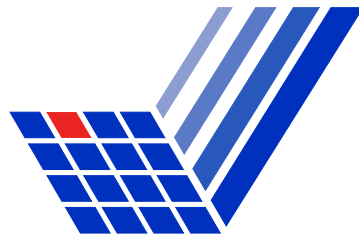


STOCHASTIK I

H.-P. SCHEFFLER

Sommersemester 1998



gesetzt von Artur A. Marczok

Diese Mitschrift wurde mit freier Software ($\text{\textit{AMS-L}\TeX}2_{\varepsilon}$, METAPOST) gesetzt. In der hier vorliegenden korrigierten Version kommt erstmals das von YOUNG RYU im CTAN zur Verfügung gestellte px -Zeichensatzpaket zum Einsatz. Es existiert auch eine PDF-Version dieses Skriptes (natürlich mit Hyperlinks). Korrekturvorschläge und Anfragen bzgl. der PDF-Version an marczok@gmx.de.

Mein besonderer Dank gilt PATRICIA RUDAWSKI und MATTHIAS DICKEL für die zahlreichen Korrekturen.

Recklinghausen, 4. August 2001

Arthur Adam Marczok

INHALTSVERZEICHNIS

0 Einführung	5
1 Die σ-Algebra und W-Maße	7
Stichprobenraum	7
σ -Algebra	8
Erzeugte σ -Algebra	9
BORELSche σ -Algebra	11
Wahrscheinlichkeitsmaß, Wahrscheinlichkeitsraum	12
Zähldichte	15
2 Der LAPLACESche W-Raum und kombinatorische W-Rechnung	19
Permutationen, Kombinationen mit/ohne Wiederholung	19
Siebformel	22
Ein-/Ausschließungsprinzip	22
3 Wahrscheinlichkeitsmaße mit Dichten auf $\mathbb{R}^d$	25
BORELSche σ -Algebra auf $\mathbb{R}^d$, BOREL-Menge	25
Maß	26
LEBESGUE(-BOREL)sches Maß	26
Meßbare Funktionen	27
Elementarfunktionen	27
Integral von Elementarfunktionen	28
Integral meßbarer Funktionen	28
Satz von BEPPO-LEVI, monotone Konvergenz	28
4 Zufallsvariablen und Unabhängigkeit	37
5 Erwartungswert und Varianz von Zufallsvariablen	47
6 Bedingte Wahrscheinlichkeiten	59
7 MARKOFF-Ketten	63

KAPITEL 0

Einführung

Aufgabe der Stochastik ist es, *Modelle* für zufällige Phänomene zu entwickeln, und damit Vorhersagen über deren Ausgang zu machen. Diese Phänomene sind außermathematisch, und ebenso ist die Frage, ob unsere Modelle diese Phänomene adäquat beschreiben, nicht mit mathematischen Methoden zu entscheiden, sondern nur durch Experimente.

Das in dieser Vorlesung behandelte mathematische Modell verwendet dabei im wesentlichen die *Mengenlehre* und (darauf aufbauend) die *Maßtheorie*. Im Gegensatz zur Analysis und linearen Algebra sind also die mathematischen Objekte der Stochastik Mengen sowie Maße, d.h. Abbildungen von Mengensystemen nach $[0, 1]$. Dies macht eine Schwierigkeit der Stochastik gegenüber der Analysis deutlich. Die andere Schwierigkeit ist das *Modellbildungsproblem*: Zu einem gegebenen „Phänomen“ muß ein mathematisches Modell (Mengensystem, W(ahrscheinlichkeits)-Maß) konstruiert werden, das dieses „Phänomen“ beschreibt.

0.1 Beispiel.

Phänomen	beobachtete Größe (Meßgröße)	mathem. Modell
(a) Werfen einer/s – Münze – Würfels	Kopf/Wappen Augenzahl	$\Omega = \{K, W\} \cong \{0, 1\}$ $\Omega = \{1, \dots, 6\}$
(b) – Wetter – Börsenschwankungen	Meßgrößen (Temp., Druck) Kurs	$\Omega = \mathbb{R}^d \quad (d \geq 1)$ $\Omega = \mathbb{R}$
(c) Ausfallraten von Produkten	Anzahl der fehlerhaften Produkte	$\Omega = \mathbb{Z}_+$

Man nennt den Raum Ω den *Stichproben-Raum*. Er ist die Grundlage des mathematischen Modells.

0.2 Beispiel. Werfen mit zwei Würfeln (rot | weiß). Meßgrößen: (a, b) mit $1 \leq a, b \leq 6$.

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

Interessiert man sich nur für die Augensumme $s = a + b$, so ist $\Omega = \{2, \dots, 12\}$ ein möglicher Stichprobenraum.

Im allgemeinen interessiert man sich nicht nur für die „Wahrscheinlichkeit“ von Elementen $i \in \Omega$, sondern für die Wahrscheinlichkeit von Teilmengen $E \subset \Omega$, den *Ereignissen*. Da mit $\omega \in \Omega$, $\{\omega\} \subset \Omega$ gilt, reicht es die „Wahrscheinlichkeit“ für Ereignisse $E \subset \Omega$ zu definieren.

$$\text{Ereignis } \Omega \supset E \mapsto P(E) \in [0, 1]$$

Wahrscheinlichkeit $\hat{=}$ Bewertung der Unsicherheit.

0.3 Beispiel. Werfen mit einem idealen Würfel: $\Omega = \{1, \dots, 6\}$. Es ist plausibel, daß $P(\{i\}) = \frac{1}{6}$ für $i \in \Omega$ gilt. Ist nun $E = \text{„gerade Zahl“} = \{2, 4, 6\}$, so ist plausibel, daß $P(E) = \frac{1}{2}$ gelten sollte.

ENDE VON KAPITEL 0

KAPITEL 1

Die σ -Algebra der Ereignisse und Wahrscheinlichkeitsmaße

In diesem Kapitel werden die zwei für die Stochastik grundlegenden Begriffe „Ereignis“ und „Wahrscheinlichkeit“ mathematisch definiert.

1.1 Definition. Als *Stichprobenraum* (oder *Merkmalmenge*) bezeichnen wir eine beliebige nicht-leere Menge $\Omega \neq \emptyset$.

Ω sollte möglichst adäquat die Ereignisse des Experiments beschreiben. Die Wahl von Ω ist nicht eindeutig; es ist aber zu hoffen, daß die Ergebnisse des Modells nicht von der Wahl von Ω abhängen. Falls Ω endlich ist, stellt die Kombinatorik eine wichtige Methode dar, die seit dem 17. Jahrhundert in der W-Rechnung zentral war. Ist Ω unendlich, so werden maßtheoretische Methoden wichtig, die KOLMOGOROFF (etwa 1930) in die W-Theorie eingeführt hat.

1.2 Beispiel.

- (1) siehe **Beispiel 0.1**
- (2) Skatspiel: Die Karten werden von 1 bis 32 durchnummeriert, etwa $1 \triangleq$ Kreuz Bube, $\dots$, $32 \triangleq$ Karo 7. Dann kann man

$$\begin{aligned}\Omega &= \{(A, B, C) \mid A, B, C \subset \{1, \dots, 32\}; A \cap B = A \cap C = B \cap C = \emptyset; \\ &\quad \#A = \#B = \#C = 10\}\end{aligned}$$

setzen und so den Stichprobenraum aller möglichen Skatspiele definieren. Dabei bezeichne für eine endliche Menge A , $\#A$ die Anzahl der Elemente von A .

1.3 Beispiel. Zusammengesetztes (wiederholtes) Experiment.

- (1) Es wird zuerst eine Münze geworfen und dann gewürfelt:

$$\begin{aligned}\Omega &= \{(W, 1), (W, 2), \dots, (W, 6), (Z, 1), \dots, (Z, 6)\} \\ &= \{W, Z\} \times \{1, \dots, 6\} \quad (\text{kartesisches Produkt})\end{aligned}$$

- (2) Das n -malige Werfen einer Münze wird durch

$$\Omega = \{W, Z\}^n = \{(\omega_1, \dots, \omega_n) \mid \omega_j \in \{W, Z\}\}$$

modelliert. Das n -Tupel $(\omega_1, \dots, \omega_n)$ symbolisiert eine Folge von Experimenten, bei der der j -te Versuch den Wert ω_j ergeben hat.

- (3) Wird das Experiment unendlich oft wiederholt, so ist

$$\Omega = \{W, Z\}^{\mathbb{N}} = \{(\omega_0, \omega_1, \dots) \mid \omega_j \in \{W, Z\}\}$$

ein geeigneter Stichprobenraum.

1.4 Konstruktion. Wir haben den Stichprobenraum Ω als Modell zur Beschreibung der möglichen Ausgänge eines Experiments eingeführt. Wir wollen aber auch komplizierte Ereignisse wie „In 100 Versuchen wurde zwischen 40 und 60 mal Zahl geworfen“ modellieren. Dies geschieht durch gewisse Teilmengen von Ω und logischen Verknüpfungen, die durch Mengenoperationen erzeugt werden:

sicheres Ereignis	$\hat{=}$	Ω
unmögliches Ereignis	$\hat{=}$	$\emptyset$
Negation eines Ereignisses	$\hat{=}$	$A^c = \Omega \setminus A$
Mindestens eines der beiden	$\hat{=}$	$A \cup B$
alle beide	$\hat{=}$	$A \cap B$
das erste, aber nicht das zweite	$\hat{=}$	$A \setminus B$
mindestens eines aus einer Folge	$\hat{=}$	$\bigcup_{n=1}^{\infty} A_n$
alle aus einer Folge	$\hat{=}$	$\bigcap_{n=1}^{\infty} A_n$

1.5 Beispiel.

- (1) Der Würfel zeigt eine ungerade Zahl $\hat{=}$

$$A_1 = \{1, 3, 5\} \subset \Omega_1 = \{1, \dots, 6\}$$

- (2) Die Augensumme zweier Würfel ist $\geq 10 \hat{=}$

$$A_2 = \{(6, 4), (4, 6), (5, 5), (6, 5), (5, 6), (6, 6)\} \subset \Omega_2 = \{1, \dots, 6\}^2$$

Ein geeignetes System von Mengen, das wir zur Beschreibung von Ereignissen verwenden, ist gegeben durch:

1.6 Definition. Ein System $\mathfrak{A} \subset \mathbb{P}(\Omega)$ von Teilmengen von Ω heißt σ -Algebra auf Ω , falls

$$(\sigma A_1) \quad \Omega \in \mathfrak{A}$$

$$(\sigma A_2) \quad A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$$

$$(\sigma A_3) \quad \text{Ist } A_n \in \mathfrak{A} \quad \forall n \geq 1 \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$$

Die $A \in \mathfrak{A}$ heißen *Ereignisse*, die $\{\omega\}$, $\omega \in \Omega$ *Elementarereignisse* (falls sie zu $\mathfrak{A}$ gehören)

1.7 Bemerkung. Die Tatsache, daß wir als System der „Ereignisse“ i. a. *nicht* die komplette Potenzmenge $\mathbb{P}(\Omega)$ verwenden, ist auf den ersten Blick befremdlich; sie ist aber praktikabel, unvermeidbar und sogar sinnvoll:

- (1) Man kann zeigen, daß auf überabzählbaren Ω es prinzipiell unmöglich ist, sinnvoll W-Maße auf ganz $\mathbb{P}(\Omega)$ zu definieren, sondern nur auf einer kleineren σ -Algebra.

1.8 Satz. Es sei $\mathfrak{A}$ eine σ -Algebra auf $\Omega \neq \emptyset$ und $A, B, A_n \in \mathfrak{A}$ für alle $n \in \mathbb{N}$. Dann gilt

$$(a) \quad \emptyset \in \mathfrak{A}, A \cap B \in \mathfrak{A}, A \setminus B \in \mathfrak{A}, A \cup B \in \mathfrak{A}$$

$$(b) \quad \bigcap_{n \in \mathbb{N}} A_n \in \mathfrak{A}$$

$$(c) \quad (i) \quad \liminf_{n \rightarrow \infty} A_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k \in \mathfrak{A}$$

$$(ii) \quad \limsup_{n \rightarrow \infty} A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k \in \mathfrak{A}$$

1.9 Bemerkung.

$$\omega \in \liminf A_n \Leftrightarrow \omega \in A_n \text{ für fast alle } n \text{ (d. h. bis auf endlich viele } n)$$

$$\omega \in \limsup A_n \Leftrightarrow \omega \in A_n \text{ für } \infty\text{-viele } n$$

BEWEIS. (a) $\emptyset = \Omega^c \in \mathfrak{A}$ wegen (σA_1) und (σA_2)

(b)

$$\begin{aligned}\bigcap_{n \in \mathbb{N}} A_n &= \left(\bigcup_{n \in \mathbb{N}} A_n^c \right)^c \in \mathfrak{A}, \text{ da } \bigcup_{n \in \mathbb{N}} A_n^c \in \mathfrak{A} \text{ nach } (\sigma A_2) \text{ und } (\sigma A_3). \\ A \cap B &= \bigcap_{n \in \mathbb{N}} A_n \in \mathfrak{A}, \text{ falls } A_0 = A, A_1 = B, A_n = \Omega \ (n \geq 2) \\ A \setminus B &= A \cap B^c \in \mathfrak{A} \\ A \cup B &= (A^c \cap B^c)^c \in \mathfrak{A}\end{aligned}$$

(c) Setzt man für $n \in \mathbb{N}$, $B_n = \bigcup_{k \geq n} A_k$, so ist das als abzählbare Vereinigung von Mengen aus $\mathfrak{A}$ in $\mathfrak{A}$ für jedes $n \in \mathbb{N}$:

$$\Rightarrow \limsup A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k = \bigcap_{n \in \mathbb{N}} B_n \in \mathfrak{A}$$

Ebenso für $\liminf A_n$!

□

Falls Ω endlich oder abzählbar unendlich ist, ist alles viel einfacher:

1.10 Satz. Sei Ω abzählbar. Die einzige σ -Algebra, die alle Elementarereignisse $\{\omega\}$ für $\omega \in \Omega$ enthält, ist $\mathfrak{A} = \mathbb{P}(\Omega)$

BEWEIS. Es reicht $\mathbb{P}(\Omega) \subset \mathfrak{A}$ zu zeigen.

Sei $A \in \mathbb{P}(\Omega)$ beliebig. Falls A abzählbar unendlich, etwa $A = \{\omega_0, \omega_1, \dots\}$, so ist

$$A = \bigcup_{n \in \mathbb{N}} \{\omega_n\} \in \mathfrak{A} \text{ nach } (\sigma A_3)$$

Falls $A = \{\omega_1, \dots, \omega_n\}$ endlich, ist

$$A = \bigcup_{j=1}^n \{\omega_j\} \in \mathfrak{A}$$

□

σ -Algebren werden häufig nicht explizit (durch Angabe *aller* zugehörigen Ereignisse), sondern implizit durch Angabe von Grundereignissen und Verwendung der Axiome (1.6) angegeben. Zum Beispiel wandelt **Satz 1.10** die implizite Forderung „alle Elementarereignisse gehören dazu“ in die explizite Angabe $\mathfrak{A} = \mathbb{P}(\Omega)$ um.

1.11 Lemma. Es sei $I \neq \emptyset$ eine Indexmenge, und für $i \in I$ sei $\mathfrak{A}_i$ eine σ -Algebra auf Ω . Dann ist auch $\bigcap_{i \in I} \mathfrak{A}_i$ eine σ -Algebra.

BEWEIS. (σA_1) : $\Omega \in \mathfrak{A}_i$ (nach (σA_1)) für alle $i \in I$. (Übung) $\Rightarrow \Omega \in \bigcap_{i \in I} \mathfrak{A}_i$. (σA_2) und (σA_3) ebenso! □

1.12 Satz und Definition. Sei $\mathfrak{E} \subset \mathbb{P}(\Omega)$ ein Mengensystem. Dann ist

$$\mathfrak{A}(\mathfrak{E}) := \bigcap_{\substack{\mathfrak{A} \text{ } \sigma\text{-Algebra} \\ \mathfrak{E} \subset \mathfrak{A}}} \mathfrak{A}$$

eine σ -Algebra, und zwar die kleinste, die $\mathfrak{E}$ enthält. Man nennt $\mathfrak{A}(\mathfrak{E})$ die **von $\mathfrak{E}$ erzeugte σ -Algebra**.

BEWEIS. Es ist $\mathbb{P}(\Omega)$ eine σ -Algebra mit $\mathfrak{E} \subset \mathbb{P}(\Omega)$. Also enthält der Durchschnitt mindestens $\mathbb{P}(\Omega)$ und (1.11) zeigt, daß $\mathfrak{A}(\mathfrak{E})$ eine σ -Algebra ist. Offenbar gilt $\mathfrak{E} \subset \mathfrak{A}(\mathfrak{E})$. Ist $\mathfrak{A}'$ irgendeine σ -Algebra, die $\mathfrak{E}$ enthält, so kommt $\mathfrak{A}'$ im Schnitt vor, also $\mathfrak{A}' \supset \bigcap \mathfrak{A} = \mathfrak{A}(\mathfrak{E})$. $\mathfrak{A}(\mathfrak{E})$ ist also die kleinste. □

1.13 Bemerkung. Der **Satz 1.10** stellt also einen Zusammenhang zwischen impliziter und expliziter Definition einer σ -Algebra dar, denn er besagt: Ω abzählbar $\Rightarrow$

$$\begin{array}{ccc} \mathfrak{A}(\text{Elementarereignisse}) & = & \mathbb{P}(\Omega) \\ \uparrow & & \uparrow \\ \text{implizit} & & \text{explizit} \end{array}$$

Das Konstruktionsprinzip der erzeugten σ -Algebra wird nun erstmals wichtig für die aus zwei Teilerperimenten zusammengesetzten Experimente, die ja als Stichprobenraum nach **Bsp. 1.3** Produktmengen haben.

1.14 Beispiel (zwei Experimente). Es werde das erste Experiment durch $(\Omega_1, \mathfrak{A}_1)$ beschrieben, sowie das zweite durch $(\Omega_2, \mathfrak{A}_2)$. Sei $\Omega = \Omega_1 \times \Omega_2$. Wir wollen mindestens die Teilmengen

$$\begin{aligned} A_1 \times \Omega_2 & \quad (\hat{=} \text{beim ersten Experiment geschieht } A_1) \quad \text{und} \\ \Omega_1 \times A_2 & \quad (\hat{=} \text{beim zweiten Experiment geschieht } A_2) \end{aligned}$$

in der σ -Algebra haben. Wir verwenden deshalb auf $\Omega = \Omega_1 \times \Omega_2$ die *Produkt σ -Algebra*

$$\mathfrak{A}_1 \otimes \mathfrak{A}_2 := \mathfrak{A}(\{A_1 \times \Omega_2, \Omega_1 \times A_2 \mid A_1 \in \mathfrak{A}_1, A_2 \in \mathfrak{A}_2\})$$

Insbesondere gehören dazu die Mengen

$$A_1 \times A_2 = (A_1 \times \Omega_2) \cap (\Omega_1 \times A_2) \quad \text{falls } A_j \in \mathfrak{A}_j$$

(1) **Spezialfall:**

Sind Ω_1, Ω_2 abzählbar, so auch $\Omega_1 \times \Omega_2$. Falls $\mathfrak{A}_1 = \mathbb{P}(\Omega_1)$, $\mathfrak{A}_2 = \mathbb{P}(\Omega_2)$, so ist jedes Elementarereignis $\{(\omega_1, \omega_2)\} \in \Omega_1 \times \Omega_2$ als

$$\{(\omega_1, \omega_2)\} = \{\omega_1\} \times \{\omega_2\} \in \mathfrak{A}_1 \otimes \mathfrak{A}_2$$

und somit nach **Satz 1.10**

$$\mathfrak{A}_1 \otimes \mathfrak{A}_2 = \mathbb{P}(\Omega_1 \times \Omega_2)$$

(2) Der zweimalige Wurf eines Würfels werde durch $\Omega = \Omega_1 \times \Omega_2$ mit $\Omega_1 = \Omega_2 = \{1, \dots, 6\}$ modelliert. Ist $\mathfrak{A}_j = \mathbb{P}(\Omega_j)$, so ist die Produkt σ -Algebra $\mathfrak{A}_1 \otimes \mathfrak{A}_2$ die Potenzmenge $\mathbb{P}(\Omega)$ und folglich z. B. :

$$\begin{aligned} \text{erster Wurf ungerade} & \hat{=} \{1, 3, 5\} \times \{1, \dots, 6\} \\ \text{Summe} \geq 10 & \hat{=} \{(6, 4), (4, 6), (5, 5), (6, 5), (5, 6), (6, 6)\} \\ \text{beide Würfel gleich} & \hat{=} \{(1, 1), (2, 2), \dots, (6, 6)\} \end{aligned}$$

1.15 Beispiel. Ein Experiment, das durch $(\Omega, \mathfrak{A})$ beschrieben wird, soll mehrmals wiederholt werden. Die Zahl der Wiederholungen werde durch eine Menge I indiziert, also etwa:

$$\begin{array}{lll} I = \{1, \dots, n\} & \text{bei} & n \text{ Wiederholungen} \\ I = \mathbb{N} & \text{bei} & \infty\text{-vielen Wiederholungen} \\ I = \mathbb{R}_+ & \text{bei} & \text{kontinuierlichem Proze\ss} \end{array}$$

I entspricht der Menge der Zeitpunkte. Stichprobenraum:

$$\Omega^I = \{(\omega_i : i \in I) \mid \omega_i \in \Omega\}$$

ω_i beschreibt das Ereignis des i -ten Experimentes bzw. der i -ten Beobachtung. Auf Ω^I soll nun eine geeignete σ -Algebra definiert werden. Dazu sollten zumindest Ereignisse gehören, die über ein bestimmtes Experiment $j \in I$ eine Aussage machen, also die Mengen

$$Z_j(A) = \{(\omega_i) \in \Omega^I \mid \omega_j \in A, \omega_i \text{ beliebig für } i \neq j\}$$

Die von dem System

$$\mathfrak{G} = \{Z_j(A) \mid j \in I, A \in \mathfrak{A}\}$$

erzeugte σ -Algebra heißt *Produkt σ -Algebra*

$$\mathfrak{A}^{\otimes I} = \mathfrak{A}(\{Z_j(A) \mid j \in I, A \in \mathfrak{A}\})$$

Falls $I = \{1, \dots, n\}$ schreiben wir auch $\mathfrak{A}^{\otimes n}$ dafür. In diesem Fall gehören die Mengen

$$A_1 \times \dots \times A_n = \bigcap_{j=1}^n Z_j(A_j)$$

zu $\mathfrak{A}^{\otimes n}$, falls alle $A_j \in \mathfrak{A}_j$.

- (1) **Bemerkung.** Ist Ω höchstens abzählbar und $\mathfrak{A} = \mathbb{P}(\Omega)$, so ist nach (1.10) auch $\mathfrak{A}^{\otimes n} = \mathbb{P}(\Omega^n)$.
- (2) **Bemerkung.** Falls I unendlich ist und Ω mehr als ein Element hat, dann ist Ω^I überabzählbar und (1.15(1)) gilt nicht.
- (3) **Beispiel.** Ein Würfel werde n -mal geworfen. Es ist $\Omega = \{1, \dots, 6\}^n$ und σ -Algebra $(\mathbb{P}\{1, \dots, 6\})^{\otimes n} = \mathbb{P}(\{1, \dots, 6\}^n)$. Das Ereignis

$$\begin{aligned} \text{„mind. eine Sechs“} &\triangleq \{(\omega_1, \dots, \omega_n) \mid \omega_j = 6 \text{ für mind. ein } j \in \{1, \dots, n\}\} \\ &= Z_1\{6\} \cup Z_2\{6\} \cup \dots \cup Z_n\{6\} \end{aligned}$$

gehört also zur σ -Algebra.

1.16 Beispiel (Die BORELSche σ -Algebra). $\mathbb{R}$ ist überabzählbar und die Potenzmenge als Menge von Ereignissen ist zu groß. Man möchte aber zumindest die Intervalle dabei haben und auch die einpunktigen Mengen.

- (1) **Definition.** Die BORELSche σ -Algebra auf $\mathbb{R}$ ist die kleinste σ -Algebra, die alle Intervalle $I = (a, b]$ mit $a < b$ enthält, also

$$\mathfrak{B}(\mathbb{R}) = \mathfrak{A}(\{(a, b] \mid a < b; a, b \in \mathbb{R}\})$$

Die Elemente von $\mathfrak{B}(\mathbb{R})$ heißen BOREL-Mengen.

- (2) **Eigenschaften.**

- (i) Jedes Elementarereignis $\{a\}$ liegt in $\mathfrak{B}(\mathbb{R})$
- (ii) $(a, b), (-\infty, b), (a, \infty) \in \mathfrak{B}(\mathbb{R})$ (offene Intervalle)
- (iii) $[a, b], (-\infty, b], [a, \infty) \in \mathfrak{B}(\mathbb{R})$ (abgeschlossene Intervalle)
- (iv) $[a, b), (a, b] \in \mathfrak{B}(\mathbb{R})$

BEWEIS. (i) Es ist $\{a\} = \bigcap_{n \geq 1} (a - \frac{1}{n}, a] \in \mathfrak{B}(\mathbb{R})$.

$$\begin{aligned} (a, b) &= (a, b] \setminus \{b\} \in \mathfrak{B}(\mathbb{R}); & [a, b) &= (a, b) \cup \{a\} \in \mathfrak{B}(\mathbb{R}) \\ (-\infty, a] &= \bigcup_{n \geq 1} (a - n, a] \in \mathfrak{B}(\mathbb{R}) \end{aligned}$$

und die übrigen gehen genauso.

□

- (3) **Beispiel.** Die Menge aller Zahlen in $[0, 1]$, die „7“ in der Dezimalentwicklung nach dem Komma haben, ist BOREL-Menge, da $A_1 = [0, 7; 0, 8)$. Genauso ist die Menge mit „7“ in der zweiten Stelle hinter dem Komma, nämlich

$$A_2 = [0, 07; 0, 08) \cup [0, 17; 0, 18) \cup \dots \cup [0, 97; 0, 98) \in \mathfrak{B}(\mathbb{R})$$

und analog $A_n \triangleq$ „7 in der n -ten Stelle“ $\in \mathfrak{B}(\mathbb{R})$.

Nach Satz 1.8 (vgl. Bem. 1.9) gehört also auch

$$\limsup_{n \rightarrow \infty} A_n \triangleq \text{„7 kommt } \infty\text{-oft vor“}$$

zu $\mathfrak{B}(\mathbb{R})$.

(4) **Satz.** Jede offene Teilmenge von $\mathbb{R}$ gehört zu $\mathfrak{B}(\mathbb{R})$.

BEWEIS. $U \subset \mathbb{R}$ offen $\Leftrightarrow \forall x \in U \exists \varepsilon > 0 : (x - \varepsilon, x + \varepsilon) \subset U$. Offensichtlich gehören die offenen Mengen $\emptyset, \mathbb{R} \in \mathfrak{B}(\mathbb{R})$. Sei nun $U \neq \emptyset$ und $U \neq \mathbb{R}$. Es ist $U \cap \mathbb{Q} = \{q_0, q_1, \dots\}$ abzählbar. Zu $q_n \in U \cap \mathbb{Q}$ wählen wir ein $\varepsilon_n > 0$, so daß $(q_n - \varepsilon_n, q_n + \varepsilon_n) \subset U$, aber $(q_n - 2\varepsilon_n, q_n + 2\varepsilon_n) \not\subset U$. [Dies geht, da $U \neq \mathbb{R}$: Wir wählen ein „maximales“ $\varepsilon > 0$ mit $(x - \varepsilon, x + \varepsilon) \subset U$ für $x \in U$.]

Dann gilt:

$$U = \bigcup_{n \in \mathbb{N}} (q_n - \varepsilon_n, q_n + \varepsilon_n)$$

„ $\supset$ “: gilt nach Definition von ε_n, q_n .

„ $\subset$ “: Sei $x \in U$ beliebig. Dann existiert genau ein $\varepsilon > 0$ mit $(x - \varepsilon, x + \varepsilon) \subset U$. Da $\mathbb{Q}$ dicht in $\mathbb{R}$ liegt, gibt es ein q_n mit $|q_n - x| < \frac{\varepsilon}{3}$.

Behauptung: $\frac{\varepsilon}{3} \leq \varepsilon_n$

BEWEIS: Annahme: $\frac{\varepsilon}{3} > \varepsilon_n$.

Sei $z \in (q_n - 2\varepsilon_n, q_n + 2\varepsilon_n)$ beliebig. Dann:

$$\begin{aligned} |z - x| &\leq |z + q_n| + |q_n - x| < 2\varepsilon_n + \frac{\varepsilon}{3} \\ &< 2\frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \end{aligned}$$

Daraus folgt aber:

$$(q_n - 2\varepsilon_n, q_n + 2\varepsilon_n) \subset (x - \varepsilon, x + \varepsilon) \subset U \quad \nexists$$

Also ist $x \in (q_n - \varepsilon_n, q_n + \varepsilon_n)$. Und folglich

$$x \in \bigcup_{n \in \mathbb{N}} (q_n - \varepsilon_n, q_n + \varepsilon_n) \in \mathfrak{B}(\mathbb{R})$$

da nach **Bsp. 1.16(2)** $(q_n - \varepsilon_n, q_n + \varepsilon_n) \in \mathfrak{B}(\mathbb{R})$ für alle $n \geq 1$. □

(5) **Korollar.** Jede abgeschlossene Teilmenge von $\mathbb{R}$ gehört zu $\mathfrak{B}(\mathbb{R})$.

BEWEIS. Ist $A \subset \mathbb{R}$ abgeschlossen, dann ist $A^c \subset \mathbb{R}$ offen. Nach **Bsp. 1.16(4)** ist dann $A^c \in \mathfrak{B}(\mathbb{R})$ und nach **Def. 1.6(σA_2)** $A = (A^c)^c \in \mathfrak{B}(\mathbb{R})$. □

Die BORELSche σ -Algebra $\mathfrak{B}(\mathbb{R})$ enthält also viele Mengen.

Nachdem wir bisher den Stichprobenraum Ω und die σ -Algebra der Ereignisse $\mathfrak{A}$ definiert haben, benötigen wir nun noch eine Abbildung $P : \mathfrak{A} \rightarrow [0, 1]$, die uns sagt „wie wahrscheinlich“ ein Ereignis $A \in \mathfrak{A}$ ist.

Dabei sollte P die folgenden „plausiblen“ Eigenschaften besitzen:

(a) $P(\Omega) = 1$ (Normierung)

(b) $P(A \cup B) = P(A) + P(B)$, falls $A \cap B = \emptyset$ (endliche Additivität)

1.17 Definition. Es sei $\mathfrak{A}$ eine σ -Algebra auf Ω . Eine Abbildung $P : \mathfrak{A} \rightarrow \mathbb{R}_+$ heißt *Wahrscheinlichkeitsmaß* auf Ω , falls gilt:

(WM₁) $P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$ für jede disjunkte Familie $(A_n)_{n \in \mathbb{N}}$ von Ereignissen $A_n \in \mathfrak{A}$

(WM₂) $P(\Omega) = 1$

[Die Familie $(A_n)_{n \in \mathbb{N}}$ heißt disjunkt, falls $A_n \cap A_m = \emptyset$ für $m \neq n$]

Das Tripel $(\Omega, \mathfrak{A}, P)$ wird als *Wahrscheinlichkeitsraum* bezeichnet.

1.18 Beispiel (LAPLACEScher W-Raum). Ist Ω eine endliche Menge, so definiert

$$P : \mathbb{P}(\Omega) \rightarrow \mathbb{R}_+ \quad P(A) := \frac{\#A}{\#\Omega}$$

ein W-Maß. Diese diskrete Gleichverteilung wird unter anderem dann verwendet, wenn aus Symmetriegründen alle Elementarereignisse als gleichwahrscheinlich angesehen werden. Wir bezeichnen die diskrete Gleichverteilung auf Ω mit $\mathcal{L}_\Omega$; $(\Omega, \mathbb{P}(\Omega), \mathcal{L}_\Omega)$ heißt LAPLACE-Raum.

1.19 Beispiel.

(1) **(fairer Münzwurf)**

Modell: $\Omega = \{W, Z\}$ mit $\mathfrak{A} = \mathbb{P}(\Omega)$. Gleichverteilung $P(A) = \frac{1}{2}\#A$ für $A \in \mathfrak{A}$.

(2) **(verfälschte Münze)**

Wappen $\hat{=}$ „1“; Zahl $\hat{=}$ „0“. Modell: $\Omega = \{0, 1\}$, $\mathfrak{A} = \mathbb{P}(\Omega)$.

Ist die Münze nicht fair, so liegt ein **BERNOULLI-Experiment** vor:

Es sei $p \in [0, 1]$. Dann ist durch $\mathcal{B}_{1,p} : \mathbb{P}(\Omega) \rightarrow \mathbb{R}_+$; $\mathcal{B}_{1,p}(\emptyset) = 0$; $\mathcal{B}_{1,p}(\Omega) = 1$; $\mathcal{B}_{1,p}\{1\} = p$; $\mathcal{B}_{1,p}\{0\} = 1 - p$ die allgemeine Form eines W-Maßes auf einer zweielementigen Menge gegeben. Für $p = \frac{1}{2}$ erhalten wir wieder die Gleichverteilung aus (1.19(1)).

(3) **(wiederholtes Werfen einer fairen Münze)**

Modell: $\Omega = \{1, 0\}^n$ mit $\mathfrak{A} = \mathbb{P}(\Omega)$ und der LAPLACE-Verteilung $P = \mathcal{L}_\Omega$. Das Ereignis „alle Würfe gleich“ $\hat{=}$ $A = \{(0, \dots, 0), (1, \dots, 1)\}$ und es gilt $P(A) = \frac{2}{2^n} = 2^{1-n}$.

(4) **(binary search)**

Aus einer angeordneten Menge von $2^n - 1$ Elementen wird ein Schlüsselwert gesucht, indem man im ersten Schritt das 2^{n-1} -te Element mit dem zu suchenden vergleicht. Besteht Gleichheit, ist man fertig; ist das zu suchende Element kleiner, so macht man mit den ersten $2^{n-1} - 1$ Elementen weiter; ist es größer, so setzt man das Intervallhalbierungsverfahren mit den $2^{n-1} - 1$ Elementen $2^{n-1} + 1, \dots, 2^n - 1$ fort.

Modell: $\Omega = \{0, 1, \dots, 2^n - 1\}$, wobei $0 \hat{=}$ „Element nicht gefunden“. Die Teilmenge $A = \{2^{n-2}, 2^{n-1}, 3 \cdot 2^{n-2}\}$ ist dann zum Beispiel das Ereignis: „man braucht höchstens zwei Suchschritte“ und $B = \{0, 1, 3, \dots, 2^n - 1\}$ ist „man braucht n Schritte“. $P = \mathcal{L}_\Omega$. Es folgt $P(B) = \frac{1}{2} + 2^{-n}$.

(5) **(Punktmaß)**

Es sei $\mathfrak{A}$ eine σ -Algebra auf Ω und $x \in \Omega$. Dann ist durch

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

ein W-Maß auf Ω definiert.

BEWEIS. Es ist $\mathcal{E}_x(\Omega) = 1$, da $x \in \Omega$ und für eine disjunkte Familie $(A_n)_n \subset \mathfrak{A}$ gilt entweder

(i) x liegt in genau einem der A_n

$$\Rightarrow x \in \bigcup_{n \in \mathbb{N}} A_n \Rightarrow \mathcal{E}_x\left(\bigcup_n A_n\right) = 1$$

und $\mathcal{E}_x(A_n) = 1$ für genau ein n .

(ii) x liegt in keinem der $A_n \Rightarrow x \notin \bigcup_n A_n$.

□

1.20 Satz. Es sei $(\Omega, \mathfrak{A}, P)$ ein W-Raum.

(a) $P(\emptyset) = 0$

(b) $\forall A, B \in \mathfrak{A}$ gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

insbesondere gilt $P(A \cup B) = P(A) + P(B)$, falls A und B disjunkt sind.

(c) Im Fall $A, B \in \mathfrak{A}$ und $A \subset B$ gilt

$$P(B \setminus A) = P(B) - P(A)$$

Insbesondere gilt $P(A) \leq P(B)$ und $P(A^c) = 1 - P(A)$.

(d) Es sei $(A_n)_n \subset \mathfrak{A}$ eine aufsteigende Folge, d. h. $A_n \subset A_{n+1} \forall n \in \mathbb{N}$. Dann gilt

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) \quad \text{Stetigkeit von unten}$$

(e) Es sei $(A_n)_n \subset \mathfrak{A}$ eine absteigende Folge, d. h. $A_n \supset A_{n+1} \forall n \in \mathbb{N}$. Dann gilt

$$P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) \quad \text{Stetigkeit von oben}$$

BEWEIS.

(a) Setzen wir $A_0 = \Omega$ und $A_n = \emptyset \forall n \geq 1$. Nach 1.17(WM₁), (WM₂) ist dann:

$$1 = P(\Omega) = P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = P(A_0) + \sum_{n \geq 1} P(\emptyset) \geq 1 + P(\emptyset)$$

$$\Rightarrow 0 \geq P(\emptyset) \geq 0.$$

(b) **Fall 1:** $A \cap B = \emptyset$

Setze $A_0 = A$, $A_1 = B$ und $A_n = \emptyset$ für $n \geq 2$. Nach Teil (a) gilt dann:

$$P(A \cup B) = P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = P(A) + P(B)$$

Fall 2: $A, B \in \mathfrak{A}$ beliebig.

Dann sind $B \setminus A$ und A , sowie $B \setminus A$ und $A \cap B$ disjunkt. Daraus folgt:

$$\begin{aligned} P(A \cup B) &= P((B \setminus A) \dot{\cup} A) = P(B \setminus A) + P(A) \\ &= P(B \setminus A) + P(A \cap B) + P(A) - P(A \cap B) \\ &= P((B \setminus A) \dot{\cup} (A \cap B)) + P(A) - P(A \cap B) \\ &= P(B) + P(A) - P(A \cap B) \end{aligned}$$

(c) Folgt sofort aus (b)

(d) Betrachte die disjunkte Folge $(B_n)_n$ mit $B_0 = A_0$; $B_{n+1} = A_{n+1} \setminus A_n$. Dann gilt

$$\begin{aligned} P(A_n) &= P\left(\bigcup_{m=0}^n B_m\right) = \sum_{m=0}^n P(B_m) \\ \Rightarrow P\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= P\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \sum_{n \in \mathbb{N}} P(B_n) \\ &= \lim_{n \rightarrow \infty} \sum_{m=0}^n P(B_m) = \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

(e) Folgt aus (c) und (d).

□

1.21 Anwendung (n -faches Würfeln). Modell: $\Omega = \{1, \dots, 6\}^n$, $\mathfrak{A} = \mathbb{P}(\Omega)$, $P = \mathcal{L}_\Omega$

$$A = \{(\omega_1, \dots, \omega_n) \mid \omega_j = 6 \text{ für mindestens ein } j \leq n\}$$

$$\hat{=} \text{mindestens eine Sechs}$$

Um $P(A)$ zu berechnen, ist es einfacher $P(A^c)$ zu berechnen und dann **Satz 1.20(c)** zu verwenden. Es ist

$$A^c = \{(\omega_1, \dots, \omega_n) \mid \omega_j \leq 5 \text{ für alle } j\}$$

Und somit

$$P(A^c) = \frac{\#A^c}{\#\Omega} = \frac{\#\{1, \dots, 5\}^n}{\#\{1, \dots, 6\}^n} = \frac{5^n}{6^n} = \left(\frac{5}{6}\right)^n$$

Nach **Satz 1.20(c)** ist dann

$$P(A) = 1 - P(A^c) = 1 - \left(\frac{5}{6}\right)^n$$

Im Fall $n = 13$ gilt $P(A) \approx 0,907$. Man muß also 13 mal würfeln, um mit der Wahrscheinlichkeit von mindestens 90% eine (oder mehrere) Sechsen zu würfeln.

Es ist oft mühsam, für alle Ereignisse $A \in \mathfrak{A}$ den Wert von $P(A)$ anzugeben. Diskrete, d. h. abzählbare W-Räume, können durch den Begriff der Zähldichte einfacher beschrieben werden.

1.22 Definition. Es sei Ω abzählbar. Eine Funktion $f : \Omega \rightarrow \mathbb{R}_+$ heißt *Zähldichte*, falls

$$\sum_{\omega \in \Omega} f(\omega) = 1$$

1.23 Satz. Es sei Ω abzählbar und $\mathfrak{A} = \mathbb{P}(\Omega)$. Zwischen W-Maßen P und Zähldichten f auf Ω besteht folgender bijektiver Zusammenhang:

(a) Ist P ein W-Maß auf Ω , so ist durch

$$f_P : \Omega \rightarrow \mathbb{R}_+, \quad f_P(\omega) := P(\{\omega\}), \quad \omega \in \Omega$$

eine Zähldichte f_P definiert

(b) Für jede Zähldichte f gibt es genau ein W-Maß P mit $f_P = f$, nämlich

$$P(A) := \sum_{\omega \in A} f(\omega)$$

BEWEIS.

(a) Es ist $f_P(\omega) = P(\{\omega\}) \geq 0$ und wegen **1.17(WM₁), (WM₂)** gilt

$$\sum_{\omega \in \Omega} f_P(\omega) = \sum_{\omega \in \Omega} P(\{\omega\}) = P\left(\bigcup_{\omega \in \Omega} \{\omega\}\right) = P(\Omega) = 1$$

(b) Es ist $P(\Omega) = \sum_{\omega \in \Omega} f(\omega)$. Ist $(A_n)_n$ eine disjunkte Familie, so gilt wegen der Assoziativität der Summation:

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{\substack{\omega \in \bigcup_{n \in \mathbb{N}} A_n \\ n \in \mathbb{N}}} f(\omega) = \sum_{n \in \mathbb{N}} \sum_{\omega \in A_n} f(\omega) = \sum_{n \in \mathbb{N}} P(A_n)$$

$\Rightarrow P$ ist W-Maß.

Ist Q ein weiteres W-Maß mit $f_Q = f$, so folgt für alle $A \subset \Omega$

$$Q(A) = Q\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} Q(\{\omega\}) = \sum_{\omega \in A} f_Q(\omega) = \sum_{\omega \in A} f(\omega) = P(A)$$

□

Es genügt also, zur vollständigen Beschreibung eines diskreten W-Maßes, seine Zähldichte anzugeben.

1.24 Beispiel.

- (1) Die Gleichverteilung $\mathcal{L}_\Omega$ auf einer endlichen Menge Ω hat die konstante Funktion

$$f(\omega) = \frac{1}{\#\Omega}$$

als Zähldichte.

- (2) Es seien $p \in [0, 1]$ und $n \geq 1$; $\Omega = \{0, \dots, n\}$, $\mathfrak{A} = \mathbb{P}(\Omega)$. Dann ist durch

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (\text{für } k \in \Omega)$$

eine Zähldichte definiert, denn $f(k) \geq 0$ und

$$\sum_{k=0}^n f(k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1$$

Das durch f definierte W-Maß $\mathcal{B}_{n,p}$ mit

$$\mathcal{B}_{n,p}\{k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

heißt *Binomialverteilung* mit Parametern n und p . Im Fall $n = 1$ erhalten wir die **BERNOULLI-Verteilung** aus **Bsp. 1.19(2)**.

- (3) Ein wichtiges W-Maß auf $\mathbb{N}$ ist die *POISSON-Verteilung* Π_λ mit Parameter $\lambda \geq 0$. Sie ist durch ihre Zähldichte

$$f(n) := \Pi_\lambda(\{n\}) := \frac{\lambda^n}{n!} e^{-\lambda} \quad (n \in \mathbb{N})$$

definiert. [Es ist $f(n) \geq 0$ und $\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = e^\lambda e^{-\lambda} = e^0 = 1$]

- (4) Die *geometrische Verteilung* $\mathcal{G}_p$ ist für $p \in (0, 1)$ das W-Maß auf $\mathbb{N}$ mit Zähldichte

$$f(n) = p(1-p)^n$$

1.25 Beispiel (kontinuierliche Gleichverteilung auf $[0, 1]$). Naiver Versuch, eine Gleichverteilung auf $\Omega = [0, 1]$ durch Zähldichte $f(\omega) = \varepsilon$ zu erhalten, schlägt fehl:

$$\text{ist } \varepsilon > 0 \quad \Rightarrow \quad P(\Omega) = \sum_{\omega \in \Omega} \varepsilon = \infty \cdot \varepsilon = \infty \neq 1$$

$$\text{ist } \varepsilon = 0 \quad \Rightarrow \quad P(\Omega) = \sum_{\omega \in \Omega} 0 = 0 \neq 1$$

Wir zeigen nun, daß unsere Probleme u. a. daher rühren, daß $\mathfrak{A} = \mathbb{P}(\Omega)$ als σ -Algebra zu groß ist. Wir formulieren den Begriff der Gleichverteilung durch die *Translationsinvarianz*, d. h. daß für alle $A \in \mathfrak{A}$, $x \in \mathbb{R}$ die verschobene Menge $A + x = \{a + x \mid a \in A\}$ im Fall $A + x \subset [0, 1]$ wieder zu $\mathfrak{A}$ gehört und $P(A + x) = P(A)$ gilt, wobei P die „Gleichverteilung“ auf $[0, 1]$ sein soll.

- (1) **Bemerkung.** Es gibt *kein* auf ganz $\mathbb{P}([0, 1])$ definiertes translationsinvariantes W-Maß P . [Beweis: VITALI, Stochastik II]

Es ist also unvermeidlich, daß wir eine kleinere σ -Algebra auf $\Omega = [0, 1]$ wählen. Wir verwenden die **BORELSche** σ -Algebra

$$\mathfrak{B}([0, 1]) = \{A \in \mathfrak{B}(\mathbb{R}) \mid A \subset [0, 1]\}$$

(Man zeige, daß dies eine σ -Algebra ist!)

- (2) **Faktum.** Es gibt genau ein W-Maß $P =: \lambda_{[0,1]}$ auf $\mathfrak{B}([0,1])$, das translationsinvariant ist. Es ist charakterisiert durch die Eigenschaft

$$\lambda_{[0,1]}((a, b]) = b - a \quad \text{für } 0 \leq a < b \leq 1$$

Es folgt:

$$\lambda_{[0,1]}(\{a\}) = \lim_{n \rightarrow \infty} \lambda_{[0,1]}(a - \frac{1}{n}, a] = 0$$

(Stetigkeit von oben!)

$$\Rightarrow \lambda_{[0,1]}([a, b]) = \lambda_{[0,1]}((a, b)) = \lambda_{[0,1]}([a, b)) = b - a$$

Die Translationsinvarianz für Intervalle kann man auch direkt ablesen:

$$\lambda_{[0,1]}((a+x, b+x]) = (b+x) - (a+x) = b-a = \lambda_{[0,1]}((a, b])$$

- (3) **Beispiel.** Wähle $\omega \in [0,1] = \Omega$ zufällig. Für $n \geq 1$ betrachten wir das Ereignis: „die n -te Nachkommastelle in der Dezimalbruchentwicklung (eindeutig, falls Periode 9 ausgeschlossen!) ist 7“.

$$A_n = \bigcup_{k=0}^{10^{n-1}-1} \left[\frac{k}{10^{n-1}} + \frac{7}{10^n}, \frac{k}{10^{n-1}} + \frac{8}{10^n} \right)$$

Es ist

$$\begin{aligned} \lambda[0,1](A_n) &= \sum_{k=0}^{10^{n-1}-1} \left[\left(\frac{k}{10^{n-1}} + \frac{8}{10^n} \right) - \left(\frac{k}{10^{n-1}} + \frac{7}{10^n} \right) \right] \\ &= 10^{n-1} \cdot \frac{1}{10^n} = \frac{1}{10} \end{aligned}$$

Wir wollen wissen, mit welcher Wahrscheinlichkeit irgendwo eine Sieben vorkommt, also die Wahrscheinlichkeit von $\bigcup_{n=1}^{\infty} A_n = A$. Deren Komplement:

$$\begin{aligned} B = A^c &= \bigcap_{n=1}^{\infty} B_n \quad \text{mit} \\ B_n &= \bigcup_{\substack{k_1, \dots, k_n \in \\ \{0, \dots, 9\} \setminus \{7\}}} \left[\frac{k_1}{10^1} + \dots + \frac{k_n}{10^n}, \frac{k_1}{10^1} + \dots + \frac{k_n+1}{10^n} \right) \\ &(\hat{=} \text{ die ersten } n \text{ Stellen sind } \neq 7) \end{aligned}$$

Es ist

$$\lambda[0,1](B_n) = \sum_{\substack{k_1, \dots, k_n \\ \neq 7}} 10^{-n} = 9^n \cdot 10^{-n} = \left(\frac{9}{10} \right)^n$$

Da offensichtlich $B_n \supset B_{n+1}$, folgt mit Stetigkeit von oben

$$\lambda[0,1](B) = \lim_{n \rightarrow \infty} \lambda[0,1](B_n) = \lim_{n \rightarrow \infty} \left(\frac{9}{10} \right)^n = 0$$

$\Rightarrow P(A) = 1$, d.h. mit Wahrscheinlichkeit 1 kommt in einer zufällig gezogenen Zahl $\omega \in [0,1]$ eine Sieben vor.

ENDE VON KAPITEL 1

KAPITEL 2

Der LAPLACESche Wahrscheinlichkeitsraum und kombinatorische Wahrscheinlichkeitsrechnung

Wird als stochastisches Modell die Gleichverteilung $\mathcal{L}_\Omega$ auf einer endlichen Menge Ω gewählt (geschieht immer dann, wenn die Elementarereignisse wegen ihrer Gleichartigkeit als gleichwahrscheinlich angesehen werden), ist die Bestimmung der Zahl der Elemente von $A \subset \Omega$ wesentlich wegen $\mathcal{L}_\Omega(A) = \frac{\#A}{\#\Omega}$. Diese Fragestellung ist Gegenstand der *Kombinatorik*. Dabei interessiert uns oft die Zahl der Möglichkeiten, wie aus einer gegebenen Menge mehrere Elemente ausgewählt werden können, etwa aus einer Urne mit/ohne Zurücklegen.

2.1 Definition. Es sei $A \neq \emptyset$ und $k \geq 1$.

- (a) (i) Jedes k -Tupel $(a_1, \dots, a_k)$ mit $a_i \in A$ heißt *k-Permutation aus A mit Wiederholung*; A^k ist die Menge der k -Permutationen aus A mit Wiederholung. Die Menge A^k beschreibt das Ziehen von Kugeln mit Zurücklegen und Reihenfolge.
- (ii) Eine *k-Permutation aus A ohne Wiederholung* ist ein k -Tupel $(a_1, \dots, a_k)$ mit $a_i \neq a_j$ für alle $i \neq j$. Die Menge aller solchen Permutationen bezeichnen wir mit $\mathcal{P}_k^A$. Sie beschreibt das Ziehen ohne Zurücklegen und in Reihenfolge.

Kommt es auf die Reihenfolge nicht an, spricht man von Kombinationen:

- (b) (i) Als *k-Kombination aus A ohne Wiederholung* bezeichnen wir eine k -elementige Teilmenge $\{a_1, \dots, a_k\} \subset A$. Die Menge dieser Kombinationen nennen wir $\mathcal{K}_k^A$.
- (ii) *k-Kombination aus A mit Wiederholung* bezeichnen wir auch als $\{a_1, \dots, a_k\}$, wobei jetzt der Fall $a_j = a_l$ nicht ausgeschlossen ist (Wir identifizieren $\{1, 1, 2\}$ mit $\{1, 2, 1\}$ und $\{2, 1, 1\}$ aber nicht mit $\{1, 2\}$). Die Menge aller k -Kombinationen mit Wiederholung nennen wir $\mathcal{M}_k^A$.

2.2 Satz. Es sei $A \neq \emptyset$, $\#A = n$ und $k \geq 1$. Dann gilt

- (a) $\#A^k = n^k$ (*k*-Permutationen mit Wiederholung)
- (b) $\#\mathcal{P}_k^A = (n)_k = n(n-1) \cdot \dots \cdot (n-k+1) = k! \binom{n}{k}$ (*k*-Permutationen ohne Wiederholung)
- (c) $\#\mathcal{K}_k^A = \binom{n}{k}$ (*k*-Kombinationen ohne Wiederholung)
- (d) $\#\mathcal{M}_k^A = \binom{n+k-1}{k}$ (*k*-Kombinationen mit Wiederholung)

BEWEIS.

- (a) $\#A^k = (\#A)^k = n^k$
- (b) Für a_1 gibt es n Möglichkeiten; für a_2 $(n-1)$ Möglichkeiten und für a_j , nachdem $a_1, \dots, a_{j-1}$ festgelegt sind, gibt es $n-j+1$ Möglichkeiten. Insgesamt also

$$n(n-1) \cdot \dots \cdot (n-k+1) = k! \binom{n}{k}$$

Möglichkeiten.

- (c) Kommt es auf die Reihenfolge nicht an, so fallen jeweils $k!$ Permutationen, die aus denselben Elementen bestehen, zu einer k -Kombination ohne Wiederholung zusammen. Deren Anzahl beträgt also

$$\frac{1}{k!} k! \binom{n}{k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- (d) Ohne Einschränkung: $A = \{1, \dots, n\}$. Jeder k -Komb. mit Wiederholung $\{a_1, \dots, a_k\}$ (wobei wir die a_i so aufschreiben können, daß $a_1 \leq a_2 \leq \dots \leq a_k$) ordnen wir eine k -Kombination ohne Wiederholung

$$f(\{a_1, \dots, a_k\}) := \{a_1 + 1, a_2 + 2, \dots, a_k + k\}$$

aus Elementen von $A' = \{2, \dots, n+k\}$ zu. f ist offensichtlich injektiv und auch surjektiv, denn das Urbild von $\{a'_1, \dots, a'_k\} \in \mathcal{K}_k^{A'}$ ist $\{a'_1 - 1, a'_2 - 2, \dots, a'_k - k\} \in \mathcal{M}_k^A$

$$\Rightarrow \#\mathcal{M}_k^A = \#\mathcal{K}_k^{A'} = \binom{\#A'}{k} = \binom{n+k-1}{k}$$

□

2.3 Beispiel.

- (1) Die Wahrscheinlichkeit, beim Skat alle 4 Buben auf die Hand zu bekommen, beträgt

$$\frac{\binom{28}{6}}{\binom{32}{10}} \approx 0,0058$$

Denn: Modell: $\Omega = \mathcal{K}_{10}^{\{1, \dots, 32\}}$ (LAPLACE-Raum) hat $\binom{32}{10}$ Elemente. Das Ereignis

$$„4 \text{ Buben}“ \triangleq A = \{\omega \in \Omega \mid 4, 12, 20, 28 \in \omega\}$$

besteht aus den Teilmengen von $\{1, \dots, 32\}$, die die 4 Buben 4, 12, 20, 28, sowie noch 6 andere aus den restlichen 28 Karten, beliebig ausgewählt, also $\#A = \binom{28}{6}$.

- (2) Aus einer Urne, in der $N = r+s$ Kugeln und zwar r rote und s schwarze liegen, werden $n \leq N$ (ohne Zurücklegen und ohne Reihenfolge) gezogen. Gesucht ist die Wahrscheinlichkeit, daß genau k davon rot sind.

$$\underbrace{1, \dots, r}_{\text{rot}}, \underbrace{r+1, \dots, r+s}_{\text{schwarz}} \quad \Omega = \mathcal{K}_n^{\{1, \dots, r+s\}} \quad \text{mit } P = \mathcal{L}_\Omega$$

$$\begin{aligned} A_k &:= \{\omega \in \Omega \mid \#(\omega \cap \{1, \dots, r\}) = k\} \\ &= \{\omega \subset \{1, \dots, r+s\} \mid \#(\omega \cap \{1, \dots, r\}) = k \wedge \#(\omega \cap \{r+1, \dots, r+s\}) = n-k\} \end{aligned}$$

Der k -elementige Teil roter Kugeln läßt sich auf genau $\binom{r}{k}$ Arten auswählen und der $(n-k)$ -elementige Teil der schwarzen auf $\binom{s}{n-k}$.

$$\Rightarrow \#A_k = \binom{r}{k} \binom{s}{n-k} \Rightarrow P(A_k) = \frac{\binom{r}{k} \binom{s}{n-k}}{\binom{r+s}{n}},$$

denn $\#\Omega = \#\mathcal{K}_n^{\{1,\dots,r+s\}} = \binom{r+s}{n}$.

Wegen $\Omega = \bigcup_{k=0}^n A_k \Rightarrow \sum_{k=0}^n P(A_k) = 1$ ist durch

$$\mathcal{H}_{n,N,r}\{k\} := \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}} \quad 0 \leq k \leq n$$

ein W-Maß auf $\{0, \dots, n\}$ definiert, die *Hypergeometrische Verteilung*.

- (3) In einer Lieferung von 10 000 Schrauben seien 2% defekt. Nimmt man 100 Schrauben heraus, so ist die Wahrscheinlichkeit für genau k defekte gerade $\mathcal{H}_{100,10\,000,200}\{k\}$.

k	0	1	2	3	4	5
$\mathcal{H}_{100,10\,000,200}\{k\}$	0,131	0,271	0,275	0,183	0,090	0,035

2.4 Beispiel. Die Mengen $A^k, \mathcal{P}_k^A, \mathcal{K}_k^A, \mathcal{M}_k^A$ können nicht nur zur Modellierung des Ziehens von k Kugeln aus einer Urne mit n Kugeln verwendet werden.

Andere Sichtweise: Verteilung von k Gegenständen auf n Zellen. Das Tupel $a_1, \dots, a_k$ beschreibt dann das Einlegen eines Gegenstandes in Zelle a_1 , eines in Zelle a_2 usw.

Belegungsmodell	Urnenmodell	mathematisches Modell
k Gegenstände auf n Zellen	k mal Ziehen aus Urne mit n Kugeln	Mengen $A = \{1, \dots, n\}$
(1) (unterscheidbare Objekte)	(1) (mit Reihenfolge)	(1) Permutationen
(a) mit Ausschließungsprinzip	(a) ohne Zurücklegen	$\Omega = \mathcal{P}_k^A, \quad \#\Omega = k! \binom{n}{k}$
(b) ohne Ausschließungsprinzip	(b) mit Zurücklegen	$\Omega = A^k, \quad \#\Omega = n^k$
(2) (nicht unterscheidbare Objekte)	(2) (ohne Reihenfolge)	(2) Kombinationen
(a) mit Ausschließungsprinzip	(a) ohne Zurücklegen	$\Omega = \mathcal{K}_k^A, \quad \#\Omega = \binom{n}{k}$
(b) ohne Ausschließungsprinzip	(b) mit Zurücklegen	$\Omega = \mathcal{M}_k^A, \quad \#\Omega = \binom{n+k-1}{k}$

- (1) Gesucht ist die Wahrscheinlichkeit, daß in einer Gruppe von k Personen ein Geburtstag mehrfach vorkommt.

Modell: $n = 365$ Zellen und k unterscheidbare Objekte; Mehrfachbelegungen sind möglich.

$$\Omega = \{1, \dots, n\}^k, \quad \mathfrak{A} = \mathbb{P}(\Omega), \quad P = \mathcal{L}_\Omega$$

Sei

$E \hat{=} \text{„ein Geburtstag kommt mehrmals vor“}$

$$= \{\omega = (\omega_1, \dots, \omega_k) \in \Omega \mid \exists i, j \in \{1, \dots, k\} : \omega_i = \omega_j\}$$

$\Rightarrow E^c \hat{=} \text{„alle Geburtstage verschieden“}$

$$= \{\omega = (\omega_1, \dots, \omega_k) \in \Omega \mid \omega_i \neq \omega_j \quad \forall i \neq j\}$$

$$= \mathcal{P}_k^{\{1, \dots, n\}}$$

$$\begin{aligned} \Rightarrow P(E) &= 1 - P(E^c) = 1 - \frac{\#\mathcal{P}_k^{\{1, \dots, n\}}}{\#\Omega} \\ &= 1 - \frac{(n)_k}{n^k} = 1 - \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{n^k} \\ &= 1 - \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right) \end{aligned}$$

Für $x \in [0, 1]$ gilt $1 - x \leq e^{-x}$

$$\begin{aligned} \Rightarrow P(E) &= 1 - \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right) \\ &\geq 1 - \exp\left(-\frac{1}{n} - \frac{2}{n} - \dots - \frac{k-1}{n}\right) = 1 - \exp\left(-k \frac{k-1}{2n}\right) \end{aligned}$$

Für $n = 365$ ist $P(E) \geq \frac{1}{2}$ für $k \geq 23$. Für $k \approx 70$ ist $P(E) \geq 0,9987$.

Ein wichtiges Hilfsmittel, das in der Kombinatorik häufig verwendet wird, aber für beliebige W-Maße gilt, ist die *Siebformel*:

2.5 Satz (SYLVESTER, POINCARÉ). Es sei $(\Omega, \mathfrak{A}, P)$ ein W-Raum und $A_1, \dots, A_n \in \mathfrak{A}$

$$P\left(\bigcup_{j=1}^n A_j\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J=k}} P\left(\bigcap_{j \in J} A_j\right)$$

BEWEIS. Entweder direkt aus der Definition des W-Maßes oder später (folgt aus Satz 2.7(b), der später bewiesen wird). $\square$

2.6 Beispiel. Auf einem Ball, an dem n Paare teilnehmen, werden alle Paare zufällig zusammengesetzt. Wie groß ist die Wahrscheinlichkeit, daß sich mindestens ein Paar wiederfindet?

Modell: $\Omega = \mathcal{P}_n^{\{1, \dots, n\}}$, $\mathfrak{A} = \mathbb{P}(\Omega)$, $P = \mathcal{L}_\Omega$

Ein Elementarereignis $\omega = (\omega_1, \dots, \omega_n)$ bedeutet dabei:

Dame 1 tanzt mit Partner von Dame ω_1

Dame 2 tanzt mit Partner von Dame ω_2 usw.

Uns interessiert die Wahrscheinlichkeit des Ereignisses

$$A = \{\omega \in \Omega \mid \omega_j = j \text{ für mindestens ein } j = 1, \dots, n\}$$

Für $J \subset \{1, \dots, n\}$ sei

$$A_J := \{\omega \in \Omega \mid \omega_j = j \ \forall j \in J\} \cong \mathcal{P}_{n-l}^{\{1, \dots, n\} \setminus J} \quad \text{falls } \#J = l$$

$$\Rightarrow P(A_J) = \frac{\#A_J}{\#\Omega} = \frac{(n-l)!}{n!} \quad \text{falls } \#J = l.$$

Nach der Siebformel gilt dann:

$$\begin{aligned} P(A) &= P\left(\bigcup_{j=1}^n A_{\{j\}}\right) = \sum_{l=1}^n (-1)^{l-1} \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J=l}} P\left(\bigcap_{j \in J} A_{\{j\}}\right) \\ &= \sum_{l=1}^n (-1)^{l-1} \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J=l}} P(A_J) = \sum_{l=1}^n (-1)^{l-1} \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J=l}} \frac{(n-l)!}{n!} \\ &= \sum_{l=1}^n (-1)^{l-1} \frac{(n-l)!}{n!} \#K_l^{\{1, \dots, n\}} = \sum_{l=1}^n (-1)^{l-1} \binom{n}{l} \frac{(n-l)!}{n!} \\ &= \sum_{l=1}^n \frac{(-1)^{l-1}}{l!} = 1 - \frac{1}{e} + \sum_{l=n+1}^{\infty} \frac{(-1)^l}{l!} \\ &\approx 0,632 \quad \text{schon für relativ kleine } n \end{aligned}$$

Die Siebformel ist ein Spezialfall der folgenden Aussage, die die Wahrscheinlichkeit dafür, daß genau bzw. mindestens k der Ereignisse $A_1, \dots, A_n$ eintreten, angibt.

2.7 Satz. Es sei $(\Omega, \mathfrak{A}, P)$ ein W-Raum und $A_1, \dots, A_n \in \mathfrak{A}$. Für $1 \leq k \leq n$ sei

$$B_k := \bigcup_{\substack{J \subset \{1, \dots, n\} \\ \#J=k}} \bigcap_{j \in J} A_j \cap \bigcap_{j \notin J} A_j^c \quad \text{„genau } k \text{ der Ereignisse treten ein“}$$

$$C_k := \bigcup_{\substack{J \subset \{1, \dots, n\} \\ \#J=k}} \bigcap_{j \in J} A_j \quad \text{„mindestens } k \text{ der Ereignisse treten ein“}$$

Es bezeichne

$$S_l := \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J=l}} P\left(\bigcap_{j \in J} A_j\right) \quad \text{für } 1 \leq k \leq n$$

Dann gilt:

$$(a) P(B_k) = \sum_{l=k}^n (-1)^{l-k} \binom{l}{k} S_l$$

$$(b) P(C_k) = \sum_{l=k}^n (-1)^{l-k} \binom{l-1}{k-1} S_l$$

BEWEIS. Der Beweis von (a) kommt später.

(b) Wir zeigen, wie (b) aus (a) folgt:

Beweis durch absteigende Induktion nach k . Für $k = n$ ist $C_n = \bigcap_{j \in \{1, \dots, n\}} A_j \Rightarrow P(C_n) = P\left(\bigcap_{j=1}^n A_j\right) = S_n$.

$k+1 \rightarrow k \geq 1$: Es ist $C_k = B_k \dot{\cup} C_{k+1}$.

$$\Rightarrow P(C_k) = P(B_k) + P(C_{k+1})$$

Nach Induktionvoraussetzung und Teil (a) folgt:

$$\begin{aligned} P(C_k) &= P(B_k) + P(C_{k+1}) \\ &= \sum_{l=k}^n (-1)^{l-k} \binom{l}{k} S_l + \sum_{l=k+1}^n (-1)^{l-k-1} \binom{l-1}{k} S_l \\ &= S_k + \sum_{l=k+1}^n (-1)^{l-k} \left(\binom{l}{k} - \binom{l-1}{k} \right) S_l \\ &= S_k + \sum_{l=k+1}^n (-1)^{l-k} \binom{l-1}{k-1} S_l \\ &= \sum_{l=k}^n (-1)^{l-k} \binom{l-1}{k-1} S_l \quad \text{da } \binom{l}{k} - \binom{l-1}{k} = \binom{l-1}{k-1} \end{aligned}$$

□

2.8 Beispiel.

(1) Wie groß ist die Wahrscheinlichkeit, beim Skat nach dem Austeilen 3 Buben zu bekommen?

Modell: $\Omega = \mathcal{K}_{10}^{\{1, \dots, 32\}}$, $\mathfrak{A} = \mathbb{P}(\Omega)$, $P = \mathcal{L}_{\Omega}$. Buben: 1, 2, 3, 4 und für $1 \leq j \leq 4$ sei

$$A_j = \{\omega \in \Omega \mid j \in \omega\} \hat{=} \text{„man bekommt Bube } j\text{“}$$

Gesucht ist dann die Wahrscheinlichkeit von $B_3 \hat{=} \text{„genau drei Buben“}$ und $C_3 \hat{=} \text{„mindestens drei Buben“}$. Es ist

$$S_3 = \sum_{\substack{J \subset \{1, \dots, 4\} \\ \#J=3}} P\left(\bigcap_{j \in J} A_j\right)$$

Das Ereignis, 3 bestimmte Buben zu bekommen, ist isomorph zu $\mathcal{K}_7^{\{1, \dots, 29\}}$ (3 Buben fest, bleiben 7 Karten aus 29). Für $J \subset \{1, \dots, 4\}$ und $\#J = 3$ gilt dann

$$P\left(\bigcap_{j \in J} A_j\right) = \frac{\#\mathcal{K}_7^{\{1, \dots, 29\}}}{\#\Omega} = \frac{\binom{29}{7}}{\binom{32}{10}} = \frac{3}{124}$$

$$\Rightarrow S_3 = \frac{3}{124} \sum_{\substack{J \subset \{1, \dots, 4\} \\ \#J=3}} 1 = \frac{3}{124} \#\mathcal{K}_3^{\{1, \dots, 4\}} = \frac{3}{124} \binom{4}{3} = \frac{12}{124} = \frac{3}{31}$$

$$S_4 = P\left(\bigcap_{j=1}^4 A_j\right) \stackrel{\text{Bsp. 2.3(1)}}{=} \frac{21}{3596}$$

$$\Rightarrow P(B_3) = \binom{3}{3} S_3 - \binom{4}{3} S_4 = \frac{66}{899} \approx 0,0734$$

$$P(C_3) = \binom{2}{2} S_3 - \binom{3}{2} S_4 = \frac{285}{3596} \approx 0,0793$$

- (2) Beim alphabetischen Sortieren einer Liste mit n Einträgen ist bei bestimmten Algorithmen (z. B. quicksort) die Rechendauer besonders groß, wenn viele Elemente richtig sortiert sind.

Modell: $\Omega = \mathcal{P}_n^{\{1, \dots, n\}}$, $\mathfrak{A} = \mathbb{P}(\Omega)$, $P = \mathcal{L}_\Omega$.

Seien

$$B_k = \{\omega \in \Omega \mid \omega_j = j \text{ für genau } k \text{ Zahlen } j \in \{1, \dots, n\}\}$$

$$A_j = \{\omega \in \Omega \mid \omega_j = j\}$$

$$A_J = \{\omega \in \Omega \mid \omega_j = j \text{ für alle } j \in J \subset \{1, \dots, n\}\}$$

In **Bsp. 2.6** haben wir gesehen, daß

$$S_l = \binom{n}{l} P(A_{\{1, \dots, l\}}) = \frac{1}{l!}$$

Nach **Satz 2.7(a)** ist dann

$$P(B_k) = \sum_{l=k}^n (-1)^{l-k} \binom{l}{k} \cdot \frac{1}{l!} = \sum_{l=k}^n (-1)^{l-k} \frac{1}{k!(l-k)!} = \frac{1}{k!} \sum_{k=0}^{n-k} \frac{(-1)^l}{l!} \approx \frac{1}{k!e}$$

ENDE VON KAPITEL 2

KAPITEL 3

Wahrscheinlichkeitsmaße mit Dichten auf $\mathbb{R}^d$

Im letzten Kapitel haben wir endliche bzw. abzählbar unendliche Wahrscheinlichkeitsräume betrachtet. Nach **Satz 1.23** sind die Wahrscheinlichkeitsmaße dann durch ihre Zähldichte eindeutig bestimmt, d. h. es gilt $P(A) = \sum_{\omega \in A} P\{\omega\}$.

Häufig treten aber auch Meßgrößen auf, die reelle Zahlen oder Vektoren im $\mathbb{R}^d$ sind. Wie in **Bsp. 1.25** gezeigt, gilt im Falle der kontinuierlichen Gleichverteilung auf $[0, 1]$, $\lambda^1\{\omega\} = 0$ für alle $\omega \in [0, 1]$. Die von uns jetzt betrachteten W-Maße auf $\mathbb{R}^d$ werden die Eigenschaft $P\{\omega\} = 0 \ \forall \omega \in \mathbb{R}^d$ besitzen (solche Maße heißen „diffus“). Gleichwohl werden auch diese Maße wieder mit Hilfe einer Funktion $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ definiert, ihrer *LEBESGUE-Dichte* im Unterschied zur Zähldichte aus **Kap. 1**. Wie in **Bsp. 1.25** müssen wir zuerst das Integral einer positiven meßbaren Funktion einführen, sowie eine kleinere σ -Algebra als die Potenzmenge. **Bsp. 1.25(1)** gilt hier entsprechend!

3.1 Definition.

(a) Für $a = (a_1, \dots, a_d)$, $b = (b_1, \dots, b_d) \in \mathbb{R}^d$ sei

$$(a, b] = \{x = (x_1, \dots, x_d) \mid a_j < x_j \leq b_j \text{ für alle } 1 \leq j \leq d\}$$

ein (halb)offener Quader.

(b) Die *BORELSche σ -Algebra auf $\mathbb{R}^d$* ist die kleinste σ -Algebra $\mathfrak{B}(\mathbb{R}^d)$, die alle halboffenen Quader $(a, b]$ mit $a, b \in \mathbb{R}^d$ enthält, d. h.

$$\mathfrak{B}(\mathbb{R}^d) = \mathfrak{A}(\{(a, b] \mid a, b \in \mathbb{R}^d\})$$

Jedes Element $E \in \mathfrak{B}(\mathbb{R}^d)$ nennen wir (BOREL)-meßbar oder *BOREL-Menge*.

(c) Für $\Omega \in \mathfrak{B}(\mathbb{R}^d)$ sei

$$\mathfrak{B}(\Omega) = \{A \in \mathfrak{B}(\mathbb{R}^d) \mid A \subset \Omega\}$$

die σ -Algebra der BORELSchen Mengen auf Ω .

3.2 Satz. $\mathfrak{B}(\mathbb{R}^d)$ enthält alle offenen und alle abgeschlossenen Teilmengen von $\mathbb{R}^d$.

BEWEIS. Es sei $A \subset \mathbb{R}^d$ offen und es sei

$$\mathfrak{M} = \{(a, b] \mid a, b \in \mathbb{Q}^d; (a, b] \subset A\}$$

Dann ist $\mathfrak{M}$ abzählbar und somit wegen **Def. 1.6(σA_3)**

$$A' := \bigcup_{Q \in \mathfrak{M}} Q \in \mathfrak{B}(\mathbb{R}^d)$$

Trivialerweise gilt $A' \subset A$. Umgekehrt: zu $x \in A \exists \varepsilon > 0$ mit

$$K_\varepsilon(x) = \{y \in \mathbb{R}^d \mid \|x - y\| < \varepsilon\} \subset A$$

In $K_\varepsilon(x)$ liegt ein $(a, b] \in \mathfrak{M}$ mit $x \in (a, b]$ und somit $x \in \bigcup_{Q \in \mathfrak{M}} Q = A'$, d. h. $A = A'$. Ist $B \subset \mathbb{R}^d$ abgeschlossen, so ist B^c offen und deshalb eine Borelmenge und wegen **Definition 1.6**(σA_2) $\Rightarrow B^c \in \mathfrak{B}(\mathbb{R}^d)$, $\xRightarrow{(\sigma A_2)} B = (B^c)^c \in \mathfrak{B}(\mathbb{R}^d)$. $\square$

3.3 Definition. Es sei $\Omega \neq \emptyset$ und $\mathfrak{A}$ eine σ -Algebra auf Ω . Eine Abbildung

$$\mu : \mathfrak{A} \rightarrow \mathbb{R}_+ \cup \{\infty\} \quad \text{mit} \quad \mu(\emptyset) = 0 \quad \text{und} \quad \mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

für jede disjunkte Folge $(A_n)_n \subset \mathfrak{A}$ heißt *Maß* auf Ω (vgl. **Def. 1.17!**)

3.4 Konvention (Rechnen mit ∞). Auf $\mathbb{R}_+ \cup \{\infty\}$ gilt:

$$\begin{aligned} a + \infty &= \infty + a = \infty && \text{für alle } a \in \mathbb{R}_+ \\ a \cdot \infty &= \infty \cdot a = \infty && \text{für alle } a > 0 \\ 0 \cdot \infty &= \infty \cdot 0 = 0 \end{aligned}$$

3.5 Bemerkung. Die für W-Maße in **Satz 1.20** bewiesenen Aussagen **(a)**, **(b)**, die Monotonieaussage $\mu(A) \leq \mu(B)$ für $A \subset B$ sowie **(d)** bleiben für allgemeine Maße gültig. Bei der Stetigkeit von oben (**Satz 1.20(e)**) ist zusätzlich $\mu(A_0) < \infty$ zu fordern.

Das für uns wichtigste Maß auf $\mathbb{R}^d$ ist das **LEBESGUE-Maß** λ^d . Die Existenz wird in Stochastik II bewiesen.

3.6 Faktum. Es gibt auf $(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d))$ ein eindeutig bestimmtes Maß λ^d mit

$$\lambda^d((a, b]) = \prod_{j=1}^d (b_j - a_j)$$

für alle Quader mit $a_j < b_j$ für $1 \leq j \leq d$. Dieses Maß λ^d wird **LEBESGUE(-BOREL)sches Maß** genannt.

3.7 Beispiel.

(1) Im Fall $d = 1$ mißt λ^1 die Gesamtlänge einer Teilmenge, λ^2 die Fläche, λ^3 das Volumen.

(2) Es sei $\Omega \in \mathfrak{B}(\mathbb{R}^d)$ mit $0 < \lambda^d(\Omega) < \infty$. Dann ist auf $(\Omega, \mathfrak{B}(\Omega))$ ein W-Maß

$$P : \mathfrak{B}(\Omega) \rightarrow [0, 1] \quad \text{durch} \quad P(A) = \frac{\lambda^d(A)}{\lambda^d(\Omega)}$$

definiert.

[(**WM**₁) folgt aus der entsprechenden Eigenschaft von λ^d und (**WM**₂) gilt wegen $P(\Omega) = \frac{\lambda^d(\Omega)}{\lambda^d(\Omega)} = 1]$

P heißt *kontinuierliche Gleichverteilung* auf Ω . Den Fall $d = 1$, $\Omega = [0, 1]$ hatten wir schon in **Beispiel 1.25** betrachtet.

(3) **Quadratische Zielscheibe**

Es sei $\Omega = (-1, 1] \times (-1, 1]$ eine quadratische Zielscheibe. Ist die Treff-Fläche durch

$$B = (-a, a] \times (-a, a]$$

mit $0 < a \leq 1$ gegeben, so ist bei gleichverteilten Treffern auf Ω die Wahrscheinlichkeit, einen Treffer in B zu landen

$$P(B) = \frac{\lambda^2(B)}{\lambda^2(\Omega)} = \frac{2a \cdot 2a}{2 \cdot 2} = a^2$$

Um eine weiter große Klasse von W-Maßen auf $\mathbb{R}^d$ zu definieren, müssen wir das Integral einer nichtnegativen integrierbaren Funktion definieren. Dazu benötigen wir:

3.8 Definition. Es sei $\mathfrak{A}$ eine σ -Algebra auf Ω . Eine Funktion $f : \Omega \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ heißt *meßbar*, wenn:

$$\forall \alpha \in \mathbb{R} : \quad \{\omega \in \Omega \mid f(\omega) < \alpha\} \in \mathfrak{A}$$

3.9 Beispiel.

(1) Im Fall $\Omega = \mathbb{R}^d$, $\mathfrak{A} = \mathfrak{B}(\mathbb{R}^d)$ ist jede stetige Funktion $f : \mathbb{R}^d \rightarrow \mathbb{R}$ meßbar.

BEWEIS. $\{\omega \in \Omega \mid f(\omega) < \alpha\} = f^{-1}((-\infty, \alpha))$ offen, da $(-\infty, \alpha) \subset \mathbb{R}$ offen und nach **Satz 3.2** somit in $\mathfrak{B}(\mathbb{R}^d)$. $\square$

(2) Sind $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ meßbar und $c \in \mathbb{R}$, so sind $c \cdot f : \omega \mapsto c \cdot f(\omega)$ und $f + g : \omega \mapsto f(\omega) + g(\omega)$ meßbar. Die Menge aller meßbaren Funktionen f ist also ein $\mathbb{R}$ -Vektorraum.

BEWEIS. Klar für $c \cdot f$. Für $f + g$ folgt dies aus

$$\{\omega \in \Omega \mid f(\omega) + g(\omega) < \alpha\} = \bigcup_{\beta \in \mathbb{Q}} \{\omega \in \Omega \mid f(\omega) < \beta\} \cap \{\omega \in \Omega \mid g(\omega) < \alpha - \beta\}$$

„ $\supset$ “ klar! „ $\subset$ “: Sei $f(\omega) + g(\omega) < \alpha$ und $\beta \in \mathbb{Q}$ mit $f(\omega) < \beta < \alpha - g(\omega)$ gewählt. Nach Voraussetzung gilt $f(\omega) < \alpha - g(\omega)$. Daraus folgt:

$$\omega \in \{\omega \in \Omega \mid f(\omega) < \beta\} \cap \{\omega \in \Omega \mid g(\omega) < \alpha - \beta\}$$

$\square$

(3) Die Indikatorfunktion $1_A : \Omega \rightarrow \mathbb{R}$ ist für $A \subset \Omega$ durch

$$1_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \in A^c \end{cases}$$

definiert. 1_A ist genau dann meßbar, wenn $A \in \mathfrak{A}$.

BEWEIS.

$$\{\omega \in \Omega \mid 1_A(\omega) < \alpha\} = \begin{cases} \emptyset & \text{falls } \alpha \leq 0 \\ A^c & \text{falls } 0 < \alpha \leq 1 \\ \Omega & \text{falls } \alpha > 1 \end{cases}$$

$\square$

(4) Jede Funktion der Gestalt

$$g = \sum_{j=1}^n \alpha_j 1_{A_j} \quad \text{mit} \quad \alpha_j \in \mathbb{R}_+, A_j \in \mathfrak{A}$$

ist nach **Bsp. 3.9(2)** und **3.9(3)** meßbar. Funktionen dieser Gestalt heißen *meßbare Elementarfunktionen*. Die Menge aller solcher Funktionen wird mit $\mathcal{E}(\mathfrak{A})$ bezeichnet. Elementarfunktionen sind gerade die meßbaren Funktionen $g \geq 0$, die nur endlich viele Werte annehmen.

(5) Es sei (f_n) eine Folge meßbarer Funktionen. Dann sind auch $\inf_{n \in \mathbb{N}} f_n$ und $\sup_{n \in \mathbb{N}} f_n$ meßbar. Denn:

$$\{\omega \in \Omega \mid \inf_{n \in \mathbb{N}} f_n(\omega) < \alpha\} = \bigcup_{n \in \mathbb{N}} \{\omega \in \Omega \mid f_n(\omega) < \alpha\} \in \mathfrak{A} \quad \text{vgl. Übungsaufgabe}$$

3.10 Definition (Integral von meßbaren Elementarfunktionen). Für $g \in \mathcal{E}(\mathfrak{A})$ und ein Maß μ auf $(\Omega, \mathfrak{A})$ definieren wir

$$\int g d\mu := \int_{\Omega} g(\omega) d\mu(\omega) := \sum_{x \in \mathbb{R}} x \cdot \mu\{\omega \in \Omega \mid g(\omega) = x\}$$

Diese Summe hat nur endlich viele Summanden!

3.11 Beispiel.

(1) Für $A \in \mathfrak{A}$ gilt

$$\int 1_A d\mu = 0 \cdot \mu(A^c) + 1 \cdot \mu(A) = \mu(A)$$

Es folgt leicht, daß die Elementarfunktion $g = \sum_{j=1}^n \alpha_j 1_{A_j}$ das Integral

$$\int g d\mu = \sum_{j=1}^n \alpha_j \mu(A_j)$$

hat. [Trivial für disjunkte A_j und dann durch Zerlegung in disjunkte Bestandteile.]

(2) $\int c \cdot g d\mu = c \cdot \int g d\mu$ für $g \in \mathcal{E}(\mathfrak{A})$, $c \geq 0$.

(3) $\int g_1 + g_2 d\mu = \int g_1 d\mu + \int g_2 d\mu$ für $g_1, g_2 \in \mathcal{E}(\mathfrak{A})$.

(4) $\int g_1 d\mu \leq \int g_2 d\mu$, falls $g_1, g_2 \in \mathcal{E}(\mathfrak{A})$ mit $g_1(\omega) \leq g_2(\omega)$ für alle $\omega \in \Omega$.

[Dies folgt trivialerweise aus der Definition und **Bsp. 3.11(1)**.]

3.12 Definition (Integral meßbarer Funktionen). Es sei $f : \Omega \rightarrow \mathbb{R}_+ \cup \{\infty\}$ meßbar. Das Integral von f bzgl. μ ist definiert durch

$$\int f d\mu := \int_{\Omega} f(\omega) d\mu(\omega) := \sup_{\substack{g \in \mathcal{E}(\mathfrak{A}) \\ g \leq f}} \int g d\mu$$

3.13 Bemerkung. Ist $f \in \mathcal{E}(\mathfrak{A})$, so ergibt sich durch die doppelte Definition von $\int f d\mu$ natürlich kein Widerspruch: Wegen **Bsp. 3.11(4)** gilt für $f \in \mathcal{E}(\mathfrak{A})$

$$\sup_{\substack{g \in \mathcal{E}(\mathfrak{A}) \\ g \leq f}} \int g d\mu \leq \int f d\mu$$

und „ $=$ “ gilt, da $g = f$ im Supremum vorkommt. Es folgt $\int f_1 d\mu \leq \int f_2 d\mu$ für meßbare f_1, f_2 mit $f_1 \leq f_2$.

Die **Definition 3.12** ist für die praktische Berechnung des Integrals meist unbrauchbar. Um eine praktikablere Berechnungsmethode im Falle $\mu = \lambda^d$ angeben zu können, benötigen wir:

3.14 Satz (BEppo-LEVI, monotone Konvergenz). Es seien $f_n : \Omega \rightarrow \mathbb{R}_+ \cup \{\infty\}$ meßbare Funktionen mit $f_n \leq f_{n+1}$ sowie $f = \sup_{n \in \mathbb{N}} f_n$. Dann gilt

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

BEWEIS. „ $\leq$ “: Es sei $n \in \mathbb{N}$ und $g \in \mathcal{E}(\mathfrak{A})$ mit $g \leq f_n$. Dann gilt $g \leq f$ und nach **Bem. 3.13** $\int g d\mu \leq \int f d\mu$. Daraus folgt

$$\int f_n d\mu \leq \int f d\mu \quad \Rightarrow \quad \lim_{n \rightarrow \infty} \int f_n d\mu \leq \int f d\mu$$

„ $\geq$ “:

1. Fall $\int f d\mu < \infty$:

Sei $\varepsilon > 0$. Wir zeigen, daß es ein $n_0 \in \mathbb{N}$ mit $\int f_{n_0} d\mu > \int f d\mu - \varepsilon$ gibt. Daraus folgt:

$$\Rightarrow \forall n \geq n_0 \quad \int f_n d\mu \geq \int f_{n_0} d\mu > \int f d\mu - \varepsilon \quad \Rightarrow \quad \text{„}\geq\text{“}.$$

Nach Definition von $\int f d\mu$ existiert ein $g \in \mathcal{E}(\mathfrak{A})$, $g \leq f$ mit

$$\int g d\mu \geq \int f d\mu - \frac{\varepsilon}{3}.$$

Da $\int g d\mu \leq \int f d\mu < \infty$, ist nach **Definition 3.10** $\mu\{\omega \mid g(\omega) > 0\} < \infty$. Wähle $\delta > 0$ mit $\delta \mu\{\omega \mid g(\omega) > 0\} < \frac{\varepsilon}{3}$ und setze $h := g \cdot 1_{\{\omega \mid g(\omega) > \delta\}}$

$$\begin{aligned} \Rightarrow g(\omega) - h(\omega) &= g(\omega) - g(\omega) \cdot 1_{\{\omega \mid g(\omega) > \delta\}} \\ &= \begin{cases} 0 & \text{falls } g(\omega) = 0 \\ g(\omega) & \text{falls } 0 < g(\omega) \leq \delta \\ 0 & \text{falls } g(\omega) > \delta \end{cases} \leq \delta 1_{\{\omega \mid g(\omega) > \delta\}} \end{aligned}$$

$$\begin{aligned} \Rightarrow \int (g - h) d\mu &\leq \delta \mu\{\omega \mid g(\omega) > 0\} < \frac{\varepsilon}{3} \\ \Rightarrow \int h d\mu &> \int g d\mu - \frac{\varepsilon}{3} \end{aligned}$$

Sei nun $A_n = \{\omega \in \Omega \mid f_n(\omega) < h(\omega)\}$. Wegen $f_n \uparrow f(\omega) \geq h(\omega)$, falls $f(\omega) > 0$, gilt $A_n \downarrow \emptyset$ und wegen der Stetigkeit von oben und

$$\mu(A_0) \leq \mu\{\omega \mid h(\omega) > 0\} \leq \mu\{\omega \mid g(\omega) > 0\} < \infty$$

folgt $\lim_{n \rightarrow \infty} \mu(A_n) = 0$. Es sei $M := \max_{\omega \in \Omega} h(\omega)$. Dann gibt es ein n_0 mit $\mu(A_{n_0}) < \frac{\varepsilon}{3M}$. Daraus folgt

$$\begin{aligned} \int f_{n_0} d\mu &\geq \int f_{n_0} 1_{\Omega \setminus A_{n_0}} d\mu \geq \int \underbrace{h(1 - 1_{A_{n_0}})}_{=1_{\Omega \setminus A_{n_0}}} d\mu \\ &= \int h d\mu - \int h 1_{A_{n_0}} d\mu > \int g d\mu - \frac{\varepsilon}{3} - \int M 1_{A_{n_0}} d\mu \\ &> \int f d\mu - \frac{2\varepsilon}{3} - M \cdot \frac{\varepsilon}{3M} = \int f d\mu - \varepsilon \end{aligned}$$

2. Fall $\int f d\mu = \infty$: Hier sind zwei Fälle möglich:

- (i) $\exists g \in \mathcal{E}(\mathfrak{A})$ mit $g \leq 1$ und $\int g d\mu = \infty \Rightarrow \exists \varepsilon > 0$, $A \in \mathfrak{A}$ mit $\varepsilon 1_A \leq f$ und $\mu(A) = \infty$. Sei $A_n = A \cap \{\omega \mid f_n(\omega) \geq \frac{\varepsilon}{2}\}$. Da $f_n \uparrow f$ folgt mit der Stetigkeit von unten $\mu(A_n) \rightarrow \infty$ und damit

$$\int f_n d\mu \geq \int \frac{\varepsilon}{2} 1_{A_n} d\mu = \frac{\varepsilon}{2} \mu(A_n) \rightarrow \infty$$

- (ii) Für alle $C > 0 \exists g \in \mathcal{E}(\mathfrak{A})$ mit $C \leq \int g d\mu < \infty$ und dann folgt wie im Fall von endlichem $\int f d\mu$ mit $A_n = \{\omega \mid f_n(\omega) > \frac{1}{2}g(\omega)\}$, daß gilt:

$$\lim_{n \rightarrow \infty} \int f_n d\mu \geq \frac{C}{2} \quad \text{für jedes } C > 0$$

□

3.15 Lemma. Für jede meßbare Funktion $f \geq 0$ existiert eine Folge $f_n \in \mathcal{E}(\mathfrak{A})$ mit $f_n \leq f_{n+1} \rightarrow f$.

BEWEIS. Setze

$$f_n(x) = \begin{cases} \frac{k}{2^n} & \text{falls } \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \leq n \\ n & \text{falls } f(x) \geq n. \end{cases}$$

□

3.16 Korollar (Eigenschaften des Integrals).

(1) $\int c \cdot f \, d\mu = c \cdot \int f \, d\mu$ für $c \geq 0$ und $f \geq 0$ meßbar.

(2) $\int f_1 + f_2 \, d\mu = \int f_1 \, d\mu + \int f_2 \, d\mu$ für $f_1, f_2 \geq 0$ meßbar.

BEWEIS. Sind nämlich $g_n \uparrow f_1, h_n \uparrow f_2$ Folgen von Elementarfunktionen nach Lemma 3.15, so gilt $(g_n + h_n) \uparrow (f_1 + f_2)$ und mit BEPPO-LEVI (3.14)

$$\begin{aligned} \int f_1 + f_2 \, d\mu &\stackrel{\text{BL}}{=} \lim_{n \rightarrow \infty} \int g_n + h_n \, d\mu = \lim_{n \rightarrow \infty} \int g_n \, d\mu + \lim_{n \rightarrow \infty} \int h_n \, d\mu \\ &\stackrel{\text{BL}}{=} \int f_1 \, d\mu + \int f_2 \, d\mu \end{aligned}$$

□

3.17 Satz. f sei eine stückweise stetige Funktion von $[a, b]$ nach $\mathbb{R}_+$. Dann gilt

$$\begin{aligned} \int 1_{[a,b]} f \, d\lambda^1 &= \int 1_{(a,b]} f \, d\lambda^1 = \int 1_{[a,b)} f \, d\lambda^1 \\ &= \int 1_{(a,b)} f \, d\lambda^1 = \int_a^b f(x) \, dx, \end{aligned}$$

wobei letzteres als RIEMANN-Integral zu verstehen ist.

BEWEIS. Die ersten vier Integrale unterscheiden sich nur um $1_{\{a\}}$ (bzw. b) und $\int 1_{\{a\}} f \, d\lambda^1 = f(a)\lambda^1\{1\} = 0$. Es reicht ohne Einschränkung, den Fall $f : [a, b] \rightarrow \mathbb{R}_+$ stetig zu betrachten (sonst Zerlegung in Teilintervalle). Für $n \geq 1$ sei $t_{n,k} = a + k \frac{b-a}{2^n}$ ($0 \leq k \leq 2^n$) eine Zerlegung von $[a, b]$ und

$$\alpha_{n,k} = \min_{x \in [t_{n,k}, t_{n,k+1}]} f(x) \quad \text{für } 0 \leq k \leq 2^n - 1.$$

Dann ist

$$f_n = \sum_{k=0}^{2^n-1} \alpha_{n,k} 1_{(t_{n,k}, t_{n,k+1}]}]$$

eine Elementarfunktion mit

$$\int f_n \, d\lambda^1 = \underbrace{\sum_{k=0}^{2^n-1} \alpha_{n,k} \frac{b-a}{2^n}}_{\text{Untersumme}} \xrightarrow{n \rightarrow \infty} \int_a^b f(x) \, dx.$$

Andererseits gilt $f_n \leq f_{n+1}$ und da f stetig ist gilt $f_n(x) \rightarrow f(x)$ für alle $x \in (a, b]$. Also ist nach BEPPO-LEVI (3.14)

$$\int 1_{(a,b]} f \, d\lambda^1 = \lim_{n \rightarrow \infty} \int f_n \, d\lambda^1 = \int_a^b f(x) \, dx$$

□

3.18 Bemerkung. Ist $f : \mathbb{R} \rightarrow \mathbb{R}_+$ stückweise stetig, so gilt

$$\int f \, d\lambda^1 = \int_{-\infty}^{\infty} f(x) \, dx.$$

Wir sind nun in der Lage, die angekündigten Maße mit Dichten zu definieren:

3.19 Satz. Es sei $f : \mathbb{R}^d \rightarrow \mathbb{R}_+ \cup \{\infty\}$ eine meßbare Funktion mit $\int f d\lambda^d = 1$. Dann wird durch

$$P(A) = \int 1_A f d\lambda^d \quad A \in \mathfrak{B}(\mathbb{R}^d)$$

ein W-Maß P auf $\mathbb{R}^d$ definiert.

BEWEIS. Es ist $P(\mathbb{R}^d) = \int 1_{\mathbb{R}^d} f d\lambda^d = \int f d\lambda^d = 1$. Die σ -Additivität folgt aus dem Satz von der monotonen Konvergenz: Ist $(A_n) \subset \mathfrak{B}(\mathbb{R}^d)$ eine disjunkte Folge und $A = \bigcup_{n \in \mathbb{N}} A_n$, so setzt man

$$f_m = 1_{\bigcup_{n=0}^m A_n} f = \sum_{n=0}^m 1_{A_n} f \quad \Rightarrow \quad f_m \uparrow 1_A \cdot f.$$

Dann folgt aus **Satz 3.14**

$$\begin{aligned} P\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= P(A) = \int 1_A f d\lambda^d \stackrel{\text{BL}}{=} \lim_{m \rightarrow \infty} \int f_m d\lambda^d \\ &= \lim_{m \rightarrow \infty} \sum_{n=0}^m \int 1_{A_n} f d\lambda^d = \lim_{m \rightarrow \infty} \sum_{n=0}^m P(A_n) = \sum_{n=0}^{\infty} P(A_n). \end{aligned}$$

□

Die Funktion f nennen wir LEBESGUE-Dichte von P . Sämtliche W-Maße, die man so erhält sind diffus, d. h. für alle $\omega \in \mathbb{R}^d$ gilt

$$P\{\omega\} = \int 1_{\{\omega\}} f d\lambda^d = f(\omega) \lambda^d\{\omega\} = 0$$

3.20 Beispiel.

(a) Es sei $\vartheta > 0$ und $f : \mathbb{R} \rightarrow \mathbb{R}_+$;

$$f(x) = \begin{cases} \vartheta e^{-\vartheta x} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad \Leftrightarrow \quad f(x) = \vartheta e^{-\vartheta x} \cdot 1_{[0, \infty)}$$

$$\Rightarrow \int f d\lambda^1 = \int_0^{\infty} \vartheta e^{-\vartheta x} dx = -e^{-\vartheta x} \Big|_0^{\infty} = 1$$

Das mit dieser Dichte definierte W-Maß heißt *Exponentialverteilung* mit Parameter ϑ und wird mit $\mathfrak{E}_{\vartheta}$ bezeichnet. Häufig wird $\mathfrak{E}_{\vartheta}$ als Modell für die Lebensdauer eines Bauteils verwendet: Für $T \geq 0$ ist

$$\mathfrak{E}_{\vartheta}([T, \infty)) = \int_T^{\infty} \vartheta e^{-\vartheta x} dx = e^{-\vartheta T}$$

die Wahrscheinlichkeit, daß das Bauteil mindestens bis zum Zeitpunkt T funktioniert.

(b) GAUSSsche Normalverteilung mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Dichte:

$$f(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Bezeichnung $\mathcal{N}_{\mu, \sigma^2}$. Zu zeigen ist noch $\int f d\lambda^1 = 1$ (Übung).

Leider läßt sich die Wahrscheinlichkeit von Intervallen nicht mehr so einfach bestimmen, da die Stammfunktion

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx$$

nicht in expliziter Form darstellbar ist. Es ist $\mathcal{N}_{0,1}((-\infty, t]) = \Phi(t)$. Eigenschaften:

$$(i) \Phi(t) = 1 - \Phi(-t) \quad (\text{Übung})$$

$$(ii) \mathcal{N}_{\mu, \sigma^2}([a, b]) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \quad (\text{Übung})$$

Es reicht daher, die Funktion Φ für Werte $t \geq 0$ zu tabellieren. Die GAUSSsche Normalverteilung ist eine der wichtigsten Verteilungen; sie tritt u. a. bei der Modellierung von Meßfehlern auf.

(c) CAUCHY-Verteilung: Es sei

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Dann gilt

$$\int f d\lambda^1 = \int_{-\infty}^{\infty} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi} \arctan x \Big|_{-\infty}^{\infty} = 1,$$

also ist f eine LEBESGUE-Dichte.

W-Maße auf $\mathbb{R}$ können auch durch den (älteren) Begriff der *Verteilungsfunktion* beschrieben werden:

3.21 Definition. Es sei P ein W-Maß auf $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. Dann heißt die Funktion

$$F_P(x) := P((-\infty, x]), \quad x \in \mathbb{R}$$

die *Verteilungsfunktion* von P .

3.22 Eigenschaften.

(a) F_P ist monoton wachsend und

$$P((a, b]) = F_P(b) - F_P(a)$$

bei $a < b$.

(b) F_P ist stetig von rechts.

BEWEIS. $x_n \downarrow x$

$$\Rightarrow \lim_{n \rightarrow \infty} F_P(x_n) = \lim_{n \rightarrow \infty} P((-\infty, x_n]) = P\left(\bigcap_{n \geq 1} (-\infty, x_n]\right) = P((-\infty, x]) = F_P(x)$$

□

(c) F_P besitzt linksseitige Grenzwerte.

BEWEIS. $x_n \uparrow x$

$$\Rightarrow \lim_{n \rightarrow \infty} F_P(x_n) = \lim_{n \rightarrow \infty} P((-\infty, x_n]) = P\left(\bigcup_{n \geq 1} (-\infty, x_n]\right) = P((-\infty, x)) \neq F_P(x)$$

□

3.23 Beispiel.

(a) Für $\vartheta > 0$ und $\mathfrak{E}_\vartheta$ gilt

$$F_{\mathfrak{E}_\vartheta}(x) = \mathfrak{E}_\vartheta((-\infty, x]) = \int_{-\infty}^x 1_{[0, \infty)}(t) \vartheta e^{-\vartheta t} dt = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\vartheta x} & x > 0 \end{cases}$$

(b) Die Verteilungsfunktion von $\mathcal{N}_{0,1}$ ist Φ ; allgemein gilt für $P := \mathcal{N}_{\mu, \sigma^2}$, $F_P(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$.

- (c) Sei $y \in \mathbb{R}$. Ein Modell für das sichere Eintreten von y ist durch den W-Raum $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \mathcal{E}_y)$ gegeben, wobei

$$\mathcal{E}_y(A) = \begin{cases} 1 & y \in A \\ 0 & y \notin A \end{cases}$$

Es gilt

$$F_{\mathcal{E}_y}(x) = \mathcal{E}_y((-\infty, x]) = \begin{cases} 0 & x < y \\ 1 & x \geq y \end{cases}$$

$\mathcal{E}_y$ heißt DIRAC-Verteilung oder Punktmaß in y .

Um auch Maße auf $\mathbb{R}^d$ für $d \geq 2$ berechnen zu können, benötigen wir ein maßtheoretisches Hilfsmittel:

3.24 Faktum (Satz von Fubini). Für jede meßbare Funktion $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ gilt:

$$\int_{\mathbb{R}^d} f(x) d\lambda^d(x) = \int_{\mathbb{R}} \int_{\mathbb{R}^{d-1}} f(y, z) d\lambda^{d-1}(y) d\lambda^1(z).$$

3.25 Beispiel.

(a) **Zweidimensionale Normalverteilung mit Korrelation ϱ :**

Es sei $-1 < \varrho < 1$ und $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)}(x_1^2 - 2\varrho x_1 x_2 + x_2^2)\right).$$

Daß $\int f d\lambda^2 = 1$ folgt aus dem Satz von Fubini:

$$\int f d\lambda^2 = \iint f(x_1, x_2) d\lambda^1(x_1) d\lambda^1(x_2).$$

Nun ist

$$\begin{aligned} \int f(x_1, x_2) d\lambda^1(x_1) &= \frac{1}{2\pi\sqrt{1-\varrho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\varrho^2)}(x_1^2 - 2\varrho x_1 x_2 + x_2^2)\right) dx_1 \\ &= \frac{1}{2\pi\sqrt{1-\varrho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x_1 - \varrho x_2)^2}{2(1-\varrho^2)} - \frac{(1-\varrho^2)x_2^2}{2(1-\varrho^2)}\right) dx_1 \\ &= \frac{1}{2\pi\sqrt{1-\varrho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\varrho^2)}(x_1 - \varrho x_2)^2\right) dx_1 \exp\left(-\frac{1}{2}x_2^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{2\pi\sqrt{1-\varrho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x_1 - \varrho x_2)^2}{2(1-\varrho^2)}\right) dx_1 \exp\left(-\frac{1}{2}x_2^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_2^2\right) \mathcal{N}_{\varrho x_2, 1-\varrho^2}(\mathbb{R}) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_2^2\right) \\ \Rightarrow \int f d\lambda^2 &= \iint f(x_1, x_2) d\lambda^1(x_1) d\lambda^1(x_2) \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_2^2\right) dx_2 = \mathcal{N}_{0,1}(\mathbb{R}) = 1 \end{aligned}$$

Wir werden später ausrechnen, daß der Korrelationskoeffizient von P mit Dichte f tatsächlich ϱ ist.

(b) **BUFFONSches Nadelproblem:**

Eine Nadel der Länge ℓ wird auf einen Boden geworfen, auf dem parallele Linien im Abstand $L \geq \ell$ gezeichnet sind. Wie groß ist die Wahrscheinlichkeit, eine Linie zu treffen?

Wir beschreiben die Lage der Nadel durch den Winkel $\varphi \in [0, \pi)$ zwischen der Nadel und den Linien und den Abstand $a \in [0, \frac{\ell}{2}]$ zwischen dem Mittelpunkt der Nadel und der nächstgelegenen Linie (vgl. Abb. 3.1).

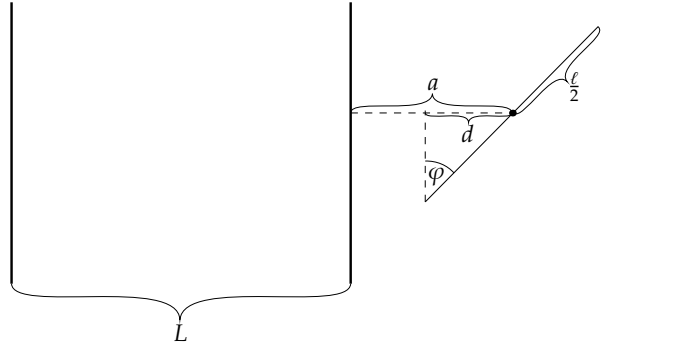


Abbildung 3.1: Zum BUFFONSchen Nadelproblem.

$$\sin \varphi = \frac{d}{\frac{\ell}{2}} \Leftrightarrow d = \frac{\ell}{2} \sin \varphi$$

Schnitt $\Leftrightarrow d \geq a \Leftrightarrow a \leq d = \frac{\ell}{2} \sin \varphi$. **Modell:** $\Omega = [0, \frac{\ell}{2}] \times [0, \pi[\subset \mathbb{R}^2$ und P sei die Gleichverteilung auf Ω , d. h.

$$P(A) = \frac{\lambda^2(A)}{\lambda^2(\Omega)} = \frac{2}{\pi L} \lambda^2(A).$$

Das Ereignis „Schnitt“ ist dann gerade

$$\begin{aligned} A &:= \left\{ (a, \varphi) \in \Omega \mid a \leq \frac{\ell}{2} \sin \varphi \right\} \\ \Rightarrow P(A) &= \frac{2}{\pi L} \lambda^2(A) = \frac{2}{\pi L} \int 1_A(x) d\lambda^2(x) \\ &= \frac{2}{\pi L} \iint 1_A(a, \varphi) d\lambda^1(a) d\lambda^1(\varphi) \\ &= \frac{2}{\pi L} \int_0^\pi \int_0^{\frac{\ell}{2} \sin \varphi} da d\varphi = \frac{2}{\pi L} \int_0^\pi \frac{\ell}{2} \sin \varphi d\varphi \\ &= \frac{2}{\pi L} \frac{\ell}{2} \cdot 2 = \frac{2\ell}{\pi L}. \end{aligned}$$

Dieses Ergebnis wurde um 1850 von dem Züricher Astronom R. WOLF benutzt, um die Zahl π auf experimentellem Weg zu bestimmen. Er führte das Experiment 5000 mal aus und bestimmte die Näherung $\pi \approx 3,1596$.

3.26 BEWEIS VON Satz (2.7)(a). Das Ereignis

$$B_k := \bigcup_{\substack{K \subset \{1, \dots, n\} \\ \#K=k}} \bigcap_{j \in K} A_j \cap \bigcap_{j \notin K} A_j^c \quad (\text{disjunkte Vereinigung})$$

hat die Indikatorfunktion

$$\begin{aligned} 1_{B_k} &= \sum_{\substack{K \subset \{1, \dots, n\} \\ \#K=k}} \prod_{j \in K} 1_{A_j} \cdot \prod_{j \notin K} (1 - 1_{A_j}) = \sum_{\substack{K \subset \{1, \dots, n\} \\ \#K=k}} \prod_{j \in K} 1_{A_j} \cdot \sum_{l=0}^{n-k} (-1)^l \sum_{\substack{L \subset \{1, \dots, n\} \setminus K \\ \#L=l}} \prod_{j \in L} 1_{A_j} \\ &= \sum_{l=0}^{n-k} (-1)^l \sum_{\substack{K \subset \{1, \dots, n\} \\ \#K=k}} \sum_{\substack{L \subset \{1, \dots, n\} \setminus K \\ \#L=l}} \prod_{j \in K \cup L} 1_{A_j} \end{aligned}$$

Es gibt $\binom{k+l}{k}$ Möglichkeiten, die Menge $J = K \cup L$ aus einer k - und einer l -elementigen Menge zusammenzusetzen (wähle aus $k+l$ Elementen k aus).

$$\begin{aligned} \Rightarrow 1_{B_k} &= \sum_{l=0}^{n-k} (-1)^l \binom{k+l}{k} \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J = k+l}} \prod_{j \in J} 1_{A_j} \\ &= \sum_{l=k}^n (-1)^{l-k} \binom{l}{k} \cdot \sum_{\substack{J \subset \{1, \dots, n\} \\ \#J = l}} 1_{\cap_{j \in J} A_j} \end{aligned}$$

Integriert man diese Identität nach P , so folgt aus den üblichen Eigenschaften (3.11) die Behauptung. $\square$

ENDE VON KAPITEL 3

KAPITEL 4

Zufallsvariablen und Unabhängigkeit

Bisher haben wir immer Wahrscheinlichkeiten von Ereignissen berechnet. Häufig interessiert man sich für Zahlen, die von einem Merkmal abhängen, wie z. B. die Augensumme beim Werfen mehrerer Würfel, und nicht für die einzelnen Würfel insgesamt. Mathematisch wird dies durch den Begriff der Zufallsvariablen beschrieben.

4.1 Definition. Es sei $(\Omega, \mathfrak{A}, P)$ ein W-Raum, $\Omega' \neq \emptyset$ eine weitere Menge und $\mathfrak{A}' \subset \mathbb{P}(\Omega')$ eine σ -Algebra. Eine Abbildung $X : \Omega \rightarrow \Omega'$ heißt *Zufallsvariable* (mit Werten in Ω') genau dann, wenn gilt:

$$\forall A' \in \mathfrak{A}' : X^{-1}(A') := \{\omega \in \Omega \mid X(\omega) \in A'\} = \{X \in A'\} \in \mathfrak{A}$$

Will man die zugehörigen σ -Algebren angeben, sagt man X ist eine $\mathfrak{A} - \mathfrak{A}'$ Zufallsvariable (ZV) oder schreibt $X : (\Omega, \mathfrak{A}) \rightarrow (\Omega', \mathfrak{A}')$.

Im Fall $X : \Omega \rightarrow \mathbb{R}$, $\mathfrak{A}' = \mathfrak{B}(\mathbb{R})$ spricht man von *reellen ZV* und im Fall $X : \Omega \rightarrow \mathbb{R}^d$, $\mathfrak{A}' = \mathfrak{B}(\mathbb{R}^d)$ von einem *Zufallsvektor*.

4.2 Interpretation. Faßt man Ω als Menge von möglichen Zuständen eines Systems auf (die i. a. nicht direkt beobachtet werden können), so kann man den Wert $X(\omega)$ als Beobachtung einer vom unbekannten zufälligen Zustand ω abhängenden Meßgröße ansehen.

4.3 Beispiel.

(a) **Würfeln mit zwei Würfeln und Berechnung der Augensumme:**

$$\Omega = \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, \dots, 6\}\}, \quad P = \mathcal{L}_\Omega, \quad \mathfrak{A} = \mathbb{P}(\Omega);$$
$$X(\omega_1, \omega_2) = \omega_1 + \omega_2$$

Man kann $\Omega' = \mathbb{N}$ oder $\{2, \dots, 12\}$ oder $\mathbb{R}$ wählen. Die Bedingung $\{X \in A'\} \in \mathfrak{A}$ ist trivialerweise erfüllt, da $\mathfrak{A} = \mathbb{P}(\Omega)$.

(b) **n -facher Münzwurf:**

Modell: $\Omega = \{0, 1\}^n$ ($1 \triangleq \text{Zahl}$, $0 \triangleq \text{Wappen}$); $\mathfrak{A} = \mathbb{P}(\Omega)$ und $P = \mathcal{L}_\Omega$. Für $j \in \{1, \dots, n\}$ sei

$$X_j : \Omega \rightarrow \{0, 1\}, \quad X_j(\omega_1, \dots, \omega_n) = \omega_j$$

das Ergebnis des j -ten Versuchs und

$$Z(\omega_1, \dots, \omega_n) = \sum_{j=1}^n \omega_j$$

die Anzahl der „Einsen“. All das sind Zufallsvariablen, da $\mathfrak{A} = \mathbb{P}(\Omega)$.

(c) **Treffer auf einer Zielscheibe:**

Zielscheibe mit Radius 1 und n Ringen gleicher Breite konzentrisch aufgemalt. Wir interessieren uns für die Zufallsvariable X , die die Nummer des getroffenen Ringes angibt.

Modell: $\Omega = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$

$$R_j = \left\{ (x, y) \in \mathbb{R}^2 \mid \frac{j}{n} \leq \sqrt{x^2 + y^2} < \frac{j+1}{n} \right\} \quad \text{für } 0 \leq j \leq n-1$$

Dann kann man $X(\omega) = j$ für $\omega \in R_j$ setzen, d. h.

$$X = \sum_{j=0}^{n-1} j 1_{R_j}$$

Dies ist eine $\mathcal{B}(\mathbb{R})$ - $\mathbb{P}\{0, \dots, n-1\}$ -ZV, denn für $A' \in \mathbb{P}\{1, \dots, n-1\}$ ist

$$\begin{aligned} X^{-1}(A') &= \bigcup_{j \in A'} R_j \in \mathcal{B}(\Omega), \text{ da} \\ R_j &= \left\{ (x, y) \mid \sqrt{x^2 + y^2} < \frac{j+1}{n} \right\} \setminus \left\{ (x, y) \mid \sqrt{x^2 + y^2} < \frac{j}{n} \right\} \\ &\in \mathcal{B}(\Omega) \text{ als Differenz offener Mengen.} \end{aligned}$$

4.4 Eigenschaften.

(1) **Lemma.** Eine Abbildung $X: \Omega \rightarrow \mathbb{R}$ ist genau dann eine reelle ZV, wenn

$$\{\omega \in \Omega \mid X(\omega) < \alpha\} \in \mathfrak{A} \quad \text{für alle } \alpha \in \mathbb{R}.$$

BEWEIS. Ist X reelle ZV, so folgt wegen $(-\infty, \alpha) \in \mathcal{B}(\mathbb{R})$, daß $\{X < \alpha\} \in \mathfrak{A}$ für alle $\alpha \in \mathbb{R}$.

Sei umgekehrt

$$\{\omega \in \Omega \mid X(\omega) < \alpha\} = X^{-1}((-\infty, \alpha)) \in \mathfrak{A}$$

für alle $\alpha \in \mathbb{R}$. Es sei

$$\mathfrak{C} := \{C \in \mathcal{B}(\mathbb{R}) \mid X^{-1}(C) \in \mathfrak{A}\}.$$

Dann ist $\mathfrak{C}$ eine σ -Algebra, denn

$X^{-1}(\mathbb{R}) = \Omega \in \mathfrak{A} \Rightarrow \mathbb{R} \in \mathfrak{C}$. Für $C \in \mathfrak{C}$ gilt wegen $X^{-1}(C^c) = [X^{-1}(C)]^c \in \mathfrak{A}$, da $X^{-1}(C) \in \mathfrak{A} \Rightarrow C^c \in \mathfrak{C}$ und schließlich:

$$\begin{aligned} \text{Ist } (C_n)_n \subset \mathfrak{C} &\Rightarrow X^{-1}(C_n) \in \mathfrak{A} \quad \forall n \\ \Rightarrow \bigcup_{n \in \mathbb{N}} X^{-1}(C_n) &= X^{-1}\left(\bigcup_{n \in \mathbb{N}} C_n\right) \in \mathfrak{A} \quad \Rightarrow \bigcup_{n \in \mathbb{N}} C_n \in \mathfrak{C} \end{aligned}$$

□

Nach Voraussetzung enthält $\mathfrak{C}$ die Intervalle $(-\infty, \alpha)$, $\alpha \in \mathbb{R}$. Da $\mathfrak{C}$ eine σ -Algebra ist, folgt leicht, daß $\mathfrak{C}$ alle Intervalle der Form $(a, b]$ mit $a, b \in \mathbb{R}$ enthält. Nun ist $\mathcal{B}(\mathbb{R})$ die kleinste σ -Algebra, die $\{(a, b]\}$ enthält. Daraus folgt dann aber $\mathcal{B}(\mathbb{R}) \subset \mathfrak{C} \subset \mathcal{B}(\mathbb{R})$. Also ist für alle $A' \in \mathcal{B}(\mathbb{R}) = \mathfrak{C} \Rightarrow X^{-1}(A') \in \mathfrak{A}$, d. h. X ist eine reelle Zufallsvariable. □

(2) Sind X, Y reelle Zufallsvariablen, so auch $X + Y$ (s. **Bsp. 3.9(2)**). Ebenso $X - Y$, $\min(X, Y)$, $\max(X, Y)$ usw.

Jede reelle Zahl $\alpha \in \mathbb{R}$ identifiziert man mit der deterministischen Zufallsvariablen (entspricht einer konstanten Funktion) $\omega \mapsto \alpha$ von Ω in $\mathbb{R}$.

(3) **Kompositionsregel.** Ist $X : (\Omega, \mathfrak{A}) \rightarrow (\Omega', \mathfrak{A}')$ und $Y : (\Omega', \mathfrak{A}') \rightarrow (\Omega'', \mathfrak{A}'')$, so ist

$$Y \circ X : (\Omega, \mathfrak{A}) \rightarrow (\Omega'', \mathfrak{A}''); \quad Y \circ X : \omega \mapsto Y(X(\omega))$$

Für $A'' \in \mathfrak{A}'' \Rightarrow A' = Y^{-1}(A'') \in \mathfrak{A}' \Rightarrow X^{-1}(A') \in \mathfrak{A} \Rightarrow \{\omega \in \Omega \mid (Y \circ X)(\omega) \in A''\} \in \mathfrak{A}$

□

(4) Die Kompositionsregel (3) wird vor allem in der Situation angewendet, wo $Y := f : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion ist. Wir haben in **Bsp. 3.9(1)** gesehen, daß jede stetige Funktion meßbar ist.
 $\Rightarrow X : \Omega \rightarrow \mathbb{R}$ Zufallsvariable und $f : \mathbb{R} \rightarrow \mathbb{R}$ stetig $\Rightarrow f(X) : \Omega \rightarrow \mathbb{R}$ Zufallsvariable.

Insbesondere sind X^m ($m \in \mathbb{N}$), e^X , $\sin X$, $\frac{1}{X^2+1}$ usw. volle Zufallsvariablen.

4.5 Beispiel (Unendlicher Münzwurf). Modell: $\Omega = \{0, 1\}^{\mathbb{N}}$, $\mathfrak{A} := (\mathbb{P}\{0, 1\})^{\otimes \mathbb{N}}$, wobei diese σ -Algebra in **Bsp. 1.15** als die kleinste σ -Algebra auf Ω definiert wurde, zu der alle Ereignisse

$$Z_j(A) = \{(\omega_0, \omega_1, \dots) \mid \omega_j \in A\} \quad \text{mit } j \in \mathbb{N}, A \in \mathbb{P}\{0, 1\}$$

(„der j -te Wurf hat die Eigenschaft A “) gehören. Sei für $j \in \mathbb{N}$:

$$X_j : \Omega \rightarrow \{0, 1\}, \quad X_j(\omega_0, \omega_1, \dots) := \omega_j$$

(„Ergebnis der j -ten Wurfs“). Wegen

$$X_j^{-1}(A) = Z_j(A) \in \mathfrak{A} \text{ für alle } A \in \mathbb{P}\{0, 1\} \quad \text{ist}$$

$$X_j : (\Omega, \mathfrak{A}) \rightarrow (\{0, 1\}, \mathbb{P}\{0, 1\})$$

eine Zufallsvariable. Wir interessieren uns für die Zeit, in der erstmals „1“ auftritt:

$$T(\omega) = \inf\{n \in \mathbb{N} \mid X_n(\omega) = 1\}, \quad T : \Omega \rightarrow \mathbb{N} \cup \{\infty\},$$

wobei $\inf \emptyset = +\infty$ gesetzt wird. Daß T eine Zufallsvariable ist, sieht man mit **Lemma 4.4(1)**. Für $\alpha > 0$ wähle $n \in \mathbb{N}$ mit $n < \alpha \leq n+1$.

$$\begin{aligned} \Rightarrow \{\omega \in \Omega \mid T(\omega) < \alpha\} &= \{\omega \in \Omega \mid T(\omega) \leq n\} \\ &= \{\omega \in \Omega \mid X_j(\omega) = 1 \text{ für mindestens ein } j \leq n\} \\ &= \bigcup_{j=0}^n \{\omega \in \Omega \mid X_j(\omega) = 1\} = \bigcup_{j=0}^n Z_j(\{1\}) \in \mathfrak{A} \end{aligned}$$

4.6 Bezeichnungen. Ist $X : (\Omega, \mathfrak{A}) \rightarrow (\Omega', \mathfrak{A}')$ eine Zufallsvariable und $A' \in \mathfrak{A}'$, so bezeichne:

$$\{X \in A'\} := \{\omega \in \Omega \mid X(\omega) \in A'\}$$

das Ereignis „der Wert von X liegt in A' “. Nach **Def. 4.1** ist $\{X \in A'\}$, also tatsächlich ein Ereignis. Falls $\{X\} \in \mathfrak{A}'$, dann ist auch

$$\{X = x\} := \{\omega \in \Omega \mid X(\omega) = x\} \in \mathfrak{A}$$

und analog ist $\{X \neq x\}$ zu definieren. Sei nun $\Omega' = \mathbb{R}$ und $\mathfrak{A}' = \mathfrak{B}(\mathbb{R})$ und $X, Y : \Omega \rightarrow \mathbb{R}$ reelle Zufallsvariablen.

$$\Rightarrow (X, Y) : \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2$$

ist ein Zufallsvektor.

Für $a < c$ und $b < d$ gilt

$$\{(X, Y) \in ((a, b), (c, d))\} = \{X \in (a, c]\} \cap \{Y \in (b, d]\} \in \mathfrak{A} \quad \text{und}$$

da $\mathfrak{B}(\mathbb{R}^2)$ nach Definition die kleinste σ -Algebra ist, die alle Rechtecke der Form $((a, b), (c, d))$ enthält, folgt wie im Beweis von **Lemma 4.4(1)**, daß $\{(X, Y) \in B\} \in \mathfrak{A}$ für alle $B \in \mathfrak{B}(\mathbb{R}^2)$. $\square$

Daraus folgt

$$\{X = Y\} = \{\omega \in \Omega \mid X(\omega) = Y(\omega)\} \in \mathfrak{A},$$

da $D = \{(x, x) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$ abgeschlossen, also $D \in \mathfrak{B}(\mathbb{R}^2)$ und $\{X = Y\} = \{(X, Y) \in D\}$. Ebenso $\{X \neq Y\}$ und $\{X < Y\} \dots$

Die Werte, die eine Zufallsvariable $X : \Omega \rightarrow \Omega'$ annimmt, treten i. a. mit unterschiedlichen Wahrscheinlichkeiten auf, die wir durch eine Wahrscheinlichkeit auf $(\Omega', \mathfrak{A}')$ beschreiben können, der Verteilung von X . Dazu benötigen wir:

4.7 Satz. Es sei $X : (\Omega, \mathfrak{A}) \rightarrow (\Omega', \mathfrak{A}')$ eine ZV auf einem W-Raum $(\Omega, \mathfrak{A}, P)$. Dann wird durch

$$P_X : \mathfrak{A}' \rightarrow \mathbb{R}_+, \quad P_X(A') = P\{X \in A'\} = P\{\omega \in \Omega \mid X(\omega) \in A'\}$$

ein W-Maß auf $(\Omega', \mathfrak{A}')$ definiert.

BEWEIS. Trivialerweise gilt $P_X(A') \geq 0$ und

$$P_X(\Omega') = P\{\omega \in \Omega \mid X(\omega) \in \Omega'\} = P(\Omega) = 1$$

Sei (A'_n) eine disjunkte Folge in $\mathfrak{A}'$

$$\Rightarrow A_n := X^{-1}(A'_n) = \{\omega \in \Omega \mid X(\omega) \in A'_n\} \in \mathfrak{A}$$

ist disjunkte Folge in $\mathfrak{A}$.

$$\begin{aligned} \Rightarrow P_X\left(\bigcup_{n \in \mathbb{N}} A'_n\right) &= P\left\{X \in \bigcup_{n \in \mathbb{N}} A'_n\right\} = P\left(\bigcup_{n \in \mathbb{N}} \{X \in A'_n\}\right) \\ &= P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{\infty} P(A_n) = \sum_{n=0}^{\infty} P_X(A'_n) \end{aligned}$$

$\square$

4.8 Definition. Das in **Satz 4.7** eingeführte Wahrscheinlichkeitsmaß P_X auf $(\Omega', \mathfrak{A}')$ heißt *Verteilung von X unter P* .

4.9 Beispiel.

- (a) Um die Verteilung der Augensumme in **Bsp. 4.3(a)** zu beschreiben, genügt es (s. **Satz 1.23**), ihre Zähldichte anzugeben. Man rechnet leicht nach:

$$\begin{aligned} P_X\{k\} &= P\{X = k\} = P\{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 = k\} \\ &= \begin{cases} \frac{k-1}{36} & \text{für } k \leq 7 \\ \frac{13-k}{36} & \text{für } k \geq 7 \end{cases} \end{aligned}$$

P_X heißt *diskrete Dreiecksverteilung* auf $(\{2, \dots, 12\}, \mathbb{P}\{2, \dots, 12\})$.

- (b) Die Zähldichte der Verteilung der Nummer des getroffenen Rings im **Bsp. 4.3(c)** ist im Fall von auf der ganzen Scheibe gleichverteilten Treffern:

$$\begin{aligned} P_X\{j\} &= P\{X = j\} = P(R_j) = \frac{\lambda^2(R_j)}{\lambda^2(\Omega)} \\ &= \frac{1}{\pi} \pi \left(\left(\frac{j+1}{n} \right)^2 - \left(\frac{j}{n} \right)^2 \right) = \frac{2j+1}{n^2} \quad \text{für } j = 0, \dots, n-1 \end{aligned}$$

- (c) (vgl. **(4.4(3))**) Ist $X : (\Omega, \mathfrak{A}) \rightarrow (\Omega', \mathfrak{A}')$ und $Y : (\Omega', \mathfrak{A}') \rightarrow (\Omega'', \mathfrak{A}'')$ und P ein W-Maß auf $(\Omega, \mathfrak{A})$, so ist die Verteilung $Y \circ X$ das W-Maß $P_{Y \circ X} = (P_X)_Y$ auf Ω'' .

(d) $X : \Omega \rightarrow (0, 1]$ mit $P_X = \lambda^1|_{(0,1]}$ und

$$Y := -\ln(X) \Rightarrow P_Y = \mathbf{E}_1 \text{ (Exponentialverteilung)}$$

BEWEIS. Sei $A \in \mathfrak{B}(\mathbb{R}_+)$ und $B := \{e^{-a} \mid a \in A\}$

$$\begin{aligned} \Rightarrow P_Y(A) &= P\{-\ln \circ X \in A\} = P\{X \in B\} = \int 1_B(x) d\lambda^1(x) \\ &= \int 1_B(e^{-u}) e^{-u} d\lambda^1(u) = \int 1_A(u) e^{-u} du = \mathbf{E}_1(A) \end{aligned}$$

□

Andere Möglichkeit über Verteilungsfunktionen:

$$\begin{aligned} F_{P_Y}(y) &= P\{-\ln(X) \leq y\} = P\{X \geq e^{-y}\} = \lambda^1([e^{-y}, 1]) \\ &= \begin{cases} 1 - e^{-y} & \text{für } y \geq 0 \\ 0 & \text{für } y < 0 \end{cases} \end{aligned}$$

und (vgl. (3.23(a))) dies heißt $F_{P_Y} = F_{\mathbf{E}_1}(y)$

(e) Es sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit $P_X = \mathcal{N}_{0,1}$. Dann besitzt X^2 eine χ_1^2 -Verteilung mit einem Freiheitsgrad, d. h. P_{X^2} hat die LEBESGUE-Dichte

$$x \mapsto \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{x}} e^{-\frac{1}{2}x} 1_{(0,\infty)}(x)$$

BEWEIS. Für $A \in \mathfrak{B}(\mathbb{R})$ sei $B := \{a \in \mathbb{R} \mid a^2 \in A\} \in \mathfrak{B}(\mathbb{R})$:

$$\begin{aligned} \Rightarrow P_{X^2}(A) &= P\{X^2 \in A\} = P\{X \in B\} = \frac{1}{\sqrt{2\pi}} \int 1_B(x) e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int 1_B(x) 1_{(0,\infty)}(x) e^{-\frac{1}{2}x^2} d\lambda^1(x) \quad (y = x^2) \\ &= \frac{1}{\sqrt{2\pi}} \int 1_B(\sqrt{y}) 1_{(0,\infty)}(y) e^{-\frac{1}{2}y} \frac{1}{2\sqrt{y}} d\lambda^1(y) \\ &= \frac{1}{\sqrt{2\pi}} \int 1_A(y) 1_{(0,\infty)}(y) \frac{1}{\sqrt{y}} e^{-\frac{1}{2}y} d\lambda^1(y) \\ &= \chi_1^2(A) \end{aligned}$$

□

Eine sehr wichtige Annahme für die Modellbildung ist die Unabhängigkeit zwischen Teilen des Experimentes, z. B. zwischen den einzelnen Würfeln einer Münze. Mathematisch wird dies durch folgende Definition modelliert:

4.10 Definition. Es sei $(\Omega, \mathfrak{A}, P)$ ein W-Raum, $I \neq \emptyset$ eine Indexmenge und $A_j \in \mathfrak{A}$ für alle $j \in I$. Die Ereignisse $(A_j)_{j \in I}$ heißen (stochastisch) unabhängig, falls

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j) \quad \text{für alle endlichen } J \subset I.$$

4.11 Beispiel.

(a) Einmal würfeln mit einem fairen Würfel:

Modell: $\Omega = \{1, \dots, 6\}$, $\mathfrak{A} = \mathbb{P}(\Omega)$, $P = \mathcal{L}_\Omega$.

$$A_1 \hat{=} \text{„gerade Zahl“} = \{2, 4, 6\}$$

$$A_2 \hat{=} \text{„mindestens 5“} = \{5, 6\}$$

A_1 und A_2 sind unabhängig, da

$$P(A_1 \cap A_2) = P\{6\} = \frac{1}{6} \quad \text{und} \quad P(A_1) = \frac{1}{2}, \quad P(A_2) = \frac{1}{3} \quad \text{und} \quad \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}$$

(b) Wenn I drei oder mehr Elemente hat, genügt es *nicht* die „paarweise Unabhängigkeit“

$$P(A_j \cap A_k) = P(A_j) \cdot P(A_k) \quad \text{für alle } j, k \in I$$

zu zeigen, um die Unabhängigkeit von $(A_j)_{j \in I}$ nachzuweisen:

$$\begin{aligned} \Omega &= \{0, 1, 2, 3\}, \quad P = \mathcal{L}_\Omega, \quad A_1 = \{0, 1\}, \quad A_2 = \{0, 2\}, \quad A_3 = \{1, 2\} \\ \Rightarrow P(A_1 \cap A_2) &= P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4} \quad \text{und} \\ P(A_i) &= \frac{1}{2} \Rightarrow P(A_i) \cdot P(A_j) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}; \quad i, j \in \{1, 2, 3\}, \end{aligned}$$

d. h. A_1, A_2, A_3 sind paarweise unabhängig, aber

$$P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0 \neq P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

Auch für Zufallsvariablen läßt sich der Begriff der Unabhängigkeit definieren, und zwar durch Zurückführung auf Ereignisse:

4.12 Definition. Es seien für $j \in I$, $X_j : (\Omega, \mathfrak{A}) \rightarrow (\Omega_j, \mathfrak{A}_j)$ Zufallsvariablen auf einem W-Raum $(\Omega, \mathfrak{A}, P)$. $(X_j)_{j \in I}$ heißen genau dann *unabhängig*, wenn gilt:

$$\forall A_j \in \mathfrak{A}_j \text{ sind die Ereignisse } (\{X_j \in A_j\})_{j \in I} \text{ unabhängig}$$

4.13 Bemerkung.

(a) $(X_j)_{j \in I}$ unabhängig $\Leftrightarrow \forall J \subset I$ endlich und $\forall A_j \in \mathfrak{A}_j$ ($j \in J$):

$$P\{X_j \in A_j \text{ für alle } j \in J\} = \prod_{j \in J} P\{X_j \in A_j\}$$

(b) Ist I eine endliche Menge, genügt es

$$P\{X_j \in A_j \text{ für alle } j \in I\} = \prod_{j \in I} P\{X_j \in A_j\}$$

für alle $A_j \in \mathfrak{A}_j$ zu zeigen, da man $A_j = \Omega_j$ für $j \notin J$ wählen kann.

Im Fall von abzählbaren Räumen Ω_j , $\mathfrak{A}_j = \mathbb{P}(\Omega_j)$ kann man das einfacher mit der Zähl-dichte nachprüfen:

4.14 Satz. Es seien für $j \in I$ Ω_j abzählbar und $\mathfrak{A}_j = \mathbb{P}(\Omega_j)$. Die Zufallsvariablen $X_j : \Omega \rightarrow \Omega_j$ sind genau dann unabhängig, falls für alle endlichen $J \subset I$

$$P\{X_j = \omega_j \text{ für alle } j \in J\} = \prod_{j \in J} P\{X_j = \omega_j\}$$

für alle $\omega_j \in \Omega_j$ ($j \in J$) gilt. Ist I endlich, reicht es dies für $J = I$ zu zeigen.

BEWEIS. „ $\Rightarrow$ “: Klar, da man $A_j = \{\omega_j\} \in \mathfrak{A}_j$ wählen kann.

„ $\Leftarrow$ “: Seien $A_j \in \mathfrak{A}_j$ für $j \in J \subset I$. Dann gilt:

$$\begin{aligned} P\{X_j \in A_j \text{ für alle } j \in J\} &= \sum_{\substack{\omega_j \in A_j; \\ (j \in J)}} P\{X_j = \omega_j \text{ für } j \in J\} \\ &= \sum_{\substack{\omega_j \in A_j; \\ (j \in J)}} \prod_{j \in J} P\{X_j = \omega_j\} = \prod_{j \in J} \sum_{\omega_j \in A_j} P\{X_j = \omega_j\} \\ &= \prod_{j \in J} P\{X_j \in A_j\} \text{ nach dem Distributivgesetz} \end{aligned}$$

□

4.15 Beispiel.(a) (vgl. **Beispiel 4.3(b)**) **n -facher Münzwurf:**

$$\Omega = \{0, 1\}^n, \quad P = \mathcal{L}_\Omega \quad \text{und} \quad \mathfrak{A} = \mathbb{P}(\Omega)$$

Die Zufallsvariablen

$$X_j : \Omega \rightarrow \{0, 1\}; \quad X_j(\omega_1, \dots, \omega_n) = \omega_j \quad (\text{Ergebnis des } j\text{-ten Wurfs})$$

sind unabhängig, denn:

$$P\{X_j = x_j\} = 2^{-n} \cdot \#\{(\omega_1, \dots, \omega_n) \mid \omega_j = x_j\} = 2^{-n} \cdot 2^{n-1} = \frac{1}{2}$$

und folglich

$$P\{X_j = x_j \text{ für } 1 \leq j \leq n\} = 2^{-n} = \prod_{j=1}^n P\{X_j = x_j\}$$

(b) Allgemeiner gilt: Sind $\Omega_1, \dots, \Omega_n$ endliche Mengen, $\Omega := \Omega_1 \times \dots \times \Omega_n$, $X_j : \Omega \rightarrow \Omega_j$ die Projektion $X_j(\omega_1, \dots, \omega_n) := \omega_j$ und $P = \mathcal{L}_\Omega$, $\mathfrak{A} = \mathbb{P}(\Omega)$, so sind die Zufallsvariablen $X_1, \dots, X_n$ unabhängig und haben die Verteilung $P_{X_j} = \mathcal{L}_{\Omega_j}$.

BEWEIS.

$$\begin{aligned} P\{X_j = x_j\} &= \frac{\#\{(\omega_1, \dots, \omega_n) \mid \omega_j = x_j\}}{\#\Omega} = \frac{\#\Omega_1 \cdots \#\Omega_{j-1} \cdot 1 \cdot \#\Omega_{j+1} \cdots \#\Omega_n}{\#\Omega_1 \cdots \#\Omega_n} \\ &= \frac{1}{\#\Omega_j} \quad \Rightarrow \quad P_{X_j} = \mathcal{L}_{\Omega_j} \end{aligned}$$

und weiterhin:

$$P\{X_j = x_j \text{ für } 1 \leq j \leq n\} = \frac{1}{\#\Omega} = \prod_{j=1}^n \frac{1}{\#\Omega_j} = \prod_{j=1}^n P\{X_j = x_j\}$$

□

(c) Sind $(X_j)_{j \in I}$ unabhängige Zufallsvariablen $X_j : \Omega \rightarrow \Omega_j$ und sind $Y_j : (\Omega_j, \mathfrak{A}_j) \rightarrow (\Omega'_j, \mathfrak{A}'_j)$ weitere Zufallsvariablen, so sind auch die Zufallsvariablen $(Y_j \circ X_j)_{j \in I}$ unabhängig.

$$\{Y_j \circ X_j \in A'_j\} = \left\{ X_j \in \underbrace{\{Y_j \in A'_j\}}_{\in \mathfrak{A}_j} \right\}$$

(d) Es sei $\Omega = [0, 1)$, $\mathfrak{A} = \mathfrak{B}(\Omega)$ und für $n \geq 1$, $\omega \in [0, 1)$ sei $X_n(\omega)$ die n -te Stelle in der Binärdarstellung von ω (eindeutig durch Ausschluß von Periode 1!).

Dann sind die $(X_n)_{n \in \mathbb{N}}$ unabhängig und für $P = \lambda^1|_\Omega$ gilt $P_{X_n} = \mathcal{L}_{\{0,1\}}$.

BEWEIS. Für $1 \leq j \leq n$ sei $X_j \in \{0, 1\}$. Dann ist

$$\{\omega \in [0, 1) \mid X_j(\omega) = x_j \text{ für } 1 \leq j \leq n\} = \{\omega \in [0, 1) \mid \text{die ersten } n \text{ Stellen nach dem Komma von } \omega \text{ sind } x_1, \dots, x_n\}$$

$$= \left[\sum_{j=1}^n x_j 2^{-j}, \sum_{j=1}^n x_j 2^{-j} + 2^{-n} \right)$$

$$\Rightarrow P\{X_j = x_j \text{ für } j = 1, \dots, n\} = 2^{-n}$$

Sei nun $J \subset \mathbb{N}$ endlich und $n = \max(J)$:

$$\begin{aligned} \Rightarrow P\{X_j = x_j \text{ für } j \in J\} &= \sum_{\substack{X_j \in \{0,1\}; \\ j \in \{1, \dots, n\} \setminus J}} P\{X_j = x_j \text{ für } 1 \leq j \leq n\} \\ &= 2^{n-\#J} \cdot 2^{-n} = 2^{-\#J} = \prod_{j \in J} 2^{-1} = \prod_{j \in J} P\{X_j = x_j\} \end{aligned}$$

□

Zur Konstruktion eines W-Raumes $(\Omega, \mathfrak{A}, P)$ und unabhängigen Zufallsvariablen $X_j : \Omega \rightarrow \Omega_j$ mit vorgegebener Verteilung $P_{X_j} = P_j$ beschränken wir uns hier auf endlich viele Zufallsvariablen mit diskreter Verteilung bzw. einer Verteilung mit Dichte.

4.16 Satz. Sind $(\Omega_j, \mathbb{P}(\Omega_j), P_j)$ abzählbare W-Räume ($1 \leq j \leq n$), so existieren ein abzählbarer W-Raum $(\Omega, \mathbb{P}(\Omega), P)$ und Zufallsvariablen $X_j : \Omega \rightarrow \Omega_j$, die unabhängig sind und die Verteilung $P_{X_j} = P_j$ haben.

BEWEIS. Sei $\Omega := \Omega_1 \times \dots \times \Omega_n$, $X_j : \Omega \rightarrow \Omega_j$ die Projektion $X_j(\omega_1, \dots, \omega_n) := \omega_j$. Die Funktion

$$(\omega_1, \dots, \omega_n) \mapsto P_1(\{\omega_1\}) \cdot \dots \cdot P_n(\{\omega_n\})$$

ist eine Zähldichte, denn

$$\sum_{(\omega_1, \dots, \omega_n)} P_1(\{\omega_1\}) \cdot \dots \cdot P_n(\{\omega_n\}) = \prod_{j \in J} \sum_{\omega_j \in \Omega_j} P_j(\{\omega_j\}) = 1^n = 1.$$

Sei P das zugehörige W-Maß auf Ω nach Satz 1.23.

$$\begin{aligned} \Rightarrow P\{X_j = \omega_j\} &= P\{(X_1, \dots, X_n) \in \Omega \mid X_j = \omega_j\} = \underbrace{\left(\sum_{X_1 \in \Omega_1} \dots \sum_{X_n \in \Omega_n} \prod_{\substack{k=1 \\ k \neq j}}^n P_k\{X_k\} \right)}_{\Omega_j \text{ fehlt}} \cdot P_j\{\omega_j\} \\ &= \left(\prod_{\substack{k=1 \\ k \neq j}}^n \sum_{X_k \in \Omega_k} P_k\{X_k\} \right) P_j\{\omega_j\} = 1^{n-1} P_j\{\omega_j\} = P_j\{\omega_j\}, \text{ d. h.} \end{aligned}$$

$P_{X_j} = P_j$. Weiter gilt für alle $(\omega_1, \dots, \omega_n) \in \Omega$

$$P\{X_j = \omega_j \text{ für } 1 \leq j \leq n\} = P\{(\omega_1, \dots, \omega_n)\} = \prod_{j=1}^n P_j\{X_j\} = \prod_{j=1}^n P\{X_j = \omega_j\}$$

und damit folgt aus Satz 4.14 die Unabhängigkeit. □

4.17 Satz. Es seien P_j W-Maße auf $\Omega_j = \mathbb{R}$ mit LEBESGUE-Dichten $f_j : \mathbb{R} \rightarrow \mathbb{R}_+$ vorgegeben ($1 \leq j \leq n$). Dann ist durch die LEBESGUE-Dichte

$$f : (x_1, \dots, x_n) \mapsto f_1(x_1) \cdot \dots \cdot f_n(x_n) \quad \text{auf } \Omega := \mathbb{R}^n$$

ein W-Maß P auf $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ definiert, für das die Zufallsvariablen

$$X_j : \Omega \rightarrow \mathbb{R} \text{ mit } X_j(\omega_1, \dots, \omega_n) := \omega_j$$

unabhängig sind und $P_{X_j} = P_j$ gilt.

BEWEIS. Seien $A_j \in \mathfrak{B}(\mathbb{R})$ für $1 \leq j \leq n$

$$\begin{aligned} \Rightarrow & \int 1_{A_1 \times \dots \times A_n}(x) \cdot f_1(x_1) \cdot \dots \cdot f_n(x_n) d\lambda^n(x) \\ &= \int \dots \int 1_{A_1}(x_1) \cdot \dots \cdot 1_{A_n}(x_n) f_1(x_1) \cdot \dots \cdot f_n(x_n) d\lambda^1(x_n) \cdot \dots \cdot d\lambda^1(x_1) \\ &= \int 1_{A_1}(x_1) \cdot f_1(x_1) d\lambda^1(x_1) \cdot \dots \cdot \int 1_{A_n}(x_n) \cdot f_n(x_n) d\lambda^1(x_n) \\ &= P_1(A_1) \cdot \dots \cdot P_n(A_n) \end{aligned}$$

nach dem Satz von FUBINI.

Setzen wir $A_j = \mathbb{R}$ für $1 \leq j \leq n$, so folgt $\int f d\lambda^n = 1$, also ist f die Dichte eines W-Raumes P auf $\mathbb{R}^n$. Nach obiger Identität folgt

$$\begin{aligned} P\{X_j \in A_j \text{ für } 1 \leq j \leq n\} &= P(\{(X_1, \dots, X_n) \in \mathbb{R}^n \mid X_j \in A_j\}) \\ &= P(A_1 \times \dots \times A_n) = \int 1_{A_1 \times \dots \times A_n}(x) f(x) d\lambda^n(x) \\ &= P_1(A_1) \cdot \dots \cdot P_n(A_n) = \prod_{j=1}^n P_j(A_j). \end{aligned}$$

Sei nun $k \leq n$ fest, $A \in \mathfrak{B}(\mathbb{R})$, $A_j = \mathbb{R}$ für $j \neq k$, $A_k = A$

$$\begin{aligned} \Rightarrow P\{X_k \in A\} &= P_{X_k}(A) = P\{X_j \in A_j \text{ für } 1 \leq j \leq n\} = P_k(A) \\ \Rightarrow P_{X_k} &= P_k \end{aligned}$$

Für beliebiges $A_j \in \mathfrak{B}(\mathbb{R})$ folgt dann

$$P\{X_j \in A_j \text{ für } 1 \leq j \leq n\} = \prod_{j=1}^n P_j(A_j) = \prod_{j=1}^n P\{X_j \in A_j\},$$

d. h. die $(X_j)_{j=1, \dots, n}$ sind unabhängig. $\square$

4.18 Definition. Das in Satz 4.16 bzw. 4.17 konstruierte W-Maß auf $\Omega_1 \times \dots \times \Omega_n$ heißt *Produktmaß* und wird mit $P_1 \otimes \dots \otimes P_n$ bezeichnet.

4.19 Anwendung.

(a) Sind $M_1, \dots, M_n \in \mathfrak{B}(\mathbb{R})$ mit $0 < \lambda^1(M_j) < \infty$, so hat die Gleichverteilung auf $\Omega = M_1 \times \dots \times M_n$ die Eigenschaft, daß die Projektionen $X_j : \Omega \rightarrow M_j$ unabhängig sind und die Verteilung von X_j die Gleichverteilung auf M_j ist.

(b) Ein Experiment habe zwei mögliche Ausgänge: $1 \triangleq$ Erfolg; $0 \triangleq$ Mißerfolg, die mit Wahrscheinlichkeiten p bzw. $1-p$ ($0 \leq p \leq 1$) auftreten. Es werde n -mal unabhängig wiederholt:

Modell: $\Omega = \{0, 1\}^n$, $\mathfrak{A} = \mathbb{P}(\Omega)$ und $P = \mathcal{B}_{1,p} \otimes \dots \otimes \mathcal{B}_{1,p}$,

wobei $\mathcal{B}_{1,p}$ die BERNOULLI-Verteilung auf $\{0, 1\}$ ist. Die Zähldichte von P ist gegeben durch

$$\Omega \ni (\omega_1, \dots, \omega_n) \mapsto \prod_{j=1}^n \mathcal{B}_{1,p}\{\omega_j\} = p^k (1-p)^{n-k},$$

wobei $k := \#\{j \leq n \mid \omega_j = 1\}$ die Anzahl der Einsen in $(\omega_1, \dots, \omega_n)$ ist.

Sei $X : \Omega \rightarrow \mathbb{N}$; $X(\omega_1, \dots, \omega_n) := \#\{j \leq n \mid \omega_j = 1\}$ die Zufallsvariable, die die Anzahl der Erfolge angibt. Da man k Erfolge auf $\binom{n}{k}$ Arten unter n Experimenten erhalten kann, ist

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Das durch diese Zähldichte definierte W-Maß auf $\mathbb{N}$ heißt *Binomialverteilung* mit Parametern n, p und wird mit $\mathcal{B}_{n,p}$ bezeichnet (vgl. Bsp 1.24(2)).

Ist $X_j : \Omega \rightarrow \{0, 1\}$ die Projektion, so gilt nach Satz 4.16, daß $P_{X_j} = \mathcal{B}_{1,p}$ und es folgt, daß $X = \sum_{j=1}^n X_j$ die Summe von n unabhängigen $\mathcal{B}_{1,p}$ -verteilten Zufallsvariablen ist.

(c) Ähnlich wie man in (b) die Binomialverteilung als Verteilung der Summe von n unabhängigen $\mathcal{B}_{1,p}$ -verteilten Zufallsvariablen erhält, kann man die *negative Binomialverteilung* aus der geometrischen Verteilung gewinnen (s. Bsp 1.24(4)).

Es seien $X_1, \dots, X_n$ n unabhängige Zufallsvariablen mit Verteilung $\mathcal{G}_p$, $p \in (0, 1]$ (d. h. $P\{X_j = k\} = p(1-p)^k$).

$$\Rightarrow P\{X_j = k_j \text{ für } 1 \leq j \leq n\} = \prod_{j=1}^n \mathcal{G}_p\{k_j\} = p(1-p)^{k_1} \cdot \dots \cdot p(1-p)^{k_n} = p^n (1-p)^k,$$

wobei $k = k_1 + \dots + k_n$. Für die Verteilung der Zufallsvariablen $X = X_1 + \dots + X_n$ gilt also:

$$\begin{aligned} P_X\{k\} &= P\{X_1 + \dots + X_n = k\} = \sum_{\substack{(k_1, \dots, k_n) \in \mathbb{N}^n \\ k_1 + \dots + k_n = k}} P\{X_j = k_j \text{ für } 1 \leq j \leq n\} \\ &= \sum_{\substack{(k_1, \dots, k_n) \in \mathbb{N}^n \\ k_1 + \dots + k_n = k}} p^n (1-p)^k =: f_n(k) p^n (1-p)^k, \end{aligned}$$

wobei $f_n(k)$ die Anzahl der Summanden, d. h. die Anzahl der $(k_1, \dots, k_n) \in \mathbb{N}^n$ mit $k_1 + \dots + k_n = k$ ist.

Da P_X ein W-Maß ist

$$\begin{aligned} \Rightarrow 1 &= \sum_{k=0}^{\infty} P_X\{k\} = \sum_{k=0}^{\infty} f_n(k) p^n (1-p)^k = 1 \quad \text{für alle } 0 < p \leq 1 \\ \Rightarrow \sum_{k=0}^{\infty} f_n(k) (1-p)^k &= p^{-n} = (1 - (1-p))^{-n} = \sum_{k=0}^{\infty} \binom{-n}{k} (1-p)^k \end{aligned}$$

nach der TAYLORSchen Formel. Koeffizientenvergleich liefert

$$f_n(k) = \binom{-n}{k} = \binom{n+k-1}{k} \quad \Rightarrow \quad P_X\{k\} = P\{X = k\} = \binom{n+k-1}{k} p^n (1-p)^k$$

für $k \geq 0$. Die Verteilung von X bezeichnen wir mit $\mathcal{NB}_{n,p}$, die *negative Binomialverteilung* mit Parametern n und p . Sie tritt als Wartezeit bis zum n -ten Erfolg in einer Folge von unabhängigen Experimenten mit Erfolgswahrscheinlichkeit p auf.

- (d) Bei der Beobachtung einer radioaktiven Substanz werden im Zeitintervall $[t_0, t_1]$ insgesamt X_1 Zerfälle beobachtet, und im Zeitintervall $[t_1, t_2]$ X_2 . Experimentell ergibt sich, daß X_1 und X_2 unabhängig sind und die POISSON-Verteilungen (vgl. Bsp 1.24(3)) Π_{λ_1} und Π_{λ_2} haben, d. h.

$$P\{X_j = k\} = \Pi_{\lambda_j}\{k\} = \frac{\lambda_j^k}{k!} e^{-\lambda_j}$$

Unter diesem Modell können wir die Verteilung der Zahl $X_1 + X_2$ der Zerfälle in $[t_0, t_2]$ bestimmen:

$$\begin{aligned} P_{X_1+X_2}\{n\} &= P\{X_1 + X_2 = n\} \\ &= \sum_{k=0}^n P\{X_1 = k, X_2 = n-k\} \\ &\stackrel{(X_1, X_2 \text{ unabh.})}{=} \sum_{k=0}^n P\{X_1 = k\} P\{X_2 = n-k\} \\ &= \sum_{k=0}^n e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \\ &= \frac{1}{n!} e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} \\ &= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!} = \Pi_{\lambda_1+\lambda_2}\{n\} \end{aligned}$$

Also hat die Zufallsvariable $X_1 + X_2$ die Verteilung $\Pi_{\lambda_1+\lambda_2}\{n\}$.

ENDE VON KAPITEL 4

KAPITEL 5

Erwartungswert und Varianz von Zufallsvariablen

Interessiert man sich schon vor der Durchführung eines Experiments für das Ergebnis einer Zufallsvariablen, so kann man dies zwar nicht exakt vorhersagen, aber immerhin einen „mittleren Wert“ ihrer Realisation. Dazu dient unter anderem der Begriff der *Erwartungswertes*. (Anwendungen: fairer Einsatz beim Glücksspiel usw.)

Vorher müssen wir noch den Integralbegriff aus **Definition 3.12** für meßbare nichtnegative Funktionen auf Funktionen mit beliebigem Vorzeichen erweitern:

Erinnerung. Es sei $X : (\Omega, \mathfrak{A}) \rightarrow (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ eine Zufallsvariable und P ein W-Maß auf $(\Omega, \mathfrak{A})$.

1. Schritt: Ist $X \geq 0$ eine Zufallsvariable, die nur endlich viele Werte annimmt, d. h. $X = \sum_{j=1}^n x_j 1_{A_j}$, wobei $x_j > 0$ und $A_j \in \mathfrak{A}$, so ist (**vgl. 3.9(4)**)

$$\int X dP := \sum_{j=1}^n x_j P(A_j)$$

2. Schritt: Ist $X : \Omega \rightarrow \mathbb{R}_+$ beliebige Zufallsvariable, so ist

$$\int X dP := \sup_{\substack{Y \leq X \\ Y \in \mathcal{E}(\mathfrak{A})}} \int Y dP \in \mathbb{R}_+ \cup \{\infty\}$$

3. Schritt: Es sei $X : \Omega \rightarrow \mathbb{R}$ Zufallsvariable. Wir zerlegen X in

$$X_+ := \max\{X, 0\} \quad \text{und} \quad X_- := \max\{-X, 0\}.$$

Dann sind X_+ und X_- nichtnegative Zufallsvariablen und nach dem 2. Schritt sind $\int X_+ dP$ und $\int X_- dP$ definiert. Weiter gilt $X = X_+ - X_-$. Wir definieren:

5.1 Definition. Eine reelle Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ heißt *integrierbar*, falls

$$\int X_+ dP < \infty \quad \text{und} \quad \int X_- dP < \infty.$$

Wir nennen für solche X

$$\int X dP := \int X_+ dP - \int X_- dP$$

das Integral von X bzgl. P . Wir bezeichnen

$$\mathcal{E}(X) := \mathcal{E}_P(X) := \int X dP$$

den Erwartungswert von X (unter P). Ist $X \geq 0$, so setzen wir $\mathcal{E}(X) = \int X dP$ auch in dem Fall, daß das Integral $+\infty$ ist. Im Fall $X \geq 0$ ist $X_+ = X$ und $X_- = 0$, also steht **Definition 5.1** nicht im Widerspruch zu **Definition 3.12**.

5.2 Beispiel.

(a) Es sei X das Ergebnis beim Würfeln. d. h.

$$\Omega = \{1, \dots, 6\}, \mathfrak{A} = \mathbb{P}(\Omega) \text{ und } P = \mathcal{L}_\Omega \text{ und } X: \Omega \rightarrow \mathbb{R}, X(\omega) = \omega.$$

Dann ist nach **Definition 3.10** das Integral der endlichwertigen Funktion

$$X = \sum_{j=1}^6 j 1_{\{j\}} \quad \text{gleich} \quad \mathcal{E}(X) = \sum_{j=1}^6 j P\{j\} = (1 + 2 + 3 + 4 + 5 + 6) \cdot \frac{1}{6} = 3,5.$$

(b) Allgemeiner sei X gleichverteilt auf einer endl. Menge $I \subset \mathbb{R}$, d. h. $P\{X = i\} = \frac{1}{\#I} \forall i \in I$. Dann ist

$$\mathcal{E}_P(X) = \sum_{i \in I} i P\{X = i\} = \frac{1}{\#I} \sum_{i \in I} i$$

der Mittelwert aller Elemente von I .

(c) Es sei X $\mathcal{B}_{1,p}$ -verteilt mit $p \in [0, 1]$, d. h. $P\{X = 0\} = 1 - p$, $P\{X = 1\} = p$. Dann gilt

$$\mathcal{E}_P(X) = 0 \cdot P\{X = 0\} + 1 \cdot P\{X = 1\} = p$$

(d) Es sei $A \in \mathfrak{A}$. Dann gilt

$$\mathcal{E}_P(1_A) = 1 \cdot P\{1_A = 1\} + 0 \cdot P\{1_A = 0\} = P(A)$$

(e) Sei $\alpha \in \mathbb{R}$ beliebig, sowie $X = \alpha = \alpha 1_\Omega$ die deterministische Zufallsvariable mit Wert α . Dann gilt

$$\mathcal{E}(X) = \alpha \cdot 1 = \alpha$$

5.3 Bemerkung. Im folgenden seien stets $X, Y: \Omega \rightarrow \mathbb{R}$ reelle Zufallsvariablen.

(a) X ist genau dann integrierbar (bzgl. P), wenn $\mathcal{E}(|X|) < \infty$.

$$[\text{Es ist } \int X_+ dP + \int X_- dP = \int |X| dP]$$

(b) Sind X und Y integrierbar, so ist auch $X + Y$ integrierbar und es gilt

$$\mathcal{E}(X + Y) = \mathcal{E}(X) + \mathcal{E}(Y) \quad (\text{Additivität!})$$

BEWEIS. Sind $X, Y \geq 0$, so wurde dies in **Kor. 3.16** gezeigt. Sind $0 \leq Y \leq X \Rightarrow 0 \leq X - Y$ und damit

$$\mathcal{E}(X) = \mathcal{E}((X - Y) + Y) = \mathcal{E}(X - Y) + \mathcal{E}(Y) \Leftrightarrow \mathcal{E}(X - Y) = \mathcal{E}(X) - \mathcal{E}(Y).$$

Nun gilt (!):

$$(X + Y)_+ \leq X_+ + Y_+ \quad (X + Y)_- \leq X_- + Y_-$$

$$X_+ + Y_+ - (X + Y)_+ = X_- + Y_- - (X + Y)_-$$

Daraus folgt:

$$\begin{aligned} \mathcal{E}(X + Y) &= \mathcal{E}((X + Y)_+) - \mathcal{E}((X + Y)_-) \\ &= \mathcal{E}(X_+ + Y_+) - \mathcal{E}(X_+ + Y_+ - (X + Y)_+) - \mathcal{E}(X_- + Y_-) \\ &\quad + \mathcal{E}(X_- + Y_- - (X + Y)_-) \\ &= \mathcal{E}(X_+) + \mathcal{E}(Y_+) - \mathcal{E}(X_-) - \mathcal{E}(Y_-) = \mathcal{E}(X) + \mathcal{E}(Y) \end{aligned}$$

□

(c) Ist X integrierbar und $c \in \mathbb{R}$, so ist auch $c \cdot X$ integrierbar und $\mathcal{E}(c \cdot X) = c \cdot \mathcal{E}(X)$.

BEWEIS. Hier nur für $c \leq 0$: (anderer Fall analog!)

$$(c \cdot X)_+ = |c| X_- \text{ und } (c \cdot X)_- = |c| X_+.$$

$$\begin{aligned} \Rightarrow \mathcal{E}(c \cdot X) &= \mathcal{E}((c \cdot X)_+) - \mathcal{E}((c \cdot X)_-) = \mathcal{E}(|c| \cdot X_-) - \mathcal{E}(|c| \cdot X_+) \\ &= |c|(\mathcal{E}(X_-) - \mathcal{E}(X_+)) = |c|(-\mathcal{E}(X)) = (-|c|)E(X) = c\mathcal{E}(X) \end{aligned}$$

□

(d) Sind X, Y integrierbar oder ≥ 0 und gilt $X \leq Y$, so folgt $\mathcal{E}(X) \leq \mathcal{E}(Y)$, denn

$$Y - X \geq 0 \Rightarrow \mathcal{E}(Y - X) \geq 0 \Rightarrow \mathcal{E}(Y) - \mathcal{E}(X) = \mathcal{E}(Y - X) \geq 0 \Rightarrow \text{Beh.}$$

(e) Ist X^2 integrierbar, so ist auch X integrierbar.

BEWEIS. Es gilt $|X| \leq \frac{1}{2}(1 + X^2)$. Aus Teil (d) folgt

$$\mathcal{E}(|X|) \leq \frac{1}{2}\mathcal{E}(1 + X^2) = \frac{1}{2} + \frac{1}{2}\mathcal{E}(X^2) < \infty$$

und die Behauptung folgt aus **Bem. 5.3(a)**.

□

Im Fall von abzählbaren W -Räumen oder Verteilungen mit Dichte gibt es einfachere Formeln als **Definition 5.1** zur Berechnung von $\mathcal{E}(X)$. Dazu benötigen wir:

5.4 Satz (Transformationsformel). Es sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum. $\mathfrak{A}'$ eine σ -Algebra auf Ω' , $X : (\Omega, \mathfrak{A}) \rightarrow (\Omega', \mathfrak{A}')$ eine Zufallsvariable und $g : \Omega' \rightarrow \mathbb{R}$ eine reelle Zufallsvariable. Dann ist $g(X) = g \circ X$ genau dann integrierbar bzgl. P , wenn g integrierbar bzgl. P_X ist und es gilt

$$\mathcal{E}_P(g \circ X) = \int g dP_X = \int g(x) dP_X(x)$$

BEWEIS.

I) g nimmt endlich viele Werte an, also $g = \sum_{j=1}^n \alpha_j 1_{A_j}$ mit $A_j \in \mathfrak{A}'$.

$$\begin{aligned} \Rightarrow \mathcal{E}_P(g \circ X) &= \sum_{j=1}^n \int \alpha_j 1_{\{X \in A_j\}} dP = \sum_{j=1}^n \alpha_j P\{X \in A_j\} \\ &= \sum_{j=1}^n \alpha_j P_X(A_j) = \int g dP_X \end{aligned}$$

II) $g \geq 0$ meßbar. Nach **Lemma 3.15** existiert eine Folge $(g_n) \subset \mathcal{E}(\mathfrak{A}')$ mit $g_n \uparrow g$. Daraus folgt $g_n \circ X \uparrow g \circ X$ und nach **BEPP-LEVI (3.14)** ist

$$\mathcal{E}_P(g \circ X) = \int g \circ X dP = \lim_{n \rightarrow \infty} \int g_n \circ X dP = \lim_{n \rightarrow \infty} \int g_n dP_X = \int g dP_X$$

Wegen **Satz 5.3(a)** folgt daraus auch die behauptete Äquivalenz.

III) g beliebige Zufallsvariable. Wegen $(g \circ X)_+ = g_+ \circ X$ und $(g \circ X)_- = g_- \circ X$ folgt

$$\mathcal{E}_P(g \circ X) = \mathcal{E}_P(g_+ \circ X) - \mathcal{E}_P(g_- \circ X) = \int g_+ dP_X - \int g_- dP_X = \int g dP_X$$

□

5.5 Folgerung.

(1) **Korollar.** Es sei X eine reelle Zufallsvariable auf einem W -Raum $(\Omega, \mathfrak{A}, P)$. Dann hängt der Erwartungswert $\mathcal{E}_P(X)$ nur von der Verteilung P_X von X ab und nicht von der speziellen Wahl des W -Raumes $(\Omega, \mathfrak{A}, P)$.

BEWEIS. Ist $X' : \Omega' \rightarrow \mathbb{R}$ eine weitere Zufallsvariable auf $(\Omega', \mathfrak{A}', P')$ mit $P'_{X'} = P_X$, so folgt aus **Satz 5.4** für $g = \text{id}_{\mathbb{R}}$:

$$\mathcal{E}_P(X) = \mathcal{E}_P(g \circ X) = \int \text{id}_{\mathbb{R}} dP_X = \int \text{id}_{\mathbb{R}} dP'_{X'} = \mathcal{E}_{P'}(X').$$

□

- (2) **Korollar.** Ist X eine Zufallsvariable, deren Werte sämtlich in einer abzählbaren Teilmenge $A \subset \mathbb{R}$ liegen, so ist X integrierbar genau dann, wenn

$$\sum_{X \in A} |x| P\{X = x\} \text{ endlich ist}$$

und es gilt

$$\mathcal{E}_P(X) = \sum_{X \in A} x P\{X = x\}$$

BEWEIS. I) Hat $X \geq 0$ nur endlich viele Werte, so folgt dies aus der Definition des Integrals.

II) Wenn $X \geq 0$ die abzählbar vielen Werte $x_0, x_1, \dots$ annimmt, also $X = \sum_{j=0}^{\infty} x_j 1_{A_j}$ mit disjunkten Mengen A_n

$$\Rightarrow X_n = \sum_{j=0}^n x_j 1_{A_j} \uparrow X$$

Mit **BEPPO-LEVI (3.14)** folgt

$$\begin{aligned} \mathcal{E}(X) &= \lim_{n \rightarrow \infty} \mathcal{E}(X_n) = \lim_{n \rightarrow \infty} \sum_{j=0}^n x_j P(A_j) \\ &= \sum_{j=0}^{\infty} x_j P\{X = x_j\} = \sum_{X \in A} x P\{X = x\}, \end{aligned}$$

da $A = \{x_0, x_1, \dots\}$. Da $x_j \geq 0 \forall j$, folgt auch die im Satz behauptete Äquivalenz.

III) Ist X eine beliebige Zufallsvariable mit abzählbar vielen Werten, so folgt die Behauptung durch Zerlegung: $X = X_+ - X_-$, $A = A_+ \cup A_-$ mit

$$A_+ = \{x \in A \mid x \geq 0\} \cup \{0\} \quad \text{und} \quad A_- = \{x \in A \mid x \leq 0\} \cup \{0\}.$$

Daraus folgt

$$\begin{aligned} \mathcal{E}(X_+) &= \sum_{X \in A_+} x \cdot P\{X_+ = x\} = \sum_{X \in A_+} x \cdot P\{X = x\} \\ \mathcal{E}(X_-) &= \sum_{X \in A_-} |x| \cdot P\{X_- = x\} = \sum_{X \in A_-} |x| \cdot P\{X = x\} \\ \Rightarrow \mathcal{E}(X) &= \mathcal{E}(X_+) - \mathcal{E}(X_-) = \sum_{X \in A_+} x \cdot P\{X = x\} - \sum_{X \in A_-} |x| \cdot P\{X = x\} \\ &= \sum_{X \in A} x \cdot P\{X = x\} \end{aligned}$$

□

- (3) **Korollar.** Es seien X eine reelle Zufallsvariable, deren Verteilung P_X eine LEBESGUE-Dichte f besitzt und $g : \mathbb{R} \rightarrow \mathbb{R}$ meßbar. Dann ist $g(X) = g \circ X$ genau dann integrierbar, wenn $\int |g(x)| f(x) d\lambda^1(x) < \infty$ und es gilt

$$\mathcal{E}_P(g(X)) = \int g(x) f(x) d\lambda^1(x).$$

BEWEIS. Es sei Q das W-Maß auf $\mathbb{R}$ mit Dichte f , d. h.

$$Q(A) = \int 1_A f d\lambda^1 \quad \text{für } A \in \mathfrak{B}(\mathbb{R}).$$

I) g nimmt nur endlich viele Werte ≥ 0 an. d. h. $g = \sum_{j=1}^n x_j 1_{A_j}$ mit $x_j \geq 0$ und $A_j \in \mathfrak{B}(\mathbb{R})$ paarweise disjunkt.

$$\begin{aligned} \Rightarrow \int g dQ &= \sum_{j=1}^n x_j Q(A_j) = \sum_{j=1}^n x_j \int 1_{A_j} \cdot f d\lambda^1 \\ &= \int \underbrace{\sum_{j=1}^n x_j 1_{A_j}(x)}_g f(x) d\lambda^1(x) = \int g(x) f(x) d\lambda^1(x). \end{aligned}$$

II) $g \geq 0$: Nach **Lemma 3.15** existiert eine Folge $(g_n) \subset \mathcal{E}(\mathfrak{B}(\mathbb{R}))$ mit $g_n \uparrow g$. Nach **BEppo-LEVI (3.14)** gilt

$$\int g dQ = \lim_{n \rightarrow \infty} \int g_n dQ = \lim_{n \rightarrow \infty} \int g_n \cdot f d\lambda^1 \stackrel{BL}{=} \int g \cdot f d\lambda^1,$$

da $g_n(x)f(x) \uparrow g(x)f(x)$ mit $n \rightarrow \infty$, da $f \geq 0$.

III) Ist $g = g_+ - g_-$ beliebig, so gilt

$$\begin{aligned} \int g dQ &= \int g_+ dQ - \int g_- dQ = \int g_+ f d\lambda^1 - \int g_- f d\lambda^1 \\ &= \int g \cdot f d\lambda^1 \end{aligned}$$

Wir wenden dies auf $Q = P_X$ an. Nach **Satz 5.4** folgt

$$\mathcal{E}_P(g \circ X) = \int g dP_X = \int g dQ = \int g(x) \cdot f(x) d\lambda^1(x)$$

□

(4) **Korollar.** Falls X integrierbar ist und seine Verteilung die Dichte f hat, so gilt

$$\mathcal{E}_P(X) = \int x f(x) d\lambda^1(x)$$

5.6 Beispiel.

(a) Es sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit geometrischer Verteilung $\mathcal{G}_p$ (wobei $p \in (0, 1]$). Dann gilt $E(X) = \frac{1}{p} - 1$.

In **Anw. 4.19(c)** hatten wir die negative Binomialverteilung $\mathcal{NB}_{2,p}$ auf $\mathbb{N}$ mit der Zähldichte

$$\mathcal{NB}_{2,p}\{k\} = \binom{k+1}{k} p^2 (1-p)^k, \quad k \in \mathbb{N}$$

eingeführt. $\Rightarrow \sum_{k=0}^{\infty} (k+1) p^2 (1-p)^k = 1$.

$$\begin{aligned} \Rightarrow \mathcal{E}(X) &= \sum_{k=0}^{\infty} k P\{X = k\} = \sum_{k=0}^{\infty} k p (1-p)^k \\ &= \frac{1}{p} \sum_{k=0}^{\infty} (k+1) p^2 (1-p)^k - \sum_{k=0}^{\infty} p (1-p)^k \\ &= \frac{1}{p} - p \frac{1}{1 - (1-p)} = \frac{1}{p} - 1 \end{aligned}$$

□

(b) Es sei X eine reelle Zufallsvariable mit Normalverteilung $\mathcal{N}_{\mu, \sigma^2}$.

1) Sei $\mu = 0$, $\sigma^2 = 1$:

$$\begin{aligned} \Rightarrow \mathcal{E}(|X|) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x| e^{-\frac{1}{2}x^2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x e^{-\frac{1}{2}x^2} dx \\ &= \frac{2}{\sqrt{2\pi}} \left(-e^{-\frac{1}{2}x^2} \right) \Big|_0^{\infty} = \frac{2}{\sqrt{2\pi}} < \infty, \end{aligned}$$

also besitzt X einen Erwartungswert. Es folgt $\mathcal{E}(X) = 0$.

2) $\mu \in \mathbb{R}$, $\sigma^2 > 0$ beliebig: Setze $X' = \frac{X-\mu}{\sigma}$. Nach **Bsp. 3.20(b)** hat X' eine $\mathcal{N}_{0,1}$ -Verteilung.

$$\Rightarrow 0 = \mathcal{E}(X') = \mathcal{E}\left(\frac{1}{\sigma}(X - \mu)\right) = \frac{1}{\sigma}\mathcal{E}(X) - \frac{\mu}{\sigma} \Leftrightarrow \mathcal{E}(X) = \mu.$$

(c) Es sei nun X eine reelle Zufallsvariable mit der CAUCHY-Verteilung γ_{α} ; die LEBESGUE-Dichte ist dann

$$x \mapsto \frac{\alpha}{\pi(\alpha^2 + x^2)} \quad \text{auf } \mathbb{R}$$

Trotz der Symmetrie um 0 ist der Erwartungswert aber nicht gleich Null. Wegen

$$\begin{aligned} \mathcal{E}(|X|) &= \int_{-\infty}^{\infty} \frac{\alpha|x|}{\pi(\alpha^2 + x^2)} dx = \frac{\alpha}{\pi} \int_0^{\infty} \frac{2x}{\alpha^2 + x^2} dx \\ &= \frac{\alpha}{\pi} \ln(\alpha^2 + x^2) \Big|_0^{\infty} = \infty \end{aligned}$$

existiert der Erwartungswert von X nicht.

(d) Es sei X eine Zufallsvariable, die $\mathcal{B}_{n,p}$ -verteilt ist mit $n \geq 1$ und $p \in [0, 1]$. Für $n = 1$ gilt nach **Bsp. 5.2(c)**: $\mathcal{E}(X) = p$. Für $n \geq 1$ gilt: $\mathcal{E}(X) = n \cdot p$.

Sind $X_1, \dots, X_n$ unabhängige $\mathcal{B}_{n,p}$ -verteilte Zufallsvariablen, so ist nach **Anw. 4.19(b)** $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$ auch $\mathcal{B}_{n,p}$ -verteilt. Aus **Korollar 5.5(1)** und **Bem. 5.3(b)** folgt dann

$$\mathcal{E}(X) = \mathcal{E}(\bar{X}) = \frac{1}{n} \sum_{j=1}^n \mathcal{E}(X_j) = \sum_{j=1}^n p = n \cdot p$$

□

Während der Erwartungswert $\mathcal{E}(X)$ eine Aussage über den mittleren Wert, den eine Zufallsvariable annimmt, macht, ist darin noch keine Information enthalten, wie weit die Zufallsvariable von diesem Wert im Mittel abweicht. Dazu braucht man

5.7 Definition. Es sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable auf einem W -Raum $(\Omega, \mathfrak{A}, P)$, für die X^2 integrierbar ist. (Man nennt X *quadratisch integrierbar*.) Die Zahl

$$\text{Var}_P(X) = \mathcal{E}_P\left(\left(X - \mathcal{E}_P(X)\right)^2\right)$$

nennt man die *Varianz* von X .

Die *Streuung* von X ist

$$\sigma(X) = \sqrt{\text{Var}_P(X)};$$

diese Größe hat die gleiche Dimension wie X .

5.8 Bemerkung.

(a) Wegen $\left(X - \mathcal{E}_P(X)\right)^2 = X^2 - 2\mathcal{E}_P(X) \cdot X + \left(\mathcal{E}_P(X)\right)^2$ und da nach **Bem. 5.3(e)** auch X integrierbar ist, ist $\text{Var}_P(X)$ wohldefiniert und ≥ 0 .

(b) Es ist $\text{Var}_P(X) = \mathcal{E}_P(X^2) - (\mathcal{E}_P(X))^2$

$$\begin{aligned}\text{Var}_P(X) &= \mathcal{E}_P\left(\left(X - \mathcal{E}_P(X)\right)^2\right) = \mathcal{E}_P\left(X^2 - 2\mathcal{E}_P(X) \cdot X + (\mathcal{E}_P(X))^2\right) \\ &= \mathcal{E}_P(X^2) - 2\mathcal{E}_P(X)\mathcal{E}_P(X) + (\mathcal{E}_P(X))^2 \\ &= \mathcal{E}_P(X^2) - (\mathcal{E}_P(X))^2\end{aligned}$$

□

(c) Wie schon der Erwartungswert, hängt Var_P nur von der Verteilung P_X von X ab. Wir schreiben kurz $\text{Var}(X)$ statt $\text{Var}_P(X)$.

5.9 Beispiel.

(a) (vgl. **Bsp. 5.2(a)**) Für die Augenzahl X eines Würfels gilt

$$\mathcal{E}(X^2) = \frac{1}{6} \sum_{j=1}^6 j^2 = \frac{91}{6}$$

und damit

$$\text{Var}(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$$

Die Streuung beträgt $\sigma(X) = \sqrt{\frac{35}{12}} \approx 1,708$.

(b) Es sei $X \mathcal{B}_{1,p}$ -verteilt

$$\Rightarrow \text{Var}(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2 = \mathcal{E}(X) - (\mathcal{E}(X))^2 = p - p^2 = p(1-p)$$

(c) Ist $X = \alpha \in \mathbb{R}$ eine deterministische Zufallsvariable, so gilt $\mathcal{E}(X^2) = \alpha^2 \Rightarrow \text{Var}(X) = 0$. Dies ist plausibel. Die Umkehrung wird in einer Übungsaufgabe gezeigt.

(d) Ist X eine beschränkte Zufallsvariable, d. h. $|X| \leq B \in \mathbb{R}_+$, so existieren Erwartungswert und Varianz von X .

(e) Für jede quadratisch integrierbare Zufallsvariable X gilt

$$\text{Var}(c \cdot X) = c^2 \cdot \text{Var}(X) \quad \text{für alle } c \in \mathbb{R}.$$

(f) Ist X eine $\mathcal{N}_{\mu, \sigma^2}$ -verteilte Zufallsvariable, so gilt $\text{Var}(X) = \sigma^2$.

Betrachte $X' = \frac{X - \mu}{\sigma}$. Dann ist $X' \mathcal{N}_{0,1}$ -verteilt. Dann gilt

$$\mathcal{E}(X'^2) = \int X'^2 dP_{X'} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}x^2} dx$$

Nun ist

$$\begin{aligned}\int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}x^2} dx &= \int_{-\infty}^{\infty} (-x)(-x) e^{-\frac{1}{2}x^2} dx \\ &= \left((-x) e^{-\frac{1}{2}x^2}\right) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-1) e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}\end{aligned}$$

Daraus folgt:

$$\text{Var}(X') = \mathcal{E}(X'^2) = 1 \quad \text{und somit}$$

$$\text{Var}(X) = \mathcal{E}((X - \mu)^2) = \mathcal{E}(\sigma^2 X'^2) = \sigma^2 \mathcal{E}(X'^2) = \sigma^2$$

□

5.10 Satz (CAUCHY-SCHWARZsche Ungleichung). Es seien X, Y quadratisch integrierbare Zufallsvariablen. Dann besitzt die Zufallsvariable $X \cdot Y$ einen Erwartungswert und es gilt

$$|\mathcal{E}(XY)| \leq \mathcal{E}(|XY|) \leq \sqrt{\mathcal{E}(X^2) \mathcal{E}(Y^2)}.$$

BEWEIS. Es sei $\lambda > 0$ beliebig. Wegen

$$X^2 - 2\lambda XY + \lambda^2 Y^2 = (X - \lambda Y)^2 \geq 0 \quad \text{folgt} \quad XY \leq \frac{1}{2\lambda} X^2 + \frac{\lambda}{2} Y^2$$

und (da diese Ungleichung auch für $-Y$ anstelle von Y gilt) es folgt

$$|XY| \leq \frac{1}{2\lambda} X^2 + \frac{\lambda}{2} Y^2.$$

Da die rechte Seite einen Erwartungswert besitzt, tut dies auch die linke Seite.

$$\Rightarrow |\mathcal{E}(XY)| \leq \mathcal{E}(|XY|) \leq \frac{1}{2\lambda} \mathcal{E}(X^2) + \frac{\lambda}{2} \mathcal{E}(Y^2)$$

1. Fall: $\mathcal{E}(Y^2) = 0$. Mit $\lambda \rightarrow +\infty$ folgt dann $\mathcal{E}(XY) = 0$ und damit die Behauptung.

2. Fall: $\mathcal{E}(Y^2) > 0$. Setze $\lambda = \sqrt{\frac{\mathcal{E}(X^2)}{\mathcal{E}(Y^2)}}$

$$\Rightarrow |\mathcal{E}(XY)| \leq \frac{1}{2\sqrt{\frac{\mathcal{E}(X^2)}{\mathcal{E}(Y^2)}}} \mathcal{E}(X^2) + \frac{\sqrt{\frac{\mathcal{E}(X^2)}{\mathcal{E}(Y^2)}}}{2} \mathcal{E}(Y^2) = \sqrt{\mathcal{E}(X^2) \mathcal{E}(Y^2)}$$

□

Satz 5.10 ermöglicht folgende Definition:

5.11 Definition. Es seien X, Y quadratisch integrierbare Zufallsvariablen. Dann heißt

$$\text{Kov}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X) \mathcal{E}(Y)$$

die *Kovarianz* von X und Y . Die Kovarianz ist ein Maß für die affin-lineare Abhängigkeit zwischen X und Y . Falls $\text{Kov}(X, Y) = 0$, so nennt man X und Y *unkorreliert*.

5.12 Satz.

(a) Es seien X, Y quadratisch integrierbare Zufallsvariablen

$$\Rightarrow |\text{Kov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}$$

(b) Sind $X_1, \dots, X_n$ quadratisch integrierbare Zufallsvariablen, dann existiert $\mathcal{E}\left(\left(\sum_{j=1}^n X_j\right)^2\right)$ und es gilt

$$\text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}(X_j) + 2 \sum_{1 \leq j < k} \text{Kov}(X_j, X_k) \quad (\text{Formel von BIENAIMÉE})$$

(c) Sind zusätzlich die Zufallsvariablen unabhängig, so sind die (X_j) unkorreliert und

$$\text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}(X_j)$$

BEWEIS.

(a) Es gilt $\text{Kov}(X, Y) = \mathcal{E}[(X - \mathcal{E}(X))(Y - \mathcal{E}(Y))]$. Aus **Satz 5.10** folgt dann

$$\begin{aligned} |\text{Kov}(X, Y)| &= \left| \mathcal{E}\left((X - \mathcal{E}(X))(Y - \mathcal{E}(Y))\right) \right| \\ &\leq \sqrt{\mathcal{E}\left((X - \mathcal{E}(X))^2\right) \mathcal{E}\left((Y - \mathcal{E}(Y))^2\right)} \\ &= \sqrt{\text{Var}(X) \text{Var}(Y)} \end{aligned}$$

(b) Die erste Behauptung folgt aus $\left(\sum_{j=0}^n X_j\right)^2 = \sum_{j,k=1}^n X_j X_k$ und **Satz 5.10**.

$$\begin{aligned} \Rightarrow \text{Var}\left(\sum_{j=1}^n X_j\right) &= \mathcal{E}\left(\left(\sum_{j=1}^n X_j\right)^2\right) - \left(\mathcal{E}\left(\sum_{j=1}^n X_j\right)\right)^2 \\ &= \sum_{j=1}^n \mathcal{E}(X_j^2) + 2 \sum_{j < k} \mathcal{E}(X_j X_k) - \sum_{j=1}^n \left(\mathcal{E}(X_j)\right)^2 - 2 \sum_{j < k} \mathcal{E}(X_j) \mathcal{E}(X_k) \\ &= \sum_{j=1}^n \text{Var}(X_j) + 2 \sum_{j < k} \text{Kov}(X_j, X_k). \end{aligned}$$

Für den Beweis von (c) brauchen wir

5.13 Lemma. Sind X, Y integrierbar und unabhängig, so besitzt $X \cdot Y$ einen Erwartungswert und es gilt

$$\mathcal{E}(X \cdot Y) = \mathcal{E}(X) \cdot \mathcal{E}(Y)$$

BEWEIS. Wir zeigen (scheinbar) etwas mehr als behauptet, nämlich

$$\mathcal{E}\left((f \circ X) \cdot (g \circ Y)\right) = \mathcal{E}(f \circ X) \mathcal{E}(g \circ Y)$$

für alle meßbaren $f, g: \mathbb{R} \rightarrow \mathbb{R}$, falls entweder $f, g \geq 0$ oder $f \circ X$ und $g \circ Y$ integrierbar sind.

I) f, g nehmen nur endlich viele Werte ≥ 0 an:

$$f(\mathbb{R}) = \{x_1, \dots, x_m\}; \quad g(\mathbb{R}) = \{y_1, \dots, y_n\}.$$

$$\begin{aligned} \mathcal{E}\left((f \circ X) \cdot (g \circ Y)\right) &= \sum_{j=1}^m \sum_{k=1}^n x_j y_k P\{f \circ X = x_j, g \circ Y = y_k\} \\ &= \sum_{j=1}^m \sum_{k=1}^n x_j y_k P\{f \circ X = x_j\} \cdot P\{g \circ Y = y_k\} \\ &= \sum_{j=1}^m x_j P\{f \circ X = x_j\} \sum_{k=1}^n y_k P\{g \circ Y = y_k\} \\ &= \mathcal{E}(f \circ X) \mathcal{E}(g \circ Y) \end{aligned}$$

II) $f, g \geq 0$. Wähle $(f_n), (g_n) \subset \mathcal{E}(\mathfrak{B}(\mathbb{R}))$ mit $f_n \uparrow f, g_n \uparrow g$. Nach **Lemma 3.15** gilt dann

$$(f_n \circ X) \cdot (g_n \circ Y) \uparrow (f \circ X) \cdot (g \circ Y)$$

und nach **BEppo-LEVI (3.14)** folgt

$$\begin{aligned} \mathcal{E}\left((f \circ X) \cdot (g \circ Y)\right) &= \lim_{n \rightarrow \infty} \mathcal{E}\left((f_n \circ X) \cdot (g_n \circ Y)\right) \\ &= \lim_{n \rightarrow \infty} \mathcal{E}\left((f_n \circ X)\right) \cdot \mathcal{E}\left((g_n \circ Y)\right) = \mathcal{E}(f \circ X) \mathcal{E}(g \circ Y) \end{aligned}$$

III) Schreibe $f = f_+ - f_-$ und $g = g_+ - g_-$ und benutze (II).

Mit $f(x) = g(x) = |x| \geq 0$ folgt $\mathcal{E}(|XY|) = \mathcal{E}(|X|) \mathcal{E}(|Y|) < \infty$, also ist $X \cdot Y$ integrierbar und mit $f(x) = g(x) = x$ folgt dann die Behauptung.

□

BEWEIS VON **Satz 5.12(c)**: Nach **Lemma 5.13** ist

$$\text{Kov}(X_j, X_k) = \mathcal{E}(X_j X_k) - \mathcal{E}(X_j) \mathcal{E}(X_k) = 0 \quad \text{für } j \neq k$$

Damit sind X_j, X_k unkorreliert und aus (b) folgt dann die Behauptung.

□

5.14 Definition. Es seien X, Y quadratisch integrierbare reelle Zufallsvariablen. Dann heißt die Zahl

$$\text{Kor}(X, Y) = \begin{cases} \frac{\text{Kov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} & , \text{ falls } \text{Var}(X) > 0, \text{Var}(Y) > 0 \\ 0 & , \text{ sonst} \end{cases}$$

die *Korrelation* von X und Y .

5.15 Bemerkung.

- (a) Wegen **Satz 5.12(a)** gilt $-1 \leq \text{Kor}(X, Y) \leq 1$.
 (b) Wegen **Satz 5.12(c)** gilt für unkorrelierte und damit erst recht für unabhängige Zufallsvariablen $\text{Kor}(X, Y) = 0$.

5.16 Beispiel. Es sei X $\mathcal{B}_{n,p}$ -verteilt mit $n \geq 1$, $p \in (0, 1)$. Dann gilt $\text{Var}(X) = np(1 - p)$.

BEWEIS. Sind $X_1, \dots, X_n$ unabhängige $\mathcal{B}_{1,p}$ -verteilte Zufallsvariablen, so hat nach **Anw. 4.19(b)** die Zufallsvariable $\bar{X} = \sum_{j=1}^n X_j$ eine $\mathcal{B}_{n,p}$ -Verteilung

$$\Rightarrow \text{Var}(\bar{X}) = \text{Var}(X) = \text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}(X_j) = np(1 - p),$$

da nach **Bsp. 5.9(b)** $\text{Var}(X_j) = p(1 - p)$ gilt. $\square$

5.17 Satz (TSCHEBYSCHEFFSche Ungleichung). Es sei X eine quadratisch integrierbare Zufallsvariable. Dann gilt für alle $\varepsilon > 0$:

$$P\{|X - \mathcal{E}(X)| \geq \varepsilon\} \leq \frac{\text{Var}(X)}{\varepsilon^2}$$

BEWEIS.

$$\begin{aligned} \text{Var}(X) &= \mathcal{E}\left((X - \mathcal{E}(X))^2\right) \geq \mathcal{E}\left(1_{\{|X - \mathcal{E}(X)| \geq \varepsilon\}} \cdot \underbrace{(X - \mathcal{E}(X))^2}_{\leq \varepsilon^2}\right) \\ &\geq \varepsilon^2 \mathcal{E}\left(1_{\{|X - \mathcal{E}(X)| \geq \varepsilon\}}\right) = \varepsilon^2 P\{|X - \mathcal{E}(X)| \geq \varepsilon\} \end{aligned}$$

Daraus folgt die Behauptung. $\square$

5.18 Bemerkung. Die obige Ungleichung präzisiert die Verwendung der Streuung $\sigma(X)$ als Maß für die Breite der Verteilung von X ; die Wahrscheinlichkeit für eine Abweichung von $\mathcal{E}(X)$ um mindestens $c \cdot \sigma(X)$ beträgt höchstens $\frac{1}{c^2}$.

Setze $\varepsilon = c \cdot \sigma(X)$

$$\Rightarrow P\{|X - \mathcal{E}(X)| \geq c \cdot \sigma(X)\} \leq \frac{\text{Var}(X)}{c^2 \cdot \sigma(X)^2} = \frac{1}{c^2}$$

$\square$

Einer der wichtigsten Sätze der Wahrscheinlichkeitsrechnung ist das von JAKOB BERNOULLI etwa 1688 entwickelte Gesetz der großen Zahl. Es macht eine Aussage zwischen dem Mittelwert $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ der ersten n identisch verteilten Messungen und dem Erwartungswert $\mathcal{E}(X_1)$.

5.19 Satz (Schwaches Gesetz der großen Zahl). Es sei $X_1, X_2, \dots$ eine Folge von quadratisch integrierbaren und unkorrelierten Zufallsvariablen mit gleichem Erwartungswert $\mu = \mathcal{E}(X_n)$ und beschränkter Varianz $\text{Var}(X_n) \leq M$ für alle $n \geq 1$. Wir setzen

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

für $n \geq 1$. Dann gilt für alle $\varepsilon > 0$

$$P\{|\bar{X}_n - \mu| \geq \varepsilon\} \rightarrow 0 \quad \text{für } n \rightarrow \infty$$

BEWEIS. Es gilt

$$\mathcal{E}(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathcal{E}(X_i) = \frac{1}{n} \cdot n\mu = \mu \quad \text{und}$$

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{\text{Bsp. 5.9(e)}}{=} \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \stackrel{\text{s. 5.12(c)}}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \leq \frac{1}{n^2} \cdot n \cdot M = \frac{M}{n}$$

Mit der **TSCHEBYSCHEFFSchen Ungleichung (5.17)** folgt dann

$$\begin{aligned} P\{|\bar{X}_n - \mu| \geq \varepsilon\} &= P\{|\bar{X}_n - \mathcal{E}(\bar{X}_n)| \geq \varepsilon\} \\ &\leq \frac{1}{\varepsilon^2} \text{Var}(\bar{X}_n) \leq \frac{M}{\varepsilon^2} \cdot \frac{1}{n} \rightarrow 0 \quad \text{für } n \rightarrow \infty \end{aligned}$$

□

5.20 Bemerkung. Die in **Satz 5.19** bewiesene Annäherung des *Stichprobenmittels* $\bar{X}_n$ an den gemeinsamen Erwartungswert der Messungen X_j bezeichnet man als *stochastische Konvergenz*:

- (1) **Definition.** Es sei $(\Omega, \mathfrak{A}, P)$ ein W-Raum und $Y_n : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen. Weiter sei $Y : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Man sagt, daß Y_n gegen Y *stochastisch konvergiert*, falls für alle $\varepsilon > 0$ gilt:

$$P\{|Y_n - Y| \geq \varepsilon\} \rightarrow 0 \quad \text{für } n \rightarrow \infty$$

Wir schreiben in diesem Fall:

$$Y_n \xrightarrow{P} Y \quad (n \rightarrow \infty)$$

- (2) **Bemerkung.**

- (a) **Satz 5.19** besagt, daß $\bar{X}_n \xrightarrow{P} \mu$ für $n \rightarrow \infty$.
 (b) In seiner ursprünglichen Form behandelt das schwache Gesetz der großen Zahl von J. BERNOULLI den folgenden Fall:

Seien $A_1, A_2, \dots \in \mathfrak{A}$ unabhängige Ereignisse mit gleicher Wahrscheinlichkeit $p := P(A_j)$ und K_n die Anzahl der Erfolge unter den ersten n Ereignissen, d. h.

$$K_n(\omega) = \#\{j \leq n \mid \omega \in A_j\} = \sum_{j=1}^n 1_{A_j}(\omega)$$

für $\omega \in \Omega$. Dann konvergiert die relative Häufigkeit $\frac{K_n}{n}$ stochastisch gegen die Erfolgswahrscheinlichkeit $p = P(A_j)$, d. h.

$$\frac{K_n}{n} \xrightarrow{P} p \quad \text{für } n \rightarrow \infty$$

Sei $X_j = 1_{A_j}$. Dann sind die X_j unabhängig und haben die Verteilung $P_{X_j} = \mathcal{B}_{1,p}$.

Nach **Satz 5.12(c)** sind die X_j unkorreliert und es gilt $\mathcal{E}(X_j) = \mathcal{E}(1_{A_j}) = P(A_j) = p$ und nach **Bsp. 5.9(b)** gilt $\text{Var}(X_j) = p(1-p) =: M$. **Satz 5.19** liefert dann die Behauptung.

□

5.21 Anwendung (Numerische Integration, Monte Carlo Methoden). Es sei $f : [0, 1] \rightarrow \mathbb{R}$ meßbar (z. B. stetig). Gesucht ist eine Näherung für $\int_0^1 f(x) d\lambda^1(x)$. Ist f „glatt“, d. h. hinreichend oft differenzierbar, so ist

$$\int_0^1 f(x) d\lambda^1(x) = \int_0^1 f(x) dx$$

und für dieses RIEMANN-Integral stellt die Numerik Quadraturformeln zur näherungsweisen Berechnung des Integrals bereit. Ist f nur stetig oder besitzt f sogar endlich viele Unstetigkeitsstellen, funktionieren die klassischen Quadraturformeln schlecht. Das schwache Gesetz der großen Zahl liefert eine andere Methode:

Es seien $X_j : \Omega \rightarrow [0, 1]$ unkorrelierte Zufallsvariablen mit Verteilung $P_{X_j} = \lambda^1|_{[0,1]}$. Dann gilt, falls $\int_0^1 f^2(x) d\lambda^1(x) < \infty$:

$$\frac{1}{n} \sum_{j=1}^n f(X_j) \xrightarrow{P} \int_0^1 f(x) d\lambda^1(x) \quad \text{für } n \rightarrow \infty$$

BEWEIS. Die Zufallsvariablen $Y_j = f(X_j)$ sind wegen der Transformationsformel 5.4

$$\mathcal{E}(Y_j^2) = \int f^2(x) dP_{X_j}(x) = \int_0^1 f^2(x) d\lambda^1(x) < \infty$$

quadratisch integrierbar und unkorreliert. Es ist

$$\mathcal{E}(Y_j) = \mathcal{E}(f(X_j)) = \int_0^1 f(x) dP_X(x) = \int_0^1 f(x) d\lambda^1(x) = \mu.$$

Weiter ist

$$\text{Var}(Y_j) = \mathcal{E}(Y_j^2) - (\mathcal{E}(Y_j))^2 = \int_0^1 f^2(x) d\lambda^1(x) - \left(\int_0^1 f(x) d\lambda^1(x) \right)^2 =: M < \infty$$

Nach Satz 5.19 gilt dann:

$$\frac{1}{n} \sum_{j=1}^n Y_j = \frac{1}{n} \sum_{j=1}^n f(X_j) \xrightarrow{P} \int_0^1 f(x) d\lambda^1(x) = \mu.$$

Für den Fehler ergibt sich aus dem Beweis von Satz 5.19 die Abschätzung:

$$P\left\{\left|\frac{1}{n} \sum_{j=1}^n f(X_j) - \int_0^1 f(x) d\lambda^1(x)\right| \geq \varepsilon\right\} \leq \frac{1}{n} \cdot \frac{1}{\varepsilon^2} \left[\int_0^1 f^2(x) d\lambda^1(x) - \left(\int_0^1 f(x) d\lambda^1(x) \right)^2 \right]$$

□

ENDE VON KAPITEL 5

KAPITEL 6

Bedingte Wahrscheinlichkeiten

Wenn wir zusätzliche Informationen über ein Experiment besitzen, ändert sich für uns die Wahrscheinlichkeit, mit der wir ein bestimmtes Ereignis einschätzen. Ein extremes Beispiel ist, wenn wir bereits wissen, daß ein Münzwurf Wappen ergeben hat, dann werden wir aufgrund dieser Information $P\{W\} = 1$ annehmen.

Oder hat man beim Skat drei Buben auf der Hand, so wird man die Wahrscheinlichkeit, daß im Skat genau ein Bube liegt, nicht mehr so hoch einschätzen, wie vor dem Aufnehmen der Karten ($\frac{7}{31} = \frac{\binom{4}{1}\binom{28}{1}}{\binom{32}{2}}$ vor dem Aufnehmen gegenüber $\frac{1}{11} = \frac{\binom{1}{1}\binom{21}{1}}{\binom{22}{2}}$ danach).

Dieses Phänomen wird durch den Begriff der bedingten Wahrscheinlichkeit beschrieben.

6.1 Definition. Es seien $(\Omega, \mathfrak{A}, P)$ ein W-Raum und $A, B \in \mathfrak{A}$ mit $P(B) > 0$. Dann heißt die Zahl

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

die *bedingte Wahrscheinlichkeit* von A unter B . Sind $X : \Omega \rightarrow \Omega_0$ und $Y : \Omega \rightarrow \Omega_1$ Zufallsvariablen, sowie A, B Ereignisse der Gestalt $A := \{X \in A'\}$ bzw. $B := \{Y \in B'\}$, so schreiben wir auch $P\{X \in A' | Y \in B'\}$.

6.2 Beispiel.

(a) Aus einem Skatspiel werden nacheinander zwei Karten gezogen (ohne Zurücklegen).

$$\begin{aligned} A &\triangleq \text{„die zweite Karte ist ein As“} \\ B &\triangleq \text{„die erste Karte ist ein As“} \\ B^c &\triangleq \text{„die erste Karte ist kein As“}. \end{aligned}$$

Dann gilt

$$P(A|B) = \frac{3}{31} \approx 0,097 \qquad P(A|B^c) = \frac{4}{31} \approx 0,129.$$

BEWEIS. Sei $\Omega = \mathcal{P}_2^{\{1, \dots, 32\}}$, $P = \mathcal{L}_\Omega$, $\mathfrak{A} = \mathbb{P}(\Omega)$. Dann ist

$$\begin{aligned} A &= \{(\omega_1, \omega_2) \in \Omega \mid \omega_2 = 1, 9, 17, 25\} \\ B &= \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 = 1, 9, 17, 25\} \\ B^c &= \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 \neq 1, 9, 17, 25\} \end{aligned}$$

Es ist

$$\begin{aligned} \#\Omega &= 32 \cdot 31 = 992, & \#B &= 4 \cdot 31 = 124, \\ \#B^c &= 28 \cdot 31 = 868, & \#(A \cap B) &= 4 \cdot 3 = 12 \quad \text{und} \\ \#(A \cap B^c) &= 28 \cdot 4 = 112 \end{aligned}$$

□

- (b) Ein Experiment, das mit Wahrscheinlichkeit $p \in (0, 1)$ gelingt, wird n mal wiederholt. Es seien S die Anzahl der Erfolge und T die Nummer des Experimentes, bei dem zum ersten Mal Erfolg eintritt ($T = n + 1$, falls kein Erfolg eintritt). Dann gilt (die plausible Tatsache)

$$P\{T = k | S = 1\} = \frac{1}{n} \quad \text{für alle } k \leq n$$

Das heißt, wenn man weiß, daß ein Erfolg eingetreten ist, dann ist der Zeitpunkt der Erfolge gleichverteilt auf allen Möglichkeiten.

Es gilt $P_S = \mathcal{B}_{n,p}$ und somit

$$P\{S = 1\} = \mathcal{B}_{n,p}\{1\} = \binom{n}{1} p^1 (1-p)^{n-1} = np(1-p)^{n-1}$$

Weiter gilt $P\{T = k, S = 1\} = p(1-p)^{n-1}$

$$\Rightarrow P\{T = k | S = 1\} = \frac{P\{T = k, S = 1\}}{P\{S = 1\}} = \frac{p(1-p)^{n-1}}{np(1-p)^{n-1}} = \frac{1}{n}$$

□

6.3 Eigenschaften. Es seien $A, B \in \mathfrak{A}$ mit $P(B) > 0$.

- (a) Durch $\mathfrak{A} \ni C \mapsto P(C|B)$ wird ein W-Maß auf $(\Omega, \mathfrak{A})$ definiert mit $P(B|B) = 1$; $P(B^c|B) = 0$.

BEWEIS. Es sei $Q(C) = P(C|B)$ für $C \in \mathfrak{A}$. Dann ist

$$Q(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = 1 \quad \text{d. h. WM}_2 \text{ aus Definition 1.17}$$

Sind C_n disjunkte Mengen aus $\mathfrak{A}$, dann sind $C_n \cap B$ disjunkt und es folgt

$$\begin{aligned} Q\left(\bigcup_{n=1}^{\infty} C_n\right) &= \frac{1}{P(B)} P\left(B \cap \bigcup_{n=1}^{\infty} C_n\right) = \frac{1}{P(B)} P\left(\bigcup_{n=1}^{\infty} (B \cap C_n)\right) = \frac{1}{P(B)} \sum_{n=1}^{\infty} P(B \cap C_n) \\ &= \sum_{n=1}^{\infty} \frac{P(B \cap C_n)}{P(B)} = \sum_{n=1}^{\infty} Q(C_n) \end{aligned}$$

□

- (b) A und B sind genau dann unabhängig, falls $P(A|B) = P(A)$ gilt.

BEWEIS. A, B unabhängig

$$\Leftrightarrow P(A \cap B) = P(A) \cdot P(B) \quad \Leftrightarrow P(A) = \frac{P(A \cap B)}{P(B)} = P(A|B)$$

□

- (c) (Umkehrformel)

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

- (d) Sind $A_1, A_2, \dots, A_n \in \mathfrak{A}$, so gilt:

$$P\left(\bigcap_{k=1}^n A_k\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1})$$

Dabei gilt diese Gleichheit auch in dem Fall, daß einige der auftretenden bedingten Wahrscheinlichkeiten nicht definiert sind, wenn man folgendes setzt: $0 \cdot \text{undefiniert} = 0$.

1. Fall $P(A_1 \cap \dots \cap A_k) > 0$ für alle $k = 1, \dots, n-1$:

$$\begin{aligned} &\Rightarrow P(A_1) \cdot P(A_2|A_1) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot \dots \cdot \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})} \\ &= P(A_1 \cap \dots \cap A_n) \end{aligned}$$

2. Fall $P(A_1 \cap \dots \cap A_k) = 0$ für ein $k = 1, \dots, n$:

$$\Rightarrow P(A_1 \cap \dots \cap A_n) \leq P(A_1 \cap \dots \cap A_k) = 0 = \text{linke Seite}$$

und die rechte Seite ist null nach obiger Konvention.

□

6.4 Satz (Formel von der totalen Wahrscheinlichkeit, Formel von BAYES). Es seien $B_1, B_2, \dots$ disjunkte Ereignisse und $A \subset \bigcup_{n=1}^{\infty} B_n$. Dann gilt

$$P(A) = \sum_{n=1}^{\infty} P(B_n) P(A|B_n) \quad \text{und (im Fall } P(A) > 0) \quad P(B_k|A) = \frac{P(B_k) P(A|B_k)}{\sum_{n=1}^{\infty} P(B_n) P(A|B_n)}$$

BEWEIS. Es ist

$$P(A) = P\left(A \cap \bigcup_{n=1}^{\infty} B_n\right) = P\left(\bigcup_{n=1}^{\infty} (A \cap B_n)\right) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(B_n) P(A|B_n);$$

diese Formel ist auch im Fall $P(B_n) = 0$ gültig, wenn man $0 \cdot \text{undefiniert} = 0$ setzt. Ist $P(A) > 0$, so gilt nach der Umkehrformel

$$P(B_k|A) = \frac{P(B_k) P(A|B_k)}{P(A)} = \frac{P(B_k) P(A|B_k)}{\sum_{n=1}^{\infty} P(B_n) P(A|B_n)}$$

□

6.5 Beispiel.

(a) **(Signalübertragung)** Über einen Datenkanal wird eine Folge von Bits übertragen. Durch Störung werden eine gesendete 0 mit Wahrscheinlichkeit ε_0 als 1 und eine gesendete 1 mit Wahrscheinlichkeit ε_1 als 0 empfangen.

Das Verhältnis der Anzahl der gesendeten Einsen zu der der Nullen betrage ϱ . Wie groß ist die Wahrscheinlichkeit, daß das übertragene Bit korrekt ist, wenn wir 0 bzw. 1 empfangen haben?

$A_i \triangleq$ „ i wurde gesendet“

$B_i \triangleq$ „ i wird empfangen“

Dann wissen wir:

$$P(B_1|A_0) = \varepsilon_0, \quad P(B_0|A_1) = \varepsilon_1, \quad \text{sowie} \quad \frac{P(A_1)}{P(A_0)} = \varrho$$

Die Formel von BAYES liefert

$$P(A_0|B_0) = \frac{P(A_0) P(B_0|A_0)}{P(A_0) P(B_0|A_0) + P(A_1) P(B_0|A_1)} = \left(1 + \frac{P(A_1)}{P(A_0)} \cdot \frac{P(B_0|A_1)}{P(B_0|A_0)}\right)^{-1}$$

Da $P(B_0|A_0) = 1 - P(B_1|A_0) = 1 - \varepsilon_0$

$$P(A_0|B_0) = \left(1 + \frac{\varrho \varepsilon_1}{1 - \varepsilon_0}\right)^{-1}$$

Analog:

$$P(A_1|B_1) = \left(1 + \frac{\varepsilon_0}{\varrho(1 - \varepsilon_1)}\right)^{-1}$$

- (b) **(Lebensdauer eines Bauteils)** Ein besonders einfaches Modell für die (zufällige) Zeit T bis zum Ausfall eines Bauteils besteht in der Annahme, daß das Teil, wenn es eine bestimmte Zeit t überstanden hat, danach mit der gleichen Wahrscheinlichkeit eine weitere Zeitperiode s durchhält, wie direkt nach dem Einbau.

Es sei $T : \Omega \rightarrow \mathbb{N}$ die Zufallsvariable, die die Lebenszeit beschreibt. Dann muß

$$P\{T \geq s + t | T \geq t\} = P\{T \geq s\} \quad \forall s, t \in \mathbb{N} \quad (*)$$

gelten. Man nennt dann die Verteilung von T *gedächtnislos*.

Behauptung. Es gilt $P_T = \mathcal{G}_p$ für ein $p \in (0, 1)$.

BEWEIS. Sei $F(t) = P\{T \geq t\}$. Dann folgt:

$$\frac{F(s+t)}{F(t)} = F(s) \quad \forall s, t \in \mathbb{N} \text{ mit } F(t) > 0.$$

Daraus folgt

$$F(s+t) = F(t)F(s) \quad \forall s, t \in \mathbb{N}.$$

Sei $p = 1 - F(1)$

$$\begin{aligned} \Rightarrow F(t) &= F(1 + (t-1)) = F(1)F(t-1) = F(1)^t \\ &= (1-p)^t \quad \forall t \in \mathbb{N} \end{aligned}$$

Wegen $\lim_{t \rightarrow \infty} F(t) = \lim_{t \rightarrow \infty} P\{T \geq t\} = 0 \Rightarrow p \in (0, 1]$.

$$\begin{aligned} \Rightarrow P_T\{t\} &= P\{T = t\} = P\{T \geq t\} - P\{T \geq t+1\} \\ &= F(t) - F(t+1) = (1-p)^t - (1-p)^{t+1} \\ &= (1-p)^t(1 - (1-p)) = p(1-p)^t = \mathcal{G}_p\{t\} \end{aligned}$$

Sei umgekehrt $P_T = \mathcal{G}_p$.

$$\begin{aligned} \Rightarrow P\{T \geq t\} &= \sum_{k=t}^{\infty} p(1-p)^k = p(1-p)^t \sum_{k=0}^{\infty} (1-p)^k \\ &= (1-p)^t p \frac{1}{1 - (1-p)} = (1-p)^t \\ \Rightarrow P\{T \geq t+s | T \geq t\} &= \frac{P\{T \geq t+s\}}{P\{T \geq t\}} \\ &= (1-p)^{(t+s)}(1-p)^{-t} = (1-p)^s = P\{T \geq s\}. \end{aligned}$$

□

ENDE VON KAPITEL 6

KAPITEL 7

MARKOFF-Ketten

In diesem Kapitel untersuchen wir zufällige zeitliche Prozesse, deren zukünftige Entwicklung zwar von der Gegenwart, nicht aber von der Vergangenheit abhängt. Dabei ist die Zeit durch die diskrete Menge $\mathbb{N}$ beschrieben und die möglichen Zustände des Prozesses liegen in einem abzählbaren Zustandsraum I .

7.1 Beispiel. Es sei $(D_1, D_2, \dots)$ eine Folge identisch verteilter unabhängiger Zufallsvariablen mit Werten in $\mathbb{Z}$ und $X_m = \sum_{j=1}^m D_j$. Wir bezeichnen einen solchen stochastischen Prozeß $(X_0, X_1, \dots)$ als *Irrfahrt* auf $I = \mathbb{Z}$. Im Spezialfall $P\{D_1 = +1\} = P\{D_1 = -1\} = \frac{1}{2}$ sprechen wir von der *einfachen symmetrischen Irrfahrt*. Es ist wegen $X_{m+1} = X_m + D_{m+1}$ klar, daß der Zustand zum Zeitpunkt $m+1$ nur vom Zustand zum Zeitpunkt m abhängt, aber nicht von der weiteren Vergangenheit, denn D_{m+1} ist nach Voraussetzung von X_m unabhängig.

7.2 Beispiel (Rechner mit $c \geq 1$ Prozessoren). Der Scheduler teilt — solange noch Prozessoren frei sind — jede ankommende task einem freien Prozessor zu, der diese dann vollständig abarbeitet. Sind keine Prozessoren mehr frei, wird die ankommende task in eine Warteschlange mit der Kapazität $m \geq 0$ gesetzt; ist auch diese voll, wird die task abgewiesen. Zur Beschreibung des Zustands des Systems verwenden wir die Zahl $i \in I := \{0, 1, \dots, c+m\}$ der tasks, die gerade bearbeitet werden oder warten. Das heißt im Fall $i < c$ werden neuankommende tasks sofort bearbeitet; im Fall $c \leq i < c+m$ in die Warteschlange gesetzt und im Fall $i = c+m$ abgewiesen. Auch hier ist plausibel, daß der Zustand X_{m+1} des Systems zum Zeitpunkt $m+1$ nur von X_m abhängt, aber nicht von der weiteren Vergangenheit.

7.3 Definition. Es sei $(X_0, X_1, \dots)$ eine Folge von Zufallsvariablen mit Werten in einer endlichen oder abzählbar unendlichen Menge I . Dann heißt $(X_0, X_1, \dots)$ eine (zeithomogene) MARKOFF-Kette mit Übergangsmatrix (oder Übergangskern) $K = (K_{ij})_{i,j \in I} \in \mathbb{M}(\mathbb{R}, I, I)$, falls für alle $m \in \mathbb{N}$

$$P\{X_{m+1} = j \mid X_0 = i_0, X_1 = i_1, \dots, X_m = i_m\} = K_{i_m j}$$

für alle $j \in I$ und alle $i_0, \dots, i_m \in I$ gilt, bei denen die bedingte Wahrscheinlichkeit definiert ist.

Die MARKOFF-Eigenschaft besteht also darin, daß der Zustand zum Zeitpunkt $m+1$ von der Vergangenheit nur über den unmittelbar vorhergehenden Zustand X_m abhängt. Die zeitliche Homogenität drückt sich dadurch aus, daß der Übertragungskern K nicht von m abhängt.

7.4 Beispiel. Bei der Irrfahrt aus **Bsp. 7.1** gilt für alle $i_0, \dots, i_m, j \in \mathbb{Z}$ (falls wir $X_0 := 0$ setzen):

$$\begin{aligned} P\{X_{m+1} = j, X_m = i_m, \dots, X_0 := i_0\} \\ &= P\{X_0 = i_0, D_1 = i_1 - i_0, D_2 = i_2 - i_1, \dots, D_{m+1} = j - i_m\} \\ &= \delta_{i_0 0} \prod_{k=1}^m P\{D_k = i_k - i_{k-1}\} \cdot P\{D_{m+1} = j - i_m\} \end{aligned}$$

Analog:

$$P\{X_m = i_m, \dots, X_0 := i_0\} = \delta_{i_0 0} \prod_{k=1}^m P\{D_k = i_k - i_{k-1}\}$$

$\Rightarrow$ Falls $P\{X_m = i_m, \dots, X_0 := i_0\} > 0$

$$\Rightarrow P\{X_{m+1} = j | X_0 = i_0, \dots, X_m = i_m\} = P\{D_{m+1} = j - i_m\} = P\{D_1 = j - i_m\} =: K_{i_m j}$$

$\Rightarrow$ Irrfahrt ist MARKOFF-Kette.

7.5 Satz. Es sei $(X_0, X_1, \dots)$ eine MARKOFF-Kette mit Übergangsmatrix K .

- (a) Es gilt $P\{X_{m+1} = j | X_m = i\} = K_{ij}$ für alle $m \in \mathbb{N}$ und $i, j \in I$ mit $P\{X_m = i\} > 0$.
- (b) Für alle $i \in I$, so daß $P\{X_m = i\} > 0$ für mindestens ein $m \in \mathbb{N}$ gilt $K_{ij} \geq 0 \forall j \in I$ und $\sum_{j \in I} K_{ij} = 1$.
- (c) Es sei $(\sigma_i, i \in I)$ der durch $\sigma_i := P\{X_0 = i\}$ definierte Zeilenvektor, der die Startverteilung beschreibt. Dann gilt für alle $m \in \mathbb{N}$

$$P\{X_m = i\} = (\sigma K^m)_i,$$

(wobei $K^m = K \cdot \dots \cdot K$ die m -te Potenz von K ist).

BEWEIS.

- (a) Mit der Formel für die totale Wahrscheinlichkeit (Satz 6.4) und Definition 7.3:

$$\begin{aligned} P\{X_{m+1} = j, X_m = i\} &= \sum_{i_0, \dots, i_{m-1} \in I} P\{X_{m+1} = j, X_m = i, X_{m-1} = i_{m-1}, \dots, X_0 = i_0\} \\ &= \sum_{i_0, \dots, i_{m-1} \in I} K_{ij} \cdot P\{X_m = i, X_{m-1} = i_{m-1}, \dots, X_0 = i_0\} \\ &= K_{ij} \cdot P\{X_m = i\}. \end{aligned}$$

- (b) Nach Fig. 6.3(a) ist $M \mapsto P\{X_{m+1} \in M | X_m = i\}$ ein W-Maß auf I , falls $P\{X_m = i\} > 0$.

- (c) Induktion nach m :

$$m = 0: K^0 = I \Rightarrow P\{X_0 = i\} = (\sigma K^0)_i = (\sigma I)_i = \sigma_i.$$

$m \rightarrow m + 1$: Mit der Formel für die totale Wahrscheinlichkeit und Teil (a) gilt:

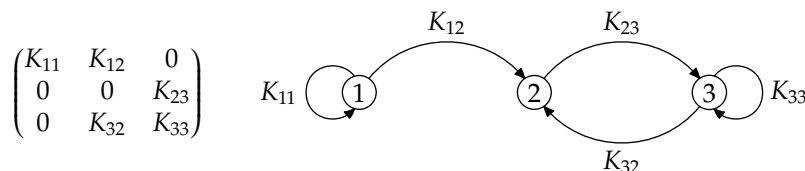
$$\begin{aligned} P\{X_{m+1} = j\} &= \sum_{i \in I} P\{X_m = i\} \cdot P\{X_{m+1} = j | X_m = i\} \\ &= \sum_{i \in I} (\sigma K^m)_i K_{ij} = (\sigma K^{m+1})_j. \end{aligned}$$

□

7.6 Bemerkung.

1. Die Übergangsmatrix $K = (K_{ij})$ läßt sich durch einen bewerteten Graphen beschreiben, wobei die Knoten gerade die Zustände $i \in I$ sind und die gerichtete Kante von i nach j die Bewertung K_{ij} bekommt. Kanten mit $K_{ij} = 0$ werden dabei weggelassen.

Beispiel: $I = \{1, 2, 3\}$



2. **Satz 7.5(c)** erlaubt es uns, die Verteilung von X_m durch den Vektor $\sigma \cdot K^m$ ihrer Zähldichte auszudrücken. Insbesondere beschreibt σ die Startverteilung des Prozesses, d.h. die Verteilung von X_0 .
3. Die Eigenschaft der Übergangsmatrix aus **Satz 7.5(b)** drückt man kurz dadurch aus, daß man K *zeilenstochastisch* nennt. Dabei kann man auch ohne die Gültigkeit der Voraussetzung $P\{X_m = i\} > 0$ für ein $m \in \mathbb{N}$ o. B. d. A. K als zeilenstochastisch annehmen, da im Fall $P\{X_m = i\} = 0 \ \forall m \in \mathbb{N}$ der Zustand $i \in I$ nie getroffen wird und wir $I \setminus \{i\}$ als Zustandsraum wählen können.
4. Sind $K, L \in M(\mathbb{R}, I, I)$ zeilenstochastisch, so ist auch $K \cdot L$ zeilenstochastisch:

Für $i, k \in I$ gilt

$$(K \cdot L)_{ik} = \sum_{j \in I} K_{ij} L_{jk} \geq 0 \quad \text{und}$$

$$\sum_{k \in I} (K \cdot L)_{ik} = \sum_{k \in I} \sum_{j \in I} K_{ij} L_{jk} = \sum_{j \in I} K_{ij} \sum_{k \in I} L_{jk} = \sum_{j \in I} K_{ij} \cdot 1 = 1$$

7.7 Beispiel. (vgl. **Bsp. 7.2**) Rechner mit $c \geq 0$ Prozessoren und Warteschlange der Länge $m \geq 0$. Wir nehmen an, daß pro Zeiteinheit die Wahrscheinlichkeit für die Beendigung einer bestimmten task $\mu > 0$ ist. Mit Wahrscheinlichkeit $\lambda \geq 0$ tritt neue task ein. Diese Wahrscheinlichkeiten seien unabhängig von der Vorgeschichte.

Weiter nehmen wir an, daß die Zeiteinheit so klein ist, daß das gleichzeitige Eintreffen bzw. Fertigwerden vernachlässigt werden kann; insbesondere soll $\lambda + c\mu < 1$ gelten.

Es bezeichne X_n den Zustand des Systems und $I = \{0, 1, \dots, c + m\}$. Die Übergangsmatrix $(K_{ij}) = k$ können wir nun mit **Satz 7.5(a)** berechnen:

Für $0 \leq i \leq c$ gilt:

$$K_{ij} = \begin{cases} i\mu & j = i - 1 \\ \lambda & j = i + 1 \\ 1 - \lambda - i\mu & j = i \\ 0 & \text{sonst} \end{cases}$$

Für $c < i < c + m$:

$$K_{ij} = \begin{cases} c\mu & j = i - 1 \\ \lambda & j = i + 1 \\ 1 - \lambda - c\mu & j = i \\ 0 & \text{sonst} \end{cases}$$

und für $i = c + m$:

$$K_{c+m, j} = \begin{cases} c\mu & j = c + m - 1 \\ 1 - c\mu & j = c + m \\ 0 & \text{sonst} \end{cases}$$

$$\Rightarrow K := (P\{X_{m+1} = j | X_m = i\})_{ij}$$

$$= \begin{pmatrix} 1 - \lambda & \lambda & 0 & \cdots & \cdots & 0 \\ \mu & 1 - \lambda - \mu & \lambda & 0 & & \vdots \\ 0 & 2\mu & 1 - \lambda + 2\mu & \lambda & 0 & \vdots \\ \vdots & \vdots & \cdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & c\mu & 1 - \lambda - c\mu & \lambda \\ 0 & 0 & \cdots & 0 & c\mu & 1 - c\mu \end{pmatrix}$$

7.8 Satz (MARKOFF (1907)). Es sei $(X_n)_{n \in \mathbb{N}}$ eine MARKOFF-Kette mit endlichem Zustandsraum I und Übergangsmatrix K . Es existiere ein $N \geq 1$, $j_0 \in I$, so daß $K_{ij_0}^N > 0$ für alle $i \in I$. Dann gibt es eine Zähldichte $\pi \in \mathbb{R}_+^I$ mit

$$P\{X_n = i\} \rightarrow \pi_i \quad \text{für alle } i \in I,$$

und π ist unabhängig von der Startverteilung.

BEWEIS. Da I endlich ist $\Rightarrow \varepsilon := \min_{i \in I} K_{ij_0}^N > 0$.

Wir betrachten den vollständigen metrischen Raum

$$X := \left\{ \pi \in \mathbb{R}_+^I \mid \sum_{i \in I} \pi_i = 1 \right\} \quad \text{mit} \quad d(\pi, \varrho) := \sum_{i \in I} |\pi_i - \varrho_i|.$$

Dann ist $\varphi : X \rightarrow X$, $\varphi(\sigma) = \sigma K$ eine Kontraktion (d. h. $d(\varphi(\sigma), \varphi(\pi)) \leq d(\sigma, \pi)$).

Wegen $\sum_{i \in I} (\pi_i - \varrho_i) = 1 - 1 = 0$ gilt für $\varphi^N(\sigma) = \sigma K^N$:

$$\begin{aligned} d(\varphi^N(\pi), \varphi^N(\varrho)) &= \sum_{j \in I} |(\pi K^N)_j - (\varrho K^N)_j| = \sum_{j \in I} \left| \sum_{i \in I} (\pi_i - \varrho_i) K_{ij}^N \right| \\ &\leq \sum_{j \in I \setminus \{j_0\}} \sum_{i \in I} |\pi_i - \varrho_i| K_{ij}^N + \left| \sum_{i \in I} (\pi_i - \varrho_i) K_{ij_0}^N \right| \\ &= \sum_{j \in I \setminus \{j_0\}} \sum_{i \in I} |\pi_i - \varrho_i| K_{ij}^N + \left| \sum_{i \in I} (\pi_i - \varrho_i) (K_{ij_0}^N - \varepsilon) \right| \\ &\leq \sum_{i \in I} |\pi_i - \varrho_i| \left(\sum_{j \in I \setminus \{j_0\}} K_{ij}^N + \varepsilon \right) = (1 - \varepsilon) d(\pi, \varrho). \end{aligned}$$

Mit dem BANACHSchen Fixpunktsatz konvergiert also $\varphi^{kN}(\sigma) = \sigma K^{kN}$ unabhängig vom Startwert σ gegen den eindeutig bestimmten Fixpunkt π von φ^N . Wegen $\varphi^N(\varphi(\pi)) = \varphi(\varphi^N(\pi)) = \varphi(\pi)$ und der Eindeutigkeit folgt $\varphi(\pi) = \pi$. Da φ Kontraktion

$$\Rightarrow d(\varphi(\sigma), \pi) = d(\varphi(\sigma), \varphi(\pi)) \leq d(\sigma, \pi)$$

und somit für $n \geq kN$:

$$d(\varphi^n(\sigma), \pi) \leq d(\varphi^{kN}(\sigma), \pi) \rightarrow 0 \quad (n \rightarrow \infty)$$

d. h. $\sigma K^n \rightarrow \pi$. □

7.9 Korollar. Unter den Voraussetzungen von Satz 7.8 gilt $K^n \rightarrow \bar{K}$, wobei

$$\bar{K} = \begin{pmatrix} - & \pi & - \\ & \vdots & \\ - & \pi & - \end{pmatrix}$$

BEWEIS. Für $l \in I$ sei $(e_l)_{l \in I} = (\delta_{li})_{i \in I}$. Dann ist e_l eine Startverteilung und nach Satz 7.8 gilt $e_l K^n \rightarrow \pi$, wobei π unabhängig von $l \in I$ ist. Für $i \in I$, $l \in I$ gilt dann:

$$(e_l K^n)_i = K_{li}^n \rightarrow \pi_i, \text{ d. h.}$$

$K^n \rightarrow \bar{K}$ und $\bar{K}_{li} \in \pi_i \forall l, i \in I$. □

7.10 Satz. Es sei $K \in \mathbb{M}(\mathbb{R}, I, I)$ eine zeilenstochastische Matrix und $\sigma \in \mathbb{R}_+^I$ eine Startverteilung, für die $\pi = \lim_{n \rightarrow \infty} \sigma K^n$ existiert. Dann ist π eine Zähldichte (d. h. $\pi_i \geq 0$; $\sum_{i \in I} \pi_i = 1$) und es gilt $\pi \cdot K = \pi$.

BEWEIS. (Hier nur für endliche I) Nach Satz 7.5(c) ist $\sigma \cdot K^n$ eine Zähldichte für alle $n \geq 1 \Rightarrow \pi$ Zähldichte. Es ist

$$\pi \cdot K = \left(\lim_{n \rightarrow \infty} \sigma K^n \right) \cdot K = \lim_{n \rightarrow \infty} \sigma K^{n+1} = \pi$$

□

7.11 Satz. Es sei π ein stationärer Vektor einer MARKOFF-Kette $(X_0, X_1, \dots)$ mit Übertragungsmatrix K und Startverteilung $P\{X_0 = i\} = \pi \ \forall i \in I$. Dann gilt

$$P\{X_m = i\} = \pi_i \text{ für alle } m \in \mathbb{N}, i \in I.$$

BEWEIS. π stationär $\Leftrightarrow \pi \cdot K = \pi$. Durch Induktion folgt $\pi \cdot K^m = \pi$. Aus Satz 7.5(c) folgt dann die Behauptung. $\square$

7.12 Beispiel. (vgl. Bsp. 7.7) Wir wenden Satz 7.8 auf die Übergangsmatrix K in (7.7) an. Da K eine Drei-Band-Matrix ist, folgt $K_{ij_0}^N > 0$ für alle $i \in I$ und jedes $j_0 \in I$ für $N = c + m - 1$. Nach (7.8) existiert also eine eindeutige stationäre Verteilung π , die nach Satz 7.10 das Gleichungssystem $\pi \cdot K = \pi$ für $\pi = (\pi_0, \dots, \pi_{c+m})$ erfüllt

$$\Rightarrow \begin{cases} \mu \pi_1 = \lambda \pi_0 \\ (k+1)\mu \pi_{k+1} = (\lambda + k\mu)\pi_k - \lambda \pi_{k-1} & \text{für } 1 \leq k \leq c-1 \\ c\mu \pi_{c+1} = (\lambda + c\mu)\pi_c - \lambda \pi_{c-1} & \text{für } c \leq k \leq c+m-1 \\ c\mu \pi_{c+m} = \lambda \pi_{c+m-1} \end{cases}$$

Sei $\varrho = \frac{\lambda}{\mu}$. Dividiert man durch $\mu > 0$, folgt

$$\pi_k = \frac{\varrho^k}{k!} \pi_0 \quad \text{für } k \leq c \quad \text{und} \quad \pi_k = \frac{\varrho^k}{c!c^k} \pi_0 \quad \text{für } c+1 \leq k \leq c+m.$$

Da $\sum_{i=0}^{c+m} \pi_i = 1$ folgt

$$\pi_0 = \left(\sum_{k=0}^{c-1} \frac{\varrho^k}{k!} + \frac{\varrho^c}{c!} \cdot \frac{1 - \left(\frac{\varrho}{c}\right)^{m+1}}{1 - \frac{\varrho}{c}} \right)^{-1}$$

Interpretation. Die Zahl $\varrho = \frac{\lambda}{\mu}$ gibt an, wieviel mal mehr tasks ankommen als ein Prozessor bearbeiten kann.

- (i) Im Überlastfall $\varrho \geq c$ ist $\pi_k \leq \pi_{k+1}$. $\Rightarrow$ Der Zustand $c+m$ hat die größte Wahrscheinlichkeit; das System lehnt wegen Überlastung also in der Regel die tasks ab. Vergrößern der Warteschlange nützt hier nichts.
- (ii) Ist $\varrho < c$, so ist der Zustand mit der höchsten Wahrscheinlichkeit π_k für $k = [\varrho] < c$ erreicht, also im Bereich, wo noch Prozessoren frei sind. Im Bereich der Warteschlangenplätze $\{c+1, \dots, c+m\}$ nehmen die Wahrscheinlichkeiten exponentiell ab, und zwar umso schneller, je kleiner $\frac{\varrho}{c}$ ist. Man kann also die Warteschlange umso kleiner dimensionieren, je kleiner dieser Quotient ist.

ENDE VON KAPITEL 7
