omega_1} \psi(s, t) f_{(S,T)}(s, t) \mu_1(ds) f_T(t) \mu_2(dt) \\ &= \int_{\Omega_1 \times N_2^C} \psi(s, t) f_{(S,T)}(s, t) [\mu_1 \otimes \mu_2](d(s, t)) < \infty \end{aligned}$$

Also ist  $\Psi \in \mathcal{L}_1(\Omega_2, \mathcal{A}_2, T(P))$ .

Dann ist nach dem Transformationssatz (Satz 6.15) auch  $\Psi \circ T \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$  und es gilt

$$\begin{aligned} \int_{T^{-1}(A_2)} \Psi \circ T dP &\stackrel{(*)}{=} \int_{A_2} \Psi(t) [T(P)](dt) \\ &\stackrel{(7.1)}{=} \int_{A_2} \Psi(t) f_T(t) \mu_2(dt) = \int_{A_2} \int_{\Omega_1} \psi(s, t) f_{(S,T)}(s, t) \mu_1(ds) \mu_2(dt) \\ &\stackrel{(**)}{=} \int_{\Omega_1 \times A_2} \psi(s, t) f_{(S,T)}(s, t) [\mu_1 \otimes \mu_2](d(s, t)) \\ &= \int_{\Omega_1 \times A_2} \psi(s, t) [(S, T)(P)](d(s, t)) \stackrel{(*)}{=} \int_{T^{-1}(A_2)} \psi \circ (S, T) dP \\ &= \int_{T^{-1}(A_2)} \psi(S, T) dP \end{aligned}$$

wobei (\*) jeweils eine Anwendung der Transformationsformel (Satz 6.15) ist und(\*\*) aus dem Satz von Fubini (Satz 9.7) folgt.

Dies zeigt (10.14) und aus (10.14) folgt sofort (10.13).

Der allgemeine Fall, bei dem  $\psi$  sowohl positive als auch negative Werte annehmen kann, folgt hierraus, indem man einfach  $\psi$  in seinen Positivteil  $\psi^+$  und seinen Negativteil  $\psi^-$  zerlegt.

**Beispiel:**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Seien

$$Y : \Omega \rightarrow \mathbb{R}$$

und

$$X : \Omega \rightarrow \mathbb{R}^k$$

zwei  $\mathcal{A}$ -messbare Zufallsvariablen.

Sei  $(Y, X)$  eine stetige Zufallsvariable, d.h.

$$(Y, X)(P) \ll \lambda^{k+1}$$

zum Beispiel sei  $(Y, X)$  multivariat normalverteilt:

$$(Y, X)(P) = \mathcal{N}_{k+1}(a, C)$$

Wir können nun die bedingten Erwartungen

$$\mathbb{E}_P[Y|X] \quad \text{und} \quad \mathbb{E}_P[Y|X=x]$$

berechnen.

Dazu setzen wir

$$S = Y, \quad T = X, \quad (\Omega_1, \mathcal{A}_1) = (\mathbb{R}, \mathfrak{B}), \quad (\Omega_2, \mathcal{A}_2) = (\mathbb{R}^k, \mathfrak{B}^{\otimes k})$$

und

$$\mu_1 = \lambda, \quad \mu_2 = \lambda^k$$

und

$$\psi(y, x) = y \quad \forall y \in \mathbb{R}, \quad \forall x \in \mathbb{R}^k$$

Die Dichte

$$f_{(Y,X)} : \mathbb{R} \times \mathbb{R}^k \rightarrow [0; \infty), \quad (y, x) \mapsto f_{(Y,X)}(y, x)$$

ist dann einfach die Normalverteilungsdichte von  $\mathcal{N}_{k+1}(a, C)$  und die Verteilungsdichte von  $X$  bzgl.  $\lambda^k$  ist

$$f_X(x) = \int_{\mathbb{R}} f_{(Y,X)}(y, x) \lambda(dy) \quad x \in \mathbb{R}$$

Die sog. „bedingte Dichte“  $f_{Y|X}$  ist dann

$$f_{Y|X}(y|x) := \frac{f_{(Y,X)}(y, x)}{f_X(x)} \quad \text{falls } f_X(x) > 0$$

und

$$f_{Y|X}(y|x) := 0 \quad \text{falls } f_X(x) = 0$$

Die bedingten Erwartungen sind schließlich

$$\mathbb{E}_P[Y|X=x] = \int_{\mathbb{R}} y f_{Y|X}(y|x) \lambda(dy)$$

und

$$\mathbb{E}_P[Y|X](\omega) = \int_{\mathbb{R}} y f_{Y|X}(y|X(\omega)) \lambda(dy)$$

## 10.5 Bedingte Wahrscheinlichkeiten

### 10.5.1 Idee: Wahrscheinlichkeiten als spezielle Erwartungswerte

Bisher haben wir uns ausschließlich mit bedingten Erwartungen beschäftigt; nun wollen wir zu bedingten Wahrscheinlichkeiten übergehen.

Normale Wahrscheinlichkeiten kann man auch als spezielle Erwartungswerte ansehen:

$$P(A) = \mathbb{E}_P[I_A]$$

Daher ist es naheliegend, bedingte Wahrscheinlichkeiten in analoger Weise als spezielle bedingte Erwartungen zu definieren:

$$P(A|\mathcal{C}) = \mathbb{E}_P[I_A|\mathcal{C}]$$

bzw.

$$P(A|T) = \mathbb{E}_P[I_A|T]$$

und

$$P(A|T = t) = \mathbb{E}_P[I_A|T = t]$$

**Achtung:** Eine so definierte bedingte Wahrscheinlichkeit ist keine Zahl zwischen 0 und 1, sondern eine Funktion

$$\omega \mapsto P(A|\mathcal{C})(\omega)$$

Für jedes  $t$  ist jedoch

$$P(A|T = t)$$

eine Zahl zwischen 0 und 1; und wie bei der faktorisierten bedingten Erwartung ergibt sich auch für  $P(A|T = t)$  ein enger Zusammenhang mit den elementaren bedingten Wahrscheinlichkeiten aus Abschnitt 10.1.1.

Die Zahl  $P(A|T = t)$  können wir tatsächlich dann als Wahrscheinlichkeit von Ereignis  $A$  unter der Bedingung  $T = t$  interpretieren.

Der Abschnitt 10.5 über bedingte Wahrscheinlichkeiten ist folgendermaßen angeordnet:

1. Im nächsten Abschnitt werden bedingte Wahrscheinlichkeiten formal in der oben beschriebenen Weise definiert.
2. Der darauffolgenden Abschnitt 10.5.3 behandelt Markov-Kerne. Markov-Kerne sind mathematische Konstrukte, die für die Wahrscheinlichkeitstheorie wichtig sind.
3. Häufig (und zwar in den wichtigsten Fällen) sind bedingte Wahrscheinlichkeiten auch Markov-Kerne. Davon handelt Abschnitt 10.5.4, in dem ein verallgemeinerter Satz von Fubini vorgestellt wird. Hier zeigt sich deutlich, was man unter bedingte Wahrscheinlichkeiten versteht.
4. Im Abschnitt 10.5.5 werden bedingte Wahrscheinlichkeiten konkret berechnet - und zwar mit Hilfe von bedingten Dichten.

## 10.5.2 Definition: Bedingte Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeiten werden (wie bereits beschrieben) über bedingte Erwartungen in naheliegender Weise definiert:

**Definition 10.21** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $\mathcal{C}$  eine Unter- $\sigma$ -Algebra von  $\mathcal{A}$ . Dann heißt für  $A \in \mathcal{A}$

$$P(A|\mathcal{C}) := \mathbb{E}_P[I_A|\mathcal{C}]$$

bedingte Wahrscheinlichkeit von  $A$  gegeben  $\mathcal{C}$  (unter  $P$ ).

**Definition 10.22** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Sei  $(\Omega', \mathcal{A}')$  ein Messraum und

$$T : \Omega \rightarrow \Omega'$$

eine  $\mathcal{A}/\mathcal{A}'$ -messbare Zufallsvariable.

Dann heißt für  $A \in \mathcal{A}$

$$P(A|T) := \mathbb{E}_P[I_A|T]$$

bedingte Wahrscheinlichkeit von  $A$  gegeben  $T$  (unter  $P$ ) und

$$t \mapsto P(A|T=t) \quad \text{mit} \quad P(A|T=t) = \mathbb{E}_P[I_A|T=t] \quad \forall t \in \Omega'$$

faktorierte bedingte Wahrscheinlichkeit von  $A$  gegeben  $T$  (unter  $P$ ).

Wie wir das von einer Wahrscheinlichkeit erwarten, gilt dann (aufgrund von Satz 10.9 tatsächlich für jedes  $A \in \mathcal{A}$ )

$$0 \leq P(A|\mathcal{C}) \leq 1 \quad P\text{-f.s.}$$

und

$$P(\emptyset|\mathcal{C}) = 0, \quad P(\Omega|\mathcal{C}) = 1 \quad P\text{-f.s.}$$

Aufgrund von Satz 10.10 ist außerdem für  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$  mit  $A_n \cap A_m = \emptyset, n \neq m$

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n \middle| \mathcal{C}\right) &= \mathbb{E}_P\left[\lim_{m \rightarrow \infty} \sum_{n=1}^m I_{A_n} \middle| \mathcal{C}\right] = \lim_{m \rightarrow \infty} \sum_{n=1}^m \mathbb{E}_P[I_{A_n}|\mathcal{C}] = \\ &= \sum_{n=1}^{\infty} P(A_n|\mathcal{C}) \quad P\text{-f.s.} \end{aligned}$$

Die Funktion

$$t \mapsto P(A|T=t)$$

wird im Folgendem mit

$$P(A|T=\cdot)$$

bezeichnet. Aus Bemerkung 10.20 folgt hieraus dann auch für jedes  $A \in \mathcal{A}$

$$0 \leq P(A|T=\cdot) \leq 1 \quad T(P)\text{-f.s.}$$

und

$$P(\emptyset|T=\cdot) = 0, \quad P(\Omega|T=\cdot) = 1 \quad T(P)\text{-f.s.}$$

und für  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$  mit  $A_n \cap A_m = \emptyset, n \neq m$

$$P\left(\bigcup_{n=1}^{\infty} A_n \mid T = \cdot\right) = \sum_{n=1}^{\infty} P(A_n \mid T = \cdot) \quad T(P)\text{-f.s.}$$

oder mit anderen Worten

Für jedes  $A \in \mathcal{A}$  ist

$$0 \leq P(A \mid T = t) \leq 1 \quad \text{für } T(P)\text{-fast alle } t \in \Omega'$$

und

$$P(\emptyset \mid T = t) = 0, \quad P(\Omega \mid T = t) = 1 \quad \text{für } T(P)\text{-fast alle } t \in \Omega'$$

und für  $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$  mit  $A_n \cap A_m = \emptyset, n \neq m$

$$P\left(\bigcup_{n=1}^{\infty} A_n \mid T = t\right) = \sum_{n=1}^{\infty} P(A_n \mid T = t) \quad \text{für } T(P)\text{-fast alle } t \in \Omega'$$

Es stellt sich daher die **Frage**, ob dann auch für  $T(P)$ -fast alle  $t \in \Omega'$

$$A \mapsto P(A \mid T = t)$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  ist.

**Dies ist im Allgemeinen nicht der Fall!!!**

**Grund:** Die  $T(P)$ -Nullmengen, auf denen die obigen Aussagen gelten, können von  $A$  abhängen.

Zum Beispiel ist für jedes  $A \in \mathcal{A}$

$$0 \leq P(A \mid T = t) \quad \text{für } T(P)\text{-fast alle } t \in \Omega'$$

Das heißt: Es gibt eine Menge  $N'_A \in \mathcal{A}'$  mit

$$[T(P)](N'_A) = 0$$

und

$$0 \leq P(A \mid T = t) \quad \forall t \in \Omega' \setminus N'_A$$

Das heißt für jedes

$$t_0 \in \Omega \setminus \left(\bigcup_{A \in \mathcal{A}} N'_A\right)$$

ist tatsächlich

$$0 \leq P(A \mid T = t_0) \quad \forall A \in \mathcal{A}$$

Weil  $\mathcal{A}$  im Allgemeinen überabzählbar ist, kann es allerdings sein, dass

$$T(P)\left(\bigcup_{A \in \mathcal{A}} N'_A\right) > 0$$

– und noch schlimmer, es kann sogar sein, dass

$$\bigcup_{A \in \mathcal{A}} N'_A = \Omega \quad \text{also} \quad \Omega \setminus \left( \bigcup_{A \in \mathcal{A}} N'_A \right) = \emptyset$$

In dem Fall, dass für jedes  $t \in \Omega'$

$$A \mapsto P(A|T = t)$$

tatsächlich ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  definiert, ist die Abbildung

$$(t, A) \mapsto P(A|T = t)$$

ein „Markov-Kern“. Diese Situation ist besonders wünschenswert und häufig auch erreichbar.

### 10.5.3 Einschub: Markov-Kerne

Markov-Kerne spielen in der Stochastik eine große Rolle, zum Beispiel als bedingte Wahrscheinlichkeiten, bei Stochastischen Prozessen oder als randomisierte Tests.

**Definition 10.23** Seien  $(\Omega, \mathcal{A})$  und  $(\Xi, \mathcal{D})$  zwei Messräume. Eine Funktion

$$\begin{aligned} K : \Xi \times \mathcal{A} &\rightarrow \mathbb{R} \\ (x, A) &\mapsto K(x, A) \end{aligned}$$

heißt *Markov-Kern*, falls

(i) für jedes  $A \in \mathcal{A}$

$$K(A, \cdot) : \Xi \rightarrow \mathbb{R}, \quad x \mapsto K(A, x)$$

eine  $\mathcal{D}$ -messbare Funktion ist und

(ii) für jedes  $x \in \Xi$

$$K(\cdot, x) : \mathcal{A} \rightarrow \mathbb{R}, \quad A \mapsto K(A, x)$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  ist.

Aus der Definition folgt sofort, dass für einen Markov-Kern  $K$

$$0 \leq K(x, A) \leq 1 \quad \forall (x, A) \in \Xi \times \mathcal{A}$$

**Beispiel: randomisierte Tests**

Sei  $(\Xi, \mathcal{D})$  ein Messraum,  $P_0$  und  $P_1$  seien zwei Wahrscheinlichkeitsmaße auf  $(\Xi, \mathcal{D})$ . Wir betrachten das Testproblem

$$H_0 : P_0 \quad \text{gegen} \quad H_1 : P_1$$

Ein Test ist eine  $\mathcal{A}$ -messbare Abbildung

$$\varphi : \Xi \rightarrow [0; 1]$$

Sei  $x \in \Xi$  die Beobachtung. Dann bedeutet

$$\varphi(x) = 1 \quad \text{Lehne } H_0 \text{ ab.}$$

und

$$\varphi(x) = 0 \quad \text{Lehne } H_0 \text{ nicht ab.}$$

und

$$\varphi(x) = 0,83$$

bedeutet

Starte einen (künstlichen) Zufallsprozess mit Verteilung

$$\text{Bin}(1; 0,83)$$

Wenn das Ergebnis dieses Zufallsprozesses gleich 1 ist, dann lehne die  $H_0$ -Hypothese ab, und wenn das Ergebnis dieses Zufallsprozesses gleich 0 ist, dann lehne die  $H_0$ -Hypothese nicht ab.

Allgemein formuliert, lautet die Testentscheidung:

Wenn  $x \in \Xi$  die Beobachtung ist, dann starte einen (künstlichen) Zufallsprozess mit Verteilung

$$\text{Bin}(1; \varphi(x))$$

Wenn das Ergebnis dieses Zufallsprozesses gleich 1 ist, dann lehne die  $H_0$ -Hypothese ab, und wenn das Ergebnis dieses Zufallsprozesses gleich 0 ist, dann lehne die  $H_0$ -Hypothese nicht ab.

(Die Fälle  $\varphi(x) = 1$  und  $\varphi(x) = 0$  sind hierbei als Spezialfälle schon enthalten.)

Ein solcher randomisierter Test  $\varphi$  läßt sich auch in einen Markov-Kern umschreiben: Setze

$$(\Omega, \mathcal{A}) = (\{0, 1\}, \mathcal{P}(\{0, 1\}))$$

und

$$K(x, A) = \text{Bin}(1; \varphi(x))(A) \quad \forall x \in \Xi, \quad \forall A \subset \{0, 1\}$$

Dann ist

$$(x, A) \mapsto K(x, A)$$

ein Markov-Kern.



**Notation 10.24** Seien  $(\Omega, \mathcal{A})$  und  $(\Xi, \mathcal{D})$  zwei Messräume und

$$K : \Xi \times \mathcal{A} \rightarrow \mathbb{R}$$

ein Markov-Kern. Dann ist für jedes  $x \in \Xi$

$$A \mapsto K(x, A)$$

ein Wahrscheinlichkeitsmaß

$$\mu_x \quad \text{mit} \quad \mu_x(A) = K(x, A) \quad \forall A \in \mathcal{A}$$

Für das Integral bzgl.  $\mu_x$

$$\int_A f(\omega) \mu_x(d\omega)$$

schreibt man auch einfach

$$\int_A f(\omega) K(x, d\omega)$$

#### 10.5.4 Verallgemeinerter Satz von Fubini

Wie wir in Abschnitt 10.5.2 gesehen haben, definiert nicht jede bedingte Wahrscheinlichkeit einen Markov-Kern

$$(x, A) \mapsto P(A | X = x)$$

Aber in für die Statistik wichtigen Fällen kann  $P(A | X = x)$  so gewählt werden ( $P(A | X = x)$  ist ja nur  $X(P)$ -fast sicher eindeutig), dass

$$(x, A) \mapsto P(A | X = x)$$

ein Markov-Kern ist.

Im Folgenden sind  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(\Omega_1, \mathcal{A}_1)$ ,  $(\Omega_2, \mathcal{A}_2)$  zwei Messräume.

Hierbei sei  $(\Omega_1, \mathcal{A}_1)$  gleich

$$(\mathbb{R}^k, \mathfrak{B}^{\otimes k}), \quad (\{1, \dots, m\}, \mathcal{P}(\{1, \dots, m\})) \quad \text{oder} \quad (\mathbb{N}, \mathcal{P}(\mathbb{N})) \quad (10.16)$$

Sei außerdem

$$Z : \Omega \rightarrow \Omega_1$$

eine  $\mathcal{A}/\mathcal{A}_1$ -messbare Zufallsvariable und

$$X : \Omega \rightarrow \Omega_2$$

eine  $\mathcal{A}/\mathcal{A}_2$ -messbare Zufallsvariable.

**Bemerkung 10.25** Man denke hierbei ruhig an eine Regression, wobei  $Z$  die Zielvariable (Response) und  $X$  die erklärende Variable (Kovariable) ist. Bedingung (10.16) besagt also nur, dass die Zielvariable ( $k$ -dimensionale) reelle Werte

$$Z = z \in \mathbb{R}^k$$

oder diskrete Werte

$$Z = z \in \{1, \dots, m\} \text{ bzw. } \mathbb{N}$$

hat.

Wir interessieren uns dann für die bedingte Verteilung von  $Z$  gegeben  $X = x$ :

$$P^{Z|X=x}$$

**Definition 10.26** Seien  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(\Omega_1, \mathcal{A}_1)$ ,  $(\Omega_2, \mathcal{A}_2)$  zwei Messräume.

Sei außerdem

$$Z : \Omega \rightarrow \Omega_1$$

eine  $\mathcal{A}/\mathcal{A}_1$ -messbare Zufallsvariable und

$$X : \Omega \rightarrow \Omega_2$$

eine  $\mathcal{A}/\mathcal{A}_2$ -messbare Zufallsvariable.

Die faktorisierte bedingte Verteilung von  $Z$  gegeben  $X$  ist definiert als

$$P^{Z|X=\cdot} : \mathcal{A}_1 \mapsto P(Z^{-1}(A_1) | X = \cdot)$$

Das heißt:  $P^{Z|X=\cdot}(A_1)$  ist die bedingte Wahrscheinlichkeit von  $Z^{-1}(A_1)$  gegeben  $X$  unter  $P$

Falls für alle  $x \in \Omega_2$

$$A_1 \mapsto P^{Z|X=x}(A_1)$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega_1, \mathcal{A}_1)$  ist, so heißt  $P^{Z|X=\cdot}$  reguläre faktorisierte bedingte Verteilung von  $Z$  gegeben  $X$

Es gilt

$$P^{Z|X=x}(A_1) = P(Z^{-1}(A_1) | X = x) = \mathbb{E}_P [I_{Z^{-1}(A_1)} | X = x]$$

und daher

$$\begin{aligned} \int_{A_2} P^{Z|X=x}(A_1) [X(P)](dx) &= P(Z^{-1}(A_1) \cap X^{-1}(A_2)) \\ &= P(Z \in A_1, X \in A_2) \end{aligned}$$

Aus den Definitionen folgt sofort:

$P^{Z|X=\cdot}$  ist genau dann eine reguläre faktorisierte bedingte Verteilung von  $Z$  gegeben  $X$ , wenn

$$(x, A_2) \mapsto P^{Z|X=x}(A_1)$$

ein Markov-Kern ist.

$P^{Z|X=x}$  läßt sich nun wirklich als die Wahrscheinlichkeitsverteilung von  $Z$  interpretieren, wenn wir wissen, dass  $X = x$  vorliegt.

Der nachfolgende Satz ist der letzte große Satz dieses Vorlesungsteils. Er beschreibt, dass

$$P^{Z|X=x}$$

genau das tut, was wir von einer bedingten Verteilung von  $Z$  gegeben  $X = x$  erwarten.

Für die Bildverteilung von  $P$  unter einer Zufallsvariablen  $Y$  schreiben wir im folgenden auch

$$P^Y \quad \text{statt} \quad Y(P)$$

**Satz 10.27 (Verallgemeinerter Satz von Fubini)**

Seien  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2)$  zwei Messräume.  $(\Omega_1, \mathcal{A}_1)$  erfülle Bedingung (10.16).

Sei außerdem

$$Z : \Omega \rightarrow \Omega_1$$

eine  $\mathcal{A}/\mathcal{A}_1$ -messbare Zufallsvariable und

$$X : \Omega \rightarrow \Omega_2$$

eine  $\mathcal{A}/\mathcal{A}_2$ -messbare Zufallsvariable.

Dann existiert eine reguläre faktorisierte bedingte Verteilung von  $Z$  gegeben  $X$

$$P^{Z|X=\cdot}$$

Für jede  $\mathcal{A}_1 \otimes \mathcal{A}_2$ -messbare Funktion

$$\psi : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$$

gilt dann

$$\mathbb{E}_P[\psi(Z, X) | X = x] = \int_{\Omega_1} \psi(z, x) P^{Z|X=x}(dz) \quad X(P)\text{-f.s.}$$

und daher auch

$$\int_{\Omega_1 \times \Omega_2} \psi(z, x) P^{(Z, X)}(d(z, x)) = \int_{\Omega_2} \int_{\Omega_1} \psi(z, x) P^{Z|X=x}(dz) P^X(dx)$$

**Bemerkungen zum Beweis:**

Der Beweis, dass eine reguläre faktorisierte bedingte Verteilung von  $Z$  gegeben  $X$  existiert benötigt ein wenig Topologie. (Es wird benutzt, dass  $(\Omega_1, \mathcal{A}_1)$  in jedem der drei zugelassenen Fällen ein sogenannter „Polnischer Raum“ ist.)

Für die Gleichung

$$\mathbb{E}_P[\psi(Z, X) | X = x] = \int_{\Omega_1} \psi(z, x) P^{Z|X=x}(dz) \quad X(P)\text{-f.s.}$$

ist es entscheidend, dass

$$A_1 \mapsto P^{Z|X=x}(A_1)$$

tatsächlich ein (Wahrscheinlichkeits-)Maß ist. Sonst wäre das Integral bzgl.  $P^{Z|X=x}$  ja auch gar nicht definiert.

Die letzte Aussage folgt dann aus

$$\begin{aligned}
 & \int_{\Omega_2} \int_{\Omega_1} \psi(z, x) P^{Z|X=x}(dz) P^X(dx) \\
 &= \int_{\Omega_2} \mathbb{E}_P[\psi(Z, X) | X = x] P^X(dx) \\
 &= \int_{\Omega} \mathbb{E}_P[\psi(Z, X) | X] dP = \int_{\Omega} \psi(Z, X) dP \\
 &= \int_{\Omega_1 \times \Omega_2} \psi(z, x) dP^{(Z, X)}
 \end{aligned}$$

□

**Beispiel 10.28** Sei

$$(\Omega_1, \mathcal{A}_1) = (\mathbb{R}, \mathfrak{B})$$

und

$$\psi(z, x) = z \quad \forall (z, x) \in \Omega_1 \times \Omega_2$$

Dann gilt also

$$\mathbb{E}_P[Z | X = x] = \int_{\Omega_1} z P^{Z|X=x}(dz) \quad X(P) \text{-f.s.}$$

### 10.5.5 Berechnung der bedingten Verteilung

In diesem Abschnitt ist  $(\Omega, \mathcal{A}, P)$  wieder ein Wahrscheinlichkeitsraum und  $(\Omega_1, \mathcal{A}_1)$  und  $(\Omega_2, \mathcal{A}_2)$  sind zwei weitere Messräume.

$$Z : \Omega \rightarrow \Omega_1$$

ist eine  $\mathcal{A}/\mathcal{A}_1$ -messbare Zufallsvariable und

$$X : \Omega \rightarrow \Omega_2$$

ist eine  $\mathcal{A}/\mathcal{A}_2$ -messbare Zufallsvariable.

Dann ist das Bildmaß  $P^{(Z, X)}$  von  $P$  unter  $(Z, X)$  ein Wahrscheinlichkeitsmaß auf

$$(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$$

Sei  $\mu_1$  ein  $\sigma$ -endliches Maß auf  $(\Omega_1, \mathcal{A}_1)$  und  $\mu_2$  ein  $\sigma$ -endliches Maß auf  $(\Omega_2, \mathcal{A}_2)$ , so dass

$$P^{(Z, X)} \ll \mu_1 \otimes \mu_2$$

Das heißt also nach dem Satz von Radon-Nikodym (Satz 7.4):

Es existiert eine Dichte

$$f_{(Z,X)} : \Omega_1 \times \Omega_2 \rightarrow [0; \infty), \quad (z, x) \mapsto f_{(Z,X)}(z, x)$$

von  $P^{(Z,X)}$  bzgl.  $\mu_1 \otimes \mu_2$ :

$$dP^{(Z,X)} = f_{(Z,X)} d(\mu_1 \otimes \mu_2)$$

Aus Abschnitt 10.4 folgt dann unmittelbar

$$P^{Z|X=x}(A_1) = \int_{A_1} f_{Z|X}(z|x) \mu_1(dz) \quad \forall A_1 \in \mathcal{A}_1$$

und daher

$$P^{Z|X=x}(dz) = f_{Z|X}(z|x) \mu_1(dz)$$

# 11 Einschub: Zufällige Maße auf $(\mathbb{R}^p, \mathfrak{B}^p)$

Betrachtet man einen Markov-Kern

$$\Omega \times \mathcal{A}' \longrightarrow \mathbb{R}$$

so kann dieser auch aufgefasst werden als eine Abbildung, die jedem  $\omega \in \Omega$  ein W-Maß auf  $(\Omega', \mathcal{A}')$  zuordnet. Ist diese Abbildung in einem geeigneten Sinn messbar, so hat man es mit einer Zufallsvariable zu tun, die Werte in einem „Raum von Maßen“ annimmt. Von einer messbaren Abbildung ist nur zu sprechen, wenn auf „der Menge der Maße“ eine geeignete  $\sigma$ -Algebra definiert ist. Diese Zusammenhänge gehen in die folgende Definition ein. Zunächst jedoch einige Begriffsbildungen:

1. Ein Maß  $\mu$  auf  $(\mathbb{R}^p, \mathfrak{B}^p)$  heißt lokal-endlich, wenn jeder Punkt aus  $x \in \mathbb{R}^p$  eine Umgebung  $U(x)$  hat, für welche gilt:  $\mu(U(x)) < \infty$ . Mit  $M$  sei die Menge aller lokal-endlichen Maße bezeichnet.
2. Ein  $A \in \mathfrak{B}^p$  heißt beschränkt, wenn  $A$  in einer  $\epsilon$ -Kugel liegt.
3. Für jedes beschränkte  $A \in \mathfrak{B}^p$  ist eine Abbildung wie folgt definiert:

$$\begin{aligned} \rho_A : M &\longrightarrow \mathbb{R}_+ \\ \mu &\longmapsto \rho_A(\mu) = \mu(A) \end{aligned}$$

4. Sei  $\mathcal{M} := \sigma(\rho_A : A \text{ beschränkt})$ . Dann ist  $(M, \mathcal{M})$  ein Messraum.

**Definition 11.1** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(M, \mathcal{M})$  der Messraum aller lokal-endlicher Maße auf  $(\mathbb{R}^p, \mathfrak{B}^p)$ ,

(i) Dann heißt jede messbare Abbildung

$$X : \Omega \longrightarrow M$$

zufälliges Maß.

(ii) Für jedes  $A \in \mathfrak{B}^p$  ist

$$X(A) : \Omega \longrightarrow \mathbb{R}_+$$

eine Zufallsvariable, genannt das „zufällige Maß der Menge  $A$ “.

(iii)

$$\begin{aligned} E_X : \mathfrak{B}^p &\longrightarrow \mathbb{R}_+ \\ A &\longmapsto \int X(A) dP \end{aligned}$$

ist ein Maß, das Intensitätsmaß von  $X$ . Ist  $E_X$  lokal-endlich, so heißt  $X$  integrierbar.

**Bemerkung 11.2**

- a) Ein Kern von  $(\Omega, \mathcal{A})$  nach  $(\mathbb{R}^p, \mathcal{B}^p)$  ist ein zufällige Maß, wenn auf  $(\mathbb{R}^p, \mathcal{B}^p)$  ein  $W$ -Maß definiert ist.
- b) die Anzahl zufälliger Ereignisse, die in einem Zeitintervall statt finden, kann als zufälliges Maß modelliert werden.
- c) Punktprozesse, wie beispielsweise der Poisson-Prozess, können als zufällige Maße modelliert werden.

## 12 Das Riemann-Stieltjes-Integral

In diesem Abschnitt wird ein weiterer Integralbegriff vorgestellt, der eng mit dem bekannten Riemann-Integral verbunden ist: das Riemann-Stieltjes-Integral. Hierbei ist das Riemann-Stieltjes-Integral eine Vorbereitung für das Itô-Integral, welches eine Bedeutung in der Theorie der stochastischen Prozesse hat. Zunächst jedoch, als Wiederholung, die „Konstruktion“ des Riemann-Integrals. Die Darstellung des Riemann- und des Riemann-Stieltjes-Integrals bezieht sich maßgeblich auf die Ausführungen von Walter (2002), Kapitel 6 S.190ff. Es wurde auch weitgehend die dort verwendete Notation übernommen. Der Vortrag der Vorlesung bezieht sich ebenfalls stark auf diese Ausführungen. Den interessierten Studierenden wird empfohlen, die angegebene Literatur zu konsultieren.

### 12.1 Treppenfunktionen

Das (bestimmte) Integral einer reellen Funktion soll auf einfache Flächenberechnung zurückgeführt werden. Deswegen betrachtet man zunächst reelle Funktionen, die in dieser Hinsicht besonders einfach sind, die sogenannten Treppenfunktionen auf einem Intervall.

**Definition 12.1** Sei  $[a; b]$  ein Intervall aus  $\mathbb{R}$ . Eine Funktion

$$T_n : [a; b] \longrightarrow \mathbb{R}$$

heißt Treppenfunktion genau dann wenn gilt: es gibt  $a = t_0 < t_1 < \dots < t_n = b$  und reelle Zahlen  $c_1, c_2, \dots, c_n$  mit  $T_n(x) = c_i$  für  $t_{i-1} \leq x < t_i$  und  $c_n = T_n(b)$ .

#### Beispiel 12.2

Der einfachste Fall einer Treppenfunktion ist eine auf einem Intervall  $[a; b]$  konstante Funktion.

Für Treppenfunktionen gestaltet sich die Berechnung der Fläche unter dieser Funktion einfach: Sie ist die Summe der einzelnen Rechtecksflächen  $c_i(x_i - x_{i-1})$ . Hierbei ist nur zu beachten, dass sich für negative  $c_i$  die Fläche „subtrahiert“.

Der oben angesprochene Umstand führt zum Begriff des Integrals einer Treppenfunktion.

**Definition 12.3** Sei  $T_n$  eine Treppenfunktion auf  $[a; b]$  mit  $T_n(x) = c_i$  für  $t_{i-1} \leq x < t_i$  und  $c_n = T_n(b)$  dann heißt der Wert

$$I(T_n) = \sum_{i=1}^n c_i(t_i - t_{i-1})$$

das bestimmte Integral der Treppenfunktion  $T_n$ .



### Beispiel 12.4

Sei

$$T_2 : [0; 1] \longrightarrow \mathbb{R}$$

mit  $t_0 = 0, t_1 = \frac{1}{2}, t_2 = 1$  und  $c_1 = \frac{1}{2}, c_2 = \frac{1}{4}$ , dann ist

$$I(T_2) = \frac{1}{2} \frac{1}{2} + \frac{1}{4} \frac{1}{2} = \frac{3}{8}$$

## 12.2 Das bestimmte Integral

Das Integral einer beliebigen, beschränkten Funktion  $f$  soll nun, falls möglich, mit Hilfe von Approximationen durch solche Treppenfunktionen bestimmt werden. Die Idee hierbei ist zunächst zu einer gegebenen Funktion zwei Treppenfunktionen anzugeben „zwischen“ denen die Funktion liegt. Dann „liegt“ auch, anschaulich gesprochen, die Fläche dieser Funktion „zwischen“ den Integralen der beiden Treppenfunktionen. Diese Intuition wird nun wie folgt präzisiert.

Sei

$$f : [a, b] \longrightarrow \mathbb{R}$$

eine reelle Funktion und  $T_n^1, T_n^2$  zwei Treppenfunktionen, für die gilt:

$$T_n^1(x) \leq f(x) \quad \text{und} \quad f(x) \leq T_n^2(x) \quad \forall x \in [a; b],$$

so heißt  $T_n^1$  *Untersumme* und  $T_n^2$  *Obersumme* von  $f$ . Man kann durch Verfeinerung der Intervalle, die den Definitionsbereich von  $T_n^1$  und  $T_n^2$  jeweils einteilen, d.h. durch „Vergrößern“ von  $n$ , Treppenfunktionen finden, die „näher“ an  $f$  sind. Die Übertragung auf die Integrale der Treppenfunktionen führt im „Grenzübergang“ zur Definition des bestimmten Integrals.

**Definition 12.5** *Eine beschränkte Funktion*

$$f : [a; b] \longrightarrow \mathbb{R}$$

heißt auf dem Intervall  $[a; b]$  *integrierbar genau dann wenn gilt:*  
*es gibt eine reelle Zahl  $I$  so, dass für alle  $\epsilon > 0$  eine Unter- und eine Obersumme  $T^1$  und  $T^2$  bzgl.  $f$  (die von  $\epsilon$  abhängen) existiert mit*

$$I - \epsilon \leq I(T^1) \leq I \leq I(T^2) \leq I + \epsilon$$

*$I$  heißt dann das bestimmte Integral der Funktion  $f$  auf  $[a; b]$  und  $f$  heißt auf  $[a; b]$  integrierbar. Man schreibt für das bestimmte Integral auch:*

$$I(f) = \int_a^b f(x) dx$$

## 12.3 Eine Alternative Vorgehensweise

Hierzu sind zunächst einige Begriffsbildungen nötig, die auch in der o.g. Literatur zu finden sind:

**Definition 12.6** *Unter einer gerichteten Menge,  $(A, \prec)$  wird eine Menge verstanden, auf welcher eine Relation mit den folgenden drei Eigenschaften definiert ist.*

- (i)  $a \prec a$  für alle  $a \in A$  (Reflexivität)
- (ii)  $a \prec b, b \prec c \Rightarrow a \prec c$  (Transitivität)
- (iii) Für  $a, b \in A$  gibt es ein  $c \in A$  mit  $a \prec c$  und  $b \prec c$

**Bemerkung 12.7** *Die natürlichen Zahlen in ihrer natürlichen  $\leq$  Relation bilden eine gerichtete Menge:  $(\mathbb{N}, \leq)$ .*

Mit Hilfe dieses Begriffs kann man einen *verallgemeinerten Limes* oder einen *Netzlimes* definieren. Hierzu ist zunächst ein „Netz“ zu definieren; dazu sei  $(X, d)$  ein metrischer Raum (Metrik  $d$  wird der einfacheren Notation halber oft unterdrückt).

**Definition 12.8** *Sei  $(A, \prec)$  eine gerichtete Menge und  $(X, d)$  ein metrischer Raum (Metrik:  $d$ ), dann heißt eine Funktion*

$$f : A \longrightarrow X$$

*ein Netz oder eine verallgemeinerte Folge. Anstelle von  $f(a)$  schreibt man, wie bei „gewöhnlichen“ Folgen,  $f_a$ .*

Man sagt auch:

„ $(f_a)_{a \in A}$  bzw.  $(f_a)$  ist ein Netz“ oder „ist ein Netz über  $A$  in  $(X, d)$ “.

Hiermit kann nun der Netzlimes definiert werden:

**Definition 12.9** *Sei  $(A, \prec)$  eine gerichtete Menge,  $(X, d)$  ein metrischer Raum und  $(f_a)_{a \in A}$  ein Netz in  $(X, d)$ , dann ist  $x_0 \in X$  der Netzlimes von  $(f_a)$  genau dann wenn*

*für jedes  $\epsilon > 0$  gibt es ein  $a_\epsilon \in A$ , so dass für alle  $a \in A$  mit  $a \succ a_\epsilon$  gilt:*

$$d(f_a, x_0) < \epsilon$$

*In Zeichen:*

$$\lim_a f_a = x_0$$

*oder in Zeichen:*

$$f_a \rightarrow x_0$$

Mit dieser Begriffsbildung kann man das Riemann-Integral ebenfalls definieren. Hierzu ist zunächst der Begriff einer Partition auf einem abgeschlossenen Intervall  $[a; b]$  zu klären, sowie die Relation „feiner als“ unter Partitionen.

**Definition 12.10**

- (i) Sei  $[a; b]$  ein abgeschlossenes Intervall in  $\mathbb{R}$  und  $a = t_0 < t_1 < \dots < t_n = b$  eine endliche Folge sog. Stützstellen, dann heißt  $Z := (t_0, \dots, t_n)$  eine Zerlegung von  $[a; b]$ .
- (ii) Seien  $Z = (t_0, \dots, t_n)$  und  $Z' = (t'_0, \dots, t'_m)$  zwei Zerlegungen von  $[a; b]$ , dann heißt  $Z'$  feiner als  $Z$ , in Zeichen  $Z \prec Z'$  genau dann wenn  $\{t_0, \dots, t_n\} \subset \{t'_0, \dots, t'_m\}$
- (iii) Sei  $Z = (t_0, \dots, t_n)$  eine Zerlegung von  $[a; b]$ , und  $\tau := (\tau_1, \dots, \tau_n)$  ein Tupel passender Zwischenpunkte, d.h.  $\tau_i \in [t_{i-1}; t_i], i = 1, \dots, n$ , dann wird durch  $(Z, \tau) \prec (Z', \tau') \implies Z \prec Z'$  eine gerichtete Menge definiert, wenn  $\tau'$  auch ein Tupel passender Zwischenpunkte ist.

**Bemerkung 12.11** Die Menge aller Zerlegungen ist bzgl.  $\prec$  eine gerichtete Menge.

**Definition 12.12 (Riemannsche Zwischensumme)**

Sei  $Z = (t_0, \dots, t_n)$  eine Zerlegung von  $[a; b]$ ,  $f$  eine beschränkte Funktion auf  $[a; b]$  und  $(\tau_1, \dots, \tau_n)$  ein Tupel passender Zwischenpunkte, d.h.  $\tau_i \in [t_{i-1}, t_i], i = 1, \dots, n$ , dann heißt

$$\sigma(Z, \tau, f) := \sum_{i=1}^n f(\tau_i)(t_i - t_{i-1})$$

Riemannsche Zwischensumme oder kurz: Zwischensumme.

**Definition 12.13 (Riemann-Integral)**

Der Netz-Limes  $\lim_Z \sigma(Z, \tau, f)$  über die Zwischensumme aus Definition 12.12 heißt, falls er existiert, das Riemann-Integral von  $f$  über  $[a; b]$ :

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

Diese Definition ist äquivalent mit der Definition über Ober- und Untersummen. Dieser Zusammenhang findet sich ausführlich dargestellt in Walter (2002), S. 146ff. Vergleiche insbesondere die Darstellung von Netzen in Walter (2002), S. 142ff.

## 12.4 Das Riemann-Stieltjes-Integral

Das Riemann-Stieltjes-Integral ist eine Verallgemeinerung des Riemann-Integrals, die wie folgt einzusehen ist: Sei wieder  $f$  eine auf  $[a; b]$  beschränkte Funktion,

$id$  die Identitätsfunktion auf diesem Intervall:  $id(x) = x$  für  $x \in [a; b]$ , die Riemannsche Zwischensumme schreibt sich dann wie folgt:

$$\sigma(Z, \tau, f, d(id)) := \sum_{i=1}^n f(\tau_i)(id(t_i) - id(t_{i-1}))$$

Die Verallgemeinerung besteht nun darin, anstelle der Identität  $id$  auch andere Funktionen  $g : [a; b] \rightarrow \mathbb{R}$  zuzulassen. Dies ergibt folgende Definitionen:

**Definition 12.14 (Riemann-Stieltjes-Summe)**

Sei  $Z = (t_0, \dots, t_n)$  eine Zerlegung von  $[a; b]$ ,  $f$  und  $g$  Funktionen auf  $[a; b]$  und  $(\tau_1, \dots, \tau_n)$  ein Tupel passender Zwischenpunkte, d.h.  $\tau_i \in [t_{i-1}; t_i]$ ,  $i = 1, \dots, n$ , dann heißt

$$\sigma(Z, \tau, f, d(g)) = \sum_{i=1}^n f(\tau_i)(g(t_i) - g(t_{i-1}))$$

Riemann-Stieltjes-Summe oder kurz: Zwischensumme.

**Definition 12.15 (Riemann-Stieltjes-Integral)**

Der Netz-Limes  $\lim_Z \sigma(Z, \tau, f, d(g))$  über die Riemann-Stieltjes-Summe aus Definition 12.14 heißt, falls er existiert, das Riemann-Stieltjes-Integral von  $f$  bezüglich  $g$  über  $[a; b]$ :

$$\int_a^b f dg = \int_a^b f(x) dg(x)$$

**Satz 12.16 (Existenz des Integrals)**

Der Netz-Limes  $\lim_Z \sigma(Z, \tau, f, d(g))$  existiert und hat den Wert  $I$ , wenn gilt: Für alle  $\epsilon > 0$  gibt es eine Zerlegung  $Z_\epsilon$  von  $[a; b]$  so, dass für alle  $(Z, \tau)$  mit  $Z \succ Z_\epsilon$  gilt

$$|I - \sigma(Z, \tau, f, dg)| < \epsilon$$

Man nennt  $f$  auch den *Integranden* und  $g$  den *Integrator*.

**Motivation:** Sei  $[0; T]$  das Zeitintervall in welchem ein Agent gewisse Anzahlen einer Aktie hält:  $z(t)$  ist die Anzahl der Aktie zum Zeitpunkt  $t$ , die gehalten wird;  $a(t)$  ist der Wert der Aktie zum Zeitpunkt  $t$ ,  $t \in [0, T]$ , dann ist

$$\int_0^T z(t) dg(t)$$

der Gesamtwert der gehaltenen Aktien über den Zeitraum  $[0; T]$  hinweg. Man verdeutliche sich dies, in dem man die Situation diskretisiert: in  $[0; T]$  wird der Aktienwert nur zu den Zeitpunkten  $0 = t_0 < t_1 < \dots < t_n = T$  festgesetzt:  $a(t_0), \dots, a(t_n)$  und die Anzahl der gehaltenen Aktien ist in den Teilintervallen  $[t_{i-1}, t_i]$ ,  $i = 1, \dots, n$  konstant  $z_i$ .

Das Itô-Integral kommt dann ins Spiel, wenn zusätzlich davon ausgegangen wird, dass die Aktienkurse und ggf. die Anzahlen „zufällig“ sind. Das Itô-Integral wird im Kapitel über Martingale besprochen werden, da dort der systematische und der pragmatische Ort dieser Begriffsbildung ist.

**Satz 12.17 (Rechenregeln)**

Für die Funktionen  $f, f_1, f_2$  (Integranden),  $g, g_1$  und  $g_2$  (Integratoren) seien die Integrale über einem Intervall  $[a; b]$

$$\int_a^b f dg$$

erklärt, sowie  $\lambda_1, \lambda_2$  reelle Zahlen. Dann gilt:

a) Linearität des Integrals

$$\int_a^b (\lambda_1 f_1 + \lambda_2 f_2) dg = \lambda_1 \int_a^b f_1 dg + \lambda_2 \int_a^b f_2 dg$$

und

$$\int_a^b f d(\lambda_1 g_1 + \lambda_2 g_2) = \int_a^b f d\lambda_1 g_1 + \int_a^b f d\lambda_2 g_2 = \lambda_1 \int_a^b f dg_1 + \lambda_2 \int_a^b f dg_2$$

b) Sei  $a < c < b$  dann gilt:

$$\int_a^b f dg = \int_a^c f dg + \int_c^b f dg$$

Eine Abschätzung des Riemann-Stieltjes-Integral geht über die sogenannte Totalvariation, die im Folgenden kurz eingeführt wird (Vergl. hierzu Walter (2002), S. 175f).

**Definition 12.18 ((Total)Variation)**

Sei  $f$  eine reelle Funktion auf  $[a; b]$  und  $Z = (t_0, \dots, t_n)$  eine Zerlegung dieses Intervalls, dann ist die Variation

$$\text{Var}(z, f) := \sum_{i=1}^n |f(t_i) - f(t_{i-1})| \quad .$$

Das Supremum über alle Zerlegungen  $Z$ :

$$V_a^b(f) := \sup_Z \text{Var}(z, f)$$

heißt Totalvariation von  $f$ .

Ist  $V_a^b(f) < \infty$ , so heißt  $f$  von beschränkter Variation oder von beschränkter Schwankung;  $V_a^a(f) := 0$ .

Eigenschaften und Zusammenhänge der Totalvariation sind im folgenden Satz festgehalten.

**Satz 12.19**

a) Jede Funktion  $f$  von beschränkter Variation auf einem Intervall  $[a; b]$  ist auf  $[a; b]$  beschränkt und es gilt  $|f(a) - f(b)| \leq V_a^b(f)$ .

b) Mit  $f, g$  von beschränkter Variation auf  $[a; b]$  und reelle Zahlen  $\lambda, \mu$  ist auch  $\lambda f + \mu g$  von beschränkter Variation. Es gilt folgende Abschätzung:

$$V_a^b(\lambda f + \mu g) \leq |\lambda| V_a^b(f) + |\mu| V_a^b(g)$$

und

$$V_a^b(fg) \leq \|f\|_\infty V_a^b(g) + \|g\|_\infty V_a^b(f).$$

Hierbei bezeichnet  $\|\cdot\|_\infty$  die Supremumsnorm.

c) Für  $a < c < b$  gilt:

$$V_a^b(f) = V_a^c(f) + V_c^b(f)$$

d) Ist  $f$  monoton auf  $[a; b]$ , so ist  $V_a^b(f) = |f(a) - f(b)|$ .

e) Für eine auf dem Intervall  $[a; b]$  stetig differenzierbare Funktion mit Ableitung  $f'$  gilt:

$$V_a^b(f) = \int_a^b |f'(t)| dt$$

Mit dieser Begriffsbildung gilt (siehe Walter (2002), S. 192):

**Satz 12.20** Sei  $f$  eine auf  $[a; b]$  stetige Funktion und  $g$  Funktion von beschränkter Variation auf  $[a; b]$ , so existiert  $\int_a^b f dg$  und es gilt die folgende Abschätzung

$$\left| \int_a^b f dg \right| \leq \|f\|_\infty V_a^b(g)$$

Zur Berechnung des Riemann-Stieltjes-Integrals kann man ein Analogon zur Partiellen Integration verwenden.

**Satz 12.21** Mit

$$\int_a^b f dg$$

existiert auch

$$\int_a^b g df$$

und es gilt

$$\int_a^b f dg + \int_a^b g df = f(b)g(b) - f(a)g(a) =: fg|_a^b$$

Es gelten die folgende Zusammenhänge mit dem (einfachen) Riemann-Integral.

**Satz 12.22**

a) Sei  $f$  eine Riemann-integrierbare Funktion auf  $[a; b]$  und  $g$  eine auf  $[a; b]$  stetig differenzierbare Funktion,  $g'$  deren Ableitung; das Riemann-Stieltjes-Integral

$$\int_a^b f dg$$

existiere. Dann gilt

$$\int_a^b f dg = \int_a^b f g' dx$$

b) Jedes Riemann-Integral

$$\int_a^b fh dt$$

lässt sich mit

$$g(t) := \int_a^t h du$$

als Riemann-Stieltjes-Integral schreiben, falls  $f$  Riemann-integrierbar und  $h$  stetig ist:

$$\int_a^b f dg$$

# 13 Gesetze der Großen Zahlen und Grenzwertsätze

Aus dem Bachelorstudium sind verschiedene Gesetze der Großen Zahlen und Grenzwertsätze bereits bekannt. Sie seien nachstehend zur Wiederholung aufgeführt. Wesentlich für Grenzwertsätze und die Gesetze der Großen Zahlen sind die verschiedenen, dort verwendeten Konvergenzbegriffe.

## 13.1 Konvergenzbegriffe

Im Folgenden sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(X_n)_{n \in \mathbb{N}}$ , abgekürzt  $(X_n)$ , eine Folge reeller Zufallsvariablen auf  $(\Omega, \mathcal{A}, P)$  sowie  $X$  eine weitere Zufallsvariable auf  $(\Omega, \mathcal{A}, P)$ .  $P_{X_n} = X_n(P)$  bzw.  $P_X = X(P)$  seien die Verteilungen der Zufallsvariablen.

### Definition 13.1 (Fast sichere Konvergenz, starke Konvergenz)

Eine Folge  $(X_n)_{n \in \mathbb{N}}$  konvergiert fast sicher (oder stark) gegen  $X$  genau dann, wenn für alle  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(\sup_{m \geq n} |X_m - X| > \epsilon) = 0$$

gilt.

Dies ist äquivalent zur folgenden Bedingung:

Die Folge  $(X_n)_{n \in \mathbb{N}}$  konvergiert *fast sicher* gegen  $X$  genau dann, wenn für alle  $\epsilon > 0$

$$P(\limsup_{n \rightarrow \infty} |X_n - X| > \epsilon) = 0$$

gilt.

Für *fast sichere Konvergenz* sagt man auch *starke Konvergenz*.

### Definition 13.2 ( $\mathcal{L}^p$ -Konvergenz, Konvergenz im $p$ -ten Mittel)

Eine Folge  $(X_n)_{n \in \mathbb{N}}$  ist  $\mathcal{L}^p$ -konvergent gegen  $X$  oder konvergiert im  $p$ -ten Mittel gegen  $X$  ( $1 \leq p < +\infty$ ) genau dann, wenn

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

gilt.

**Bemerkung 13.3** Aus der Konvergenz im  $p$ -ten Mittel folgt die Konvergenz im 1-ten Mittel. Hier spricht man von „Konvergenz im Mittel“.



**Definition 13.4 (Stochastische Konvergenz, Konvergenz nach Wahrscheinlichkeit)**

Eine Folge  $(X_n)_{n \in \mathbb{N}}$  konvergiert stochastisch oder nach Wahrscheinlichkeit gegen  $X$  genau dann, wenn für alle  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0$$

gilt.

Hierzu ist äquivalent:

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

**Notation 13.5** Für eine Folge von Zufallsvariablen  $(X_n)_{n \in \mathbb{N}}$  die stochastisch gegen  $X$  konvergiert, schreibt man auch

$$\text{plim}_{n \rightarrow \infty} X_n = X .$$

**Definition 13.6 (Schwache Konvergenz, Konvergenz in Verteilung)**

Eine Folge  $(P_{X_n})_{n \in \mathbb{N}}$  der Verteilungen der Folge  $(X_n)_{n \in \mathbb{N}}$  konvergiert schwach gegen die Verteilung  $P_X$  von  $X$  genau dann, wenn für alle stetigen und beschränkten Funktionen  $f$

$$\lim_{n \rightarrow \infty} \mathbb{E}_P(f \circ X_n) = \mathbb{E}_P(f \circ X)$$

gilt.

Es gilt der folgende Satz:

**Satz 13.7** Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge reeller Zufallsvariablen und  $X$  eine reelle Zufallsvariable auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  und konvergiert  $(X_n)_{n \in \mathbb{N}}$  stochastisch gegen  $X$ , dann konvergiert die Folge der Verteilungen  $(P_{X_n})_{n \in \mathbb{N}}$  schwach gegen  $P_X$ . Ist  $X$   $P$ -fast sicher konstant, so gilt hiervon auch die Umkehrung.

**Beweis:**  $\rightsquigarrow$  Bauer (2002), S. 36. □

**Bemerkung 13.8** Es gelten folgende Implikationen für  $p > q \geq 1$ :

$$\begin{array}{ccc} \mathcal{L}^p\text{-Konvergenz} & \implies & \mathcal{L}^q\text{-Konvergenz} \\ & & \downarrow \\ \text{fast sichere Konvergenz} & \implies & \text{stochastische Konvergenz} \\ & & \downarrow \\ & & \text{schwache Konvergenz} \end{array}$$

## 13.2 Gesetze der Großen Zahlen

Zunächst einige Bezeichnungsweisen,  $(X_n)_{n \in \mathbb{N}}$  bezeichne wieder eine Folge reeller Zufallsvariablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ . Mit  $\mathbb{E}(X_n)$  bzw.  $\mathbb{V}(X_n)$  werden deren Erwartungswerte bzw. Varianzen (falls existent) bezeichnet. Ferner sei definiert:

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

Dies ergibt dann  $(S_n)_{n \in \mathbb{N}}$  als die Folge der Summen der jeweils ersten  $n$  Zufallsvariablen.

### Definition 13.9 (Schwachtes Gesetz der Großen Zahlen)

Eine Folge  $(X_n)_{n \in \mathbb{N}}$  genügt dem Schwachen Gesetz der Großen Zahlen, falls die Folge

$$\frac{1}{n} S_n$$

stochastisch (bzw. nach Wahrscheinlichkeit) gegen 0 konvergiert.

In Zeichen:

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} S_n = 0$$

### Definition 13.10 (Starkes Gesetz der Großen Zahlen)

Eine Folge  $(X_n)_{n \in \mathbb{N}}$  genügt dem Starken Gesetz der Großen Zahlen, falls die Folge

$$\frac{1}{n} S_n$$

fast sicher gegen 0 konvergiert.

In Zeichen:

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n = 0 \quad P\text{-f.s.}$$

Hinsichtlich der schwachen Konvergenz von Folgen gilt der folgende Satz:

**Satz 13.11** Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge paarweise unkorrelierter Zufallsvariablen mit  $\mathbb{E}(X_n) = 0$  und  $\mathbb{V}(X_n) < \infty$  für alle  $n \in \mathbb{N}$ . Ist zudem  $(a_n)$  eine Folge positiver reeller Zahlen, für welche gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{a_n^2} \sum_{i=1}^n \mathbb{V}(X_i) = 0$$

so folgt

$$\text{plim}_{n \rightarrow \infty} \frac{1}{a_n} \sum_{i=1}^n X_i = 0,$$

d.h. die Folge  $\left( \frac{1}{a_n} \sum_{i=1}^n X_i \right)_{n \in \mathbb{N}}$  konvergiert stochastisch gegen 0.

**Beweis:**  $\rightsquigarrow$  Übung

□

**Bemerkung 13.12** Die Bedingung ist mit  $a_n = n$  sicher erfüllt, wenn die Zufallsvariablen identische verteilt sind.

**Beweis:**  $\rightsquigarrow$  Übung □

An folgenden Beispiel wird verdeutlicht, dass die Bedingung hinreichend, aber nicht notwendig ist.

**Beispiel 13.13** Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge von unabhängigen und identisch Cauchy-verteilten Zufallsvariablen mit Skalenparameter 1. Dann konvergiert die Folge

$$S_n = \frac{1}{a_n} \sum_{i=1}^n X_i$$

stochastisch gegen 0, falls gilt

$$\lim_{n \rightarrow \infty} \frac{a_n}{n} = 0$$

Die Bedingung kann aber gar nicht erfüllt sein, da für Cauchy-Verteilungen keine Erwartungswerte existieren.

### 13.3 Die Kolmogorov'schen Sätze

Diese Sätze geben Antwort auf die Frage, unter welchen Bedingungen eine Folge von Zufallsvariablen dem *Starken Gesetz der Großen Zahlen* genügt.

**Satz 13.14** Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge unabhängiger und integrierbarer reeller Zufallsvariablen mit

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{V}(X_n) < \infty,$$

so genügt die Folge  $(X_n)_{n \in \mathbb{N}}$  dem *Starken Gesetz der Großen Zahlen*.

**Satz 13.15** Jede Folge  $(X_n)_{n \in \mathbb{N}}$  unabhängiger quadrat-integrierbarer reeller und identisch verteilter Zufallsvariablen genügt dem *Starken Gesetz der Großen Zahlen*.

### 13.4 Der Zentrale Grenzwertsatz

Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge unabhängiger reeller quadrat-integrierbarer Zufallsvariablen mit  $\mathbb{V}(X_n) > 0$ . Die Zufallsvariable

$$S_n := \frac{\sum_{i=1}^n (X_i - E(X_i))}{\sigma(X_1 + \dots + X_n)}$$

heißt standardisierte Summe und  $\sigma(X_1 + \dots + X_n)$  sei die Standardabweichung der Summe  $X_1 + \dots + X_n$ .

**Definition 13.16 (Zentraler Grenzwertsatz)**

Für die Folge  $(X_n)_{n \in \mathbb{N}}$  gilt der Zentrale Grenzwertsatz falls die Folge der Verteilungen  $(S_n(P))_{n \in \mathbb{N}}$  der standardisierten Summen schwach gegen die  $\mathcal{N}(0, 1)$ -Verteilung konvergiert.

Die Gültigkeit des Zentralen Grenzwertsatzes ist an Bedingungen geknüpft. Wichtige Bedingungen sind: die Lindeberg-Bedingung und die Fellersche Bedingung.

**Definition 13.17 (Lindeberg-Bedingung, Fellersche Bedingung)**

Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge unabhängiger reeller quadrat-integrierbarer Zufallsvariablen mit  $\mathbb{V}(X_n) > 0$ . Sei ferner abkürzend

$$\begin{aligned}\sigma_n &:= \sigma(X_n) = (\mathbb{V}(X_n))^{1/2} \\ s_n &:= \sigma(X_1 + \dots + X_n) = (\mathbb{V}(X_1) + \dots + \mathbb{V}(X_n))^{1/2} \\ \eta_n &:= \mathbb{E}(X_n)\end{aligned}$$

und

$$L_n(\epsilon) := \frac{1}{s_n^2} \sum_{i=1}^n \int_{|x-\eta_i| \geq \epsilon s_n} (x - \eta_i)^2 [X_i(P)](dx).$$

Die Folge  $(X_n)_{n \in \mathbb{N}}$  genügt der Lindeberg-Bedingung genau dann, wenn

$$\lim_{n \rightarrow \infty} L_n(\epsilon) = 0 \quad \forall \epsilon > 0.$$

Die Folge  $(X_n)_{n \in \mathbb{N}}$  genügt der Fellerschen Bedingung genau dann, wenn

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{\sigma_i}{s_n} = 0.$$

**Definition 13.18 (Asymptotische Vernachlässigbarkeit)**

Eine Familie  $(X_{ni})_{i=1, \dots, k_n, n=1, 2, \dots}$  heißt asymptotisch vernachlässigbar, wenn gilt:

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq k_n} P(|X_{ni}| \geq \epsilon) = 0 \quad \forall \epsilon > 0.$$

Es gilt der folgende Satz von Lindeberg-Feller.

**Satz 13.19** Für jede Folge unabhängiger quadratisch integrierbarer Zufallsvariablen  $(X_n)_{n \in \mathbb{N}}$  mit Varianzen  $\mathbb{V}(X_n) > 0$  sind die folgenden Aussagen gleichwertig:

- Es gilt der Zentrale Grenzwertsatz und die Folge genügt der Fellerschen Bedingung.
- Es gilt der Zentrale Grenzwertsatz und die Folge  $(X_{ni})$  ist asymptotisch vernachlässigbar.
- Die Folge genügt der Lindeberg-Bedingung.

**Bemerkung 13.20** Der Satz von de Moivre-Laplace ist ein Spezialfall hiervon.

# Literaturverzeichnis

Heinz Bauer. *Maß- und Integrationstheorie*. de Gruyter Lehrbuch. de Gruyter, Berlin, 2. Auflage, 1992. ISBN 3-11-013625-2.

Heinz Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter Lehrbuch. de Gruyter, Berlin, 5. durchges. u. verb. Auflage, 2002. ISBN 3-11-017236-4.

Jørgen Hoffmann-Jørgensen. *Probability with a view toward statistics*, Band 1. Chapman & Hall, New York, 1994a. ISBN 0-41-205221-0.

Jørgen Hoffmann-Jørgensen. *Probability with a view toward statistics*, Band 2. Chapman & Hall, New York, 1994b. ISBN 0-41-205231-8.

Andrej N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Julius Springer, 1933.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1997. doi: 10.1017/CBO9780511802256.

Wolfgang Walter. *Analysis 2*. Springer Lehrbuch. Springer-Verlag, Berlin Heidelberg, 5. Auflage, 2002. ISBN 3-540-42953-0.

Die Kapitel 1 bis 10 orientieren sich hauptsächlich an den Kapiteln I bis III in Bauer (1992).