

Wahrscheinlichkeiten

Manfred Jäger-Ambrożewicz

www.mathfred.de

www.mathstat.de

6. Mai 2025

Das Skript ist teilweise fragmentarisch und hat bestimmt etliche Mängel/Fehler. Auch die Reihenfolge/Gliederung ist noch nicht final. Über Hinweise würde ich mich freuen; z.B. per Email an jaegera at htw-berlin Punkt de.

Das Skript genügt (noch?) nicht den Anforderungen, die an eine wissenschaftliche Arbeit gestellt werden, denn die Quellen zu vielen Aussagen werden (noch) nicht vor Ort angegeben.

Inhaltsverzeichnis

1	Basis der Wahrscheinlichkeitstheorie	5
1.1	Grundlegendes: Axiome für Ereignisse deren Wahrscheinlichkeiten	5
1.2	Erste Regeln	13
1.3	Laplace	15
1.4	Bedingte Wahrscheinlichkeiten und Unabhän- gigkeit	16
1.5	Unendlich oft	26
1.6	Exkurs Kombinatorik	29
2	Diskrete Modelle	38
2.1	Allgemeins	38
2.2	Dirac und Laplace	44
2.3	Bernoulli, Binomial und Multinomial	45
2.4	Hypergeometrisch	48
2.5	Belegungsmodelle, Maxwell-Boltzmann, Bose- Einstein, Fermi-Dirac	49
2.6	Poisson	55
2.7	Geometrisch und Pascal	57
3	Standardmodelle mit Dichten	59
3.1	Überabzählbar ist indiskret	59
3.2	Verteilung mit Dichte	60
3.3	Gleich-, Normal und Exponentialverteilung	64

3.4	Transformationen	70
3.5	Gamma, Chi und Beta-Verteilung	71
4	Verteilungs- und Quantilfunktionen	74
4.1	Verteilungsfunktion	74
4.2	Quantile, Quantilsfunktionen und Verallgemeinerte Inverse	78
4.3	Risikomessung	83
5	Lebesgue-Stieltjes Integral - Erwartungswert bezüglich \mathbb{P} bzw. F	99
5.1	Grundlagen und Maßtheoretische Induktion	99
5.2	Konvergenzsätze	105
5.3	Linearität und Monotonie	106
5.4	Dreiecksungleichung und Cauchy-Bunjakowski-Schwarz-Ungleichung	107
5.5	Varianz	108
5.6	Ungleichungen: Markow, Chebycehv, Jensen	110
6	Zufallsvektoren	112
7	Konvergenz von Folgen von Zufallsvariablen	130
8	Gesetze der großen Zahlen	146
9	Zentraler Grenzwertsatz	161
10	Maßtheorie	177
10.1	Einführung zur Maßtheorie	177
10.2	Mengensysteme	178
10.3	Inhalt, Prämaß, Maß	180
10.4	Fortsetzungssätze	183
10.5	Maße mit F definieren	187
10.6	Vollständigkeit	191

11 Messbare Abbildungen	193
11.1 Grundlegende Begriffe	193
11.2 Treppenfunktionen	204
11.3 $\bar{\mathbb{R}}$ und numerische Funktionen	205
12 Integration	207
13 Bedingte Erwartung – bedingte Wahrscheinlichkeit	217
13.1 Bedingte Erwartungen gegeben B	217
13.2 Radon-Nikodym	218
13.3 Bedingte Erwartung bezüglich \mathcal{F}	220
13.4 Bedingte Wahrscheinlichkeit und bedingte Verteilung	234
13.5 Totale Wahrscheinlichen	242
14 Erzeugende Funktionen	244
15 Charakteristische Funktionen	245
Quellenverzeichnis	254

1 Basis der Wahrscheinlichkeitstheorie

1.1 Grundlegendes: Axiome für Ereignisse deren Wahrscheinlichkeiten

In diesem Kapitel werden einige Grundbegriffe und dazugehörige Resultate eingeführt. Die Erläuterungen, Definitionen und Ergebnisse sind so grundlegend, dass sie sich in nahezu jeder Quelle finden, z.B. in Billingsley [2], Blitzstein und Hwang [4], Georgii [13], Henze [18] und Krenzel [22]. Aus meiner Sicht eignen sich vor allem Henze [18] und Blitzstein und Hwang [4].

1.1.1 Bemerkung: i.) In der Wahrscheinlichkeitstheorie wird ein Rahmen für die systematische mathematische Analyse von *unsicheren Sachverhalten* entwickelt. Für solche Sach-

verhalte verwenden wir den Begriff **Experiment**.

Die **Menge der möglichen Ergebnisse** des Experiments bezeichnen wir mit

$\Omega \neq \emptyset$ ist die Menge der Ergebnisse.

Ich stelle mir unter Ω regelmäßig die Menge der möglichen **präzisen Protokolle** vor, die ein gewissenhafter Protokollant anfertigen könnte.

Für ein bestimmtes **Ergebnis** $\omega \in \Omega$ verwenden wir typischerweise das Symbol

ω ist ein typisches Ergebnis.

Für **bestimmte Teilmengen**

$$A \subset \Omega$$

verwenden wir den Begriff **Ereignis**. Ereignisse sind also „zusammengesetzte“ Ergebnisse; Ereignisse sind in einer Menge zusammengefasste Ergebnisse.

ii.) Im Mittelpunkt der Wahrscheinlichkeitstheorie steht (natürlich) der Begriff der Wahrscheinlichkeit. Es sei $A \subset \Omega$ ein **Ereignis**. Wir geben zwei denkbare **Interpretationen** der Wahrscheinlichkeit $\mathbb{P}(A)$ von A an (vgl. z.B. Georgii (2015, Seite 14)).

- **Frequentistische Interpretation:** Wir stellen uns vor, dass ein Experiment *zu gleichen Bedingungen* unendlich oft wiederholt wird. Dann ist $\mathbb{P}(A)$ die relative Häufigkeit (Frequenz) mit der das Ereignis A *typischerweise* eintritt.
- **Grad der Gewissheit:** Die Wahrscheinlichkeit von A ist der **Grad der Gewissheit** mit der A eintritt. Wenn ich auf das Eintreten des Ereignisses A wetten würde, dann würde ich mich dabei an dem Wert $\mathbb{P}(A)$ orientieren.

c.) Für beide Interpretationen (sowie für andere Interpretationen) lassen sich gute Gründe bzw. passende Beispiele angeben, leider aber auch jeweils unpassende. Die Mathematik (in engeren Sinn) *funktioniert* unabhängig von der Interpretation, so dass wir uns mit der Interpretation *vorläufig* nicht beschäftigen müssen. Wir verwenden nämlich den sogenannten **axiomatischen Ansatz**, der die *vernünftigen Regeln für den Umgang mit Wahrscheinlichkeiten festlegt*, ohne eine Definition für Wahrscheinlichkeit anzugeben.¹ Bezogen auf die Wahrscheinlichkeitstheorie geht dieses Vorgehen auf den Mathematiker Andrei Kolmogorow zurück.²

Sie werden aber später im Studium und insbesondere in der

¹Wahrscheinlichkeit ist dann ein sogenannter undefinierter Grundbegriff vergleichbar zu Punkt in der Geometrie.

²Quelle

Statistik feststellen, dass die Interpretationen des Begriffs der Wahrscheinlichkeit für die Entwicklung bzw. die Auswahl mathematischer Modelle ausschlaggebend sein kann; insbesondere für die Unterscheidung zwischen frequentistischer und bayesianischer Statistik.

1.1.2 Beispiel: Wir betrachten das Experiment, das darin besteht, einen Würfel (mit sechs Seiten) zu werfen. Die **möglichen Ergebnisse** erfassen wir durch

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Dabei erfasst $\omega = 3$ den Sachverhalt, dass der Würfel so liegt, dass die Seite mit drei Augen oben liegt. Wir wählen die Potenzmenge $\mathcal{P}(\Omega)$ als **Menge der Ereignisse**. Das Ereignis „eine gerade Zahl wurde gewürfelt“ wird durch $A = \{2, 4, 6\}$ repräsentiert. Das Ereignis „eine 1 oder eine 6 wurde gewürfelt, wird durch $B = \{1\} \cup \{6\} = \{1, 6\}$ repräsentiert. Wir beobachten hier, dass es eine Entsprechung zwischen der (normal-)sprachlichen Verwendung des Wortes „oder“ und der Verwendung der Verknüpfung \cup aus der Mengenlehre gibt. Das Ereignis „eine gerade Zahl und eine Zahl größer als 2“ wird also durch $\{4, 6\} = \{2, 4, 6\} \cap \{3, 4, 5, 6\}$ repräsentiert. Die Verknüpfung \cap hat also ihre Entsprechung im „und“. Das Ereignis „eine gerade Zahl wurde nicht gewürfelt“ wird durch $\{1, 3, 5\} = \{2, 4, 6\}^c$ repräsentiert. Das „nicht“

aus der Umgangssprache gehört also zum c der Mengenlehre.

1.1.3 Bemerkung: Die im vorhergehenden Beispiel gefundenen Entsprechungen gelten allgemein. Zu den bekannten Verknüpfungen von Aussagen gehören ebenfalls bekannte Verknüpfungen von Mengen (also Ereignissen).

- Wir haben die Paare:
 - \cup gehört zum *oder*
 - \cap gehört zum *und*
 - c gehört zum *nicht*

1.1.4 Beispiel: Angenommen Sie würfeln so lange, bis Sie schließlich eine 6 würfeln. Wir wollen davon ausgehen, dass Sie unter Umständen unendlich lange würfeln, also trotz andauernder Versuche immer eine von 6 verschiedene Zahl würfeln. Etwaige praktische Einwände gegen die Unmöglichkeit ∞ -vieler Würfe dürfen wir als Mathematiker*innen ignorieren. Die Ergebnismenge ist dann $\Omega = \{\infty, 1, 2, 3, \dots\}$. An diesen Beispiel erkennen wir, dass wir auch Mengen Ω mit unendlich vielen Elementen betrachten sollen.

1.1.5 Definition: 1.) Es sei $\Omega \neq \emptyset$ eine Menge. Ein Teil-

mengensystem $\mathcal{F} \subset \mathcal{P}(\Omega)$ heißt **σ -Algebra** auf Ω , falls:

- i.) $\Omega \in \mathcal{F}$.
- ii.) $A \in \mathcal{F}$ impliziert $A^c := \Omega \setminus A \in \mathcal{F}$.
- iii.) $A_i \in \mathcal{F}, i \in \mathbb{N}$ impliziert $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

2.) Es sei $\Omega \neq \emptyset$ eine Menge und \mathcal{F} eine σ -Algebra auf Ω . Dann heißt das Paar (Ω, \mathcal{F}) **messbarer Raum**.

3.) Es sei (Ω, \mathcal{F}) ein messbarer Raum. Eine Abbildung

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$$

heißt **Wahrscheinlichkeitsmaß** auf (Ω, \mathcal{F}) , falls:

- i.) $\mathbb{P}(\Omega) = 1$.
- ii.) Für jede Folge $A_i, i \in \mathbb{N}$ von paarweise disjunkten Ereignissen $A_i \in \mathcal{F}$ gilt

$$\mathbb{P} \left(\biguplus_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

4.) Es sei (Ω, \mathcal{F}) ein messbarer Raum und \mathbb{P} ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{F}) . Dann heißt das Triple $(\Omega, \mathcal{F}, \mathbb{P})$ **Wahrscheinlichkeitsraum**.

► Wäre $\Omega = \emptyset$, dann einerseits $\mathbb{P}(\emptyset) = \mathbb{P}(\Omega) = 1$. Andererseits $\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \uplus \emptyset \uplus \emptyset \dots) = \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \dots$. Dann muss also andererseits $\mathbb{P}(\emptyset) = 0$ gelten. Die Voraussetzung $\Omega \neq \emptyset$

ist also notwendig, um den Widerspruch $0 = 1$ zu vermeiden.

1.1.6 Bemerkung: In der einführenden Bemerkung (1.1.1) haben wir (nahezu) beiläufig angemerkt, dass wir **nur bestimmte** Teilmengen von Ω **Ereignisse nennen**. Wir werden in der Tat nur die Elemente $A \in \mathcal{F}$ einer σ -Algebra \mathcal{F} als Ereignis bezeichnen. Dafür gibt es zwei Gründe, von denen wir den ersten erst später – wenn wir etwas Maßtheorie besprochen haben – erläutern werden. Es gibt jedoch einen elementaren Grund, warum wir nicht alle Teilmengen als Ereignisse auffassen.

Wir verwenden die σ -Algebra \mathcal{F} zur Erfassung des **Informationsstands**³ (eventuell Zwischenstands) des Beobachters des Experiments nach dem Ende des Experiments. Für den Fall des Würfels: Die σ -Algebra $\mathcal{F} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$ erfasst, dass der Beobachter lediglich beobachtet, ob eine gerade Zahl oder eine ungerade Zahl gewürfelt wurde.



Um das zu veranschaulichen, betrachten wir den abgebildeten **abgeklebten Würfel** erklären.⁴ Die ungeraden Zahlen sind rot abgeklebt und die geraden Zahlen blau. Man kann

³Wir beachten dabei den Hinweis von Billingsley (1995, S. 57 ff), dass wir streng zwischen mathematischen Aussagen und deren Interpretation unterscheiden müssen.

⁴Diese Illustration ist originär- :-).

die Zahlen unter dem Klebeband erkennen, wir sollen/wollen uns aber vorstellen, dass wir die Zahlen unter dem Klebeband nicht erkennen können. Wir wissen lediglich, dass sich unter rot eine ungerade und unter blau eine gerade Augenzahl verbirgt.

Angenommen eine 3 wurde gewürfelt. Der Beobachter sieht *rot*. Der Beobachter mit der **Teilinformation** \mathcal{F} weiß nicht, ob eine 3, eine 1 bzw. eine 5 gewürfelt wurde. Er weiß, dass $A = \{1, 3, 5\}$ eingetreten ist. Die $\omega \in A$ sind alle rot abgeklebt und nicht unterscheidbar.

1.1.7 Bemerkung: Die folgende Übersicht fasst noch mal die Begriffe zusammen

- Ω ist die **Menge der möglichen Ergebnisse (Protokolle)**. Die Elemente von Ω werden mit ω bezeichnet. Ein ω ist also **ein mögliches Ergebnis (Protokoll)**. Wenn man WT mit Blick auf die Anwendung in der Statistik betreibt, dann nennt man Ω manchmal auch den **Stichprobenraum**.
- Das Teilmengensystem \mathcal{F} der messbaren Mengen bildet eine **σ -Algebra**. Die Mengen $A \in \mathcal{F}$ heißen **Ereignisse**; und nur diese Teilmengen $A \in \mathcal{F}$ von Ω heißen Ereignisse. A ist also die Zusammenfassung von Ergebnissen zu einer Menge.

- Auf der Menge \mathcal{F} definieren wir das **Wahrscheinlichkeitsmaß** \mathbb{P} . Der Definitionsbereich von \mathbb{P} ist also \mathcal{F} und nicht Ω . Die Zielmenge von \mathbb{P} ist $[0, 1]$. Also

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1].$$

Die charakteristischen Eigenschaften eines Wahrscheinlichkeitsmaßes sind:

- i.) $\mathbb{P}(\Omega) = 1$
- ii.) \mathbb{P} ist σ -additiv.

1.2 Erste Regeln

1.2.1 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Dann gilt:

- 1.) $\mathbb{P}(\emptyset) = 0$.
- 2.) Für disjunkte⁵ $A, B \in \mathcal{F}$ gilt:

$$\mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B).$$

⁵Disjunkt bedeutet, dass sich die Ereignisse ausschließen. Also $A \cap B = \emptyset$

3.) Für alle $A \in \mathcal{F}$ gilt

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

4.) Für alle $A, B \in \mathcal{F}$ mit $A \subset B$ gilt:

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A).$$

5.) Für alle $A, B \in \mathcal{F}$ gilt:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

6.) Für alle $A, B \in \mathcal{F}$ gilt:

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c).$$

7.) Für $A, B \in \mathcal{F}$ mit $A \subset B$ gilt

$$\mathbb{P}(A) \leq \mathbb{P}(B).$$

Beweis: Zu 1.) Es gilt $\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \uplus \emptyset \uplus \dots) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$.
Wäre $\mathbb{P}(\emptyset) = \varepsilon > 0$, dann würde $\sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$ divergieren und nicht gleich ε . Nur für $\mathbb{P}(\emptyset) = 0$ kann $\mathbb{P}(\emptyset) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$ gelten.

Zu 2.) Es gilt $\mathbb{P}(A \cup B \cup \emptyset \cup \emptyset \dots) = \mathbb{P}(A) + \mathbb{P}(B) + 0 + 0 + \dots = \mathbb{P}(A) + \mathbb{P}(B)$.

Zu 3.) Es gilt $A \uplus A^c = \Omega$. Also $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \uplus A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$.

Zu 4.) Es gilt $A \uplus (B \setminus A) = B$ (beachte: $A \subset B$ wird genutzt!). Also $\mathbb{P}(B) = \mathbb{P}(A \uplus (B \setminus A)) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$.

Zu 5.) Es gilt $A \cup B = (A \setminus (A \cap B)) \uplus (B \setminus (A \cap B)) \uplus (A \cap B)$
 . DIY

Zu 6.) Es gilt $A = A \cap \Omega = A \cap (B \uplus B^c) = (A \cap B) \uplus (A \cap B^c)$.
 Also $\mathbb{P}(A) = \mathbb{P}((A \cap B) \uplus (A \cap B^c)) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$.

1.2.2 Satz: Für jede Folge $A_i \in \mathcal{F}, i \in \mathbb{N}$ von Ereignissen in (Ω, \mathcal{F}) gilt die Sub- σ -Additivität

$$\mathbb{P} \left(\bigcup_{i \in \mathbb{N}} A_i \right) \leq \sum_{i \in \mathbb{N}} \mathbb{P}(A_i).$$

1.3 Laplace

1.3.1 Definition: Ein $(\Omega, \mathcal{F}, \mathbb{P})$ heißt **Laplace Experiment** oder **Laplace Wahrscheinlichkeitsraum**, falls:

- i.) $\Omega \neq \emptyset$ ist eine endliche Menge mit n Elementen.
- ii.) $\mathcal{F} = \mathcal{P}(\Omega)$.

iii.) Für alle $A \in \mathcal{F}$ ist

$$\mathbb{P}(A) = \frac{\#(A)}{\#(\Omega)} = \frac{\#(A)}{n} = \frac{\text{Anzahl der für } A \text{ günstigen Ergebnisse}}{\text{Anzahl aller Ergebnisse}}.$$

1.4 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

1.4.1 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$ mit $\mathbb{P}(B) > 0$. Dann heißt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

die **bedingte Wahrscheinlichkeit von A gegeben B** .

1.4.2 Satz: Es sei $B \in \mathcal{F}$ ein Ereignis mit $\mathbb{P}(B) > 0$. Dann ist die Funktion

$$\mathbb{P}^B : \begin{cases} \mathcal{F} \rightarrow [0, 1] \\ A \mapsto \mathbb{P}(A|B) \end{cases}$$

ein (weiteres) Wahrscheinlichkeitsmaß auf (Ω, \mathcal{F}) .

1.4.3 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$. A und B heißen **unabhängig**, falls

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

1.4.4 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$ mit $\mathbb{P}(B) > 0$. Dann gilt: A und B sind genau dann unabhängig, wenn $\mathbb{P}(A|B) = \mathbb{P}(A)$ gilt.

1.4.5 Bemerkung: Die Konzepte disjunkt bzw. unabhängig werden von Anfängern leider oft verwechselt. Dabei bedeuten die Konzepte etwas grundsätzlich verschiedenes. Unabhängig bedeutet, dass wir die Wahrscheinlichkeitseinschätzung für A bei Eintreten von B nicht revidieren. Wenn A und B disjunkt sind, dann revidiert man die Wahrscheinlichkeitseinschätzung aber sehr wohl (jedenfalls typischerweise). Wenn das Eintreten von B , das Eintreten von A ausschließt, so ist $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(\emptyset)/\mathbb{P}(B) = 0$. Wenn also $\mathbb{P}(A) > 0$ ist, dann wird die Wahrscheinlichkeitseinschätzung *radikal* revidiert. Die Eigenschaft der Disjunktheit bzw. des sich Ausschließens, stellt also **eine extreme Form der Abhängigkeit dar**.

1.4.6 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$. Wenn A und B unabhängig sind, dann sind auch A und B^c unabhängig.

1.4.7 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und I eine abzählbare Indexmenge. Eine Folge $B_i \in \mathcal{F}, i \in I$ heißt **Nullmengenlose abzählbare Partition**

von Ω , falls:

- i.) Die Menge B_i sind paarweise disjunkt.
- ii.) Für alle $i \in I$ gilt $\mathbb{P}(B_i) > 0$.
- iii.) Es gilt

$$\Omega = \bigsqcup_{i \in I} B_i.$$

1.4.8 Satz von der totalen Wahrscheinlichkeit oder Cocktailformel oder Fallunterscheidungsformel: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $B_i, i \in I$ eine **Nullmengenlose abzählbare Partition** von Ω . Dann gilt für alle $A \in \mathcal{F}$

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

[Wie viel Vitamin C ist in B_i und wie viel B_i ist in A .]

1.4.9 Satz von Bayes: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $B_i, i \in I$ eine **Nullmengenlose abzählbare Partition** von Ω . Dann gilt für alle $A \in \mathcal{F}$ mit $\mathbb{P}(A) > 0$

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)} = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\mathbb{P}(A)}.$$

Für $\Omega = B \uplus B^c$ erhalten wir insbesondere

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}.$$

1.4.10 Definiton: Es sei I eine nichtleere Indexmenge und $A_i \in \mathcal{F}, i \in I$ eine Familie von Ereignissen.

i.) Die Familie $A_i \in \mathcal{F}, i \in I$ heißt **unabhängig**, wenn für alle endlichen $J \subset I$

$$\mathbb{P} \left(\bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbb{P}(A_j).$$

ii.) Die Ereignisse heißen **paarweise unabhängig**, wenn für alle $i, j \in I, i \neq j$ die Ereignisse A_i und A_j unabhängig sind.

1.4.11 Bemerkung: Für den Nachweis, dass A, B, C unabhängig sind, müssen die folgenden Gleichungen verifiziert werden:

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C),$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

$$\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C),$$

$$\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$$

Für die Unabhängigkeit ist weder $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ noch die paarweise Unabhängigkeit $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C), \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$ hinreichend.

► Vgl. die entsprechenden Aufgaben ► Für das Konzept der bedingten Unabhängigkeit Vgl. Blitzstein und Hwang [4]

1.4.12 Satz (Multiplikationsformel): Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, A_2, \dots, A_n \in \mathcal{F}$ mit $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Dann gilt

$$\begin{aligned} & \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}) \cdot \\ & \quad \mathbb{P}(A_{n-1} | A_1 \cap \dots \cap A_{n-2}) \cdot \\ & \quad \cdot \dots \\ & \quad \cdot \mathbb{P}(A_2 | A_1) \dots \\ & \quad \cdot \mathbb{P}(A_1) \end{aligned}$$

1.4.13 Satz (Unabhängige Kopplung / Produktexperimente, diskreter Fall, Krenzel [22, S. 27f]): Gegeben sind für $n \geq 2$

- n Experimente $(\Omega_1, \mathcal{P}(\Omega_1), \mathbb{P}_1), \dots, (\Omega_n, \mathcal{P}(\Omega_n), \mathbb{P}_n)$ mit abzählbaren nicht-leeren Mengen Ω_i .

Auf diesen *separaten* Räumen können wir keine gemeinsame Wahrscheinlichkeitsmodellierung betreiben. Wir wollen **einen gemeinsamen Raum** für diese Experimente definieren, **der deren unabhängige Durchführung erfasst**. Man spricht von der **unabhängigen Kopplung**.

Wir definieren

$$\begin{aligned}\Omega &= \Omega_1 \times \dots \times \Omega_n, \\ \mathbb{P}(\{\omega\}) &= \prod_{i=1}^n \mathbb{P}_i(\{\omega_i\}), \quad \omega = (\omega_1, \dots, \omega_n), \\ \mathbb{P}(A) &= \sum_{\omega \in A} \mathbb{P}(\omega).\end{aligned}$$

Dann ist $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein Wahrscheinlichkeitsraum, d.h. durch die Vorgabe der Ereigniswahrscheinlichkeiten ergibt sich ein Wahrscheinlichkeitsmaß.

Für diesen Wahrscheinlichkeitsraum gilt dann

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i=1}^n \{\Omega_1 \times \dots \times A_i \times \dots \times \Omega_n\}\right) \\ = \prod_{i=1}^n \mathbb{P}(\Omega_1 \times \dots \times A_i \times \dots \times \Omega_n)\end{aligned}$$

oder mit der Notation (S^{A_i} für Streifen A_i)

$$S^{A_i} = \Omega_1 \times \dots \times A_i \times \dots \times \Omega_n$$

die Gleichung

$$\mathbb{P}\left(\bigcap_{i=1}^n S^{A_i}\right) = \prod_{i=1}^n \mathbb{P}(S^{A_i})$$

Also: Die Wahrscheinlichkeit $\mathbb{P}(\bigcap_{i=1}^n S^{A_i})$, dass im 1-ten Teil-experiment A_1 **und** im 2-ten Teilexperiment in A_2 **und** ...,

die n -ten Teilerperiment A_n eintritt, ist gleich dem **Produkt** $\prod_{i=1}^n \mathbb{P}(A_i \text{ im } i\text{-ten Teilerperiment})$ der Wahrscheinlichkeiten $\mathbb{P}(A_i \text{ im } i\text{-ten Teilerperiment})$. Da man dabei auch $A_i = \Omega_i, i \notin J, J \subset \{1, \dots, n\}$ wählen kann, sind die **Teilergebnisse** $\Omega_1 \times \Omega_2 \times \dots \times A_i \times \dots \times \Omega_n$ **unabhängig** (im Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$) im Sinn der Definition (1.4.10). Diese Beobachtung rechtfertigt (endgültig) die Bezeichnung **unabhängige Kopplung** für dieses Wahrscheinlichkeitsmodell. In der Tat erhält man *nur so* einen Wahrscheinlichkeitsraum, so dass

- $\mathbb{P}(\Omega_1 \times \Omega_2 \times \dots \times A_i \times \dots \times \Omega_n) = \mathbb{P}_i(A_i)$ und
- $\Omega_1 \times \Omega_2 \times \dots \times A_i \times \dots \times \Omega_n$ bilden eine unabhängige Familie.

Der so definierte Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ heißt **Produktwahrscheinlichkeitsraum** der Experimente $(\Omega_i, \mathcal{P}(\Omega_i), \mathbb{P}_i)$ und beschreibt die **unabhängige Kopplung** der Experimente $(\Omega_i, \mathcal{P}(\Omega_i), \mathbb{P}_i)$.

Die obigen Aussagen/Behauptungen werden im folgenden er-

läutert/belegt. Für $A_i \in \Omega_i$ ist

$$\begin{aligned}
 \mathbb{P}(A_1 \times A_2 \times \dots \times A_n) &= \sum_{\omega \in A_1 \times A_2 \times \dots \times A_n} \mathbb{P}(\{\omega\}) \\
 &= \sum_{\omega \in A_1 \times A_2 \times \dots \times A_n} \mathbb{P}_1(\{\omega_1\}) \cdot \dots \cdot \mathbb{P}_n(\{\omega_n\}) \\
 &= \sum_{\omega_1 \in A_1} \dots \sum_{\omega_n \in A_n} \mathbb{P}_1(\{\omega_1\}) \cdot \dots \cdot \mathbb{P}_n(\{\omega_n\}) \\
 &= \left(\sum_{\omega_1 \in A_1} \mathbb{P}_1(\{\omega_1\}) \right) \cdot \dots \cdot \left(\sum_{\omega_n \in A_n} \mathbb{P}_n(\{\omega_n\}) \right) \\
 &= \prod_{i=1}^n \mathbb{P}_i(A_i).
 \end{aligned}$$

Also

$$\mathbb{P}(A_1 \times A_2 \times \dots \times A_n) = \prod_{i=1}^n \mathbb{P}_i(A_i).$$

Diese Gleichung *sieht so aus wie* eine entsprechende Gleichung aus der Definition der Unabhängigkeit. **Aber Vorsicht:** Auf der rechten Seite stehen die \mathbb{P}_i und auf der linken Seite \mathbb{P} ; anders als in der Definition für Unabhängigkeit.

Wir beobachten jetzt (so dass links und rechts \mathbb{P} stehen wird):

$$\begin{aligned}
 \mathbb{P}\left(\bigcap_{i=1}^n \{\Omega_1 \times \dots \times A_i \times \dots \times \Omega_n\}\right) &= \mathbb{P}(A_1 \times \dots \times A_n) \\
 &= \prod_{i=1}^n \mathbb{P}_i(A_i) \\
 &= \prod_{i=1}^n \mathbb{P}_1(\Omega_1) \cdot \dots \cdot \mathbb{P}_i(A_i) \cdot \dots \cdot \mathbb{P}_n(\Omega_n) \\
 &= \prod_{i=1}^n \mathbb{P}(\Omega_1 \times \dots \times A_i \times \dots \times \Omega_n)
 \end{aligned}$$

Also

$$\begin{aligned}
 &\mathbb{P}\left(\bigcap_{i=1}^n \{\Omega_1 \times \dots \times A_i \times \dots \times \Omega_n\}\right) \\
 &= \prod_{i=1}^n \mathbb{P}(\Omega_1 \times \dots \times A_i \times \dots \times \Omega_n)
 \end{aligned}$$

Beachte, dass es $\mathbb{P}(A_i)$ nicht definiert ist; deshalb die unübersichtliche rechte Seite.

1.4.14 Satz (Konstruktion mehrstufiger Modelle / Bedingte Kopplung): Gegeben sind für $n \geq 2$

- n abzählbare nicht-leere Ergebnismengen $\Omega_1, \dots, \Omega_n$.
- Eine Zähldichte \mathbb{P}_1 auf Ω_1 .
- Für alle $k = 2, \dots, n$ und beliebige $\omega_i \in \Omega_i, i = 1, \dots, k -$

1 seien (bedingte) **Zähldichten** $\omega_k \mapsto \mathbb{P}_k(\omega_k | \omega_1, \dots, \omega_{k-1})$ auf Ω_k gegeben.

Wir definieren:

$$\begin{aligned}\Omega &= \Omega_1 \times \dots \times \Omega_n, \\ \mathbb{P}(\omega) &= \mathbb{P}_1(\omega_1) \cdot \mathbb{P}_2(\omega_2 | \omega_1) \cdot \dots \cdot \mathbb{P}_n(\omega_n | \omega_1, \dots, \omega_{n-1}), \\ \mathbb{P}(A) &= \sum_{\omega \in A} \mathbb{P}(\omega).\end{aligned}$$

Dann ist $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Dieser Wahrscheinlichkeitsraum heißt **mehrstufige bedingte Kopplung** der Räume $(\Omega_k, \mathcal{P}(\Omega_k), \mathbb{P}_k)$, $k = 1, \dots, n$.

$(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ist der einzige Wahrscheinlichkeitsraum mit:

in Worten
...

$$\mathbb{P}(\{\eta_1\} \times \Omega_2 \times \dots \times \Omega_n) = \mathbb{P}_1(\eta_1) \text{ und}$$

$$\begin{aligned}\mathbb{P}(\Omega_1 \times \Omega_2 \times \dots \times \{\eta_j\} \times \dots \times \Omega_n \mid \{\eta_1\} \times \Omega_2 \times \dots \times \Omega_n \cap \\ \dots \cap \Omega_1 \times \Omega_2 \times \dots \times \{\eta_{j-1}\} \times \dots \times \Omega_n) \\ = \mathbb{P}_j(\eta_j | \eta_1, \dots, \eta_{j-1}) \text{ falls}\end{aligned}$$

$$\mathbb{P}(\{\eta_1\} \times \Omega_2 \times \dots \times \Omega_n \cap \dots \cap \Omega_1 \times \Omega_2 \times \dots \times \{\eta_{j-1}\} \times \dots \times \Omega_n) > 0$$

► Wir beachten, dass die Ergebnismengen für z.B. die zweite Stufe für alle Ergebnisse auf der ersten Stufe gleich sind (obwohl sich die Wahrscheinlichkeiten natürlich ändern können). Das sieht nach einer merklichen Einschränkung der Allgemeinheit aus. Wir werden in Beispielen sehen, dass dem nicht

so ist (durch das Hinzufügen von Ergebnissen mit Wahrscheinlichkeit Null).

1.5 Unendlich oft

1.5.1 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_i \in \mathcal{F}, i \in \mathbb{N}$ und $A \in \mathcal{F}$. Wir definieren den **von-unten-Limes** $A_i \nearrow A$, wenn für alle $i \in \mathbb{N}$ $A_i \subset A_{i+1}$ und $A = \bigcup_{i \in \mathbb{N}} A_i$. Wir definieren den **von-oben-Limes** $A_i \searrow A$, wenn für alle $i \in \mathbb{N}$ $A_i \supset A_{i+1}$ und $A = \bigcap_{i \in \mathbb{N}} A_i$.

In beiden Fällen schreiben wir dann $A = \lim A_i$, wenn die *Art* des Limes (also \searrow bzw. \nearrow) aus dem Zusammenhang eindeutig hervor geht.

1.5.2 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_i \in \mathcal{F}, i \in \mathbb{N}$ und $A \in \mathcal{F}$. Dann gilt:

- i.) $A_i \nearrow A$ impliziert $\mathbb{P}(A_i) \rightarrow \mathbb{P}(A)$ (also $\mathbb{P}(\lim A_i) = \lim \mathbb{P}(A_i)$) (Stetigkeit von unten)
- ii.) $A_i \searrow A$ impliziert $\mathbb{P}(A_i) \rightarrow \mathbb{P}(A)$ (also $\mathbb{P}(\lim A_i) = \lim \mathbb{P}(A_i)$) (Stetigkeit von oben).

Beweis: Henze [18, S. 24].

1.5.3 Bemerkung: Die **Stetigkeitseigenschaft ist keine technische Randnotiz**. Angenommen wir betrachten das

Experiment des fortgesetzten fairen Würfeln. Wir interessieren uns für das Ereignis, dass wir keine 6 würfeln. A_i sei das Ereignis, dass beim i -ten Wurf keine 6 gewürfelt wird. Es gilt natürlich $\mathbb{P}(A_i) = 5/6$. Wir wollen auch annehmen, dass die Ereignisse A_i unabhängig sind. Dann gilt $n \in \mathbb{N}$

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \left(\frac{5}{6}\right)^n.$$

Für $n \rightarrow \infty$ konvergiert diese Wahrscheinlichkeit gegen Null. Bedeutet das, dass die Wahrscheinlichkeit

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \mathbb{P}(\text{"nie eine 6"})$$

nie eine 6 zu würfeln, Null ist? Wegen der bewiesenen Stetigkeit können wir uns davon überzeugen: Wir definieren $B_n = \bigcap_{i=1}^n A_i$, $B = \bigcap_{i=1}^{\infty} A_i$. Es gilt $B_n \supset B_{n+1}$ und $B = \bigcap_{i \in \mathbb{N}} B_i$.

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} B_i\right) = \mathbb{P}(\lim B_i) = \lim \mathbb{P}(B_i) = 0.$$

Es ist *beruhigend*, dass die unterstellten Eigenschaften eines Wahrscheinlichkeitsmaßes tatsächlich die Stetigkeit von Maß implizieren.

1.5.4 Satz (Stetigkeit und σ -additiv): Es sei \mathcal{F} eine σ -Algebra und $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ eine Abbildung mit $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$ und $\mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$, $A, B \in \mathcal{F}$ disjunkt.

Dann gilt: \mathbb{P} ist genau dann σ -additiv, wenn \mathbb{P} stetig ist.

Beweis:

1.5.5 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_i \in \mathcal{F}, i \in \mathbb{N}$. Dann definieren wir:

$$\limsup A_i := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega \mid \text{für unendlich viele } i \text{ gilt } \omega \in A_i\}$$

= „unendlich viele der Ereignisse A_i treten ein“

=: ∞ -oft A_i

$$\liminf A_i := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega \mid \text{für alle bis auf endlich viele } i \text{ ist } \omega \in A_i\}$$

= „alle bis auf endlich viele der Ereignisse A_i treten ein“

=: schließlich alle A_i

1.5.6 Satz (Erstes Lemma von Borel & Cantelli): Wenn $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$ konvergiert, dann $\mathbb{P}(\limsup A_i) = 0$. M.a.W: Wenn $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$ konvergiert, dann ist die Wahrscheinlichkeit, dass unendlich viele A_i eintreten, Null.

Beweis: Vgl. Henze [18, S. 64].

1.5.7 Satz (Zweites Lemma von Borel & Cantelli): Wenn A_i eine **unabhängige** Folge von Ereignissen ist und $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$ divergiert, dann $\mathbb{P}(\limsup A_i) = 1$. M.a.W: Wenn

für unabhängige A_i die Reihe $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$ divergiert, dann ist die Wahrscheinlichkeit, dass unendlich viele A_i eintreten, Eins.

Beweis: Vgl. Henze [18, S. 64].

1.5.8 Satz (Ein einfaches 0-1 Gesetz): Wenn A_i eine unabhängige Folge von Ereignissen ist, dann gilt $\mathbb{P}(\limsup A_i) = 1$ oder $\mathbb{P}(\limsup A_i) = 0$.

1.6 Exkurs Kombinatorik

Meine Favoriten zum Thema Kombinatorik sind Mazur [25] und Biggs [3].

1.6.1 Bemerkung: Insbesondere für die Analyse von Laplace Experimenten benötigt man Ergebnisse aus der Kombinatorik. Wir betrachten die folgenden Regeln:

1. Die Anzahl der Möglichkeiten n unterschiedliche Objekte anzuordnen ist $n!$. Hier bezeichnet

$$n! = n \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1$$

n **Fakultät.**

Für den ersten Platz gibt es n Möglichkeiten. Wenn der

erste Platz bestimmt ist, dann gibt es für den zweiten Platz $(n - 1)$ Möglichkeiten

2. Wir betrachten wieder die Anzahl der Möglichkeiten n Objekten anzuordnen. Allerdings können wir bei den Anordnungen n_1 Objekte der ersten Art, n_2 Objekte der zweiten Art, ..., n_K Objekte der k -ten Art nicht unterscheiden. Die Anzahl derartiger unterscheidbarer Anordnungen ist $\frac{n!}{n_1!n_2!\dots n_k!}$; der Koeffizient

$$\binom{n}{n_1, \dots, n_K} := \frac{n!}{n_1!n_2!\dots n_K!}$$

heißt **Multinomialkoeffizient**. Beachte, dass $\sum_{k=1}^K n_k = n$ gilt.

Angenommen wir haben 4 rote Kugel, 3 grüne Kugel und 2 blaue Kugeln. Zunächst stellen wir uns vor, dass wir die Kugel markieren, so dass wir die roten Kugel, die grünen Kugel und die blauen Kugel doch unterscheiden können. Wir haben also 9 unterscheidbare Kugeln. Wir wissen schon, dass es $9!$ mögliche Anordnungen gibt. Angenommen wir haben die Kugeln in einer konkreten Reihenfolge angeordnet. Wenn wir jetzt die roten Kugeln untereinander neu anordnen, dann erhalten wir bei nicht unterscheidbaren Kugeln **keine** neue Anordnung. Wir müssen für diese Zuviel-Zählung korri-

gieren. Das geschieht indem man durch $4!3!2!$ dividiert.

Für den Multinomialkoeffizienten gibt es eine weitere **Interpretation**. Wir wollen n Bälle **färben** und haben K Farben. Dann ist $\binom{n}{n_1, \dots, n_K}$ die Anzahl der Möglichkeiten n Kugeln so zu färben, dass n_1 die Farbe 1, ..., n_K die Farbe K bekommen. Noch eine sehr ähnliche **Interpretation**: Es gibt K (unterscheidbare) Behälter auf die n Bälle **verteilt** werden sollen; vgl. Mazur [25, Seite 142] oder Biggs [3, Chapter 12].⁶

Warum ist das so? Wir stellen uns vor, dass die Behälter die Farbe flüssig enthalten und nebeneinander aufgereiht sind. Wir stellen uns auch vor, dass die Kugel nummeriert sind. Wir legen jetzt die Kugeln so vor den Behältern abm dass n_1 vor dem ersten Behälter liegen, n_2 vor dem zweiten Behälter liegen, Es gibt für die Anordnung $n!$ Möglichkeiten. Bezogen auf die Färbung bzw. Verteilung der Kugel im ersten bzw. in den Behälter, kommt es aber nicht auf die Anordnung der n_1 Kugeln vor dem ersten Behälter an. Analog für die Kugeln vor den anderen Behältern. Wir müssen demnach wieder korrigieren und erreichen das, in dem wir durch $n_1!n_2! \cdot \dots \cdot n_K!$ dividieren.

⁶Wenn die Behälter nicht unterschieden werden können, dann erhält man Partitionen (in Mengen/Teams). Hier besteht eine Verwechslungsgefahr; vgl. Mazur [25, 145] oder Biggs [3, Chapter 12].

3. **Partitionen** mit Vorgaben. Wenn die Behälter nicht unterschieden werden können, dann erhält man Partitionen (in Mengen/Teams). Hier besteht eine Verwechslungsgefahr; vgl. Mazur [25, 145] oder Biggs [3, Chapter 12].

$$\frac{\binom{n}{n_1, \dots, n_K}}{p_1! \cdot \dots \cdot p_n!}$$

Dabei bezeichnet p_1 die Anzahl der Teams mit einer Kugel, p_2 die Anzahl der Teams mit zwei Kugeln,

Warum ist das so? Warum muss man durch $p_1! \cdot \dots \cdot p_n!$ dividieren? Dazu stellen wir uns vor, dass die Behälter vorab so sortiert sind, dass $n_1 \leq n_2 \leq \dots \leq n_K$ gilt. Zuerst kommen die p_1 Behälter, die nur einen Ball erhalten sollen. Dann die p_2 Behälter, die zwei Bälle enthalten sollen, Dann ordnen wir die Bälle vor den Behältern wie eben an. Wenn man die Behälter unterscheiden kann, dann erhalten wir $\binom{n}{n_1, \dots, n_K}$ mögliche Anordnungen, so dass der erste Behälter n_1 , der zweite n_2 Bälle ... enthält. Jetzt können wir für alle j die Anordnung Blöcke mit p_j Bälle ändern und erhalten eine gleichartige Partition! Es gibt $p_1! \cdot \dots \cdot p_n!$ Anordnungen solche Umordnungen.

Es gilt

$$\frac{\binom{n}{n_1, \dots, n_K}}{p_1! \cdot \dots \cdot p_n!} = \frac{n!}{\prod_{j=1}^n (j!)^{p_j} p_j!}$$

Diese Zahl gibt an, wie viele Partitionen/Team-Einteilungen von n Spielern es gibt, so dass es p_j Team mit j Spielern gibt.

1.6.2 Bemerkung (Objekte/Subjekte ziehen): Wir ziehen k Objekte/Subjekte aus einer Gesamtheit von n Objekten/Subjekten. Wir müssen dann beachten, ob die gezogenen Objekte zurück gelegt werden oder nicht. Und wir müssen beachten, ob die Reihenfolge für das Zählen relevant ist oder nicht.

1. Es gibt n^k **geordnete Stichproben mit Zurücklegen** der Größe k aus einer Grundgesamtheit von n Objekten.
2. Es gibt $(n)_k := n(n-1)(n-2)\dots(n-k+1) = \frac{n!}{(n-k)!}$ **geordnete Stichproben ohne Zurücklegen** der Größe k aus einer Grundgesamtheit von n Objekten. Man nennt

$$(n)_k := n \cdot (n-1) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

fallende Faktorielle.

3. Es gibt $\binom{n}{k}$ **ungeordnete** Stichproben **ohne Zurücklegen** der Größe k aus einer Grundgesamtheit von n Objekten. Man nennt

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

Binomialkoeffizienten.

4. Es gibt $\binom{n+k-1}{k}$ **ungeordnete** Stichproben **mit Zurücklegen** der Größe k aus einer Grundgesamtheit von n Objekten. Man nennt

$$\binom{\binom{n}{k}}{k} := \binom{n+k-1}{k}$$

Multimengenkoeffizient.

Die Stichprobenmodelle sind zum Vergleich noch mal in der folgenden Tabelle angegeben.

	mit Wiederholung	ohne Wiederholung
geordnet	n^k	$(n)_k$
ungeordnet	$\binom{\binom{n}{k}}{k} := \binom{n+k-1}{k}$	$\binom{n}{k}$

1.6.3 Bemerkung (Objekte/Subjekte verteilen): Es gibt noch eine andere⁷ *Interpretation* für die Formeln in der vorgehenden Tabelle: **Verteilung von k Bällen auf n Behälter.**

⁷Die wir teilweise in der vorhergehenden Bemerkung betrachtet haben.

Manchmal ist diese Auslegung praktischer. Wegen der Verwechslungsgefahr wähle ich Bälle, die auf Behälter verteilt werden und Kugeln, die gezogen werden. Es ist in der Tat auch anschaulich, sich die Behälter als Farbtöpfe vorzustellen. Die Verwechslungsgefahr besteht insbesondere bezüglich n und k . k ist die Anzahl der Objekte/Subjekte mit denen etwas *gemacht* wird (verteilt bzw. gezogen).

Bei dieser Interpretation werden k Bälle auf n Behälter verteilt. Die Behälter sind in allen vier Varianten unterscheidbar – z.B. farblich oder durch Nummern. Die Bälle u.U. nicht.

	mit Mehrfachbelegung	ohne Mehrfachbeleg.
unterschiedliche Kugeln	n^k	$(n)_k$
nicht-unterscheidbare Kugeln	$\binom{n}{k}$	$\binom{n}{k}$

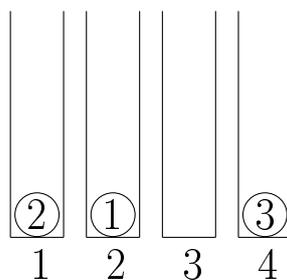
1.6.4 Bemerkung (Ziehen und Verteilen): Wir besprechen kurz den **Zusammenhang der beiden Perspektiven**. Die Bälle seien nummeriert (also unterscheidbar) und entsprechend seien auch die Kugeln nummeriert. Die Bijektion zwischen Ziehung und Verteilung ist dann die folgenden: Wenn wir zuerst die Kugel mit der Nummer l ziehen, dann legen wir den Ball mit der Nummer 1 in den l -ten Behälter.

Also:

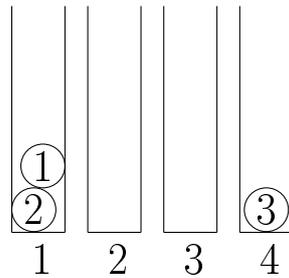
zufällig Kugel l bei Zug 1 \leftrightarrow **Ball 1 in zufällig Behälter l**
zufällig Kugel m bei Zug 2 \leftrightarrow **Ball 2 in zufällig Behälter m**

Eselsbrücke für den Übergang von Ziehungen auf Belegungen: Man *zieht* die Behälter für die nummerierten bzw. nicht-nummerierten Bälle.

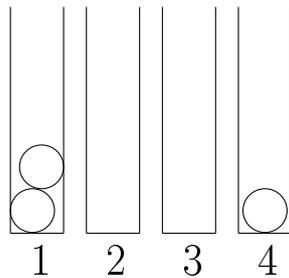
Beispiele: i.) Die folgende Abbildung zeigt eine Verteilung von 3 verschiedenen Bällen auf 4 Behälter, wobei in jeden Behälter maximal ein Ball liegt. Die Bälle sind nummeriert also unterschiedlich. Die Behälter sind ebenfalls unterschiedlich. Gemäß der Tabelle gibt es $(4)_3 = 4 \cdot 3 \cdot 2 = 24$ *solche* Verteilungen. Warum? DIY!



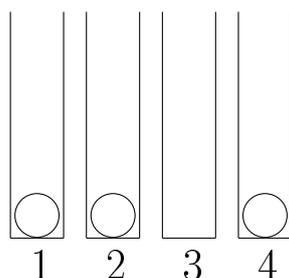
ii.) Die folgende Abbildung zeigt eine Verteilung von 3 verschiedenen Bällen auf 4 Behälter, wobei im ersten Behälter 2 Kugeln liegen. Gemäß der Tabelle gibt es $4^3 = 64$ *solche* Verteilungen. Warum? DIY!



iii.) Die folgende Abbildung zeigt eine Verteilung von 3 nicht-unterscheidbaren Bällen (nicht-unterscheidbar, da nicht nummeriert) auf 4 Behälter, wobei im ersten Behälter 2 Bälle liegen. Gemäß der Tabelle gibt es $\binom{4}{3} = \binom{4+3-1}{3} = \binom{6}{3} = 20$ *solche* Verteilungen. Warum? DIY!



iv.) Die folgende Abbildung zeigt eine Verteilung von 3 nicht-unterscheidbaren Bällen (nicht-unterscheidbar, da nicht nummeriert) auf 4 Behälter, wobei in jedem Behälter nur maximal ein Ball liegt. Gemäß der Tabelle gibt es $\binom{4}{3} = \binom{4}{3} = 4$ *solche* Verteilungen. Warum? DIY!



2 Diskrete Modelle

2.1 Allgemeins

Die im Folgenden erläuterten Modelle werden in vielen Quellen diskutiert. Wir verwenden Goergii [13], Henze [18], Tappe [31] und Dobrow und Wagamann [34]

2.1.1 Definition: Ein Wahrscheinlichkeitsmaß \mathbb{P} heißt **diskrete Verteilung oder diskretes Wahrscheinlichkeitsmaß** auf (Ω, \mathcal{F}) , falls

- i.) Ω eine nicht-leere abzählbare Menge und
- ii.) $\mathcal{F} = \mathcal{P}(\Omega)$ ist.

P.S. Abzählbar bedeutet endlich oder abzählbar unendlich.

2.1.2 Definition: Es sei Ω eine nicht-leere abzählbare Menge und $p_j \in [0, 1], j \in \Omega$ seien reelle Zahlen im Intervall $[0, 1]$

mit

$$\sum_{j \in \Omega} p_j = 1.$$

Dann heißt $(p_j)_{j \in \Omega}$ **Zähldichte** oder **Wahrscheinlichkeitsfunktion**.

Eine Zähldichte/Wahrscheinlichkeitsfunktion ist (noch) kein Wahrscheinlichkeitsmaß! Insbesondere stimmen die Definitionsbereiche nicht überein. Die Wahrscheinlichkeitsfunktion liefert nur Wahrscheinlichkeiten für die **Ergebnisse**; das Wahrscheinlichkeitsmaß für alle **Ereignisse**. Aber aus einer Wahrscheinlichkeitsfunktion ergibt sich – wie der folgende Satz zeigt – ein Wahrscheinlichkeitsmaß.

2.1.3 Satz: Es sei Ω eine nicht-leere abzählbare Menge und $(p_k)_{k \in \Omega}$ eine Zähldichte auf $(\Omega, \mathcal{P}(\Omega))$. Dann gibt es genau ein diskretes Wahrscheinlichkeitsmaß \mathbb{P} auf $(\Omega, \mathcal{P}(\Omega))$ mit $\mathbb{P}(\{\omega_k\}) = p_k, k \in \Omega$.

Für dieses Wahrscheinlichkeitsmaß gilt

$$A \mapsto \mathbb{P}(A) = \sum_{\omega_k \in A} p_k, \quad A \in \mathcal{P}(\Omega)$$

Wir nennen \mathbb{P} das durch die Zähldichte $(p_k)_{k \in \Omega}$ definierte Wahrscheinlichkeitsmaß.

2.1.4 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum (nicht notwendigerweise diskret) und $X : \Omega \rightarrow \mathbb{R}$ eine Abbildung. X heißt **diskrete Zufallsvariable**, falls:

- i.) $\text{Bild}(X) = X(\Omega) = \{X(\omega) | \omega \in \Omega\}$ ist abzählbar. $X(\Omega)$ heißt Bild von X .
- ii.) Für alle $A \subset X(\Omega)$ ist $X^{-1}(A) \in \mathcal{F}$, d.h. X ist **messbar**.

2.1.5 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable. Dann heißt

$$\mathbb{P}^X : \begin{cases} \mathcal{P}(X(\Omega)) \rightarrow [0, 1] \\ A \mapsto \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) \end{cases}$$

die **Verteilung** von X . Dabei ist $\mathbb{P}^X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A))$ die Wahrscheinlichkeit das X einen Wert in der Menge A annimmt.

2.1.6 Bemerkung: Wir wollen $\mathbb{P}^X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$ berechnen können. Das geht aber nur, wenn $X^{-1}(A) \in \mathcal{F}$ gilt, denn nur für Mengen in \mathcal{F} ist \mathbb{P} definiert. Aus diesen Grund fordern wir bei Zufallsvariablen die Messbarkeit.

2.1.7 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable mit Vertei-

lung \mathbb{P}^X . Dann ist \mathbb{P}^X ein diskretes Wahrscheinlichkeitsmaß auf $(X(\Omega), \mathcal{P}(X(\Omega)))$.

2.1.8 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable mit Verteilung \mathbb{P}^X . Wir schreiben dann

$$X \sim \mathbb{P}^X.$$

2.1.9 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable mit Verteilung \mathbb{P}^X . Es sei ferner $p(x) = \mathbb{P}^X(\{x\})$, $x \in \text{Bild}(X)$ die Zähldichte für \mathbb{P}^X . Dann definieren wir

$$\mathbb{E}(X) = \sum_{x \in \text{Bild}(X)} p(x) \cdot x,$$

falls die Reihe – so es eine ist – absolut konvergiert. Wir sagen dann, dass der Erwartungswert von X existiert.

2.1.10 Satz (Gesetz des unbewussten Statistikers): Es gilt:

$$\mathbb{E}(g(X)) = \sum_{x \in \text{Bild}(X)} p(x)g(x).$$

Beweis: Vgl. Grimmett und Stirzaker [15, Seite 228]. Wir

setzen $Y = g(X)$ und beobachten

$$\mathbb{P}(g(X) = y) = \sum_{x:g(x)=y} \mathbb{P}(X = x).$$

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{y \in Y(\Omega)} \mathbb{P}(Y = y)y \\ &= \sum_{y \in Y(\Omega)} \mathbb{P}(g(X) = y)y \\ &= \sum_{y \in Y(\Omega)} \sum_{x:g(x)=y} \mathbb{P}(X = x)y \\ &= \sum_{y \in Y(\Omega)} \sum_{x:g(x)=y} \mathbb{P}(X = x)g(x) \\ &= \sum_{x \in X(\Omega)} \mathbb{P}(X = x)g(x) \end{aligned}$$

2.1.11 Satz: Es gilt:

- i.) $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$. Der Erwartungswert ist ein linearer Operator.
- ii.) $X \leq Y \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y)$. Der Erwartungswert ist ein monotoner Operator.

Beweis: Vgl. Grimmett und Stirzaker [14, Seite 55].

2.1.12 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable, so dass

der Erwartungswerte von X und von $(X - \mathbb{E}(X))^2$ existiert. Dann definieren wir

$$\mathbb{V}(X) := \sigma_X^2 := \mathbb{E}((X - \mathbb{E}(X))^2)$$

$\mathbb{V}(X)$ heißt die **Varianz** der Zufallsvariable X . Wir sagen, dass die Varianz von X existiert. $\sigma_X = \sqrt{\mathbb{V}(X)}$ heißt die **Standardabweichung** von X .

2.1.13 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable, so dass der Erwartungswerte von X und von $(X - \mathbb{E}(X))^2$ existiert. Dann

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

► Der Erwartungswert und die Varianz sind sehr wichtige Konzepte mit vielen schönen Eigenschaften, die wir genauer untersuchen, wenn wir den geeigneten allgemeinen Integrationsbegriff eingeführt haben. Die Eigenschaften kann man auch direkt analysieren, das ist jedoch mühsamer.

2.1.14 Bemerkung: Auf der Formelsammlung sind die Erwartungswerte und die Varianzen für die diskreten Standardmodelle angegeben.

2.2 Dirac und Laplace

2.2.1 Definition: Es sei Ω eine nicht-leere abzählbare Menge und $k^* \in \Omega$. Dann ist

$$p_k = \begin{cases} 1 & \text{für } k = k^* \\ 0 & \text{sonst} \end{cases}$$

eine Zähldichte. Das durch diese Zähldichte definierte diskrete Wahrscheinlichkeitsmaß heißt das im Punkt k^* konzentrierte **Dirac-Wahrscheinlichkeitsmaß**.

2.2.2 Definition: Es sei Ω eine endliche Menge mit $n \in \mathbb{N}$ Elementen. Dann ist $p_k = \frac{1}{n}, k = 1, \dots, n$ eine Zähldichte. Das durch diese Zähldichte definierte diskrete Wahrscheinlichkeitsmaß heißt **Laplace-Wahrscheinlichkeitsmaß** und $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein **Laplace-Experiment**.

Ein Experiment $(\Omega, \mathcal{F}, \mathbb{P})$ heißt also **Laplace Experiment** oder **Laplace Wahrscheinlichkeitsraum**, falls:

- i.) $\Omega \neq \emptyset$ ist eine endliche Menge mit $n \in \mathbb{N}$ Elementen.
- ii.) $\mathcal{F} = \mathcal{P}(\Omega)$.
- iii.) Für alle $A \in \mathcal{F}$ ist

$$\mathbb{P}(A) = \frac{\#(A)}{\#(\Omega)} = \frac{\#(A)}{n} = \frac{\text{Anzahl der für } A \text{ günstigen Ergebnisse}}{\text{Anzahl aller Ergebnisse}}.$$

2.3 Bernoulli, Binomial und Multinomial

2.3.1 Definition: Es sei $\Omega = \{\circ, \square\}$ und $p \in (0, 1)$. Dann ist $p(\circ) = p, p(\square) = 1 - p$ eine Zähldichte. Das durch diese Zähldichte definierte diskrete Wahrscheinlichkeitsmaß heißt **Bernoulli-Wahrscheinlichkeitsmaß** mit Erfolgswahrscheinlichkeit p . Man nennt $(\Omega, \mathcal{F}, \mathbb{P})$ ein **Bernoulli-Experiment**. Das Ergebnis \circ nennt man “Erfolg” und das Ergebnis \square “Misserfolg”. p wird **Erfolgswahrscheinlichkeit** genannt.

Anstatt $\Omega = \{\circ, \square\}$ wird typischerweise $\Omega = \{1, 0\}$ verwendet, wobei 1 für Erfolg steht. In der Finanzmathematik wird regelmäßig $\Omega = \{1, -1\}$, wobei 1 wieder für Erfolg steht.

2.3.2 Definition und Satz: Es sei $\Omega = \{0, 1, 2, \dots, n\}$ und $p \in (0, 1)$. Dann ist

$$p_k = \binom{n}{k} p^k (1 - p)^{n-k}$$

eine Zähldichte. Das durch diese Zähldichte definierte diskrete Wahrscheinlichkeitsmaß heißt **Binomialverteilung** mit **Erfolgswahrscheinlichkeit** p und n Wiederholungen.

Für den Beweis von $\sum_{i=1}^n p_i = 1$ benötigt man den Binomi-

schen Lehrsatz:

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

und eine magische Null $1 = 1 - p + p$.

2.3.3 Bemerkung: Die Binominalverteilung hat extrem viele Anwendungen. Man stellt sich dabei die unabhängige n -fache Wiederholung eines Bernoulli-Experiments vor. p_k ist dann die Wahrscheinlichkeit bei n unabhängigen Wiederholungen **genau k Erfolge und $n - k$ Misserfolge** zu beobachten.

2.3.4 Satz (Additionsgesetz für die Binomial-Verteilung):

Wenn X eine Binomial-Verteilung mit m Wiederholungen und Erfolgswahrscheinlichkeit p und Y eine Binomial-Verteilung mit n Wiederholungen und Erfolgswahrscheinlichkeit p hat, dann hat $X + Y$ eine Binomial Verteilung mit $m + n$ Wiederholungen und Erfolgswahrscheinlichkeit p .

Beweis: Vgl. Henze [18, S. 94].

2.3.5 Definition und Satz: Es sei $\Omega = \{(n_1, n_2, \dots, n_K) \in \mathbb{N}_0^K \mid \sum_{j=1}^K n_j = n\}$ und $p_i \in (0, 1), i = 1, \dots, K$, wobei $\sum_{i=1}^K p_i =$

1 gilt. Dann ist

$$p((n_1, n_2, \dots, n_K)) = \binom{n}{n_1, n_2, \dots, n_K} p_1^{n_1} \cdot p_2^{n_2} \cdot \dots \cdot p_K^{n_K}$$

eine Zähldichte. Das durch diese Zähldichte definierte diskrete Wahrscheinlichkeitsmaß heißt **Multinomialverteilung**.

Für den Nachweis, dass sich die Wahrscheinlichkeiten zu Eins addieren, benötigt man die Verallgemeinerung des Binomischen Lehrsatzes für Multinomialkoeffizienten (vgl. Mazur [25, S. 143]).

2.3.6 Bemerkung: Die Multinomialverteilung ist eine Verallgemeinerung der Binomialverteilung. Hier wird ein Experiment mit K **möglichen Ergebnissen** n -mal **unabhängig wiederholt**. Bei jeder Wiederholung ist die Wahrscheinlichkeit, das Ergebnis $i \in \{1, \dots, K\}$ zu erhalten, gleich p_i . Der Multinomialkoeffizient tritt an die Stelle des Binomialkoeffizienten.

$p((n_1, n_2, \dots, n_K))$ ist dann die Wahrscheinlichkeit n_1 mal das Ergebnis 1, n_2 mal das Ergebnis 2, ... und n_K das Ergebnis K zu erhalten.

Die Multinomialverteilung kann man beispielsweise auf die zufällige **Verteilung** von n Teilchen auf K Zustände anwenden; oder auf die zufällige **Färbung** von n Bälle in K Farben.

Für jeden Ball ist dabei die Wahrscheinlichkeit, in der i -ten Farbe gefärbt zu werden, gleich p_i . $p((n_1, n_2, \dots, n_K))$ erfasst dabei nicht, welche der Bälle wie gefärbt werden, sondern nur wie viele Bälle wie gefärbt sind: n_1 Bälle in der Farbe 1, n_2 in der Farbe 2

Wir können uns hier auch folgendes **Ziehungsmodell** vorstellen: In einer Urne liegen verschieden farbige Kugeln; K verschiedene Farben. p_i bezeichnet den Anteil der i -farbigen Kugeln. Wir ziehen n Kugeln mit zurücklegen.

2.4 Hypergeometrisch

2.4.1 Satz und Definition: Es seien $N, M, k \in \mathbb{N}$ mit $M \leq N$ und $k \leq N$ und $\Omega = \{0, 1, \dots, k\}$. Dann ist $p_l = \frac{\binom{M}{l} \binom{N-M}{k-l}}{\binom{N}{k}}$ eine Zähldichte. Das durch diese Zähldichte definierte diskrete Wahrscheinlichkeitsmaß heißt **hypergeometrische Verteilung** mit Parametern N, M, k .

2.4.2 Bemerkung: Die hypergeometrische Verteilung wird für die Modellierung des gleichzeitigen/**ungeordneten Ziehens ohne Zurücklegen von k Kugeln** aus einer Urne mit N Kugel, von denen M Kugel **grün** bzw. $N - M$ **blau** sind, verwendet. p_l ist dann **die Wahrscheinlichkeit l grüne Kugeln zu ziehen**.

Wir wollen nachrechnen, dass sich die **genannte Wahrscheinlichkeiten** ergibt, wenn man die Kugeln **nacheinander zunächst geordnet ohne Zurücklegen** zieht und dann die Reihenfolge ignoriert. Es gibt $(N)_k$ mögliche geordnete Ziehungen von k Kugel aus einer Urne mit N Kugeln. Das erklärt den Nenner in der folgenden abgesetzten Gleichung. Wir betrachten dann die Anzahl der Möglichkeiten l der M grünen Kugel bzw. $k - l$ der $N - M$ blauen Kugel zu ziehen; wobei wir die Reihenfolge/Nummer zunächst ignorieren. Für solche Auswahlen gibt es $\binom{M}{l} \binom{N-M}{k-l}$ Möglichkeiten. Für jede dieser Auswahlen gibt es $k!$ Anordnungsmöglichkeiten.

Jetzt setzen wir alles zusammen und beobachten schließlich

$$\frac{\binom{M}{l} \binom{N-M}{k-l} \cdot k!}{(N)_k} = \frac{\binom{M}{l} \binom{N-M}{k-l}}{\frac{N!}{(N-k)!k!}} = \frac{\binom{M}{l} \binom{N-M}{k-l}}{\binom{N}{k}}.$$

2.5 Belegungsmodelle, Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac

2.5.1 Bemerkung: Nützlich ist die hypergeometrische Verteilung z.B. in der **Qualitätskontrolle**. Ein Produzent erhält eine Lieferung von 10000 Bauteilen und will die Qualität prüfen. Es ist zu teuer alle Bauteile zu überprüfen. Man entnimmt eine Stichprobe von z.B. 30 Bauteilen und prüft diese.

Die Situation gleicht dann offenbar der Entnahme von Kugel (die jeweils eine von zwei Farben haben) aus einer Schale.

2.5.2 Bemerkung (Belegungsmodelle): Belegungsmodelle modellieren die **Verteilung von Kugeln auf Behälter**¹. Zwei Aspekte muss man beachten: **Beobachtungstiefe** und **Restriktionen**.

Die mögliche **Verringerung der Beobachtungstiefe** bei Belegungen besteht darin, dass/ob man die Nummern auf den Kugeln ignoriert. Die Nummern auf den Kugeln werden *unsichtbar*. Relevant ist nicht, welche Kugeln in den Behältern liegen, sondern wie viele. Die mögliche **Restriktion** ist gegebenenfalls, dass maximal eine Kugel in einem Behälter liegen darf.

Bei **maximaler Beobachtungstiefe** und **ohne Restriktion** ist die Ergebnismenge die Menge der n -Tupel mit Einträgen aus $\{1, \dots, K\}$. n Kugeln werden auf die Behälter verteilt. K ist die Anzahl Behälter/Farben und n die Anzahl der Kugeln. Das Ergebnis wird als n -Tupel (x_1, \dots, x_n) protokolliert. Dabei bezeigt $x_i = k$ den Behälter k in den die Kugel i gelegt.

$$\Omega_1 = \{(x_1, \dots, x_n) | x_i \in \{1, \dots, K\}, i = 1, \dots, n\}$$

¹Die Behälter sind in dieser Bemerkung stets unterscheidbar, also mit Nummern oder Farben versehen.

und alle Ergebnisse haben die Wahrscheinlichkeit (jedenfalls ist das eine *natürliche/denkbar*e Voraussetzung)

$$\mathbb{P}_1(\{(x_1, \dots, x_n)\}) = \left(\frac{1}{K}\right)^n = \frac{1}{K^n}.$$

Wir stellen uns jetzt vor, dass die Nummern auf den Kugeln verschwinden. Das **Experiment soll dabei unverändert** bleiben. Es soll sich **nur die Beobachtbarkeit ändern**. Bestimmte Ergebnisse, die wir bei genauer Beobachtungstiefe unterscheiden (beobachten) können, sind dann nicht mehr unterscheidbar und werden zu einem neuen Ergebnis **zusammengelegt**. Durch den Übergang von nummeriert zu nicht-nummeriert werden bestimmte Elemente sozusagen *identifiziert*. Aus nummerierten Belegungen werden nicht-nummerierte Belegungen. Angenommen wir betrachten 5 Kugeln und 7 Behälter. Die Ergebnisse $(2, 2, 2, 2, 3)$, $(2, 2, 2, 3, 2)$, $(2, 2, 3, 2, 2)$, $(2, 3, 2, 2, 2)$ und $(3, 2, 2, 2, 2)$ beispielsweise werden zu $\langle 2, 2, 2, 2, 3 \rangle$.² Bei verringerter Beobachtungstiefe beobachten wir, dass eine Kugel im dritten Behälter liegt und vier Kugeln im zweiten Behälter. Wir beobachten jedoch nicht, welche Kugel in den zweiten Behälter gelegt wird. Die anderen Behälter sind leer. Bei genauer Beobachtungstiefe können wir z.B. $(2, 2, 2, 2, 3)$ und $(2, 2, 2, 3, 2)$ unterscheiden können. Bei $(2, 2, 2, 2, 3)$ liegt die Kugel 5 im Behälter 3; bei $(2, 2, 2, 3, 2)$ die Kugel 4 im Behälter 3. Die Anzahl der Ergebnisse, die jeweils zusam-

²Wir verwenden \langle, \rangle für Multimengen.

mengelegt werden entspricht dem Multinomialkoeffizienten: $\binom{5}{0,4,1,0,0,0,0} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{4! \cdot 1!} = 5$. Die Zahl der Ergebnisse, die zusammengelegt werden, variiert: Das Ergebnis $\langle 2, 2, 2, 2, 2 \rangle$ ergibt sich genau aus einem Ergebnis $(2, 2, 2, 2, 2)$. Das Ergebnis $\langle 2, 2, 2, 2, 3 \rangle$ ergibt sich aus fünf: $(2, 2, 2, 2, 3)$, $(2, 2, 2, 3, 2)$, $(2, 2, 3, 2, 2)$, $(2, 3, 2, 2, 2)$ und $(3, 2, 2, 2, 2)$. Wie viele Ergebnisse zusammengelegt werden, wird durch den Multinomialkoeffizienten angegeben.

Durch dieses Zusammenlegen erhalten wir dementsprechend auf der Menge der Multimengen $\Omega_4 = \{ \langle x_1, \dots, x_n \rangle \mid x_i \in \{1, \dots, K\} \}$ eine³ **Multinomialverteilung**.

$$\mathbb{P}_4(\langle 2, 2, 2, 2, 3 \rangle) = \binom{5}{0, 4, 1, 0, 0, 0, 0} \cdot \left(\frac{1}{K}\right)^5,$$

wobei

$$\Omega_4 = \{ \langle x_1, \dots, x_n \rangle \mid x_i \in \{1, \dots, K\} \}.$$

Es gilt $|\Omega_4| = \binom{K}{n}$; Ω_4 hat $\binom{K}{n} = \binom{K+n-1}{n}$ Elemente.

In der Physik kennt man diese Verteilung auch als **Maxwell-Boltzmann-Statistik**⁴ (siehe z.B. Feller [10, S. 39, 41f]).

Wir erhalten die *Konsistenz* zur geordneten Ziehung mit Zu-

³Wir haben vorne schon Multinomialverteilungen allgemein betrachtet. Hier ist $p_i = 1/K$ für $i = 1, \dots, K$.

⁴Mathematiker wählen nicht den Begriff Statistik.

rücklegen aus der Formel (vgl. Mazur [25, S. 143]) aus

$$K^n = \sum_{\substack{(t_1, \dots, t_K) \\ t_1 + \dots + t_K = n}} \binom{n}{t_1, \dots, t_K},$$

wobei die Summe über alle (t_1, \dots, t_K) mit $t_1 + \dots + t_K = n$ läuft.

Auf Ω_4 wird durch die Multinomialverteilung (durch das Zusammenlegung) **kein Laplace Experiment definiert, sondern ein sogenanntes Maxwell-Boltzmann-Experiment.**

Wenn man allen n -Multimengen die Wahrscheinlichkeit

$$\frac{1}{\binom{K}{n}}$$

zuweist, dann ergibt sich ein Laplace-Experiment auf Ω_4 .

Die Verteilung wird in diesem Zusammenhang auch **Bose-Einstein-Statistik** genannt (siehe z.B. Feller [10, S. 41f]).

Wir erhalten ein weiteres Belegungsmodell, wenn wir **Mehrfachbelegungen** ausschließen. Dann erhalten wir ein **Laplace-Experiment** auf den n -**Tupeln** $\Omega_2 \subset \Omega_1$ mit paarweise unterschiedlichen Einträgen

$$\Omega_2 = \{(x_1, \dots, x_n) \mid x_i \in \{1, \dots, K\}, x_i \neq x_j, i \neq j\}$$

mit Wahrscheinlichkeiten (für alle Ergebnisse gleich)

$$\mathbb{P}_2(\{(x_1, \dots, x_n)\}) = \frac{1}{(K)_n}.$$

Wenn wir dabei auch noch die Nummern nicht beobachten/beachten, dann betrachten wir auf der Menge der n -**Mengen**

$$\Omega_3 = \{\{x_1, \dots, x_n\} \mid x_i \in \{1, \dots, K\}, x_i \neq x_j, i \neq j\}$$

die Wahrscheinlichkeiten (für alle Ergebnisse gleich)

$$\mathbb{P}_3(\{x_1, \dots, x_n\}) = \frac{1}{\binom{K}{n}}.$$

Diese Verteilung wird in der Physik **Fermi-Dirac-Statistik** genannt; vgl. Feller [10, S. 41] und Blitzstein und Hwang [4, Seite 18f]. Bei Übergang von Ω_2 auf Ω_3 entsteht diesmal ein **Laplace-Experiment**, denn immer eine gleiche Anzahl von Ereignissen kann man – wenn die Nummern verschwinden – nicht mehr unterscheiden (nämlich $n!$). Also rechnen wir nach

$$\frac{(K)_n}{n!} = \frac{K!}{(K-n)!n!} = \frac{K!}{(K-n)!K!} = \binom{K}{n}.$$

2.6 Poisson

2.6.1 Satz und Definition: Es sei $\Omega = \{0, 1, 2, 3, \dots\} = \mathbb{N}_0$ und $\lambda > 0$. Dann ist

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}$$

eine Zähldichte. Das durch diese Zähldichte definierte diskrete Wahrscheinlichkeitsmaß heißt **Poisson-Verteilung** mit Parameter λ . λ wird manchmal **Eintrittsrate** oder **Ankunftsrate** genannt.

2.6.2 Satz und Definition: Wir betrachten ein Experiment, bei dem über einen bestimmten Zeitraum (z.B. 90 Minuten) ein Ergebnis (Erfolg/Tor) mit einer *homogenen/konstanten* Rate λ eintritt. Weiter nehmen an: Wenn wir die 90 Minuten in n gleiche Teilzeiträume zerlegen, dann tritt das Ereignis mit der Rate $\frac{\lambda}{n}$ ein. Wenn das n hinreichend groß ist, dann kann man sich vorstellen, dass das Ereignis keinmal oder nur einmal im kurzen Zeitintervall nach Teilung durch n eintritt. In den kleinen Zeitintervallen haben wir dann Bernoulli-Experimente von denen wir auch annehmen wollen, dass Sie unabhängig sind. Wenn wir die *Erfolge* über 90 Minuten hinweg zählen, können wir somit das Experiment als **Binominal-Experiment mit n Wiederholung und $p = \frac{\lambda}{n}$ auffassen.**

Es gilt (vgl. z.B. Wagaman und Dobrow [34, S. 108f]):

$$\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

Da $\left(\frac{\lambda}{n}\right)$ klein ist, spricht man vom **Gesetz der kleinen Zahlen** (law of rare events).

- ▶ Die Anzahl der Soldaten, die durch einem Pferdetritt getötet werden.
- ▶ Die Anzahl der Tore in einem Fußballspiel.

2.6.3 Satz (Additionsgesetz für die Poisson-Verteilung):

Wenn X eine Poisson-Verteilung mit Parameter λ_X und Y eine Poisson-Verteilung mit Parameter λ_Y hat, dann hat $X+Y$ eine Poisson-Verteilung mit Parameter $\lambda_X + \lambda_Y$.

Beweis: Henze [18, S. 97]

- ▶ Tore der Heimteams und der Auswärtsteams. Tore insgesamt.

2.7 Geometrisch und Pascal

2.7.1 Satz und Definition: Es sei $\Omega = \{0, 1, 2, 3, \dots\} = \mathbb{N}_0$ und $p \in (0, 1)$. Dann ist

$$p_k = (1 - p)^k p$$

eine Zähldichte. Das durch diese Zähldichte definierte diskrete Wahrscheinlichkeitsmaß heißt **geometrische Verteilung** mit Parameter p .

2.7.2 Bemerkung: Die geometrische Verteilung verwendet man für das idealisierte Experiment des **Wartens** auf den ersten Erfolg. Ein Bernoulli-Experiment wird so lange unabhängig wiederholt bis der erste Erfolg eintritt. Dann erfasst p_k die Wahrscheinlichkeit von zunächst k Misserfolgen bis zum ersten Erfolg.

2.7.3 Definition und Satz: Es sei $\Omega = \{0, 1, 2, 3, \dots\} = \mathbb{N}_0$, $p \in (0, 1)$ und $n \in \mathbb{N}$. Dann ist

$$p_k = \binom{n + k - 1}{k} (1 - p)^k p^n$$

eine Zähldichte. Das durch diese Zähldichte definierte diskrete Wahrscheinlichkeitsmaß heißt **Pascal-Verteilung** oder **negative Binomialverteilung** mit Parametern n und p .

2.7.4 Bemerkung: Wir betrachten die unabhängige Wiederholung eines Bernoulli-Experiments bis erstmals n Erfolge eingetreten sind p_k ist dann die Wahrscheinlichkeit $n + k$ Versuchen erstmals n Erfolge eingetreten.

3 Standardmodelle mit Dichten

3.1 Überabzählbar ist indiskret

3.1.1 Bemerkung: Es gibt **keinen** Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit den folgenden Eigenschaften:

- i.) Ω ist **überabzählbar unendlich**,
- ii.) \mathcal{F} enthält alle ein-elementigen Mengen $\{\omega\}, \omega \in \Omega$,
- iii.) $\mathbb{P}(\{\omega\}) > 0$ für alle $\omega \in \Omega$.

Bei diskreten Wahrscheinlichkeitsräumen konnten wir auf Basis der Zähldichte/ Wahrscheinlichkeitsfunktion $p_j = \mathbb{P}(\omega_j)$ die Wahrscheinlichkeit für ein Ereignis $A \subset \Omega$ durch Summation ermitteln: $\mathbb{P}(A) = \sum_{\omega_j \in A} \mathbb{P}(\omega_j) = \sum_{\omega_j \in A} p_j$.

Wenn Ω überabzählbar unendlich ist, dann geht das so nicht mehr. Wir können weder mit Wahrscheinlichkeitsfunktionen

noch mit der Summation arbeiten. Wir benötigen **Dichten** und **Integrale**.

Für den Widerspruchsbeweis betrachtet man

$$\Omega = \bigcup_{k=1}^{\infty} \left\{ \omega \mid \frac{1}{k+1} < \mathbb{P}(\omega) \leq \frac{1}{k} \right\} = \bigcup_{k=1}^{\infty} A_k$$

Mindestens eine der Mengen A_k muss unendlich viele Elemente enthalten, sonst wäre Ω abzählbar; sagen wir A_k . Dann wäre $\mathbb{P}(A_k) > 1$.

3.2 Verteilung mit Dichte

3.2.1 Definition: Es sei $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b] : a \leq b, a, b \in \mathbb{R}\})$ die kleinste σ -Algebra, die alle Loravalle $(a, b]$ enthält, d.h. alle links offenen und rechts abgeschlossenen Intervalle. Die Mengen $B \in \mathcal{B}(\mathbb{R})$ heißen **Borelmengen** (in \mathbb{R}). Jede andere σ -Algebra, die die **Loravalle** enthält, umfasst $\mathcal{B}(\mathbb{R})$. Man sagt, dass \mathcal{B} von den Loravallen erzeugt wird. Das Konzept der „kleinsten“ σ -Algebra wird später genauer betrachtet.

3.2.2 Definition: Es sei $f : \mathbb{R} \rightarrow \mathbb{R}$ integrierbare¹ Funktion. f heißt **Dichte**, falls:

¹Was das genau bedeutet wird erst später präzise definiert. Wir gehen hier *naiv* vor.

- i.) $f \geq 0$,
 ii.) $\int_{-\infty}^{\infty} f(x)dx = 1$.

3.2.3 Definition und Satz: Es sei $f : \mathbb{R} \rightarrow \mathbb{R}$ eine Dichte. Dann existiert genau ein Wahrscheinlichkeitsmaß \mathbb{P} auf dem $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit

$$\mathbb{P}((-\infty, x]) = \int_{-\infty}^x f(z)dz, x \in \mathbb{R}.$$

In diesem Fall sagen wir: f ist die Dichte von \mathbb{P} .

3.2.4 Satz: Es sei $(\mathbb{R}, \mathcal{B}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und f die Dichte von \mathbb{P} . Dann gilt für alle $a, b \in \mathbb{R}, a < b$:

$$\begin{aligned} \mathbb{P}(\{a\}) &= 0 \\ \mathbb{P}((-\infty, a]) &= \mathbb{P}((-\infty, a)) = \int_{-\infty}^a f(x)dx \\ \mathbb{P}((a, \infty)) &= \mathbb{P}([a, \infty)) = \int_a^{\infty} f(x)dx \\ \mathbb{P}((a, b)) &= \mathbb{P}((a, b]) = \mathbb{P}([a, b)) = \mathbb{P}([a, b]) = \int_a^b f(x)dx \end{aligned}$$

3.2.5 Definition und Satz: i.) Es sei (Ω, \mathcal{F}) ein messbarer Raum und $X : \Omega \rightarrow \mathbb{R}$ mit $X^{-1}(B) \in \mathcal{F}$ für alle $B \in \mathcal{B}(\mathbb{R})$. Dann heißt X eine reellwertige Zufallsvariable. Die Eigenschaft $X^{-1}(B) \in \mathcal{F}$ für alle $B \in \mathcal{B}(\mathbb{R})$ heißt **Messbarkeit** von X . [Erinnerung $X^{-1}(B) = \{\omega | X(\omega) \in B\}$]

ii.) Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Es sei $X : \Omega \rightarrow \mathbb{R}$ eine **Zufallsvariable**. Dann heißt

$$\mathbb{P}^X = \begin{cases} \mathcal{B}(\mathbb{R}) \rightarrow [0, 1] \\ B \mapsto \mathbb{P}(X^{-1}(B)) =: \mathbb{P}(X \in B) \end{cases}$$

die **Verteilung von X** und wir schreiben $X \sim \mathbb{P}^X$.

Ist $F^X(x) = \mathbb{P}^X((-\infty, x]) = \mathbb{P}(X \leq x)$ die Verteilungsfunktion von X , dann schreiben wir auch $X \sim F^X$.

\mathbb{P}^X ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}^X)$ ist ein Wahrscheinlichkeitsraum.

iii.) Es sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Wenn f die Dichte von \mathbb{P}^X ist, dann heißt f auch die Dichte von X und wir schreiben $X \sim f$.

3.2.6 Definition: Es sei X eine Zufallsvariable mit Dichte f . Wenn das folgende Integral

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

existiert, dann heißt $\mathbb{E}(X)$ der **Erwartungswert** von X .

3.2.7 Satz: Es gilt:

- i.) $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$. Der Erwartungswert ist ein **linearer** Operator.

ii.) $X \leq Y \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y)$. Der Erwartungswert ist ein **monotoner** Operator.

3.2.8 Definition: Es sei X eine Zufallsvariable mit Dichte f , so dass die Erwartungswerte von X und $(X - \mathbb{E}(X))^2$ existieren. Dann heißt

$$\sigma_X^2 = \mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

die **Varianz** von X . σ_X heißt **Standardabweichung** von X .

► Der Erwartungswert und die Varianz sind sehr wichtige Konzepte mit vielen schönen Eigenschaften, die wir genauer untersuchen, wenn wir den geeigneten allgemeinen Integrationsbegriff eingeführt haben. Die Eigenschaften kann man auch direkt analysieren, das ist jedoch mühsamer.

3.2.9 Bemerkung: Die folgenden schönen Eigenschaften werden hier aus dem oben genannten Grund ohne Beweis und ohne detaillierte Angaben der Voraussetzungen angegeben.

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2, \\ \mathbb{V}(aX + b) &= a^2\mathbb{V}(X) = a\mathbb{V}a.\end{aligned}$$

Wenn X und Y unabhängig (eigentlich reicht un korreliert),

dann

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y).$$

3.3 Gleich-, Normal und Exponentialverteilung

3.3.1 Satz und Definition: Es sei $a < b$ und

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{falls } x \in (a, b) \\ 0 & \text{sonst} \end{cases}$$

f ist eine Dichte. Das Wahrscheinlichkeitsmaß \mathbb{P} mit dieser Dichte heißt **Gleichverteilung** auf dem Intervall (a, b) .

3.3.2 Satz und Definition: Es sei $\mu \in \mathbb{R}, \sigma > 0$ und

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

$\phi_{\mu, \sigma}(x)$ ist eine Dichte. Das Wahrscheinlichkeitsmaß \mathbb{P} mit dieser Dichte heißt **Normalverteilung** oder **Gauß-Verteilung** mit den Parametern μ und σ . Für $\mu = 0, \sigma = 1$ erhalten wir

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

In diesem Fall heißt \mathbb{P} die **Standardnormalverteilung**.

Siehe Henze [18, S. 127, 137].

3.3.3 Bemerkung: Es wird sich zeigen, dass die Normalverteilung eine Sonderrolle in der Wahrscheinlichkeitstheorie einnimmt. Die soll hier schon angedeutet werden. Wir betrachten dazu eine **Monte-Carlo Studie** im R Code Demo_N_ZgS.R.

Wir betrachten jeweils für jeden MC-Durchlauf n unabhängige Wiederholungen eines Bernoulli-Experiments mit Erfolgswahrscheinlichkeit p . Wir erhalten also eine Binomialverteilung für die Anzahl der Erfolge. Dieses Experiment – d.h. die n Wiederholungen – wiederholen wir m mal. Wir betrachten dann die relativen Häufigkeiten der Anzahl der Erfolge jeweils bei den m Durchläufen und vergleichen diese mit der Binomialverteilung und mit einer *angepassten* Normalverteilung. Für *großes* m erkennt man kaum einen Unterschied zwischen den relativen Häufigkeiten, der Binomialverteilung bzw. der **Normalverteilung**. Unser Experiment hat also zunächst nichts mit der Normalverteilung zu tun. Im *Grenzübergang* erhalten wir aber die Normalverteilung. **SEHR BEMERKENSWERT!** Mit diesem Zusammenhängen beschäftigen wir uns genauer, wenn wir den Zentralen Grenzwertsatz behandeln.

3.3.4 Satz und Definition: Es sei $\mu \in \mathbb{R}, \sigma > 0$. Dann ist

$$f(x) = \begin{cases} \frac{1}{x} \phi_{\mu, \sigma}(\ln(x)) & \text{falls } x > 0 \\ 0 & \text{sonst} \end{cases}$$

eine Dichte. Das Wahrscheinlichkeitsmaß \mathbb{P} mit dieser Dichte heißt **logarithmische Normalverteilung** (bzw. **Log-Normalverteilung**) mit den Parametern μ und σ .

3.3.5 Bemerkung: Es sei $X \sim N(\mu, \sigma)$. Wir werden später nachweisen, dass dann $Y = e^X$ log-normalverteilt ist. Anders formuliert: Ist Y log-normalverteilt, dann ist $\ln(Y)$ normalverteilt.

Siehe Henze [18, S. 159].

3.3.6 Satz und Definition: Es sei $\lambda > 0$. Dann ist

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

eine Dichte. Das Wahrscheinlichkeitsmaß \mathbb{P} mit dieser Dichte heißt **Exponentialverteilung** mit Parameter λ .

3.3.7 Bemerkung: Die **Exponentialverteilung** ist das stetige Pendant zur diskreten **geometrischen Verteilung**. Dementsprechend wird sie auch für die Modellierung von **Wartezeiten** verwendet. Bei der geometrischen Verteilung wird die

Zahl der Perioden bis zum ersten Erfolg betrachtet. Bei der Exponentialverteilung die Dauer in Zeiteinheiten bis zum ersten Erfolg betrachtet. Man kann ein Experiment mit Exponentialverteilung als Grenzwert von Experimenten mit geometrischen Verteilungen auffassen. Wir betrachten dazu die Wahrscheinlichkeit $\mathbb{P}^{\text{EXP}(\lambda)}((t, \infty))$, d.h. der erste Erfolg tritt nach t ein. Wir beobachten

$$\mathbb{P}([0, t]) = \int_0^t \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^t = 1 - e^{-\lambda t}$$

Also ist die Wahrscheinlichkeit des ersten Erfolgs nach t gleich

$$\mathbb{P}((t, \infty)) = e^{-\lambda t}.$$

Wir werden jetzt nachweisen, dass sich diese Wahrscheinlichkeit näherungsweise für ein Experiment mit geometrischer Verteilung ergibt. Zudem wird diese Approximation genauer, wenn n größer wird.

Wir betrachten Epochen E_i der Länge $\frac{t}{n}$, wobei o.B.d.A. $\mathbb{N} \ni n > t$ gelten soll. Wir zerlegen die nicht-negative Achse in Intervalle/Epochen der Länge $\frac{t}{n}$.

$$\begin{aligned} \mathbb{R}_{>0} &= \left(0, 1 \cdot \frac{t}{n}\right] \cup \left(1 \cdot \frac{t}{n}, 2 \cdot \frac{t}{n}\right] \cup \dots \cup \left((n-1) \cdot \frac{t}{n}, n \cdot \frac{t}{n}\right] \cup \dots \\ &= E_1 \cup E_2 \cup \dots \cup E_n \cup \dots \end{aligned}$$

Wir modellieren die zufällige **Anzahl** X der Epochen, die

ohne Erfolg vergehen. Wir unterstellen, dass X geometrisch verteilt mit Erfolgswahrscheinlichkeit p ist und setzen $W := X \cdot \frac{t}{n}$. Dann erfasst W die Zeit bis zum ersten Erfolg; wobei die Zeit nur in diskreten Schritten erfasst wird.

Wir betrachten

$$W > t \Leftrightarrow X \cdot \frac{t}{n} > t \Leftrightarrow X > n.$$

Wir betrachten die Wahrscheinlichkeit, dass der Erfolg irgendwann nach mehr als n Perioden eintritt. Also

$$\mathbb{P}^{\text{geo}(n)}(W > t) = \mathbb{P}(X > n) = (1 - p)^n.$$

Wir wählen jetzt $p = \lambda \frac{t}{n}$; wenn $\lambda \frac{t}{n}$ nicht kleiner 1 ist, dann müssen wir ein hinreichend großes n wählen. Diese Wahrscheinlichkeit ist proportional zur Länge der Epochen $\frac{t}{n}$, wobei der Proportionalitätsfaktor λ ist. Der Proportionalitätsfaktor erfasst die **Intensität** mit der der Erfolg eintritt. Dann ist

$$\begin{aligned} \mathbb{P}^{\text{geo}(n)}(W > t) &= \mathbb{P}(X > n) = (1 - p)^n \\ &= \left(1 - \frac{\lambda t}{n}\right)^n \rightarrow e^{-\lambda t} = \mathbb{P}^{\text{EXP}(\lambda)}((t, \infty)). \end{aligned}$$

Also erhalten wir in der Tat die Exponentialverteilung, wenn wir eine Sequenz von mit n indizierten Experimenten mit geometrischer Verteilung betrachten.

3.3.8 Bemerkung: Eine bemerkenswerte Eigenschaft der Exponentialverteilung ist die sogenannte **Gedächtnislosigkeit**

$$\mathbb{P}(X \geq t + s | X \geq s) = \mathbb{P}(X \geq t).$$

3.3.9 Definitionen: Es sei τ_1, τ_2, \dots eine Folge von unabhängigen Zufallsvariablen mit $\tau_i \sim \text{Exp}(\lambda)$. Die **Ankunftszeit** für n Sprünge ($n \in \mathbb{N}$) ist

$$S_n = \sum_{k=1}^n \tau_k.$$

Die τ_i heißen dann **Zwischenankunftszeiten**.

Für $t \geq 0$ zählt die **Zufallsvariable**

$$N_t = \begin{cases} 0 & \text{falls } 0 \leq t < S_1 \\ k & \text{falls } S_k \leq t < S_{k+1} \end{cases}$$

die Sprünge (Ankünfte) bis t . Wir bemerken, dass die Pfade von $N_t, t \geq 0$ von **rechts-stetig** sind.

Man sagt, dass N_t ein Poisson Prozess mit **Intensität** λ ist.

3.3.10 Satz: N_t hat eine Poisson Verteilung:

$$\mathbb{P}(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad k = 0, 1, \dots$$

Beweis: vgl. Shreve [30]

3.4 Transformationen

► Die folgenden beiden Sätze beschäftigen sich mit **Transformationen Zufallsvariablen mit Dichten**. Transformationen von Zufallsvariablen ergeben sich sehr oft; insbesondere werden viele in der Statistik zentralen Verteilungen durch Transformation anderer - z.B. der Normalverteilung – Verteilungen definiert. Eine ausführliche Darstellung einschließlich vieler hier nicht erläuteter technischer Einzelheiten findet man in Tappe [31, Kapitel 8]. Die Beispiele aus Tappe [31, Kapitel 8] bearbeiten wir als Übungsaufgaben. Eine weitere besonders interessante Quelle ist Blitzstein und Hwang [4]

3.4.1 Satz: Es sei X ein zufälliges Ereignis mit Werten im Intervall $I = (a, b)$, $a < b$ und Dichte f_X ; also $X \sim f_X$. Es sei ferner $g : (a, b) \rightarrow \mathbb{R}$ eine streng monoton wachsende C^1 -Funktion; mit h bezeichnen wir die Inverse von g auf $g(I)$. Dann ist

$$f_T(t) = f_X(h(t))|h'(t)|$$

die Dichte von $T = g(X)$, d.h. $T \sim f_T$. Beweis: Tappe [31, Seite 180f]

3.4.2 Korrelar: Es sei $X \sim f_X$ und $a, b \in \mathbb{R}$, $a \neq 0$. Dann

hat die Transformation $Y = aX + b$ die Dichte

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Beweis: Tappe [31, Seite 181]

3.4.3 Satz: Es sei X eine Zufallsvariable mit Dichte $f_X \sim X$ und mit Werten in $I = I_1 \uplus I_2$, wobei die I_i disjunkte offene reelle Intervalle sind. Ferner sei $g : I \rightarrow \mathbb{R}$ eine Funktion mit $g \in C^1(I_i)$ mit $g'(x) \neq 0$ für $x \in I_i$. Dann ist

$$f_T(t) = \mathbb{1}_{g(I_1)}(t) f_X(h_1(t)) |h_1'(t)| + \mathbb{1}_{g(I_2)}(t) f_X(h_2(t)) |h_2'(t)|$$

die Dichte von $T = g(X)$, wobei $h_k := g_k^{-1}$, $g_k = g|_{I_k}$

3.5 Gamma, Chi und Beta-Verteilung

► Das Skript zu den folgenden vier Verteilungen ist noch sehr fragmentarisch. Vergleiche für die **Gamma**-Verteilung, die **Chi-Quadrat**-Verteilung, die t -Verteilung und die **Beta-Verteilung** die entsprechenden Abschnitte in Blitzstein und Hwang [4] und Henze [18].

3.5.1 Definition: Wir definieren die **Gamma-Funktion**

$$\Gamma(a) = \int_0^{\infty} x^a \frac{e^{-x}}{x} dx$$

Für alle $a > 0$ gilt $\Gamma(a + 1) = a\Gamma(a)$ und insbesondere für $n \in \mathbb{N}$ gilt $\Gamma(n) = n!$.

Siehe Blitzstein und Hwang [4, Seite 357]

3.5.2 Satz und Definition: Es seien $\alpha, \lambda > 0$. Dann ist

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{falls } x > 0 \\ 0 & \text{sonst} \end{cases}$$

eine Dichte. Das Wahrscheinlichkeitsmaß \mathbb{P} mit dieser Dichte heißt **Gammaverteilung** mit Parametern λ, α (siehe z.B. Tappe [31, S. 59], Henze [18, S. 157]) und Blitzstein [4, Seite 357].

Bei $\lambda = 1, \alpha > 0$ spricht man von einer **Standard-Gamma-Verteilung**.

Bei $\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$ mit $n \in \mathbb{N}$ spricht man von einer **Chi-Quadrat-Verteilung** mit n Freiheitsgraden (FG) (siehe Georgii [13, S.277]).

3.5.3 Bemerkung: Die Gammaverteilung mit Parameter $0 < \lambda, n, \alpha = n \in \mathbb{N}$ beschreibt die *Wartezeit* bis n Erfolge bzw. Schäden eingetreten sind (siehe Georgii [13, Seite 47] oder Blitzstein [4, Seite 361]). Für $n = 1$ erhält man die Exponentialverteilung. In der Tat gilt: Wenn $X_1, X_2, \dots, X_n \sim$

i.u. $\text{Exp}(\lambda)$ sind, dann gilt

$$X_1 + X_2 + \dots + X_n \sim \text{Gamma}(n, \lambda).$$

3.5.4 Bemerkung: Wenn $X_i \sim N(0, 1), i = 1, \dots, k$, dann ist $X_1^2 + X_2^2 + \dots + X_k^2 \sim \chi^2$ mit k Freiheitsgraden. Also eine Gammaverteilung mit $\alpha = \frac{k}{2}, \lambda = \frac{1}{2}$. Die Verteilung ist in der **Statistik** sehr oft relevant.

Siehe Blitzstein [4, Seite 477 ff] und Henze [18, Seite 157].

3.5.5 Definition: t -Verteilung Henze [18, Seite 247] und Blitzstein [4, Seite 477 ff]

3.5.6 Definition: Beta-Verteilung Blitzstein [4, Seite 351 ff]

4 Verteilungs- und Quantilfunktionen

Verteilungsfunktionen werden in allen Bücher zur Wahrscheinlichkeitstheorie behandelt; also auch in Henze [18]. Quantilfunktionen werden dort auch behandelt und vergleichsweise detailliert in Föllmer und Schied [12] untersucht. Quantile und Quantilfunktionen sind insbesondere für die **Finanzmathematik (Risikomanagement)** von SEHR großer Bedeutung. In der Tat ist das bekannteste Risikmaß – der Value at Risk – ein Quantil.

4.1 Verteilungsfunktion

4.1.1 Definition: Eine Funktion $F : \mathbb{R} \rightarrow [0, 1]$ heißt **Verteilungsfunktion**, falls:

- F ist nicht fallend, d.h. $x \leq y$ impliziert $F(x) \leq F(y)$.
- F ist von rechts stetig, d.h. $\lim_{y \rightarrow x, y > x} F(y) = F(x)$.
- $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$.

4.1.2 Satz: Es sei \mathbb{P} ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$.

Dann ist

$$F = \begin{cases} \mathbb{R} \rightarrow [0, 1] \\ x \mapsto \mathbb{P}((-\infty, x]) \end{cases}$$

eine **Verteilungsfunktion**. $F(x) = \mathbb{P}((-\infty, x])$ ist also die Wahrscheinlichkeit, beim Experiment $(\Omega, \mathcal{B}, \mathbb{P})$ ein Ergebnis kleiner oder gleich x zu beobachten.

4.1.3 Satz: Es sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Verteilungsfunktion. Dann gibt es genau ein Wahrscheinlichkeitsmaß \mathbb{P} auf $(\mathbb{R}, \mathcal{B})$, so dass $\mathbb{P}((-\infty, x]) = F(x)$ für alle $x \in \mathbb{R}$ gilt.

4.1.4 Satz: i.) Es sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Verteilungsfunktion. Dann existiert für alle $x \in \mathbb{R}$ der linksseitige Grenzwert

$$\lim_{y \rightarrow x, y < x} F(y) =: F(x-).$$

Für diesen Grenzwert schreiben wir $F(x-)$.

Beweis: Vgl. Rudin [27, Seite 95]

ii.) Es sei F eine Verteilungsfunktion, dann ist die Menge der Unstetigkeitsstellen höchstens abzählbar unendlich.

4.1.5 Bemerkung: Die Unstetigkeitsstellen einer Verteilungsfunktion sind nicht notwendigerweise *isoliert* (vgl. Rudin [27, Seite 97]).

4.1.6 Satz: Es sei \mathbb{P} ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ und $F : \mathbb{R} \rightarrow [0, 1]$, die durch \mathbb{P} wie in (4.1.2) definierte Verteilungsfunktion mit $F(x) = \mathbb{P}((-\infty, x])$. Dann gilt für alle $x, y \in \mathbb{R}, x < y$:

i.) $F(x) = \mathbb{P}((-\infty, x])$

ii.) $F(x-) = \mathbb{P}((-\infty, x))$

iii.) $F(x) = 1 - \mathbb{P}((x, \infty))$

iv.) $F(x-) = 1 - \mathbb{P}([x, \infty))$

v.) $\mathbb{P}(\{x\}) = F(x) - F(x-)$ [Sprunghöhe bei x]

vi.) $\mathbb{P}((x, y]) = F(y) - F(x)$

vii.) $\mathbb{P}([x, y]) = F(y) - F(x-)$

viii.) $\mathbb{P}((x, y)) = F(y-) - F(x)$

ix.) $\mathbb{P}([x, y)) = F(y-) - F(x-)$

4.1.7 Satz: Es sei $f : \mathbb{R} \rightarrow \mathbb{R}$ eine Dichte. Dann ist $F :$

$\mathbb{R} \rightarrow [0, 1]$ mit

$$F(x) = \int_{-\infty}^x f(z) dz$$

eine Verteilungsfunktion.

Man spricht von der Verteilungsfunktion F mit der **Dichte** f .

4.1.8 Satz: Es sei $\Omega \subset \mathbb{R}$ höchstens abzählbar, $\mathcal{F} = \mathcal{P}(\Omega)$ und \mathbb{P} ein **diskretes** Wahrscheinlichkeitsmaß auf (Ω, \mathcal{F}) mit Zähldichte p . Dann ist $F : \mathbb{R} \rightarrow [0, 1]$ mit

$$F(x) = \sum_{k \in (-\infty, x] \cap \Omega} p_k$$

eine Verteilungsfunktion. Das ist eine Treppenfunktion mit Stufen an den Stellen k mit Stufenhöhe p_k .

4.1.9 Bemerkung: Es gibt auch Verteilungsfunktionen, die weder eine Dichte haben noch diskret sind: Verteilungsfunktionen mit Sprungstellen, die aber keine (reinen) Treppenfunktionen sind.

4.2 Quantile, Quantilsfunktionen und Verallgemeinerte Inverse

4.2.1 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, X eine reellwertige Zufallsvariable mit Verteilungsfunktion F und $\lambda \in (0, 1)$. Eine reelle Zahl q heißt λ -**Quantil** der Zufallsvariable X , falls

$$\begin{aligned} \mathbb{P}(X \leq q) &\geq \lambda \quad \text{und} \\ \mathbb{P}(X \geq q) &\geq 1 - \lambda . \end{aligned}$$

In Worten: (i) Die Wahrscheinlichkeit, dass X die Grenze q nicht überschreitet ist mindestens λ **und** (ii) die Wahrscheinlichkeit, dass X die Grenze q nicht unterschreitet ist mindestens $1 - \lambda$.

► Das ist vielleicht eine/die Gelegenheit in z.B. Fahrmeir et al. [11, Seite 59 ff] die Abschnitte zu empirischen Quantilen nachzulesen.

4.2.2 Bemerkung: Es gilt (wir beginnen mit der zweiten Bedingung aus der Definition)

$$\begin{aligned} 1 - \lambda &\leq \mathbb{P}(X \geq q) \\ \Leftrightarrow 1 - \mathbb{P}(X \geq q) &\leq \lambda \\ \Leftrightarrow F(q-) = \mathbb{P}(X < q) &\leq \lambda \end{aligned}$$

Dementsprechend ist q ein λ -**Quantil** der Zufallsvariable X , falls

$$F(q) = \mathbb{P}(X \leq q) \geq \lambda \text{ und} \\ F(q-) = \mathbb{P}(X < q) \leq \lambda.$$

4.2.3 Satz: Es sei X eine Zufallsvariable und $X \sim F$. Es sei $\lambda \in (0, 1)$. $q \in \mathbb{R}$ ist genau dann ein λ -Quantil von X , wenn

$$F(q-) \leq \lambda \leq F(q+) = F(q)$$

ist.

4.2.4 Definition: Es sei F eine Verteilungsfunktion. Dann heißt $(0, 1) \ni \lambda \mapsto q^+(\lambda)$ mit

$$q^+(\lambda) = \inf\{q \in \mathbb{R} : F(q) > \lambda\} \\ = \sup\{q \in \mathbb{R} : F(q) \leq \lambda\}$$

die **obere Quantilsfunktion** von F und $(0, 1) \ni \lambda \mapsto q^-(\lambda)$

$$q^-(\lambda) = \sup\{q \in \mathbb{R} : F(q) < \lambda\} \\ = \inf\{q \in \mathbb{R} : F(q) \geq \lambda\}$$

die **untere Quantilsfunktion** von F .

Insbesondere im **Risikomanagement** nennt man die untere

re Quantilsfunktion auch **Verallgemeinerte Inverse** und schreibt

$$F^{\leftarrow}(\lambda) = q^{-}(\lambda)$$

4.2.5 Bemerkung: Wie sieht die Menge $A := \{q \in \mathbb{R} : F(q) \geq \lambda\}$ aus? Wenn $q_1 \in A$ ist und $q_2 > q_1$ gilt, dann ist $F(q_2) \geq F(q_1) \geq \lambda$. Also ist auch $q_2 \in A$. Also ist A ein Intervall der Form (z, ∞) oder der Form $[z, \infty)$. In der Tat hat das Intervall wegen der von-rechts-Stetigkeit von F die Form $[z, \infty)$. Also gilt

$$\begin{aligned} q^{-}(\lambda) &= \inf\{q \in \mathbb{R} : F(q) \geq \lambda\} \\ &= \min\{q \in \mathbb{R} : F(q) \geq \lambda\}. \end{aligned}$$

Wir können also anstatt des Infimums das Minimum bilden. Das Minimum wird angenommen: Es gibt also ein q^* mit der Eigenschaft $F(q^*) \geq \lambda$ und für alle $q < q^*$ gilt $F(q) < \lambda$.

Beweis: Angenommen $z = \inf\{q \in \mathbb{R} : F(q) \geq \lambda\}$, aber $F(z) \geq \lambda$ gilt nicht. Also $F(z) < \lambda$. Wir betrachten eine Folge (q_i) mit $q_i > z$ und $q_i \rightarrow z$ (eine Folge die von links gegen z konvergiert). Wegen der Stetigkeit von rechts für F gilt $F(q_i) \rightarrow F(z)$. Dann muss es – da $F(q) < \lambda$ – ein j mit $F(q_j) < \lambda$ und $q_j > z$ geben. Das ist jedoch ein Widerspruch, denn für alle q im Intervall (z, ∞) gilt $F(q) \geq \lambda$.

4.2.6 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine reellwertige Zufallsvariable mit Verteilungsfunktion F .

(1) Dann ist q^- nicht-fallend und von links stetig und

(2) q^+ nicht-fallend von rechts stetig.

(3) Es gilt $q^-(\lambda) \leq q^+(\lambda)$.

Beweis: Siehe Föllmer und Schied [12, S. 538f]

4.2.7 Satz: Es sei F eine Verteilungsfunktion. Für jedes $\lambda \in (0, 1)$ ist **die Menge der λ -Quantile das abgeschlossene Intervall $[q^-(\lambda), q^+(\lambda)]$.**

4.2.8 Bemerkung: Wenn F eine strikt monotone stetige Verteilungsfunktion ist, dann ist für alle $\lambda \in (0, 1)$:

$$q^+(\lambda) = q^-(\lambda) = F^{\leftarrow}(\lambda).$$

4.2.9 Lemma: Es sei F eine Verteilungsfunktion und

$$F^{\leftarrow}(\lambda) = q^-(\lambda) = \inf\{x | F(x) \geq \lambda\} = \min\{x | F(x) \geq \lambda\}$$

die untere Quantilsfunktion bzw. **eine Verallgemeinerte Inverse** von F . Dann gilt für $x \in \mathbb{R}$, $\lambda \in (0, 1)$

$$F(x) \geq \lambda \Leftrightarrow x \geq F^{\leftarrow}(\lambda).$$

Beweis (Siehe Henze [18, 151]): Für \Rightarrow müssen wir nur bemerken, dass $\tilde{x} = F^{\leftarrow}(\lambda) = \min\{x | F(x) \geq \lambda\}$ gemäß Definition das kleinste x mit $F(x) \geq \lambda$ ist. Also $x \geq \tilde{x} = F^{\leftarrow}(\lambda)$.

Für \Leftarrow nehmen wir an, dass $x \geq F^{\leftarrow}(\lambda) = \tilde{x}$ gibt, aber $F(x) < \lambda$ ist. Wir betrachten eine Folge (x_i) mit $x_i > x$ und $x_i \rightarrow x$. Wegen der Stetigkeit von rechts von F gilt $F(x_i) \rightarrow F(x)$. Aus der Konvergenz $F(x_i) \rightarrow F(x)$ und $F(x) < \lambda$ folgt, dass es ein $x_j > x \geq \tilde{x}$ mit $F(x_j) < \lambda$ gibt. Es gäbe also $x_j > x \geq F^{\leftarrow}(\lambda)$ mit $F(x_j) < \lambda$. Das ist aber ein Widerspruch, denn aus $x_j \geq \tilde{x}$ folgt $F(x_j) \geq F(\tilde{x}) = \lambda$.

4.2.10 Lemma: Es sei F eine Verteilungsfunktion und $U \sim \text{Unif}((0,1))$. Dann hat die Zufallsvariable

$$X = F^{\leftarrow}(U)$$

die Verteilungsfunktion F ; $X \sim F$.

Beweis: Siehe Henze [18, 153]. Es gilt gemäß des obigen Lemmas

$$\begin{aligned} \mathbb{P}(X \leq x) &\leq \mathbb{P}(F^{\leftarrow}(U) \leq x) && X = F^{\leftarrow}(U) \\ &= \mathbb{P}(U \leq F(x)) && \text{Lemma} \\ &= F(x). && \text{VF von Unif} \end{aligned}$$

► Auf der Grundlage des vorhergehenden Lemmas, kann man Zufallszahlen gemäß einer Verteilung F erzeugen, wenn man Zufallszahlen gemäß einer Gleichverteilung erzeugen kann.

4.2.11 Lemma: Es sei X eine Zufallsvariable mit Verteilungsfunktion F . Ferner sei F stetig. Dann gilt $F(X) \sim \text{Unif}([0, 1])$. Also für $u \in [0, 1]$

$$\mathbb{P}(F(X) \leq u) = u$$

Beweis: Siehe Henze [18, 153].

4.3 Risikomessung

► Der Rest des Paragraphen ist als Vertiefung bzw. Anwendung für Risikomanager gedacht und noch arg fragmentarisch.

4.3.1 Terminologie und Einführung: Im folgenden stellen wir uns vor, dass ein Investor oder Manager das Risiko einer Vermögensposition analysiert. Den Wert dieser Vermögensposition bezeichnen wir mit V . Wir stellen uns ferner vor, dass der Wert in mindestens zwei Zeitpunkten betrachtet wird. Wir verwenden $t = 0$ für den Zeitpunkt zu dem der Vermögenswert fix und bekannt ist. Wir bezeichnen diesen

Wert mit V_0 . Den Wert V der Vermögensposition zu einem zukünftigen Zeitpunkt sehen wir als eine Zufallsvariable an, dessen Wert uns für den Zeitpunkt $t = 1$ interessiert. Wir verzichten manchmal bei der Variable V für den Zeitpunkt $t = 1$ auf den Index, d.h. wir schreiben einfach V anstatt V_1 . Wir untersuchen meistens nicht V , sondern die **Veränderung** $G = V - V_0$ oder den **Verlust** $L = V_0 - V$. Trotzdem sprechen wir vom Value at Risk der Vermögensposition. G bezeichnet einen Gewinn (der natürlich auch negativ sein kann). Da wir Risikoprüfung betreiben, werden wir – der Literatur folgenden – oft die Variable Verlust $L = -G$ betrachten.

4.3.2 Definition: Es sei V ein Vermögenswert und $G = V - V_0$ und $L = -G$. Dann ist der **Value-at-Risk (V@R)** von V zur **Risikotoleranz** λ (eine kleine Zahl, z.B. $\lambda = 0.01$) die reelle Zahl

$$\begin{aligned} \text{V@R}_\lambda(G) &= -q_{FG}^+(\lambda) \\ &= q_{F-G}^-(1 - \lambda) \\ &= q_{FL}^-(1 - \lambda). \end{aligned}$$

Wenn wir Verluste betrachtet (α nahe 1, z.B. $\alpha = 0.99$)

$$\begin{aligned} \text{V@R}_\alpha(L) &= q_{F_L}^-(\alpha) \\ &= F_L^{\leftarrow}(\alpha) \\ &= \inf\{x | F_L(x) \geq \alpha\} \\ &= \min\{x | F_L(x) \geq \alpha\} \end{aligned}$$

dabei heißt α das **Sicherheitsniveau (Konfidenzniveau)**.

Die Abbildung

$$F^{\leftarrow}(\lambda) = \inf\{x | F(x) \geq \lambda\} = \min\{x | F(x) \geq \lambda\}$$

die **Verallgemeinerte Inverse** von F

4.3.3 Satz: Es gilt

$$\text{V@R}_\lambda(G) = \inf\{m | \mathbb{P}(G + m < 0) \leq \lambda\}.$$

bzw.

$$\text{V@R}_\alpha(L) = \inf\{m | \mathbb{P}(L - m > 0) \leq 1 - \alpha\}$$

Beweis: Es gilt gemäß Definition

$$\text{V@R}_\lambda(G) = q_{F_L}^-(1 - \lambda) = \inf\{m \in \mathbb{R} : F_L(m) \geq 1 - \lambda\}.$$

Wir zeigen, dass $F_L(m) \geq 1 - \lambda$ gdw $\mathbb{P}(G + m < 0) \leq \lambda$.

Dann muss natürlich auch $\inf\{m \in \mathbb{R} : F_L(m) \geq 1 - \lambda\} =$

$\inf\{m \mid \mathbb{P}(G + m < 0) \leq \lambda\}$ gelten.

Für m gilt $F_L(m) = \mathbb{P}(L \leq m)$. Also gilt für m

$$\begin{aligned} F_L(m) &\geq 1 - \lambda \\ \Leftrightarrow \mathbb{P}(L \leq m) &\geq 1 - \lambda \\ \Leftrightarrow \lambda &\geq 1 - \mathbb{P}(L \leq m) \end{aligned}$$

und

$$1 - \mathbb{P}(L \leq m) = \mathbb{P}(L > m) = \mathbb{P}(G < -m) = \mathbb{P}(G + m < 0).$$

Also gilt für m

$$F_L(m) \geq 1 - \lambda \Leftrightarrow \mathbb{P}(G + m < 0) \leq \lambda.$$

Damit haben wir den ersten Teil des Satzes bewiesen.

Für den zweiten Teil beobachten wir

$$\begin{aligned} G + m &< 0 \\ \Leftrightarrow -L + m &< 0 \\ \Leftrightarrow m &< L \\ \Leftrightarrow 0 &< L - m. \end{aligned}$$

Also auch

$$\mathbb{P}(G + m < 0) = \mathbb{P}(0 < L - m).$$

Also auch

$$\inf\{m \mid \mathbb{P}(G + m < 0) \leq \lambda\} = \inf\{m \mid \mathbb{P}(0 < L - m) \leq \lambda\}.$$

4.3.4 Bemerkung: (1) Wir reden vom Value at Risk von V , obwohl die Definition auf G bzw. L Bezug nimmt. Wir werden auch vom Value at Risk von G oder von L sprechen. Gemeint ist das negative des oberen λ -Quantil der Zufallsvariable G bzw. das untere $\alpha = (1 - \lambda)$ -Quantil des Verlustes L .

(2) Der Value at Risk wird in den Einheiten gemessen, in denen die Zufallsvariable G gemessen wird, d.h. in der Regel in **Geldeinheiten**.

(3) Typischerweise wird das für die Risikoprüfung relevante λ -Quantil $q_{FG}^+(\lambda)$ von G negativ sein.

(4) Der Value at Risk ist wegen der Multiplikation mit -1 so definiert, dass er die Grundlage für eine Kapitalanforderung ist; das wird weiter unten noch verdeutlicht.

(5) Wenn man anstatt des Zuwachses G die Verlustvariable L betrachtet, dann sind für den Übergang **drei Anpassung** nötig: (i) Anstatt der oberen Quantilsfunktion betrachtet man die untere Quantilsfunktion. (ii) Anstatt der Risikotoleranz λ betrachtet man das **Sicherheitsniveau** $\alpha = 1 - \lambda$. (iii) Die Multiplikation mit -1 entfällt.

(6) Es ist $G + m = -L + m$. Also ist $G + m < 0$ genau

dann, wenn $m < L$. Wenn wir m als **Eigenkapital** auffassen, dann erkennen wir, dass $V@R_\lambda(G)$ der kleinste Eigenkapitalwert ist, der ausreicht eine Insolvenz – hier definiert als Verluste größer als das Eigenkapital – mit Wahrscheinlichkeit λ zu vermeiden. Man kann also den $V@R$ im Kontext der **Eigenkapitalregulierung** unmittelbar anwenden. Der $V@R$ ist auch aus diesem Grund populär.

4.3.5 Beispiel: Es sei $G \sim N(\mu, \sigma^2)$ Normalverteilt und die Risikotoleranz $\lambda \in (0, 1)$.

$$V@R_\lambda(G) = -\mathbb{E}(G) - \Phi^{-1}(\lambda) \cdot \sigma \quad (4.1)$$

$$= -\mu - \Phi^{-1}(\lambda) \cdot \sigma \quad (4.2)$$

Für $\lambda = 0.01$ erhalten wir $\Phi^{-1}(0.01) = -2.33$, so dass $V@R_\lambda(X) = -\mu + 2.33 \cdot \sigma$. Für $\lambda = 0.0001$ ist $V@R_\lambda(X) = -\mu + 3.09 \cdot \sigma$. Praktisch: Wenn wir den Erwartungswert und die Varianz von X geschätzt haben und die Annahme der Normalverteilung angemessen ist, dann können wir für den Fall einer Normalverteilung den $V@R$ leicht ermitteln.

4.3.6 Satz: Der Vermögenswert V sei eine Zufallsvariable und $V_0 \neq 0$ eine reelle Zahl. Ferner sei $R = \frac{V-V_0}{V_0}$ eine Zufallsvariable mit endlicher Varianz σ^2 und Erwartungswert μ und die Verteilungsfunktion $F_{R^{sz}}$ der Zufallsvariable $R^{sz} = \frac{R-\mu}{\sigma}$

sei strikt monoton steigend und stetig. Dann gilt

$$V@R_\lambda(V) = -(\mu + \sigma \cdot F_{R^{sz}}^{-1}(\lambda))V_0.$$

Beweis: Da die Verteilungsfunktion annahmegemäß strikt monoton steigend und stetig ist gilt für den V@R die Gleichung

$$\lambda = \mathbb{P}(V - V_0 \leq -V@R_\lambda(V))$$

Dann folgt

$$\begin{aligned} \lambda &= \mathbb{P}\left(\frac{R - \mu}{\sigma} \leq -\frac{V@R_{t,\lambda}(V)}{V_0\sigma} - \frac{\mu}{\sigma}\right) \\ &= \mathbb{P}\left(R^{sz} \leq -\frac{V@R_{t,\lambda}(V)}{V_0\sigma} - \frac{\mu}{\sigma}\right) \end{aligned}$$

Also

$$\begin{aligned} F_{R^{sz}}^{-1}(\lambda) &= -\frac{V@R_{t,\lambda}(V)}{V_0\sigma} - \frac{\mu}{\sigma} \\ \Rightarrow V@R_\lambda(V) &= -(\mu + \sigma \cdot F_{R^{sz}}^{-1}(\lambda))V_0 \end{aligned}$$

4.3.7 Bemerkung: Der Betrachtungszeitpunkt sei $\tau - 1$. Wir wollen eine Risikoeinschätzung für τ ermitteln. In der Praxis geht man regelmäßig so vor.

i.) Zunächst erstellt man für die Rendite ein Zeitreihenmodell der Form

$$R_t = \mu_t + \sigma_t z_t, t \in \mathbb{Z},$$

wobei $\mu_t = \mathbb{E}_{t-1}(R_t)$, $\mathbb{V}_{t-1}(R_t) = \sigma_t^2$. Ferner wird angenommen, dass $(z_t) \sim \text{iid}(0, 1)$ mit Verteilungsfunktion $F_{R^{sz}}$ ist ((z_t) ist also striktes **weißes Rauschen**).

ii.) Wir verwenden den Satz (4.3.6) und schätzen den aktuellen Value at Risk mit

$$\text{V@R}_{\lambda, \tau}(V) = -(\mu_\tau + \sigma_\tau \cdot F_{R^{sz}}^{-1}(\lambda))V_{\tau-1}.$$

Bei dieser Methoden muss man also zwei statistische Probleme lösen:

- Ein Zeitreihenmodell $R_t = \mu_t + \sigma_t z_t$ schätzen; mindestens muss man μ_t und σ_t schätzen.
- Die Verteilungsfunktion $F_{R^{sz}}$ schätzen.

4.3.8 Bemerkung: Value at Risk ist nicht unumstritten. Zwei Eigenschaften sind problematisch:

- V@R ist im Allgemeinen nicht sub-additiv.
- V@R beachtet nicht, wie die Verteilung von X links von V@R gestaltet ist.

4.3.9 Behauptung: Es sei X eine Zufallsvariable mit Ver-

teilungsfunktion $F = F^X$. Dann gilt

$$\mathbb{E}(X) = \int X dF = \int_0^1 F^{\leftarrow}(u) du.$$

Beweis: (den Beweis können wir eigentlich noch nicht würdigen, da wir Integration bezüglich dF und den Umgang mit solchen Lebesgue-Stieltjes-Integralen noch nicht kennen).

Es sei U eine Zufallsvariable mit $U \sim \text{Unif}((0,1))$. Dann hat (auch) die Zufallsvariable $Y = F^{\leftarrow}(U)$ die Verteilungsfunktion F . Also gilt

$$\begin{aligned} \mathbb{E}(X) &= \int Y dF \\ &= \int F^{\leftarrow}(U) dF^U \\ &= \int F^{\leftarrow}(u) du. \end{aligned}$$

4.3.10 Beispiel: Wir betrachten $X \sim \text{Bernoulli}$ mit $\text{Bild}(X) = \{-b, a\}$ für $a, b > 0$ und $\mathbb{P}(X = a) = p$. Dann ist für $\lambda \in (0, 1)$

$$F^{\leftarrow}(\lambda) = \begin{cases} -b & \text{falls } \lambda \leq 1 - p \\ a & \text{falls } \lambda > 1 - p \end{cases}$$

Es gilt einerseits

$$\mathbb{E}(X) = pa - (1 - p)b.$$

Andererseits

$$\begin{aligned}
 \int F^{\leftarrow}(u)du &= \int_0^{1-p} (-b)du + \int_{1-p}^1 a du \\
 &= (-b)(1-p) + a(1 - (1-p)) \\
 &= (-b)(1-p) + ap \\
 &= pa - (1-p)b.
 \end{aligned}$$

4.3.11 Bemerkung: Es sei X eine Zufallsvariable mit strikt monoton wachsender C^1 -Verteilungsfunktion F und $f = F'$. Dann ist F invertierbar. Wir betrachten die Substitution $u = F(x)$, $x = F^{-1}(u)$ und verwenden die Substitutionsregel der Integration $du = F'(x)dx = f(x)dx$

$$\mathbb{E}(X) = \int x f(x)dx = \int F^{-1}(u)du.$$

4.3.12 Definition: Es sei L eine Zufallsvariable mit $L \sim F$ und $\mathbb{E}(|L|) < \infty$. Wir definieren

$$\begin{aligned}
 \text{ES}_\alpha(L) &= \frac{1}{1-\alpha} \int_\alpha^1 F^{\leftarrow}(u)du \\
 \text{TVaR}_\alpha(L) &= \mathbb{E}(L|L > F^{\leftarrow}(\alpha))
 \end{aligned}$$

ES steht für **Expected Shortfall** und TVaR für **Tail-Value-at-Risk**.

► Die folgende Bemerkung stellt Resultate über den Zusam-

menhang zwischen ES und TVaR zusammen. TVaR ist anschaulicher, jedoch nicht sub-additiv. ES ist sub-additiv.

4.3.13 Bemerkung (McNeil et al. [7, Seite 283]): Es sei L eine Verteilungsfunktion und $L \sim F$. Dann gilt

$$\text{ES}_\alpha(L) = \frac{1}{1-\alpha} \mathbb{E} [[L - F^{\leftarrow}(\alpha)]^+] + F^{\leftarrow}(\alpha)$$

bzw.

$$\begin{aligned} \text{ES}_\alpha(L) &= \frac{1}{1-\alpha} [\mathbb{E} [\mathbb{1}_{L > F^{\leftarrow}(\alpha)} \cdot L] + F^{\leftarrow}(\alpha)(1-\alpha - \mathbb{P}(L > F^{\leftarrow}(\alpha)))] \\ &= \mathbb{E}(L|L > F^{\leftarrow}(\alpha)) \cdot \gamma + F^{\leftarrow}(\alpha) \cdot (1-\gamma) \\ &= \text{TVaR}_\alpha(L) \cdot \gamma + \text{VaR}_\alpha(L) \cdot (1-\gamma). \end{aligned}$$

Dabei ist

$$\begin{aligned} \gamma &= \frac{\mathbb{P}(L > F^{\leftarrow}(\alpha))}{1-\alpha} \\ &= \frac{1 - \mathbb{P}(L \leq F^{\leftarrow}(\alpha))}{1-\alpha} = \frac{1 - F(F^{\leftarrow}(\alpha))}{1-\alpha}. \end{aligned}$$

Ferner gilt

$$\mathbb{E} [[L - F^{\leftarrow}(\alpha)]^+] = \mathbb{E} [\mathbb{1}_{L > F^{\leftarrow}(\alpha)} \cdot L] - F^{\leftarrow}(\alpha) \cdot \mathbb{P}(L > F^{\leftarrow}(\alpha)).$$

Wenn F^L an der Stelle $F^{\leftarrow}(\alpha)$ stetig ist, dann gilt

$$\text{ES} = \mathbb{E}(L|L > F^{\leftarrow}(\alpha)) = \text{TVaR}.$$

Beweis: Wir beachten

$$\begin{aligned}
 \frac{1}{1-\alpha} \mathbb{E} \left[[L - F^{\leftarrow}(\alpha)]^+ \right] &= \frac{1}{1-\alpha} \int_0^1 [F^{\leftarrow}(u) - F^{\leftarrow}(\alpha)]^+ du \\
 &= \frac{1}{1-\alpha} \int_{\alpha}^1 [F^{\leftarrow}(u) - F^{\leftarrow}(\alpha)]^+ du \\
 &= \frac{1}{1-\alpha} \int_{\alpha}^1 F^{\leftarrow}(u) du - \frac{F^{\leftarrow}(\alpha)}{1-\alpha} \int_{\alpha}^1 1 du \\
 &= \frac{1}{1-\alpha} \int_{\alpha}^1 F^{\leftarrow}(u) du - F^{\leftarrow}(\alpha).
 \end{aligned}$$

Also

$$\begin{aligned}
 \text{ES}_{\alpha}(X) &= \frac{1}{1-\alpha} \int_{\alpha}^1 F^{\leftarrow}(u) du \\
 &= \frac{1}{1-\alpha} \mathbb{E} \left[[L - F^{\leftarrow}(\alpha)]^+ \right] + F^{\leftarrow}(\alpha)
 \end{aligned}$$

Weiterhin beachten wir

$$\begin{aligned}
 \mathbb{E} \left[[L - F^{\leftarrow}(\alpha)]^+ \right] &= \mathbb{E} \left[\mathbb{1}_{L > F^{\leftarrow}(\alpha)} \cdot (L - F^{\leftarrow}(\alpha)) \right] \\
 &= \mathbb{E} \left[\mathbb{1}_{L > F^{\leftarrow}(\alpha)} \cdot L \right] - \mathbb{E} \left[\mathbb{1}_{L > F^{\leftarrow}(\alpha)} \cdot F^{\leftarrow}(\alpha) \right] \\
 &= \mathbb{E} \left[\mathbb{1}_{L > F^{\leftarrow}(\alpha)} \cdot L \right] - F^{\leftarrow}(\alpha) \cdot \mathbb{P}(L > F^{\leftarrow}(\alpha))
 \end{aligned}$$

Also weiterhin

$$\begin{aligned}
 \text{ES}_\alpha(X) &= \frac{1}{1-\alpha} \mathbb{E} \left[[L - F^{\leftarrow}(\alpha)]^+ \right] + F^{\leftarrow}(\alpha) \\
 &= \frac{\mathbb{E} \left[\mathbb{1}_{L > F^{\leftarrow}(\alpha)} \cdot L \right] - F^{\leftarrow}(\alpha) \cdot \mathbb{P}(L > F^{\leftarrow}(\alpha))}{1-\alpha} + F^{\leftarrow}(\alpha) \\
 &= \frac{\mathbb{E} \left[\mathbb{1}_{L > F^{\leftarrow}(\alpha)} \cdot L \right] - F^{\leftarrow}(\alpha) \cdot \mathbb{P}(L > F^{\leftarrow}(\alpha))}{1-\alpha} + F^{\leftarrow}(\alpha) \\
 &= \frac{\mathbb{E} \left[\mathbb{1}_{L > F^{\leftarrow}(\alpha)} \cdot L \right] - F^{\leftarrow}(\alpha) \mathbb{P}(L > F^{\leftarrow}(\alpha)) + (1-\alpha)F^{\leftarrow}(\alpha)}{1-\alpha}
 \end{aligned}$$

4.3.14 Bemerkung: Man kann den ES als Erwartungswert einer Zufallsvariable eines zwei-stufigen Experiments auffassen: Zunächst mit Wahrscheinlichkeit $1 - \gamma$ der fixe Wert $F^{\leftarrow}(\alpha)$ oder mit Wahrscheinlichkeit γ die Zufallsvariable L mit Verteilung unter der Bedingung $L > F^{\leftarrow}(\alpha)$. Dann erhalten wir den Erwartungswert

$$\text{ES} = \gamma \cdot \mathbb{E}(L | L > F^{\leftarrow}(\alpha)) + (1 - \gamma) \cdot F^{\leftarrow}(\alpha)$$

mit

$$\gamma = \frac{\mathbb{P}(L > F^{\leftarrow}(\alpha))}{1 - \alpha}.$$

4.3.15 Definition und Behauptung (Rockafellar und Uryasev [26, Seite 1449]): Es sei F eine Verteilungsfunk-

tion. Wir definieren

$$F^{\alpha\text{-Tail}}(x) = \begin{cases} 0 & x < F^{\leftarrow}(\alpha) \\ \frac{F(x) - \alpha}{1 - \alpha} & x \geq F^{\leftarrow}(\alpha) \end{cases}$$

$F^{\alpha\text{-Tail}}(x)$ ist eine Verteilungsfunktion.

4.3.16 Satz (Rockafellar und Uryasev [26, Seite 1448]):

Es sei L eine Verteilungsfunktion und $L \sim F$. Es gilt

$$\begin{aligned} \text{TVaR} &= \int L dF^{\alpha\text{-Tail}} \\ &= \mathbb{E}(L | L \text{ im } \alpha\text{-Tail}). \end{aligned}$$

4.3.17 Beispiel: Es sei $\text{Bild}(L) = \{\dots, 2, 3, 4\}$ mit $\mathbb{P}(L = 2) = \mathbb{P}(L = 3) = \mathbb{P}(L = 4) = 0.02$. Es sei $\alpha = 0.95$. Dann ist $F^{\leftarrow}(\alpha) = 2$. Es ist $\mathbb{P}(L > F^{\leftarrow}(\alpha)) = 0.96$ und $\mathbb{P}(L = F^{\leftarrow}(\alpha)) = 0.02$.

Dann gilt einerseits

$$\begin{aligned} \int L dF^{\alpha\text{-Tail}} &= 2 \cdot \frac{0.96 - 0.95}{1 - 0.95} + 3 \cdot \frac{0.98 - 0.96}{1 - 0.95} + 4 \cdot \frac{1 - 0.98}{1 - 0.95} \\ &= 2 \cdot \frac{0.01}{0.05} + 3 \cdot \frac{0.02}{0.05} + 4 \cdot \frac{0.02}{0.05} \\ &= 2 \cdot \frac{1}{5} + 3 \cdot \frac{2}{5} + 4 \cdot \frac{2}{5} = \frac{2 + 6 + 8}{5} \\ &= \frac{16}{5} = 3.2 \end{aligned}$$

Andererseits gilt

$$F^{\leftarrow}(u) = \begin{cases} 4 & \text{für } 0.98 \leq u \\ 3 & \text{für } 0.96 \leq u < 0.98 \\ 2 & \text{für } 0.94 \leq u < 0.96 \\ \dots & \dots \dots \end{cases}$$

Also

$$\begin{aligned} & \frac{1}{1-\alpha} \int_{0.95}^1 F^{\leftarrow}(u) du \\ &= \frac{1}{0.05} (2[x]_{0.95}^{0.96} + 3[x]_{0.96}^{0.98} + 4[x]_{0.98}^1) \\ &= \frac{1}{0.05} (2 \cdot (0.96 - 0.95) + 3 \cdot (0.98 - 0.96) + 4 \cdot (1 - 0.98)) \\ &= \frac{1}{0.05} (2 \cdot 0.01 + 3 \cdot 0.02 + 4 \cdot 0.02) = \frac{0.16}{0.05} \\ &= 3.2 \end{aligned}$$

Es gibt noch eine dritte Möglichkeit TVaR berechnen. Es ist

$$\gamma = \frac{\mathbb{P}(L > F^{\leftarrow}(\alpha))}{1-\alpha} = \frac{0.04}{0.05} = \frac{4}{5}$$

und weiterhin

$$\begin{aligned} & (1 - \gamma) \cdot F^{\leftarrow}(\alpha) + \gamma \cdot \mathbb{E}(L|L > F^{\leftarrow}(\alpha)) \\ &= \frac{1}{5} \cdot 2 + \frac{4}{5} \cdot \left(3 \cdot \frac{\mathbb{P}(L = 3)}{\mathbb{P}(L > 2)} + 4 \cdot \frac{\mathbb{P}(L = 4)}{\mathbb{P}(L > 2)} \right) \\ &= \frac{1}{5} \cdot 2 + \frac{4}{5} \cdot \left(3 \cdot \frac{0.02}{0.04} + 4 \cdot \frac{0.02}{0.04} \right) \\ &= \frac{2}{5} + \frac{4}{5} \cdot \left(3 \cdot \frac{2}{4} + 4 \cdot \frac{2}{4} \right) \\ &= \frac{2}{5} + \frac{4}{5} \cdot \left(\frac{6}{4} + \frac{8}{4} \right) = \frac{2}{5} + \frac{4}{5} \cdot \left(\frac{14}{4} \right) = \frac{16}{5} \\ &= 3.2 \end{aligned}$$

5 Lebesgue-Stieltjes

Integral -

Erwartungswert

bezüglich \mathbb{P} bzw. F

5.1 Grundlagen und Maßtheoretische Induktion

5.1.1 Definition und Satz: i.) Es seien $(\Omega_1, \mathcal{F}_1)$ und $(\Omega_2, \mathcal{F}_2)$ messbare Räume. Eine Abbildung $X : \Omega_1 \rightarrow \Omega_2$ heißt **messbar**, falls $X^{-1}(B) \in \mathcal{F}_1$ für alle $B \in \mathcal{F}_2$. Das Urbild jeder \mathcal{F}_2 -Menge ist eine \mathcal{F}_1 -Menge.

ii.) Es sei $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(\Omega_2, \mathcal{F}_2)$ ein messbarer Raum und $X : \Omega_1 \rightarrow \Omega_2$ messbar. Dann heißt

X eine **Zufallsvariable**.

iii.) Es sei $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(\Omega_2, \mathcal{F}_2)$ ein messbarer Raum und $X : \Omega_1 \rightarrow \Omega_2$ eine Zufallsvariable. Dann heißt die Abbildung

$$\mathbb{P}^X = \begin{cases} \mathcal{F}_2 \rightarrow [0, 1] \\ B \mapsto \mathbb{P}(X^{-1}(B)) =: \mathbb{P}(X \in B) \end{cases}$$

die **Bildmaß von X** . \mathbb{P}^X ist ein Wahrscheinlichkeitsmaß auf $(\Omega_2, \mathcal{F}_2)$.

Insbesondere falls $(\Omega_2, \mathcal{F}_2) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist, nennen wir \mathbb{P}^X auch die **Verteilung** von X .

iv.) Es sei $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega_1 \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine reellwertige Zufallsvariable. Dann heißt $F^X : \mathbb{R} \rightarrow [0, 1]$

$$F^X(x) = \mathbb{P}^X((-\infty, x]) = \mathbb{P}(X \leq x)$$

die **Verteilungsfunktion von X** . $F^X(x)$ entspricht also der Wahrscheinlichkeit, dass X die Grenze x nicht überschreitet.

► **Beispiel:** Es sei $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F}_1 = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$ und $\Omega_2 = \{0, 1\}$, $\mathcal{F}_2 = \{\emptyset, \Omega_2, \{0\}, \{1\}\}$. Es sei $X = \mathbb{1}_{\{1,2,3\}}$. X ist nicht messbar. Für die Auswertung von X benötigt man Informationen, die in \mathcal{F}_1 nicht verfügbar sind.

5.1.2 Definition: i.) Es sei $\Omega \neq \emptyset$ eine Menge und $A \subset \Omega$. Die Abbildung $\mathbb{1}_A : \Omega \rightarrow \mathbb{R}$ mit

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{für } \omega \in A \\ 0 & \text{für } \omega \in \Omega \setminus A \end{cases}$$

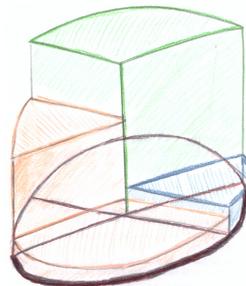
heißt **Indikatorfunktion** der Menge A .

ii.) Es sei $\Omega \neq \emptyset$ eine Menge. Eine Familie $(A_i)_{i \in I}$ von Teilmengen von Ω heißt **endliche Zerlegung** von Ω , falls:

- Für ein $n \in \mathbb{N}$ ist $I = \{1, \dots, n\}$ eine endliche Menge mit n Elementen.
- Es gilt $\Omega = \bigsqcup_{i=1, \dots, n} A_i$.

iii.) Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ heißt **elementar**, wenn es eine Zerlegung $(A_i)_{i \in I}$ von Ω mit $A_i \in \mathcal{F}, i \in I$ sowie reelle Zahlen $\alpha_i \in \mathbb{R}, i \in I$ gibt, so dass

$$X(\omega) = \sum_{i=1, \dots, n} \alpha_i \mathbb{1}_{A_i}(\omega).$$



5.1.3 Definition: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine elementare Zufallsvariable mit

$$X(\omega) = \sum_{i=1, \dots, n} \alpha_i \mathbb{1}_{A_i}(\omega).$$

Dann heißt

$$\mathbb{E}(X) = \sum_{i=1, \dots, n} \alpha_i \mathbb{P}(A_i)$$

der **Erwartungswert** von X (oder das **Lebesgue-Integral** von X bezüglich \mathbb{P}).

5.1.4 Definition: Es sei $X : \Omega \rightarrow \bar{\mathbb{R}}_{\geq 0}$ eine **numerische** nicht-negative Zufallsvariable. Dann heißt

$$\mathbb{E}(X) = \sup\{\mathbb{E}(Y) \mid X \geq Y, Y \text{ elementare ZV}\}$$

der **Erwartungswert** von X .

Zu beachten ist, dass $X(\omega) = \infty$ bzw. $\mathbb{E}(X) = \infty$ zugelassen sind. Die folgenden Regeln werden dabei unterstellt: $\infty \cdot 0 = 0 \cdot \infty = 0$, $\infty \cdot \infty = \infty$ und für alle $x > 0$ gilt $\infty \cdot x = x \cdot \infty = \infty$.

Wir verzichten hier auf die Erläuterung *technischer* Aspekte, die sich einerseits dadurch ergeben, dass wir $\bar{\mathbb{R}}$ anstatt \mathbb{R} verwenden. Wir können andererseits andere technische Aspekte vermeiden, die sich ergeben würde, wenn wir den Erwartungswert

tungswert nur für reellwertige Zufallsvariablen definieren würden. In $\bar{\mathbb{R}}$ beispielsweise hat jede nicht-leere Teilmenge ein Supremum. Für Details vergleiche Elstrodt [8] oder Henze [18, Seite 320].

5.1.5 Satz: Es sei $X : \Omega \rightarrow \bar{\mathbb{R}}_{\geq 0}$ eine numerische nicht-negative Zufallsvariable. Dann gilt

- i.) Es gibt eine Folge $(X_i)_{i \in \mathbb{N}}$ von elementaren nicht-negativen Zufallsvariablen mit $X_i \nearrow X$.
- ii.) Für jede Folge $(X_i)_{i \in \mathbb{N}}$ von elementaren nicht-negativen Zufallsvariablen mit $X_i \nearrow X$ gilt

$$\mathbb{E}(X_i) \nearrow \mathbb{E}(X).$$

Beweishinweis: Für $n \in \mathbb{N}$ betrachten wir die Zerlegung

$$A_{j,n} = \begin{cases} \left\{ \frac{j}{2^n} \leq X < \frac{j+1}{2^n} \right\} & \text{für } j = 0, \dots, n2^n - 1 \\ \{X \geq n\} & \text{für } j = n2^n \end{cases}$$

sowie die Funktion

$$X_n := \sum_{j=0}^{n2^n} \frac{j}{2^n} \mathbb{I}_{A_{j,n}}.$$

Beachte, dass die Ordinate in Intervalle zerlegt wird und von diesen Intervalle die Urbilder gebildet werden.

Vgl. Henze [18, Seite 328] für Details.

5.1.6 Bemerkung: Das bemerkenswerte an ii.) des vorhergehenden Satzes ist, dass sich für jede Folge der gleiche Limes ergibt. Das Integral ist also unabhängig von der Wahl der (monotonen) Folge.

5.1.7 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

i.) [**Linearität** von \mathbb{E}] Für zwei nicht-negative numerische Zufallsvariablen X, Y und $\alpha, \beta \in \mathbb{R}$ gilt

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y).$$

ii.) [**Monotonie** von \mathbb{E}] Für zwei nicht-negative numerische Zufallsvariablen X, Y mit $X \leq Y$ gilt

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

5.1.8 Definition: Es sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Es sei $X^+ = \max\{X, 0\}$, $X^- = \max\{0, -X\}$; also $X = X^+ - X^-$. X heißt **integrierbar** bezüglich \mathbb{P} , falls $\mathbb{E}(X^+) < \infty$ und $\mathbb{E}(X^-) < \infty$ gilt. In diesem Fall schreiben wir $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ (oder kurz $X \in \mathcal{L}^1$, wenn Missverständnisse ausgeschlossen sind). Wenn X integrierbar ist, dann definieren wir

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-).$$

Für $X \in \mathcal{L}^1$ sind auch die folgenden Schreibweisen üblich

$$\mathbb{E}(X) = \mathbb{E}^{\mathbb{P}}(X) = \int_{\Omega} X d\mathbb{P} = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega)$$

5.1.9 Satz: Es sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Dann ist äquivalent:

i.) $X \in \mathcal{L}^1$.

ii.) $|X| \in \mathcal{L}^1$.

iii.) $X^+, X^- \in \mathcal{L}^1$.

iv.) Es gibt $Y \in \mathcal{L}^1, Y \geq 0$ mit $|X| \leq Y$. Y heißt integrierbare **Majorante**.

5.1.10 Bemerkung: Wenn \mathbf{X} beschränkt ist, dann ist \mathbf{X} integrierbar. Insbesondere ist die Dirichlet Funktion χ auf dem Intervall $(0, 1]$ integrierbar und es gilt

$$\mathbb{E}(\chi) = 0$$

5.2 Konvergenzsätze

5.2.1 Satz von der majorisierten Konvergenz: Es sei $X_n : \Omega \rightarrow \bar{\mathbb{R}}$ eine Folge von Zufallsvariablen, so dass $X_n(\omega)$ mit Wahrscheinlichkeit 1 konvergiert. Ferner sei $g \in \mathcal{L}^1$ eine nicht-negative Zufallsvariable mit $|X_n| \leq g$ mit Wahrscheinlichkeit 1.

lichkeit 1. Dann gilt:

i.) Für alle $n \in \mathbb{N}$ ist $X_n \in \mathcal{L}^1$ und es gibt $X \in \mathcal{L}^1$ mit $X_n \rightarrow X$ mit Wahrscheinlichkeit 1.

ii.) Für jedes $X \in \mathcal{L}^1$ mit $X_n \rightarrow X$ mit Wahrscheinlichkeit 1 gilt

$$\lim_{n \rightarrow \infty} \int X_n d\mathbb{P} = \int X d\mathbb{P} = \mathbb{E}(X) = \int \left(\lim_{n \rightarrow \infty} X_n \right) d\mathbb{P}$$

und

$$\lim_{n \rightarrow \infty} \int |X_n - X| d\mathbb{P} = 0.$$

5.3 Linearität und Monotonie

5.3.1 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

i.) (**Linearität** von \mathbb{E}) Für zwei Zufallsvariablen $X, Y \in \mathcal{L}^1$ und $\alpha, \beta \in \mathbb{R}$ gilt

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y).$$

ii.) (**Monotonie** von \mathbb{E}) Für zwei Zufallsvariablen $X, Y \in \mathcal{L}^1$ mit $X \leq Y$ gilt

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

5.4 Dreiecksungleichung und Cauchy-Bunjakowski-Schwarz-Ungleichung

5.4.1 Satz (Dreiecksungleichung): Für eine Zufallsvariable $X \in \mathcal{L}^1$ gilt

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$$

► Blitzstein und Hwang [4, Theorem 4.4.2, Seite 164] nennen das folgende Resultat Fundamentalbrücke zwischen Wahrscheinlichkeit und Erwartung.

5.4.2 Satz: Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{F}$. Dann gilt $\mathbb{P}(A) = \mathbb{E}(1_A)$.

5.4.3 Definition: Eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ heißt **quadrat-integrierbar**, falls $\mathbb{E}(X^2) < \infty$ ist. In diesem Fall schreiben wir $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ [oder kurz $X \in \mathcal{L}^2$, falls Missverständnisse ausgeschlossen werden können].

5.4.4 Satz (Cauchy-Bunjakowski-Schwarz-Ungleichung): Es sei $X, Y \in \mathcal{L}^2$. Dann $XY \in \mathcal{L}^1$ und

$$\begin{aligned}(\mathbb{E}(XY))^2 &\leq \mathbb{E}(X^2)\mathbb{E}(Y^2) \\ |\mathbb{E}(XY)| &\leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}\end{aligned}$$

Beweis: Vgl. Tappe [31, S. 122]

5.4.5 Satz: $\mathcal{L}^2(\mathbb{P}) \subset \mathcal{L}^1(\mathbb{P})$. Also, wenn eine Zufallsvariable quadrat-integrierbar ist, dann ist sie auch integrierbar (bezüglich eines Wahrscheinlichkeitsmaßes).

5.4.6 Bemerkung: Der Satz gilt für Wahrscheinlichkeitsmaße, jedoch i.A. nicht für beliebige Maße.

5.5 Varianz

5.5.1 Definition: Es sei $X \in \mathcal{L}^2$. Dann heißt

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

die **Varianz** von X .

5.5.2 Satz: Es sei $X \in \mathcal{L}^2$. Dann gilt:

- i.) $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
- ii.) $\mathbb{V}(\alpha X + \beta) = \alpha^2 \mathbb{V}(X)$.

5.5.3 Satz: Es sei $X : \Omega \rightarrow [0, \infty]$ eine nicht-negative numerische Zufallsvariable. Dann gilt

$$\mathbb{E}(X) = 0 \quad \Leftrightarrow \quad X \equiv 0 \quad \mathbb{P}\text{-f.ü.}$$

$X \equiv 0$ \mathbb{P} -f.ü. bedeutet $\mathbb{P}(X \equiv 0) = 1$.

Beweis: Vgl. Tappe [31, S. 111]. Wir betrachten für die Hinrichtung

$$\frac{1}{n} \mathbb{P} \left(X > \frac{1}{n} \right) = \frac{1}{n} \mathbb{E} \left[\mathbb{1}_{X > \frac{1}{n}} \right] = \mathbb{E} \left[\frac{1}{n} \mathbb{1}_{X > \frac{1}{n}} \right] \leq \mathbb{E}(X) = 0$$

Wegen der Stetigkeit von \mathbb{P} gilt

$$\mathbb{P}(X > 0) = \lim \mathbb{P} \left(X > \frac{1}{n} \right) = 0.$$

Für die Rückrichtung: Es sei $\mathbb{P}(X > 0) = 0$. Dann

$$\begin{aligned} 0 \leq \mathbb{E}(X \mathbb{1}_{X \leq n}) &= \mathbb{E}(X \mathbb{1}_{0 < X \leq n}) \leq \mathbb{E}(n \mathbb{1}_{0 < X \leq n}) \\ &= n \mathbb{P}(0 < X \leq n) = 0 \end{aligned}$$

Also $\mathbb{E}(X \mathbb{1}_{X \leq n}) = 0$. Es gilt $\lim X \mathbb{1}_{X \leq n} = X$. Dann aus dem Satz über die monotone Konvergenz

$$\mathbb{E}(X) = \mathbb{E} \left(\lim_{n \rightarrow \infty} X \mathbb{1}_{X \leq n} \right) = \lim_{n \rightarrow \infty} \mathbb{E}(\mathbb{1}_{X \leq n}) = 0$$

5.5.4 Bemerkung: $X \equiv 0$ **fast sicher** bedeutet $\mathbb{P}(\{\omega \in \Omega \mid X(\omega) = 0\}) = \mathbb{P}(X = 0) = 1$.

5.5.5 Beispiel: Für die Dirichlet Funktion χ gilt $\mathbb{E}(\chi) = 0$, obwohl χ unendlich oft den Wert 1 annimmt.

5.5.6 Satz: Es sei $X \in \mathcal{L}^2$. Dann gilt

$$\mathbb{V}(X) = 0 \quad \Leftrightarrow \quad X \equiv \mathbb{E}(X) \quad \mathbb{P}\text{-f.ü..}$$

5.6 Ungleichungen: Markow, Chebycehv, Jensen

5.6.1 Satz (Allgemeine Markow-Ungleichung): Es sei $a > 0$ und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable sowie $h : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$, sodass $h \circ X \in \mathcal{L}^1$. Dann gilt

$$\mathbb{P}(h(X) \geq a) \leq \frac{\mathbb{E}(h(X))}{a}.$$

Beweis: Es gilt

$$\begin{aligned} \mathbb{E}(h(X)) &= \mathbb{E}(\mathbb{1}_{h(X) \geq a} h(X) + \mathbb{1}_{h(X) < a} h(X)) \\ &\geq a \mathbb{P}(h(X) \geq a) \end{aligned}$$

5.6.2 Satz (Ungleichung von Markow): Es sei $X \in \mathcal{L}^1$. Dann gilt

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

5.6.3 Satz (Ungleichung von Chebycehv): Es sei $X \in \mathcal{L}^2$. Dann gilt

i.) Für alle $a > 0$

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(X^2)}{a^2}.$$

ii.) Für alle $a > 0$

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2},$$

mit $\mu = \mathbb{E}(X)$ und $\sigma^2 = \mathbb{V}(X)$.

5.6.4 Satz (Ungleichung von Jensen): Es sei $I \subset \mathbb{R}$ ein Intervall, $g : I \rightarrow \mathbb{R}$ eine konvexe Funktion, $X \in \mathcal{L}^1$ eine integrierbare Zufallsvariable mit $\text{Bild}(X) \subset I$ und $g(X) \in \mathcal{L}^1$. Dann gilt

$$g(\mathbb{E}(X)) \leq \mathbb{E}(g(X)).$$

6 Zufallsvektoren

6.0.1 Definition: i.) Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine σ -Algebra $\mathcal{G} \subset \mathcal{F}$ heißt **Sub- σ -Algebra**.

ii.) Es sei I eine Indexmenge und $\mathcal{G}_i, i \in I$ eine Familie von Sub- σ -Algebren. Die Familie $\mathcal{G}_i, i \in I$ heißt **unabhängig**, falls für alle endlichen Teilmengen $J \subset I$ und $A_j \in \mathcal{G}_j, j \in J$

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j)$$

gilt.

iii.) Es sei I eine Indexmenge und $(\Omega_i, \mathcal{F}_i), i \in I$ messbare Räume und $X_i : \Omega \rightarrow \Omega_i$ Zufallsvariablen. Die Zufallsvariablen $(X_i), i \in I$ heißen **unabhängig**, falls die $\sigma(X_i) = X_i^{-1}(\mathcal{F}_i)$ unabhängig sind.

6.0.2 Bemerkung: Beachte, dass alle Zufallsvariablen im vorhergehenden Satz den gleichen Definitionsbereich jedoch möglicherweise unterschiedliche Zielmengen haben.

► Die Definition für unabhängige Zufallsvariablen ist *arg abstrakt*. Der folgende Satz enthält eine anschauliche äquivalente Charakterisierung.

6.0.3 Satz: Es seien $X : \Omega \rightarrow \Omega_1$ und $Y : \Omega \rightarrow \Omega_2$ Zufallsvariablen. Dann ist äquivalent:

- i.) X und Y sind unabhängig.
- ii.) Für alle $A \in \mathcal{F}_1$ und $B \in \mathcal{F}_2$ gilt

$$\mathbb{P}(X \in A \text{ und } Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

- iii.) Für beliebige messbare Abbildungen $f : \Omega_1 \rightarrow \Omega'_1$ und $g : \Omega_2 \rightarrow \Omega'_2$ mit Werten in messbaren Räumen $(\Omega'_1, \mathcal{F}'_1)$ bzw. $(\Omega'_2, \mathcal{F}'_2)$ sind die Zufallsvariablen $f \circ X$ und $g \circ Y$ unabhängig.

Also: Die Unabhängigkeit bleibt bei jeder messbaren Transformation erhalten!

Beweis: Siehe Tappe [31, Seite 152].

6.0.4 Satz: Es seien $X : \Omega \rightarrow \Omega_1$ und $Y : \Omega \rightarrow \Omega_2$ Zufallsvariablen. Dann ist äquivalent:

- i.) X und Y sind unabhängig.
- ii.) Für zwei beliebige messbare Funktionen $f : \Omega_1 \rightarrow \mathbb{R}$ und $g : \Omega_2 \rightarrow \mathbb{R}$ mit $f \circ X \in \mathcal{L}^1$ und $g \circ Y \in \mathcal{L}^1$ gilt

ii.) ist in Tappe etwas anders. Beweis angeben!

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).$$

[dann ist insbesondere $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$]

Beweis: Vgl. Schilling [28, Seite 152] für den Beweis der schwierigen Richtung i.) \Rightarrow ii.)

ii.) \Rightarrow i.). Angenommen $A \in \mathcal{F}_1, B \in \mathcal{F}_2$. Dann gilt

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \mathbb{E}(\mathbb{1}_A(X)\mathbb{1}_B(Y)) = \mathbb{E}(\mathbb{1}_A(X))\mathbb{E}(\mathbb{1}_B(Y)) \\ &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \end{aligned}$$

6.0.5 Bemerkung: Es seien $X : \Omega \rightarrow \mathbb{R}$ und $Y : \Omega \rightarrow \mathbb{R}$ **diskrete** Zufallsvariablen mit Zähldichten $p(x) = \mathbb{P}^X(\{x\}), x \in X(\Omega)$ bzw. $q(y) = \mathbb{P}^Y(\{y\}), y \in Y(\Omega)$. Dann sind X und Y genau dann unabhängig, wenn für alle $x \in X(\Omega)$ und $y \in Y(\Omega)$

$$\mathbb{P}(X = x \text{ und } Y = y) = p(x)q(y)$$

gilt.

6.0.6 Definiton: i.) Es sei $\mathcal{B}(\mathbb{R}^2) = \sigma(\{(a_1, b_1] \times (a_2, b_2] : a_1 \leq b_1, a_2 \leq b_2, a_1, a_2, b_1, b_2 \in \bar{\mathbb{R}}\})$. Die Mengen $B \in \mathcal{B}(\mathbb{R}^2)$ heißen **Borelmengen** (in \mathbb{R}^2).

ii.) Es seien $X : \Omega \rightarrow \mathbb{R}$ und $Y : \Omega \rightarrow \mathbb{R}$ reellwertige Zufalls-

variablen. Die Abbildung

$$\mathbb{P}^{(X,Y)}(B) = \begin{cases} \mathcal{B}(\mathbb{R}^2) \rightarrow [0, 1] \\ B \mapsto \mathbb{P}((X, Y) \in B) \end{cases}$$

heißt die **gemeinsame Verteilung** von X und Y . $\mathbb{P}^{(X,Y)}(B)$ ist also die Wahrscheinlichkeit ein Ergebnis im 2-dim Bereich $B \subset \mathbb{R}^2$ zu beobachten, wobei $B \in \mathcal{B}(\mathbb{R}^2)$ eine Borelmenge ist.

ii.) Die Abbildung

$$F^{(X,Y)} = \begin{cases} \mathbb{R}^2 \rightarrow [0, 1] \\ (x, y) \mapsto \mathbb{P}(X \leq x \wedge Y \leq y) = \mathbb{P}^{(X,Y)}((-\infty, x] \times (-\infty, y]) \end{cases}$$

heißt **gemeinsame Verteilungsfunktion** der Zufallsvariablen X, Y . $F(x, y)$ ist also die Wahrscheinlichkeit, dass X einen Wert kleiner gleich x **und** Y einen Wert kleiner gleich y annimmt.

iii.) Die Verteilungen \mathbb{P}^X und \mathbb{P}^Y von X bzw. Y nennt man in diesem Zusammenhang auch **Randverteilungen**.

6.0.7 Satz: Es seien $X : \Omega \rightarrow \mathbb{R}$ und $Y : \Omega \rightarrow \mathbb{R}$ reellwertige Zufallsvariablen. Dann ist äquivalent:

i.) X und Y sind unabhängig.

ii.) Für alle $x, y \in \mathbb{R}$ gilt $F^{(X,Y)}(x, y) = F^X(x)F^Y(y)$.

6.0.8 Definition: Es sei F die gemeinsame Verteilungsfunktion der Zufallsvariablen $X, Y : \Omega \rightarrow \mathbb{R}$. Wenn es eine Funktion $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ mit

$$F(a, b) = \int_{-\infty}^b \int_{-\infty}^a f(x, y) dx dy \quad \text{für } a, b \in \mathbb{R}$$
$$F(\infty, \infty) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$$

gibt, dann heißt f die **gemeinsame Dichte** der Zufallsvariablen X und Y .

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

heißt die **Randdichte** von X und

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

heißt die **Randdichte** von Y .

6.0.9 Satz: Es seien $X, Y : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen mit gemeinsamer Dichte f . X und Y sind genau dann unabhängig, wenn für fast alle x, y gilt:

$$f(x, y) = f_X(x)f_Y(y).$$

▷ Bisher wurde Unabhängigkeit untersucht. Wie ist es mit der **Nicht-Unabhängigkeit (Abhängigkeit)**? Da fängt man

am besten mit der **linearen** Nicht-Unabhängigkeit (Abhängigkeit) an. ... Kovarianz! \triangleleft

6.0.10 Definition: Es seien $X, Y \in \mathcal{L}^2$. Dann heißt

$$\begin{aligned}\operatorname{cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\ &= \mathbb{E}((X - \mu_X)(Y - \mu_Y))\end{aligned}$$

die **Kovarianz** von X, Y . Wenn $\operatorname{cov}(X, Y) = 0$ gilt, dann heißen X und Y **unkorreliert**.

Natürlich $\operatorname{cov}(X, X) = \mathbb{V}(X)$.

6.0.11 Satz: Für $X, Y \in \mathcal{L}^2$ gilt

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mu_X\mu_Y.$$

6.0.12 Satz: Es seien $X, Y \in \mathcal{L}^2$. Wenn X und Y unabhängig sind, dann sind X und Y unkorreliert.

6.0.13 Bemerkung: Während die Unabhängigkeit bei **jeder** messbaren **Transformation** erhalten bleibt (siehe vorne), bleibt – wie der vorhergehende Satz zeigt – die Unkorreliertheit i.A. **nur bei linearen/affinen Transformationen** erhalten.

6.0.14 Satz: Es seien $X, Y \in \mathcal{L}^2$. Wenn X und Y unkorreliert sind, dann sind auch $aX + b$ und $cY + d$ unkorreliert.

6.0.15 Satz: Für $X, Y \in \mathcal{L}^2$ und $a, b, c, d \in \mathbb{R}$ gilt

$$\text{cov}(aX + b, cY + d) = ac \cdot \text{cov}(X, Y)$$

6.0.16 Bemerkung: Es sei $X \sim N(0, 1)$ und $Y = X^2$. Natürlich sind X und Y nicht unabhängig. Aber:

$$\begin{aligned}\text{cov}(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(X X^2) \\ &= \mathbb{E}(X^3) \\ &= 0\end{aligned}$$

Dieses Beispiel funktioniert natürlich auch für andere Verteilungen als die Normalverteilung.

6.0.17 Satz: Die Abbildung

$$\text{cov} : \mathcal{L}^2 \times \mathcal{L}^2 \rightarrow \mathbb{R}$$

ist eine symmetrische, bilineare und positiv semi-definite Abbildung/Operator, d.h.

i.) Für alle $Y \in \mathcal{L}^2$ ist

$$\text{cov}(\cdot, Y) : \mathcal{L}^2 \rightarrow \mathbb{R}$$

linear. Also gilt für alle $a_1, a_2 \in \mathbb{R}$ und $X_1, X_2 \in \mathcal{L}^2$

$$\text{cov}(a_1 X_1 + a_2 X_2, Y) = a_1 \text{cov}(X_1, Y) + a_2 \text{cov}(X_2, Y).$$

Analog ist für alle $X \in \mathcal{L}^2$

$$\text{cov}(X, \cdot) : \mathcal{L}^2 \rightarrow \mathbb{R}$$

linear.

ii.) Für alle $X, Y \in \mathcal{L}^2$ gilt $\text{cov}(X, Y) = \text{cov}(Y, X)$.

iii.) Für alle $X \in \mathcal{L}^2$ gilt $\text{cov}(X, X) \geq 0$.

6.0.18 Satz: Wenn die Zufallsvariablen $X_1, X_2, \dots, X_n \in \mathcal{L}^2$ paarweise unkorreliert sind, dann gilt:

$$\mathbb{V} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \mathbb{V}(X_i).$$

6.0.19 Satz: Es seien $X_1, X_2 \in \mathcal{L}^2$. Dann gilt:

$$\mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2) + 2 \text{cov}(X_1, X_2)$$

$$\mathbb{V}(X_1 - X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2) - 2 \text{cov}(X_1, X_2)$$

6.0.20 Definition: Es seien $X_1, X_2, \dots, X_n \in \mathcal{L}^2$. Dann heißt die $n \times n$ Matrix

$$\Sigma_X = (\text{cov}(X_i, X_j))_{i=1, \dots, n, j=1, \dots, n}$$

die **Kovarianzmatrix** von des Zufallsvektor X .

▷ Vorsicht, obwohl auf der Hauptdiagonalen der Matrix Σ_X

die Varianzen $\mathbb{V}(X_i) = \sigma_i^2$ von X_i stehen, verwenden wir für die Matrix der Ko-Varianzen die Notation Σ_X und nicht Σ_X^2 .

6.0.21 Satz: Die Kovarianzmatrix ist symmetrisch und positiv-semidefinit, d.h. $\Sigma_X = (\Sigma_X)^T$ und für alle $\mathbf{b} \in \mathbb{R}^n$ gilt

$$\mathbf{b}^T \Sigma_X \mathbf{b} = \langle \Sigma_X \mathbf{b}, \mathbf{b} \rangle \geq 0.$$

6.0.22 Satz (Sandwich-Formel): Es seien $X_1, X_2, \dots, X_n \in \mathcal{L}^2$, $\mathbf{X}^T = (X_1, X_2, \dots, X_n)$, $\mathbf{A} \in M(m, n, \mathbb{R})$ und $\mathbf{b} \in \mathbb{R}^m$. Dann gilt für die affine Transformation $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$

$$\Sigma_{\mathbf{Y}} = \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T.$$

6.0.23 Satz (Cauchy-Bunjakowski-Schwarz-Ungleichung):

Es seien $X, Y \in \mathcal{L}^2$.

i.) Dann $XY \in \mathcal{L}^1$ und

$$\begin{aligned} (\mathbb{E}(XY))^2 &\leq \mathbb{E}(X^2)\mathbb{E}(Y^2) \\ |\mathbb{E}(XY)| &\leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)} \end{aligned}$$

ii.) Ferner

$$\begin{aligned} |\text{cov}(X, Y)| &\leq \sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)} = \text{std}(X)\text{std}(Y) \\ (\text{cov}(X, Y))^2 &\leq \mathbb{V}(X)\mathbb{V}(Y) \end{aligned}$$

6.0.24 Defintion: Es seien $X, Y \in \mathcal{L}^2$ mit $\mathbb{V}(X), \mathbb{V}(Y) >$

0. Dann heißt

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\text{std}(X) \cdot \text{std}(Y)}$$

der **Korrelationskoeffizient nach von Bravais und Pearson** von X und Y .

6.0.25 Satz: Es seien $X, Y \in \mathcal{L}^2$ mit $\mathbb{V}(X), \mathbb{V}(Y) > 0$.

Dann gilt

i.) $-1 \leq \rho(X, Y) \leq 1$.

ii.) Es gilt $\rho(X, Y) = 1$ genau dann, wenn es $a, b \in \mathbb{R}, a > 0$ mit $\mathbb{P}(Y = aX + b) = 1$ gibt.

iii.) Es gilt $\rho(X, Y) = -1$ genau dann, wenn es $a, b \in \mathbb{R}, a < 0$ mit $\mathbb{P}(Y = aX + b) = 1$ gibt.

iv.) X, Y unabhängig impliziert $\rho(X, Y) = 0$.

v.) Für alle a, b, c, d gilt

$$\rho(aX + b, cY + d) = \frac{ac}{|ac|} \rho(X, Y).$$

6.0.26 Bemerkung: Wenn ρ nahe 1 ist, dann ist **der lineare Zusammenhang eng**. Die etwaigen Beobachtungen liegen nahe an einer Geraden. Man kann an ρ die **Steigung dieser Geraden nicht ablesen**.

Copulas

6.0.27 Definition: Eine **d -dimensionale Copula** ist die gemeinsame Verteilungsfunktion C eines d -dimensionalen Zufallsvektors $\mathbf{U} = (U_1, \dots, U_d)^\top$ mit $U_i \sim \text{Unif}([0, 1])$ für $i = 1, \dots, d$.

6.0.28 Satz: Eine Abbildung $C : [0, 1]^d \rightarrow [0, 1]$ ist genau dann eine Copula, falls

- $C(u_1, \dots, u_d) = 0$ falls $u_i = 0$ für (mindestens) ein i .
- $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ für $i = 1, \dots, d$ und $u_i \in [0, 1]$.
- Für alle $(a_1, \dots, a_d), (b_1, \dots, b_d) \in [0, 1]^d$ mit $a_i \leq b_i$

$$0 \leq \sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1+\dots+i_d} C(u_{1,i_1}, \dots, u_{d,i_d}) \leq 1,$$

wobei $u_{j,1} = a_j$ und $u_{j,2} = b_j$ für $j = 1, \dots, d$.

Wenn für einen Zufallsvektor $\mathbf{U} \sim C$ gilt, dann muss natürlich $\mathbb{P}(a_1 \leq U_1 \leq b_1, \dots, \mathbb{P}(a_1 \leq U_1 \leq b_1)) \geq 0$ gelten. Die Rechteckgleichung stellt sicher, dass das auch der Fall ist.

6.0.29 Bemerkung:

- Man nennt: Für alle $(a_1, \dots, a_d), (b_1, \dots, b_d) \in [0, 1]^d$ mit $a_i \leq b_i$

$$0 \leq \sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1+\dots+i_d} C(u_{1,i_1}, \dots, u_{d,i_d}) \leq 0$$

wobei $u_{j,1} = a_j$ und $u_{j,2} = b_j$ für $j = 1, \dots, d$

die d -dimensionale **Rechteckgleichung**.

- Die Rechteckgleichung für $d = 2$: Für $0 \leq a_k \leq b_k$, $k = 1, 2$ gilt

$$0 \leq C(b_1, b_2) + C(a_1, a_2) - C(a_1, b_2) - C(b_1, a_2) \leq 1$$

Vgl. Henze Seite 133, 317

6.0.30 Satz: Es sei $\mathbf{X} = (X_1, \dots, X_d)^\top \in \mathcal{F}(F_1, \dots, F_d)$ ein Zufallsvektor mit gemeinsamer Verteilungsfunktion $F^{\mathbf{X}}$. Dann **existiert** ein Zufallsvektor $\mathbf{U} = (U_1, \dots, U_d)^\top$ mit $U_i \sim \text{Unif}([0, 1])$ mit gemeinsamer Verteilungsfunktion $C = C_{\mathbf{X}}$, so dass

$$F^{\mathbf{X}}(t_1, \dots, t_d) = C(F_1(t_1), \dots, F_d(t_d)) \quad \forall t_i \in \mathbb{R}.$$

Sind die Funktionen F_i stetig, dann ist C **eindeutig**. In der Tat

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)).$$

Die rechte Seite $C(F_1(t_1), \dots, F_d(t_d))$ beschreibt die Zerlegung der gemeinsamen Verteilung $F^{\mathbf{X}}$ in Ränder F_1, \dots, F_d und Copula C .

6.0.31 Satz: Sind Verteilungsfunktionen F_1, \dots, F_d und eine Copula C gegeben, so existiert ein Zufallsvektor $\mathbf{X} = (X_1, \dots, X_d)^\top \in \mathcal{F}(F_1, \dots, F_d)$ mit

$$F^{\mathbf{X}}(t_1, \dots, t_d) = C(F_1(t_1), \dots, F_d(t_d)).$$

6.0.32 Bemerkung: Was bedeutet die Gleichung

$$F^{\mathbf{X}}(t_1, t_2) = C(F_1(t_1), F_2(t_2))?$$

Wir definieren $u_i = F_i(t_i) = \mathbb{P}(X_i \leq t_i)$. u_i ist also die Wahrscheinlichkeit, dass X_i nicht über der Schwelle t_i ist.

Wir wollen

$$\begin{aligned} F^{\mathbf{X}}(x_1, x_2) &= \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) \\ &= C(F_1(x_1), F_2(x_2)) \\ &= C(\mathbb{P}(X_1 \leq x_1), \mathbb{P}(X_2 \leq x_2)) \\ &= C(u_1, u_2) \end{aligned}$$

interpretieren!

Nehmen wir zur Illustration an, dass $t_1 = 1.89$ cm, $u_1 = 0.90$ und $t_2 = 1.64$ cm, $u_2 = 0.95$ und $F^{\mathbf{X}}(t_1, t_2) = C(F_1(t_1), F_2(t_2)) =$

0.04. Auf Basis der Zahlen $t_1 = 1.89$ und $t_2 = 1.64$ können wir nicht erkennen, ob es sich um gemeinsam relativ große Werte handelt. Die gemeinsame Wahrscheinlichkeit, dass $X_1 \leq 1.89$ und $X_2 \leq 1.64$, ist (nur) für 0.04. Aber was bedeutet das? Ist $X_1 \leq 1.89$ und $X_2 \leq 1.64$ bemerkenswert? Schon, denn die Wahrscheinlichkeit größere Werte für X_1 bzw. X_2 zuhalten, sind klein: $\mathbb{P}(X_1 > t_1) = 1 - 0.9 = 0.1$ bzw. $\mathbb{P}(X_2 > t_2) = 1 - 0.95 = 0.05$. Der Wert $C(0.9, 0.95) = 0.045$ besagt, dass die Wahrscheinlichkeit für ein gemeinsames Übertreffen der 90%-Schwelle (für X_1) bzw. der 95%-Schwelle (für X_2), immerhin 0.045 ist. Anstatt die Schwellen in cm zu messen, werden die Schwellen in Wahrscheinlichkeiten gemessen. In diesem Sinn ermöglichen Copulas einen Wechsel der Einheiten. Aus “größer als 1.64” wird “unwahrscheinlich” (nämlich ein Ereignis mit einer Wahrscheinlichkeit 0.05).

Für den Fall stetiger F_i erhalten wir eine besonders einfache Interpretation, für diesen **Wechsel der Einheiten**: $C(u_1, u_2)$ entspricht der **Wahrscheinlichkeit**, dass X_1, X_2 unter den u_1 - bzw. u_2 -Quantilen bleibt.

Die Wahrscheinlichkeit, dass X_1 ihr 0.95-Quantil und X_2 sein 0.9-Quantil übertrifft, ist immerhin 0.045.

6.0.33 Satz: Es sei $\mathbf{X} = (X_1, \dots, X_d)^T$ ein Zufallsvektor mit stetigen Verteilungsfunktionen F_i und Copula C . Ferner seien

T_1, \dots, T_d strikt monoton wachsende Transformationen. Dann ist C auch eine $(T_1(X_1), \dots, T_d(X_d))^T$.

Beweis: Vgl. McNeil et al. [7, Seite 224].

6.0.34 Bemerkung: Copulas ermöglichen eine **Zerlegung** in die Randverteilungen und der Zusammenhangstruktur

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

Die Zerlegung erkennt man einerseits direkt an der Gleichung. Die Argumente von $C(\cdot, \dots, \cdot)$ sind separat die univariaten Verteilungen F_i ; sonst nichts.

Andererseits ist auch der letzte Satz erhellend. Die Copula bleibt erhalten, wenn wir die Zufallsvariable monoton transformieren. ...

Das erleichtert die Arbeitsteilung. ...

6.0.35 Bemerkung: $\exists F$ eine mVF. Im allgemeinen gibt es ∞ -viele Copula C mit

$$F^{\mathbf{X}}(t_1, \dots, t_d) = C(F_1(t_1), \dots, F_d(t_d)) \quad \forall t_i \in \mathbb{R}.$$

Wenn die Ränder steig sind, dann ist die Copula eindeutig

und es gilt:

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)).$$

Die Ursache für die Nicht-Eindeutigkeit kann man sich an einen einfachen bivariaten Beispiel klarmachen (vgl. McNeil [7, Seite 224])). Die gemeinsame Dichte für \mathbf{X} sei wie in der folgenden Tabelle

Tab1	$X_1 = 2$	$X_2 = 4$	X_1
$X_1 = 2$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$
$X_2 = 4$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{5}{8}$
X_2	$\frac{3}{8}$	$\frac{5}{8}$	

Wir wissen aus dem Existenzsatz, dass es ein C mit

$$F^{\mathbf{X}}(t_1, t_2) = C(F_1(t_1), F_2(t_2)) \quad \forall t_i \in \mathbb{R}.$$

gibt. Es gilt $\text{Bild}(F_i) = \{0, \frac{3}{8}, 1\}$. Die obige Funktionalgleichung für C liefert hier nur eine einzige Restriktion für C ; nämlich $C(\frac{3}{8}, \frac{3}{8}) = \frac{1}{8}$. Diese Restriktion entsteht an der Stelle $t_1 = 2 = t_2$. Dort ist einerseits $F^{\mathbf{X}}(2, 2) = \frac{1}{8}$ und $F_1(2) = \frac{3}{8} = F_2(2)$. Wenn wir jetzt z.B. $t_1 = 3 = t_2$ betrachten, dann erhalten wir wieder $F^{\mathbf{X}}(3, 3) = \frac{1}{8}$ und $F_1(3) = \frac{3}{8} = F_2(3)$. Also wieder: $C(\frac{3}{8}, \frac{3}{8}) = \frac{1}{8}$. Wir können aus der Gleichung $F^{\mathbf{X}}(t_1, t_2) = C(F_1(t_1), F_2(t_2))$ nur an den Stellen $\text{Bild } F_1 \times \text{Bild } F_2$ ermitteln. Den Wert der Copula an der Stelle $u_1 = \frac{1}{2} = u_2$ können

(zwar nicht beliebig, wir müssen insb. die Rechteckungleichung beachten) *wählen*.

Angenommen wir wählen zwei Copula C^1 und C^2 , die $F^{\mathbf{X}}(t_1, t_2) = C(F_1(t_1), F_2(t_2))$ erfüllen, und simulieren dann gemäß

- $(U_1^i, U_2^i) \sim C^i$
- $\mathbf{X}^i = (F^{\leftarrow}(U_1), F^{\leftarrow}(U_2))$

Gilt dann $F^{\mathbf{X}^1} = F^{\mathbf{X}^2}$.

6.0.36 Bemerkung:

- Betrachte eine Stichprobe $\mathbf{X}_i = (x_{i1}, \dots, x_{id})^T, i = 1, \dots, n$ des Zufallsvektors \mathbf{X} .
- Es sei $u_{ij} = \frac{r_{ij}}{n+1}$. Dabei ist r_{ij} der Rang von x_{ij} in $x_{kj}, k = 1, \dots, n$.
- Die u_{ij} erfassen die Abhängigkeitsstruktur.
- Die Ränder werden auf die Gleichverteilung standardisiert.

Transformationsformel für multivariate Dichten $d > 1$

6.0.37 Satz: Es sei $S \subset \mathbb{R}^n$ eine offene Menge und \mathbf{X} ein Zufallsvektor mit gemeinsamer Dichte $f_{\mathbf{X}} : \mathbb{R}^n \rightarrow \mathbb{R}$. Ferner sei $g : S \rightarrow \mathbb{R}^n$ eine injektive Abbildung mit $g \in C^1$ und

$$\det J_g(\mathbf{x}) \neq 0$$

für alle $\mathbf{x} \in S$. Dabei bezeichnet

$$J_g(\mathbf{x}) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial g_2}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial g_2}{\partial x_n}(\mathbf{x}) \\ \cdots & \cdots & \cdots \\ \frac{\partial g_n}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial g_n}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

die Jacobi-Matrix von g an der Stelle \mathbf{x} . Dann ist die Umkehrabbildung $h : g(S) \rightarrow \mathbb{R}^n$ stetig differenzierbar und $\mathbf{Y} = g(\mathbf{X})$ hat die Dichte

$$f_{\mathbf{Y}} = f_{\mathbf{X}}(h(\mathbf{y}))|\det J_g(\mathbf{x})|$$

Beweis: Tappe [31, Seite 188f]

7 Konvergenz von Folgen von Zufallsvariablen

7.0.1 Bemerkung: Es sei (x_i) eine Folge reeller Zahlen und x eine reelle Zahl. Wir sagen: (x_i) konvergiert gegen x , falls es für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$ gibt, so dass $|x_i - x| < \varepsilon$ für alle $i \geq N$ gilt. Egal wie klein wir die ε -Umgebung wählen, wir können ein N finden, so dass die Folge **ab N komplett in der ε -Umgebung von x liegt**.

Wir wollen nun das Konzept der *Konvergenz* auf Folgen von Zufallsvariablen (anstatt von reellen Zahlen) übertragen. Dafür gibt es mindestens vier wichtige Formen. Es liegt auf der Hand, dass der Konvergenzbegriff komplexer wird, denn Zufallsvariablen sind Abbildungen und nicht bloß Zahlen. Da ist der Begriff der Umgebung/Nähe deutlich *schwieriger* zu erfassen.

7.0.2 Defintion: Es sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable und $X_i, i \in \mathbb{N}$ eine Folge von Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}$.

i.) $(X_i) \rightarrow X$ **fast sicher** $:\Leftrightarrow \mathbb{P}(X_i \rightarrow X) = \mathbb{P}(\{\omega \in \Omega | X_i(\omega) \rightarrow X(\omega)\}) = 1$. In diesem Fall schreiben wir

$$\text{f.s.-lim } X_i = X \text{ oder } X_i \xrightarrow{\text{f.s.}} X$$

ii.) Es sei $p = 1$ oder $p = 2$. $(X_i) \rightarrow X$ **im p-ten Mittel** $:\Leftrightarrow X_i \in \mathcal{L}^p, i \in \mathbb{N}$ und $\mathbb{E}(|X_i - X|^p) \rightarrow 0$ für $i \rightarrow \infty$. In diesem Fall schreibt man

$$\mathcal{L}^p\text{-lim } X_i = X \text{ oder } X_i \xrightarrow{\mathcal{L}^p} X$$

iii.) $(X_i) \rightarrow X$ **nach Wahrscheinlichkeit** $:\Leftrightarrow$ Für alle $\varepsilon > 0$ gilt: für $i \rightarrow \infty$ gilt $\mathbb{P}(|X_i - X| > \varepsilon) \rightarrow 0$. In diesem Fall schreibt man

$$\text{n.W.-lim } X_i = X \text{ oder } X_i \xrightarrow{\text{n.W.}} X$$

iv.) $(X_i) \rightarrow X$ **in Verteilung** $:\Leftrightarrow$ Für alle $x \in \mathbb{R}$, in denen F^X stetig ist, gilt $F^{X_i}(x) \rightarrow F^X(x)$. In diesem Fall schreiben wir

$$\text{i.V.-lim } X_i = X \text{ oder } X_i \xrightarrow{\text{i.V.}} X$$

7.0.3 Satz (Eindeutigkeit): Wenn n.W.- $\lim X_i = X$ und n.W.- $\lim X_i = Y$, dann $\mathbb{P}(X = Y) = 1$.

Beweis: Es sei $\varepsilon > 0$. Es gilt

$$\begin{aligned} \mathbb{P}(|X - Y| > \varepsilon) &= \mathbb{P}(|X - X_n + X_n - Y| > \varepsilon) \\ &\leq \mathbb{P}(|X - X_n| + |X_n - Y| > \varepsilon) \\ &\leq \mathbb{P}\left(|X - X_n| > \frac{\varepsilon}{2} \text{ oder } |X_n - Y| > \frac{\varepsilon}{2}\right) \\ &\leq \mathbb{P}\left(|X - X_n| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - Y| > \frac{\varepsilon}{2}\right) \rightarrow 0 \end{aligned}$$

Die dritte Zeile ist erklärungsbedürftig. Wir müssen zeigen, dass für ω mit $|X - X_n| + |X_n - Y| > \varepsilon$ auch $|X - X_n| > \frac{\varepsilon}{2}$ oder $|X_n - Y| > \frac{\varepsilon}{2}$ gilt. Wenn $|X - X_n| > \frac{\varepsilon}{2}$ oder $|X_n - Y| > \frac{\varepsilon}{2}$ nicht gilt, dann gilt $|X - X_n| \leq \frac{\varepsilon}{2}$ und $|X_n - Y| \leq \frac{\varepsilon}{2}$. Dann gilt $|X - X_n| + |X_n - Y| \leq \varepsilon$. Also gilt ω mit $|X - X_n| + |X_n - Y| > \varepsilon$ nicht.

Es folgt $\mathbb{P}(|X - Y| > \varepsilon) = 0$. Also $\mathbb{P}(X = Y) = 1$.

7.0.4 Bemerkung: Wir werden gleich sehen, dass die f.s.-Konvergenz die n.W.-Konvergenz impliziert und ebenso die \mathcal{L}^p -Konvergenz die n.W.-Konvergenz impliziert. **Also gilt die obige Eindeutigkeit auch für die f.s.- und die \mathcal{L}^p Grenzwerte.**

7.0.5 Bemerkung: i.) Es gelte $X = \text{f.s.-}\lim X_i$. Wenn wir

zufällig ein ω (einen *Pfad*) ziehen, dann ist das mit Wahrscheinlichkeit 1 ein Pfad mit Eigenschaft: Der gesamte Pfad $X_i(\omega)$ bleibt ab einer Stelle N in einer beliebig gewählten ε -Umgebung von $X(\omega)$ [also für alle $i \geq N$ gilt $|X_i(\omega) - X(\omega)| < \varepsilon$]. Mit Wahrscheinlichkeit 1 erhalten wir also die Konvergenz in \mathbb{R} .

Gilt *nur* $X = \text{n.W.} \text{-} \lim X_i$, dann kann es sogar sein, dass die pfadweise Konvergenz für **keinen** Pfad gilt. Es kann also sein, dass für alle Pfade gilt: Es gibt ein $\varepsilon > 0$, so dass es für alle $N \in \mathbb{N}$ noch Stellen $k \geq N$ gibt, so dass $|X_k(\omega) - X(\omega)| > \varepsilon$ gilt.

ii.) Bei der i.V.-Konvergenz werden nur x betrachtet, in denen F^X stetig ist. Warum das zweckmäßig ist, wird in einer Übungsaufgabe behandelt.

7.0.6 Satz: Für $\varepsilon > 0$, $n \in \mathbb{N}$ definieren wir

$$A_n(\varepsilon) = \{|X_n - X| > \varepsilon\},$$

$$B_n(\varepsilon) = \left\{ \bigcup_{m \geq n} A_m(\varepsilon) \right\}.$$

Dann ist äquivalent

i.) f.s.- $\lim X_i = X$.

ii.) Für alle $\varepsilon > 0$ gilt $\mathbb{P}(B_n(\varepsilon)) \rightarrow 0, n \rightarrow \infty$.

$A_n(\varepsilon)$ umfasst ε -Abweichungen in der Periode n

$B_n(\varepsilon)$ umfasst ε -Abweichungen in der Periode n oder danach.

Beweis: Der folgende Beweis ist aus Grimmett und Stirzaker [14, Seite 356f].¹

Wir stellen zunächst ein paar Vorüberlegungen, die uns dann die eigentliche Argumentation erleichtern.

(1) Die Menge $B_n(\varepsilon)$ werden monoton nicht-größer: $B_{n_2}(\varepsilon) \subset B_{n_1}(\varepsilon)$ für $n_2 \geq n_1$. Also kann man den Grenzwert als Schnitt definieren:

$$A(\varepsilon) := \lim_{n \rightarrow \infty} B_n(\varepsilon) := \bigcap_{n \in \mathbb{N}} B_n(\varepsilon) \quad .$$

Es gilt: $\omega \in A(\varepsilon)$ gilt genau dann, wenn für alle n gibt es $m \geq n$ mit $|X_m(\omega) - X(\omega)| > \varepsilon$.

(2) Wir definieren die Ergebnisse in denen punktweise Konvergenz vorliegt:

$$C = \{\omega \in \Omega \mid X_n(\omega) \rightarrow X(\omega)\}.$$

Dann gilt: $\omega \in C$ gdw für alle $\varepsilon > 0$ gilt $\omega \notin A(\varepsilon)$. C^c ist die Menge der ω bei denen keine Konvergenz vorliegt. Dann ist

$$C^c = \bigcup_{\varepsilon > 0} A(\varepsilon).$$

¹Vgl. auch Chung [5, S. 69]

Jetzt können wir die folgenden Äquivalenz zeigen:

$$\mathbb{P}(C^c) = 0 \Leftrightarrow \forall \varepsilon > 0 : \mathbb{P}(A(\varepsilon)) = 0.$$

i.) Dann gilt: Wenn $\mathbb{P}(C^c) = 0$ gilt, dann muss für alle $\varepsilon > 0$ auch $\mathbb{P}(A(\varepsilon)) = 0$ gelten.

ii.) Umgekehrt gilt auch: Gilt für alle $\varepsilon > 0$ auch $\mathbb{P}(A(\varepsilon)) = 0$, dann ist auch $\mathbb{P}(C^c) = 0$. In der Tat

$$\begin{aligned} \mathbb{P}(C^c) &= \mathbb{P}\left(\bigcup_{\varepsilon > 0} A(\varepsilon)\right) \\ &= \mathbb{P}\left(\bigcup_{m=1}^{\infty} A(1/m)\right) \quad \text{denn } A(\varepsilon) \text{ ist monoton in } \varepsilon \\ &\leq \sum_{m=1}^{\infty} \mathbb{P}(A(1/m)) = 0 \end{aligned}$$

Zusammen:

f.s.-Konvergenz gilt $\Leftrightarrow \mathbb{P}(C^c) = 0 \Leftrightarrow \forall \varepsilon > 0 : \mathbb{P}(A(\varepsilon)) = 0$.

(3) Wir setzen jetzt die obigen Argumente und **die Stetigkeit von \mathbb{P} ein** und erhalten das gewünschte Resultat

$$\begin{aligned}
& X_n \xrightarrow{\text{f.s.}} X \\
& \Leftrightarrow \mathbb{P}(C^c) = 0 \\
& \Leftrightarrow \forall \varepsilon > 0 : \mathbb{P}(A(\varepsilon)) = 0 \\
& \Leftrightarrow \forall \varepsilon > 0 : \mathbb{P}(\lim_{n \rightarrow \infty} B_n(\varepsilon)) = 0 \\
& \Leftrightarrow \forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \mathbb{P}(B_n(\varepsilon)) = 0.
\end{aligned}$$

7.0.7 Satz: $(X_i) \rightarrow X$ **fast sicher** gilt genau dann, wenn für alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sup_{m \geq n} |X_m - X| > \varepsilon) = 0$$

gilt.

Beweis: Siehe Henze [18, Seite 196].

7.0.8 Bemerkung: Die f.s.-Konvergenz ist äquivalent zu $\mathbb{P}(B_n(\varepsilon)) \rightarrow 0, n \rightarrow \infty$. Die n.W.-Konvergenz ist äquivalent zu $\mathbb{P}(A_n(\varepsilon)) \rightarrow 0, n \rightarrow \infty$

7.0.9 Satz: Für $\varepsilon > 0, n \in \mathbb{N}$ definieren wir wie oben

$$A_n(\varepsilon) = \{|X_n - X| > \varepsilon\}.$$

Dann gilt: Wenn für alle $\varepsilon > 0$ die Reihe $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n(\varepsilon))$ konvergiert, dann gilt f.s.- $\lim X_i = X$.

Das $\mathbb{P}(A_n(\varepsilon))$ eine Nullfolge ist, ist zunächst nicht hinreichend für die f.s. Konvergenz. Wenn die Konvergenz so schnell ist, dass $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n(\varepsilon))$ konvergiert, dann gilt die f.s.-Konvergenz.

Beweis: Siehe Grimmett und Stirzaker [14, Seite 356f].

7.0.10 Satz: f.s.- $\lim X_i = X$ impliziert n.W.- $\lim X_i = X$.

Beweis: Folgt mit Satz 7.0.6.

7.0.11 Satz: \mathcal{L}^p - $\lim X_i = X$ impliziert n.W.- $\lim X_i = X$.

Beweis: Siehe Grimmett und Stirzaker [14, Seite 356].

7.0.12 Satz: n.W.- $\lim X_i = X$ impliziert i.V.- $\lim X_i = X$.

Beweis: Siehe Grimmett und Stirzaker [14, Seite 355].

7.0.13 Bemerkung: i.) Die Rückrichtungen zu den Implikationen der vorhergehenden Sätze gelten im Allgemeinen nicht. Wir werden/haben für jede Rückrichtung ein Beispiel für die Nichtgültigkeit angeben.

ii.) Im Allgemeinen kann man weder von der Gültigkeit der f.s.-Konvergenz auf die Gültigkeit der \mathcal{L}^p -Konvergenz schließen, noch von der Gültigkeit der \mathcal{L}^p -Konvergenz auf die Gül-

tigkeit der f.s.-Konvergenz.

iii.) Es gelten “bedingte” Rückrichtungen.

7.0.14 Beispiel (wandernde Türme, z.B. Vgl. Henze [18, S. 197f]): Es sei $(\Omega, \mathcal{A}, \mathbb{P}) = ((0, 1], \mathcal{B}((0, 1]), \lambda_{(0,1]})$. Für $m = 1, 2, 3, \dots$ und $k = 1, 2, 3, \dots, 2^m$ definieren wir

$$B_{m,k} = ((k-1)2^{-m}, k2^{-m}] \text{ und } X_{2^m+k-2} = \mathbb{1}_{B_{m,k}}.$$

Wenn man die obigen Konstruktionsanweisungen ausschreibt, dann erhält man

$$\begin{aligned} X_0 &= \mathbb{1}_{(0, \frac{1}{2}]}, X_1 = \mathbb{1}_{(\frac{1}{2}, 1]}, X_2 = \mathbb{1}_{(0, \frac{1}{4}]}, X_3 = \mathbb{1}_{(\frac{1}{4}, \frac{2}{4}]}, X_4 = \mathbb{1}_{(\frac{2}{4}, \frac{3}{4}]}, \\ X_5 &= \mathbb{1}_{(\frac{3}{4}, 1]}, X_6 = \mathbb{1}_{(0, \frac{1}{8}]}, X_7 = \mathbb{1}_{(\frac{1}{8}, \frac{2}{8}]}, \dots \end{aligned}$$

Dann:

- i.) Es gilt n.W.- $\lim X_i = 0$.
- ii.) Es gilt \mathcal{L}^p - $\lim X_i = 0$.
- iii.) Es gilt nicht f.s.- $\lim X_i = 0$. In der Tat ist die Wahrscheinlichkeit einen konvergenten Pfad mit ziehen Null!

Wir beobachten : (1) Also folgt aus der n.W. Konvergenz i.A. die f.s.-Konvergenz nicht.

(2) Zudem folgt aus der \mathcal{L}^p -Konvergenz i.A. die f.s.-Konvergenz nicht.

7.0.15 Beispiel: Es sei X_i eine Folge von unabhängigen Zufallsvariablen mit

$$X_i = \begin{cases} i^3 & \text{mit der Wahrscheinlichkeit } \frac{1}{i^2} \\ 0 & \text{mit der Wahrscheinlichkeit } 1 - \frac{1}{i^2}. \end{cases}$$

Dann gilt n.W.- $\lim X_i = 0$, jedoch gilt $\mathbb{E}(|X_i|) = \mathbb{E}(X_i) = i \rightarrow \infty$. Also gilt die Rückrichtung von Satz (7.0.11) nicht: Aus der n.W.-Konvergenz folgt i.A. die \mathcal{L}^P -Konvergenz nicht.

7.0.16 Beispiel: Es sei X eine Bernoulli Zufallsvariable mit $p = 1/2$. Es sei ferner $X_i = X$ für alle $i \in \mathbb{N}$ und $Y = 1 - X$. Dann gilt i.V.- $\lim X_i = Y$, jedoch gilt n.W.- $\lim X_i = Y$ nicht. Also gilt die Rückrichtung von Satz (7.0.12) nicht.

7.0.17 Beispiel: Es sei X_i eine Folge von unabhängigen Zufallsvariablen mit

$$X_i = \begin{cases} i^3 & \text{mit der Wahrscheinlichkeit } \frac{1}{i^2} \\ 0 & \text{mit der Wahrscheinlichkeit } 1 - \frac{1}{i^2}. \end{cases}$$

i.) Betrachte $A_n(\varepsilon) = \{|X_n - 0| > \varepsilon\}$. Dann ist $\mathbb{P}(A_n(\varepsilon)) < 1/n^2$. Da die Reihe $\sum_{n=1}^{\infty} 1/n^2$ konvergiert, konvergiert auch $\sum_{n=1}^{\infty} \mathbb{P}(A_n(\varepsilon))$. Mit Satz (7.0.9) folgt f.s.- $\lim X_i = 0$.

ii.) Wir haben schon erläutert, dass \mathcal{L}^1 - $\lim X_i = 0$ **nicht** gilt. Also: Im Allgemeinen impliziert die f.s.-Konvergenz die \mathcal{L}^1 -Konvergenz nicht.

Das ist bemerkenswert: Obwohl für fast alle Pfade $X_i(\omega) \rightarrow 0$ gilt, divergiert der *Durchschnitt*² $\mathbb{E}(X_i)$.

7.0.18 Beispiel: Es sei X_i eine Folge von unabhängigen Zufallsvariablen mit

$$X_i = \begin{cases} 1 & \text{mit der Wahrscheinlichkeit } \frac{1}{i} \\ 0 & \text{mit der Wahrscheinlichkeit } 1 - \frac{1}{i}. \end{cases}$$

Dann gilt \mathcal{L}^1 - $\lim X_i = 0$, jedoch gilt f.s.- $\lim X_i = 0$ nicht. Im Allgemeinen impliziert die \mathcal{L}^1 -Konvergenz die f.s.-Konvergenz also **nicht**.

Beweis: Die Wahrscheinlichkeit für $X_i = 0$ konvergiert gegen 1. Allerdings ist die Konvergenz zu *langsam*.

Für den Nachweis, dass f.s.- $\lim X_i = 0$ nicht gilt, folgen wir Grimmett und Stirzacker [14, Seite 357].

Es sei

$$\begin{aligned} B_n(\varepsilon) &= \{\exists m \geq n \text{ mit } |X_m(\omega) - 0| > \varepsilon\} \\ &= \{\exists m \geq n \text{ mit } X_m(\omega) = 1\}. \end{aligned}$$

Wir zeigen, dass für alle n und $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P}(B_n(\varepsilon)) = 1.$$

²Eigentlich ist das kein Durchschnitt ...

In der Tat

$$\begin{aligned}
\mathbb{P}(B_n(\varepsilon)) &= \mathbb{P}\left(\bigcup_{m \geq n} A_m(\varepsilon)\right) \\
&= 1 - \mathbb{P}\left(\left[\bigcup_{m \geq n} A_m(\varepsilon)\right]^c\right) \\
&= 1 - \mathbb{P}\left(\left[\bigcap_{m \geq n} A_m^c(\varepsilon)\right]\right) \\
&= 1 - \mathbb{P}\left(\lim_{r \rightarrow \infty} \left[\bigcap_{r \geq m \geq n} A_m^c\right]\right) \\
&= 1 - \lim_{r \rightarrow \infty} \mathbb{P}\left(\left[\bigcap_{r \geq m \geq n} A_m^c(\varepsilon)\right]\right) \\
&= 1 - \lim_{r \rightarrow \infty} \prod_{r \geq m \geq n} \mathbb{P}(A_m^c(\varepsilon)) \\
&= 1 - \lim_{r \rightarrow \infty} \prod_{r \geq m \geq n} \left(1 - \frac{1}{m}\right) \\
&= 1 - \lim_{r \rightarrow \infty} \prod_{r \geq m \geq n} \frac{m-1}{m} \\
&= 1 - \lim_{r \rightarrow \infty} \frac{n-1}{n} \cdot \frac{n+1-1}{n+1} \cdot \dots \cdot \frac{r-1}{r} \\
&= 1 - \lim_{r \rightarrow \infty} \frac{n-1}{r} \\
&= 1.
\end{aligned}$$

7.0.19 Bemerkung: Das Beispiel (7.0.17) zeigt, wie **verzwickt** es mit der Konvergenz von Zufallsvariablen sein kann.

Gemäß der starken Form der fast sicheren Konvergenz liegt Konvergenz vor, trotzdem konvergiert der Erwartungswert $\mathbb{E}(X_i)$ nicht gegen $\mathbb{E}(X)$. Obwohl mit Wahrscheinlichkeit 1 also $X_i(\omega) \rightarrow X(\omega)$ gilt, gilt $\mathbb{E}(X_i) \rightarrow \mathbb{E}(X)$ nicht. Die naheliegende Vertauschungsregel $\mathbb{E}(\text{f.s.-lim } X_i) = \lim \mathbb{E}(X_i)$ gilt also i.A. nicht.

Wenn es eine integrierbare Majorante für die Zufallsvariablen X_i gibt, dann kann man die Grenzwerte gemäß Satz (5.2.1) vertauschen. Eine genaue Zusatzannahme ist die Annahme der gleichgradigen Integrierbarkeit.

7.0.20 Definition: Es sei X_i eine Folge von \mathcal{L}^1 -Zufallsvariablen. Die Familie (X_i) heißt **gleichgradig integrierbar**, falls

$$\lim_{A \rightarrow \infty} \sup_{i \geq 1} \mathbb{E}(\mathbb{1}_{|X_i| \geq A} \cdot |X_i|) = 0.$$

7.0.21 Satz: Es sei $X_i, i \in \mathbb{N}$ eine Folge von \mathcal{L}^1 -Zufallsvariablen, X eine \mathcal{L}^1 -Zufallsvariable und $X = \text{n.W.-lim}_{i \rightarrow \infty} X_i$. Dann ist äquivalent:

- i.) (X_i) ist gleichgradig integrierbar.
- ii.) $\mathcal{L}^1\text{-lim } X_i = X$.

Beweis: Vgl. Durrett [6, Seite 245] bzw. Grimmet und Stirzaker [14, Seite 396]

7.0.22 Satz: Es gilt:

i.) Aus $\mathcal{L}^1\text{-}\lim_{n \rightarrow \infty} X_i = X$ folgt $\lim_{n \rightarrow \infty} \mathbb{E}(X_i) = \mathbb{E}(X)$
und $\lim_{n \rightarrow \infty} \mathbb{E}(|X_i|) = \mathbb{E}(|X|)$.

ii.) Aus $\mathcal{L}^2\text{-}\lim_{n \rightarrow \infty} X_i = X$ folgt $\lim_{n \rightarrow \infty} \mathbb{E}(X_i^2) = \mathbb{E}(X^2)$

Beweis: Der erste Teil von i.) sieht fast selbstverständlich aus. Wir haben aber schon gesehen, dass wir bei ZV vorsichtig sein müssen. Für den Beweis verwendet man die **Dreiecksungleichung**

$$|\mathbb{E}(X_i - X)| \leq \mathbb{E}(|X_i - X|).$$

Gemäß Definition der \mathcal{L}^1 -Konvergenz gilt $\mathbb{E}(|X_i - X|) \rightarrow 0$, d.h. die rechte Seite konvergiert gegen 0. Also konvergiert auch die linke Seite gegen 0.

Für den zweiten Teil von i.) nutzen wir die umgekehrte Dreiecksungleichung: Für $x, y \in \mathbb{R}$ gilt

$$||x| - |y|| \leq |x - y|.$$

Dann folgt aus der **Monotonie** des Erwartungswertes:

$$\mathbb{E}(|X_i| - |X|) \leq \mathbb{E}(|X_i - X|).$$

Gemäß Definition der \mathcal{L}^1 -Konvergenz gilt $\mathbb{E}(|X_i - X|) \rightarrow 0$. Also konvergiert auch die linke Seite gegen 0.

Für den Beweis von ii.) beachten wir den Hinweis in Jacod und Protter [19, Seite 149]

7.0.23 Satz: Es gilt:

- i.) f.s.-lim $X_i = X$ und f.s.-lim $Y_i = Y$ impliziert f.s.-lim $X_i + Y_i = X + Y$.
- ii.) \mathcal{L}^p -lim $X_i = X$ und \mathcal{L}^p -lim $Y_i = Y$ impliziert \mathcal{L}^p -lim $X_i + Y_i = X + Y$.
- iii.) n.W.-lim $X_i = X$ und n.W.-lim $Y_i = Y$ impliziert n.W.-lim $X_i + Y_i = X + Y$.

Für die i.V.-Konvergenz gilt die Implikation im Allgemeinen nicht.

Beweis: Vgl. Grimmett und Stirzaker [15, Ex. 7.11.2] und Henze [18, Seite 210]

7.0.24 Satz (Sluzki): Es gilt:

- i.) Aus i.V.-lim $X_i = X$ und n.W.-lim $|X_i - Y_i| = 0$ folgt i.V.-lim $Y_i = X$.
- i.a) Aus i.V.-lim $X_i = X$ und n.W.-lim $Y_i = a$ folgt i.V.-lim $X_i + Y_i = X + a$.
- ii.) Aus i.V.-lim $X_i = X$ und n.W.-lim $Y_i = c$ folgt i.V.-lim

$$X_i Y_i = cX.$$

- iii.) Aus $\text{i.V.-lim } X_i = X$ und $\text{n.W.-lim } Y_i = c, c \neq 0$ folgt
 $\text{i.V.-lim } X_i/Y_i = X/c$.

Beweis: Vgl. Henze [18, Seite 209]

7.0.25 Satz: Es sei $f : \mathbb{R} \rightarrow \mathbb{R}$ eine stetige Funktion.

- i.) Aus $\text{i.V.-lim } X_i = X$ folgt $\text{i.V.-lim } f(X_i) = f(X)$.
- ii.) Aus $\text{n.W.-lim } X_i = X$ folgt $\text{n.W.-lim } f(X_i) = f(X)$.
- iii.) Aus $\text{f.s.-lim } X_i = X$ folgt $\text{f.s.-lim } f(X_i) = f(X)$.

Beweis: Siehe Jacod und Protter [19, S. 146] für ii.) und iii.)
und Grimmett und Stirzaker [14, S. 360] für i.)

8 Gesetze der großen Zahlen

8.0.1 Definition: Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von Zufallsvariablen mit $\mathbb{E}(X_i) = \mu$ für alle $i \in \mathbb{N}$. Man sagt, dass (X_i) dem **schwachen Gesetz der großen Zahlen** genügt, falls

$$\text{n.W.-lim } \frac{1}{n} \sum_{i=1}^n X_i = \mu.$$

Der Durchschnitt konvergiert n.W. gegen den Erwartungswert.

8.0.2 Definition: Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von Zufallsvariablen mit $\mathbb{E}(X_i) = \mu$ für alle $i \in \mathbb{N}$. Man sagt, dass (X_i) dem **starken Gesetz der großen Zahlen** genügt, falls

$$\text{f.s.-lim} \frac{1}{n} \sum_{i=1}^n X_i = \mu.$$

8.0.3 Beispiel: Es sei $\mathbb{E}(X) = \mu$ und $X_i = X$. Dann

$$\text{f.s.-lim} \frac{1}{n} \sum_{i=1}^n X_i = X.$$

Also gilt das GgZ hier i.A. nicht.

8.0.4 Bemerkung: Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von Zufallsvariablen mit $\mathbb{E}(X_i) = \mu$ für alle $i \in \mathbb{N}$. Es ist naheliegend die Stichprobenfunktion $Y_N = \frac{1}{N} \sum_{i=1}^N X_i$ als Schätzfunktion für den Parameter μ zu verwenden. Es gilt immerhin für alle $N \in \mathbb{N}$, dass

$$\mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N X_i \right] = \frac{N\mu}{N} = \mu.$$

Hier haben wir folgende Situation: Wir entnehmen eine zufällige Stichprobe (X_1, \dots, X_N) der Größe N und betrachten den Erwartungswert von $Y_N = \frac{1}{N} \sum_{i=1}^N X_i$. Die Tatsache, dass $\mathbb{E}(Y_N) = \mu$ gilt, nennt man **Erwartungstreue** der Stichprobenfunktion Y_N . In den vorhergehenden Definitionen betrachten wir das Verhalten von $Y_N = \frac{1}{N} \sum_{i=1}^N X_i$ für beliebig groß werdendes N . Wenn (8.0.1) gilt, dann sagt (Y_N) ist

(**schwach**) **konsistent für** μ . Wenn (8.0.2) gilt, dann sagt (Y_N) ist **stark konsistent für** μ .

8.0.5 Satz: Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von unkorrelierten \mathcal{L}^2 -Zufallsvariablen mit $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma^2$ für alle $i \in \mathbb{N}$.

$$\text{n.W.-lim } \frac{1}{n} \sum_{i=1}^n X_i = \mu$$

und

$$\mathcal{L}^2\text{-lim } \frac{1}{n} \sum_{i=1}^n X_i = \mu.$$

Beweis: Es sei $Y_n = \frac{1}{n} \sum_{i=1}^n X_n$. Aus der Ungleichung von Chebychev (5.6.3) folgt

$$\mathbb{P}(|Y_n - \mu| \geq \varepsilon) \leq \frac{\mathbb{V}(Y_n)}{\varepsilon^2}.$$

Ferner folgt:

$$\mathbb{V}(Y_n) = \mathbb{V} \left(\frac{1}{n} \sum_{i=1}^n X_n \right) = \frac{1}{n^2} \sigma^2 n = \frac{\sigma^2}{n}.$$

Somit gilt:

$$\mathbb{P}(|Y_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \text{ für } n \rightarrow \infty,$$

also die n.W.-Konvergenz.

Es gilt

$$\mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N X_i \right] = \mu.$$

Ferner gilt:

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n X_i - \mu \right)^2 \right] &= \mathbb{V} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \\ &= \mathbb{V} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \rightarrow 0 \end{aligned}$$

für $n \rightarrow \infty$.

Wir wissen schon, dass aus der \mathcal{L}^2 -Konvergenz die n.W.-Konvergenz folgt. Auf den Beweis der n.W.-Konvergenz hätten wir eigentlich verzichten können. \square

Wir geben noch ohne Beweis die allgemeine Aussage an, gemäß derer unter den obigen Bedingungen sogar die f.s.-Konvergenz gilt. Für den Beweis vergleiche z.B. Chung [5, Theorem 5.1.2] oder Krenzel [22, Satz 12.4]

8.0.6 Satz (Rajchman¹): Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von **unkorrelierten quadrat-integrierbaren \mathcal{L}^2 -Zufallsvariablen** mit $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma^2$ für alle

¹https://de.wikipedia.org/wiki/Aleksander_Rajchman

$i \in \mathbb{N}$. Dann gilt

$$\text{f.s.-lim} \frac{1}{n} \sum_{i=1}^n X_i = \mu$$

und

$$\mathcal{L}^2\text{-lim} \frac{1}{n} \sum_{i=1}^n X_i = \mu.$$

Beweis: Krengel [22] oder Chung [5]

8.0.7 Satz (Etemadi): Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von paarweise unabhängigen und identisch verteilten \mathcal{L}^1 -Zufallsvariablen. Dann gilt

$$\text{f.s.-lim} \frac{1}{n} \sum_{i=1}^n X_i = \mu$$

und

$$\mathcal{L}^1\text{-lim} \frac{1}{n} \sum_{i=1}^n X_i = \mu.$$

Beweis: Vgl. Billingsley [2, Section 22] oder Klenke [21, Satz 5.17] oder Henze [18, Seite 201] für den Beweis der f.s.-Konvergenz. Die \mathcal{L}^1 -Konvergenz folgt dann wegen i.i.d. relativ einfach; bekanntlich folgt die \mathcal{L}^1 -Konvergenz i.a. nicht aus der f.s.-Konvergenz.

8.0.8 Bemerkung: Wir betrachten ein Beispiel einer Folge von Zufallsvariablen für die das schwache Gesetz der großen Zahlen gilt, das starke Gesetz jedoch nicht; vgl. Grinstead und Snell [16, Seite 314f, Nr. 16].

Wir betrachten eine Folge unabhängiger Zufallsvariablen (X_i) mit

$$\mathbb{P}(X_i = i) = \frac{1}{2i \log i}, \quad \mathbb{P}(X_i = -i) = \frac{1}{2i \log i} \quad \text{und}$$
$$\mathbb{P}(X_i = 0) = 1 - \frac{1}{i \log i}$$

8.0.9 Bemerkung: i.) Gesetze der großen Zahlen sind von **grundsätzlicher Bedeutung** für die Wahrscheinlichkeitstheorie. Angenommen wir betrachten die fortgesetzte unabhängige Wiederholungen eines Experiments, dessen Ergebnis Erfolg oder Misserfolg sein kann – also $\Omega = \{\text{Erfolg}, \text{Misserfolg}\}$ – und bei dem die Erfolgswahrscheinlichkeit $p = \mathbb{P}(\{\text{Erfolg}\})$ beträgt. Es sei X_i („Erfolg bei der i -ten Wiederholung“) = 1. Dann ist $S_n = \sum_{i=1}^n X_i$ die Anzahl der Erfolge nach n Wiederholungen. Der *gesunde Menschenverstand* besagt, dass relative Häufigkeit S_n/n gegen p „konvergiert“. In der Tat, wird eine entsprechende Aussage oft als „Definition“ von Wahrscheinlichkeit angesehen.² Gemäß des Satzes von Kolmogorow bzw. Etemati gilt die Konvergenzaussage in der Tat fast

²Vgl. den Wikipediaeintrag <https://de.wikipedia.org/wiki/Wahrscheinlichkeit> unter Häufigkeitsprinzip – statistische Wahrscheinlichkeitsauffassung.

sicher bzw. in \mathcal{L}^1 (also in den denkbar stärksten Formen). Das Gesetz ist jedoch hier nicht Ausgangspunkt (Axiom/Definition), sondern das Resultat der Überlegungen! Der Ausgangspunkt – das Axiomensystem nach Kolmogoroff – ist also auch diesbezüglich gut gewählt.

ii.) Für unabhängig wiederholte Bernoulli Experimente geht das schwache Gesetz der großen Zahlen auf Jakob Bernoulli zurück (veröffentlicht 1713). Das entsprechende starke Gesetz hat Borel (veröffentlicht 1909) bewiesen.

iii.) Für eine Folge von unabhängigen und identisch verteilten \mathcal{L}^1 -Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ hat Kolmogorow das Gesetz der großen Zahlen 1930 bewiesen. Nasrollah Etemadi [9] hat die *Verallgemeinerung* (paarweise unabhängig) 1981 *elementar* bewiesen.

iv.) Für unkorrelierte Zufallsvariablen deren Varianz es eine gemeinsame Schranke gibt, gilt gemäß des Satzes von Rajchman auch das starke Gesetz der großen Zahl (vgl. auch die nächste Ordnungsnummer). In diesem Sinn ist also nicht der Unterschied zwischen Unkorreliertheit und Unabhängigkeit für die Gültigkeit des schwachen anstatt des starken Gesetzes ausschlagend. \triangleleft

8.0.10 Bemerkung (vgl. Chung [5, S. 107]): Wir betrachten zunächst den Fall $X_i \in \mathcal{L}^2$. Wir erlauben, dass die Zufallsvariablen X_i abhängig sind und unterschiedliche Varianzen $\sigma_i^2 = \mathbb{V}(X_i) = \mathbb{E}(X_i^2)$ haben. Ohne wesentliche

Einschränkung der Allgemeinheit betrachten wir zentrierte Zufallsvariablen $\mathbb{E}(X_i) = 0$.

Wir definieren die Partialsummen

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

und untersuchen die Konvergenz von $\frac{S_n}{n}$. Für die \mathcal{L}^2 -Konvergenz gilt

$$\mathcal{L}^2\text{-}\lim \frac{S_n}{n} = 0 \Leftrightarrow \lim \mathbb{E} \left(\left(\frac{S_n}{n} \right)^2 \right) = 0 \Leftrightarrow \mathbb{E} (S_n^2) = o(n^2).$$

Mit wachsenden n wird $\mathbb{E} (S_n^2)$ anwachsen. Wenn dieses Wachstums hinreichend langsam ist, dann gilt die \mathcal{L}^2 -Konvergenz. Aus der \mathcal{L}^2 -Konvergenz folgt bekanntlich die Konvergenz nach Wahrscheinlichkeit.

Also: Wenn $\mathbb{E} (S_n^2)$ hinreichend langsam wächst, dann gilt das schwache Gesetz der großen Zahlen. Wir erhalten die Implikation

$$\mathbb{E} (S_n^2) = o(n^2) \Rightarrow \text{n.W.-}\lim \frac{S_n}{n} = 0.$$

Uns interessiert also die Größenordnung von $\mathbb{E} (S_n^2)$. Wir beachten dazu

$$\mathbb{E}(S_n^2) = \sum_{i=1}^n \mathbb{E}(X_i^2) + \sum_{j \neq k} \mathbb{E}(X_j X_k)$$

und betrachten zwei Fälle:

- Gilt $\mathbb{E}(X_j X_k) = 0, i \neq j$ (man sagt, dass die Zufallsvariablen X_i orthogonal³ sind) und $\sigma_i^2 \leq M, i \in \mathbb{N}$, so gilt

$$\begin{aligned}\mathbb{E}(S_n^2) &= \sum_{i=1}^n \mathbb{E}(X_i^2) \leq nM \\ \Rightarrow \frac{\mathbb{E}(S_n^2)}{n^2} &\leq \frac{M}{n}\end{aligned}$$

Also $\mathbb{E}(S_n^2) = o(n^2)$ und es folgt das schwache Gesetz der großen Zahlen.

- Wenn wir lediglich voraussetzen, dass die Varianzen und Kovarianzen eine gemeinsame Schranke $\text{cov}(X_i, X_j) = \mathbb{E}(X_i X_j) \leq M$ haben, dann folgt im allgemeinen das Gesetz der großen Zahlen jedoch nicht. Dazu betrachten wir

$$\mathbb{E}(S_n^2) = \sum_{i=1}^n \mathbb{E}(X_i^2) + \sum_{j \neq k} \mathbb{E}(X_j X_k).$$

Auf der rechten Seite stehen n^2 Terme. Die rechte Seite

³Gilt $\mathbb{E}(X_i) = 0$, dann $\mathbb{E}(X_i X_j) = \text{cov}(X_i, X_j)$.

ist dann $O(n^2)$, d.h.

$$\begin{aligned}\mathbb{E}(S_n^2) &\leq Mn^2 \\ \Rightarrow \frac{\mathbb{E}(S_n^2)}{n^2} &\leq M.\end{aligned}$$

Im Allgemeinen folgt so $\mathbb{E}(S_n^2) = o(n^2)$ nicht; dazu müsste $\frac{\mathbb{E}(S_n^2)}{n^2}$ nicht nur beschränkt sein, sondern sogar gegen Null gehen.

Wenn wir uns die Gleichung

$$\mathbb{E}(S_n^2) = \sum_{i=1}^n \mathbb{E}(X_i^2) + \sum_{j \neq k} \mathbb{E}(X_j X_k).$$

nochmal ansehen, dann erkennen wir, dass sich für den Fall $\mathbb{E}(X_j X_k) \neq 0$ i.A. zu *viele Terme* ergeben. Gilt $\mathbb{E}(X_j X_k) = 0$, dann sind es n Terme; für $\mathbb{E}(X_j X_k) \neq 0$ jedoch n^2 .

Der folgende Satz fasst die obige Bemerkung zusammen. Anschließend entwickeln wir Argumente für die Gültigkeit des Gesetzes der großen Zahlen für abhängige Zufallsvariablen. Solche Sätze nennt man **Ergodensätze**.

8.0.11 Satz: Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von **unkorrelierten quadrat-integrierbaren \mathcal{L}^2 -Zufallsvariablen** mit $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma_i^2 \leq M$ für alle $i \in \mathbb{N}$. Dann

gilt

$$\begin{aligned}\mathcal{L}^2\text{-lim} \frac{1}{n} \sum_{i=1}^n X_i &= \mu \\ \text{n.W.-lim} \frac{1}{n} \sum_{i=1}^n X_i &= \mu \\ \text{f.s.-lim} \frac{1}{n} \sum_{i=1}^n X_i &= \mu\end{aligned}$$

Die Verallgemeinerung gegenüber Satz (8.0.5) besteht darin, dass die Zufallsvariablen unterschiedliche Varianzen haben dürfen. Unter den genannten Voraussetzungen gilt – wie in der dritten Gleichung angegeben – sogar das starke Gesetz der großen Zahlen. Siehe Chung [5, Seite 108]

Für korrelierte Zufallsvariablen deren zweite Momente eine gemeinsame Schranke haben – also $\mathbb{E}(X_i X_j) \leq M$ –, gilt somit das Gesetz der großen Zahlen **im Allgemeinen nicht**. Solche Gesetze gelten jedoch unter als nächstes erläuterten passenden Bedingungen. Diese stellen sicher, dass Abhängigkeiten hinreichend *schwach* sind.

GgZ bei Abhängigkeiten

8.0.12 Definition: Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von \mathcal{L}^2 -Zufallsvariablen. (X_i) heißt **schwach stationärer**

stochastischer Prozess, falls

- $\mathbb{E}(X_i) = \mu$ für alle $i \in \mathbb{N}$. Alle Zufallsvariablen haben den gleichen Erwartungswert.
- $\text{cov}(X_i, X_j) = \gamma(i - j)$ für alle $i, j \in \mathbb{N}$. Die Kovarianz $\text{cov}(X_i, X_j)$ von X_i und X_j hängt nur von der Differenz $i - j$ ab.

8.0.13 Bemerkung: Es sei $X_i = X$ für eine Zufallsvariable X mit $\mathbb{E}(X) = \mu$, $\mathbb{V}(X) = \sigma^2$. Dann gilt $\mathbb{E}(X_i) = \mu$ und $\gamma(i - j) = \text{cov}(X_i, X_j) = \text{cov}(X, X) = \sigma^2$. Also ist X_i stationär. Das Gesetz der großen Zahlen gilt natürlich nicht:

$$\text{n.W.-lim} \frac{1}{n} \sum_{i=1}^n X_i = X \neq \mu.$$

8.0.14 Satz: Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ ein schwach stationärer Prozess. Wenn $\gamma_k \rightarrow 0, k \rightarrow \infty$, dann

$$\begin{aligned} \mathcal{L}^2\text{-lim} \frac{1}{n} \sum_{i=1}^n X_i &= \mu \\ \text{n.W.-lim} \frac{1}{n} \sum_{i=1}^n X_i &= \mu \end{aligned}$$

(Man sagt, dass X_i mittelwert-ergodisch ist.)

Beweis: Karlin und Taylor [20, Seite 476] beweisen eine stärkere Aussage. Der Beweis von dort ist auf die hier betrachtet

Situation angepasst. Wir unterstellen o.B.d.A., dass $\mathbb{E}(X_i) = 0$. Wir definieren $\text{cov}(X_i, X_j) = \gamma(i-j) = \gamma_h, h = i-j$. Dann folgt

$$\begin{aligned} \mathbb{E}(S_n^2) &= \sum_{i=1}^n \mathbb{E}(X_i^2) + 2 \sum_{0 \leq i < k \leq n} \mathbb{E}(X_i X_k) \\ &= n\sigma^2 + 2 \sum_{j=1}^n \sum_{k=j+1}^n \gamma(k-j) \\ &= n\sigma^2 + 2 \sum_{j=1}^{n-1} (n-j)\gamma_j = n\sigma^2 + 2n \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \gamma_j \end{aligned}$$

Wir erhalten als Bedingung für die \mathcal{L}^2 -Konvergenz:

$$\frac{1}{n} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \gamma_j \rightarrow 0.$$

Wir weisen jetzt nach, dass $\gamma_j \rightarrow 0$ dafür hinreichend ist.

Zunächst wählen wir m so groß, so dass $|\gamma_i| \leq \frac{\varepsilon}{2}$ für alle $i \geq m$. Das geht, denn $\gamma_i \rightarrow 0$. Dann gilt für $n-1 \geq m$ bzw. für $n \geq m+1$

$$\frac{1}{n} \sum_{j=m}^{n-1} |\gamma_j| \leq \frac{1}{n} \sum_{j=m}^{n-1} \frac{\varepsilon}{2} \leq \frac{1}{n} n \frac{\varepsilon}{2} = \frac{\varepsilon}{2}.$$

Jetzt wählen wir $n \geq m+1$ (m bleibt fest), so dass

$$\frac{1}{n} \sum_{j=1}^{m-1} |\gamma_j| \leq \frac{\varepsilon}{2}.$$

Das geht, denn m ist fest.

Jetzt setzen wir die beiden Abschätzungen zusammen:

$$\frac{1}{n} \sum_{j=1}^{n-1} |\gamma_j| = \frac{1}{n} \sum_{j=1}^{m-1} |\gamma_j| + \frac{1}{n} \sum_{j=m}^{n-1} |\gamma_j| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Also

$$\frac{1}{n} \sum_{j=1}^{n-1} |\gamma_j| \rightarrow 0.$$

Wir erhalten die gewünschte Konvergenz

$$\left| \frac{1}{n} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \gamma_j \right| \leq \frac{1}{n} \sum_{j=1}^{n-1} \left| \left(1 - \frac{j}{n}\right) \gamma_j \right| \leq \frac{1}{n} \sum_{j=1}^{n-1} |\gamma_j| \rightarrow 0$$

Hauptsatz der Statistik – Glivenko-Cantelli

8.0.15 Definition: Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Familie von identisch und unabhängig verteilten Zufallsvariablen.

Dann heißt

$$F_n(x, \omega) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(-\infty, x]}(X_i(\omega)) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i(\omega) \leq x}$$

die **empirische Verteilungsfunktion** von X_1, \dots, X_n .

Beachte: Für festes x ist $\omega \mapsto F_n(x, \omega)$ eine Zufallsvariable!

8.0.16 Satz: Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit Verteilungsfunktion F . Dann gilt für alle $x \in \mathbb{R}$

$$\text{f.s.-} \lim_{n \rightarrow \infty} F_n(x) = F(x)$$

Es gilt sogar (**Satz von Glivenko-Cantelli**)

$$\text{f.s.-} \lim_{n \rightarrow \infty} \left(\sup_{x \in \mathbb{R}} |F_n(x, \omega) - F(x)| \right) = 0.$$

Beweis: Vgl. Henze [18, Seite 277]

9 Zentraler Grenzwertsatz

9.0.1 Satz (Lindeberg-Levy): Es sei $X_i : \Omega \rightarrow \mathbb{R}, i \in \mathbb{N}$ eine Folge von unabhängig und identisch verteilten \mathcal{L}^2 -Zufallsvariablen mit $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma^2 > 0$ für alle $i \in \mathbb{N}$. Es sei

$$S_n = \sum_{k=1}^n X_k$$
$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - \mathbb{E}(S_n)}{\text{std}(S_n)}$$

Dann konvergiert S_n^* für $n \rightarrow \infty$ in Verteilung gegen eine Zufallsvariable mit Standardnormalverteilung; es gilt also

$$\text{i.V.-lim } S_n^* = Z, Z \sim \mathcal{N}(0, 1),$$

dabei hat Z die Verteilungsfunktion

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz.$$

Beweis: Siehe Henze [18, S. 215f]

9.0.2 Satz (ZGS-MW-Approximation¹): Es sei $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma^2$. Für großes N ist

$$\bar{X}_N := \frac{1}{N} (X_1 + \dots + X_N)$$

approximativ $\mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right)$:

$$\bar{X}_N \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right).$$

Beweis: Zunächst beobachten wir

$$S_N^* = \frac{S_N - N\mu}{\sigma\sqrt{N}} = \frac{N(\bar{X}_N - \mu)}{\sigma\sqrt{N}} = \sqrt{N} \left(\frac{\bar{X}_N - \mu}{\sigma} \right).$$

Gemäß ZGS ist $S_N^* = \sqrt{N} \left(\frac{\bar{X}_N - \mu}{\sigma} \right)$ näherungsweise $\mathcal{N}(0, 1)$. Also insbesondere $\mathbb{E}(S_N^*) = 0$ und $\mathbb{V}(S_N^*) = 1$. Ferner gilt

$$\bar{X}_N = \sigma S_N^* \frac{1}{\sqrt{N}} + \mu$$

Dann folgt $\mathbb{E}(\bar{X}_N) = \mu$ und

$$\mathbb{V}(\bar{X}_N) = \mathbb{V}\left(\sigma S_N^* \frac{1}{\sqrt{N}} + \mu\right) = \sigma^2 \frac{1}{N}.$$

► Vielleicht ist die Aussage $\bar{X}_N \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right)$ insbesondere für Anwender noch interessanter als die, die man ZGS nennt. Da die Varianz aber von N abhängt, ist sie *unbequemer*.

¹MW steht für Mittelwert und ZGS für Zentraler Grenzwertsatz.

9.0.3 Bemerkung: Gemäß GgZ gilt

$$\left(\frac{\bar{X}_N - \mu}{\sigma} \right) \rightarrow 0.$$

Gemäß ZGS: Wenn man mit \sqrt{N} dann ergibt sich eine nicht-triviale Verteilung

$$\sqrt{N} \left(\frac{\bar{X}_N - \mu}{\sigma} \right) \rightarrow Z, Z \sim \mathcal{N}(0, 1)$$

9.0.4 Bemerkung: Wir beachten

$$\mathbb{V}(\bar{X}_N - \mu) = \mathbb{V}(\bar{X}_N) = \frac{\sigma^2}{N}$$

also auch

$$\text{sd}(\bar{X}_N - \mu) = \frac{\sigma}{\sqrt{N}}$$

Die Standardabweichung wird also für größeres N immer kleiner. **Der MW liegt immer näher am Erwartungswert. Allerdings ist die Konvergenz langsam;** nämlich nur von Ordnung $N^{1/2}$.

9.0.5 Bemerkung: i.) Egal welches reellwertige Experiment (mit endlicher Varianz) unabhängig wiederholt wird, für hinreichend viele Wiederholungen entspricht die **Verteilung der standardisierten Summe der Zufallsergebnisse** näherungsweise der Verteilung der **Standardnormalverteilung**.

ii.) Wenn sich das Ergebnis eines Zufallsexperiments durch unabhängige additive Überlagerung einer großen Zahl von Zufallsgrößen ergibt, dann erhalten wir näherungsweise normalverteilte Beobachtungen.

iii.) Salopp (vgl. Henze [18, 216]): Eine Summe S_n aus vielen unabhängig und identisch verteilten Summanden *vergisst* – bis auf Erwartungswert und Varianz vergisst – im Limes für $n \rightarrow \infty$ die Verteilung der einzelnen Summanden.

9.0.6 Bemerkung: Regelmäßig verwendet man den Zentralen Grenzwertsatz als Grundlage für Approximationen.

i.) Die erste Variante betrachtet $S_n = \sum_{k=1}^n X_k$.

S_n^* ist für großes n näherungsweise $\mathcal{N}(0, 1)$. Wir können also die Wahrscheinlichkeiten $\mathbb{P}(S_n^* \leq k) \approx \Phi(k)$ gemäß der Standardnormalverteilung bestimmen. Ferner

$$\begin{aligned} S_n^* &\leq k \\ \Leftrightarrow \frac{S_n - n\mu}{\sigma\sqrt{n}} &\leq k \\ \Leftrightarrow S_n - n\mu &\leq k\sigma\sqrt{n} \\ \Leftrightarrow S_n &\leq n\mu + k\sigma\sqrt{n} \end{aligned}$$

Wir erhalten dann die nützliche Formel

$$\mathbb{P}(S_n \leq n\mu + k\sigma\sqrt{n}) \approx \Phi(k).$$

Analog für um Null symmetrische Intervalle

$$\begin{aligned}
 & -k \leq S_n^* \leq k \\
 \Leftrightarrow & -k \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq k \\
 \Leftrightarrow & -k\sigma\sqrt{n} \leq S_n - n\mu \leq k\sigma\sqrt{n} \\
 \Leftrightarrow & n\mu - k\sigma\sqrt{n} \leq S_n \leq n\mu + k\sigma\sqrt{n}
 \end{aligned}$$

bzw.

$$\begin{aligned}
 & -k \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq k \\
 \Leftrightarrow & k\sigma\sqrt{n} \geq n\mu - S_n \geq -k\sigma\sqrt{n} \\
 \Leftrightarrow & S_n + k\sigma\sqrt{n} \geq n\mu \geq S_n - k\sigma\sqrt{n}
 \end{aligned}$$

Gemäß ZGS gilt dann

$$\begin{aligned}
 \mathbb{P}(S_n + k\sigma\sqrt{n} \geq n\mu \geq S_n - k\sigma\sqrt{n}) & \approx 2\Phi(k) - 1 \text{ bzw} \\
 \mathbb{P}(n\mu - k\sigma\sqrt{n} \leq S_n \leq n\mu + k\sigma\sqrt{n}) & \approx 2\Phi(k) - 1
 \end{aligned}$$

wobei wir $\Phi(k) - \Phi(-k) = 2\Phi(k) - 1$ verwenden.

ii.) Die zweite (eigentlich äquivalente) Variante betrachtet stattdessen den **Durchschnitt**. Es gilt

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{n(\frac{1}{n}S_n - \mu)}{\sigma\sqrt{n}} = \frac{\frac{1}{n}S_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

Also ist $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$

Wir beobachten (analog zu oben) um symmetrische Intervalle

$$\begin{aligned} -k &\leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq k \\ \Leftrightarrow -k\sigma/\sqrt{n} &\leq \bar{X}_n - \mu \leq k\sigma/\sqrt{n} \\ \Leftrightarrow \mu - k\sigma/\sqrt{n} &\leq \bar{X}_n \leq \mu + k\sigma/\sqrt{n} \end{aligned}$$

bzw.

$$\begin{aligned} -k &\leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq k \\ \Leftrightarrow k\sigma/\sqrt{n} &\geq \mu - \bar{X}_n \geq -k\sigma/\sqrt{n} \\ \Leftrightarrow \bar{X}_n - k\sigma/\sqrt{n} &\geq \mu \geq \bar{X}_n + k\sigma/\sqrt{n} \end{aligned}$$

Gemäß ZGS gilt

$$\mathbb{P}(\bar{X}_n - k\sigma/\sqrt{n} \geq \mu \geq \bar{X}_n + k\sigma/\sqrt{n}) \approx 2\Phi(k) - 1 \text{ bzw.}$$

$$\mathbb{P}(\mu - k\sigma/\sqrt{n} \leq \bar{X}_n \leq \mu + k\sigma/\sqrt{n}) \approx 2\Phi(k) - 1$$

9.0.7 Satz von Moivre-Laplace: Es sei (X_i) eine Folge unabhängiger Bernoulli-verteilter Zufallsvariablen und

$$S_n = \sum_{i=1, \dots, n} X_i$$

und

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}. \quad (9.1)$$

Dann gilt

$$\text{i.V.-} \lim S_n^* = Z, Z \sim \mathcal{N}(0, 1). \quad (9.2)$$

9.0.8 Bemerkung: Der Satz von Moivre-Laplace ist zwar einerseits lediglich ein Spezialfall des Satzes von Lindeberg-Levy. Andererseits wahrscheinlich aber die am häufigsten verwendete Variante; vgl. Blitzstein und Hwang [4, S. 475]. Praktisch verwendet man dabei die sogenannte *Stetigkeitskorrektur*.

Also, wenn $Y \sim \text{Binomial}(n, p)$, dann ist $Y = X_1 + \dots + X_n$ mit $X_i \sim \text{iiBernuolli}(p)$. Dann gilt die Approximation

$$Y \simeq \mathcal{N}(np, np(1-p))$$

Dann

$$\begin{aligned} \mathbb{P}(Y = k) &= \mathbb{P}\left(k - \frac{1}{2} < Y < k + \frac{1}{2}\right) \\ &\approx \Phi\left(\frac{k + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - 1/2 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

9.0.9 Bemerkung: Wir betrachten eine Umfrage, die ermitteln soll, wie viele Personen die Alternative A der Alternative B vorziehen; vgl. Grinsead und Snell [16, S. 333]. Wir fassen jede Antwort als das Ergebnis eines **Bernoulli-Experiments** mit Erfolgswahrscheinlichkeit p auf. *Erfolg*

bedeutet, dass wir zufällig jemanden *gezogen/angerufen* haben, der/die die Alternative A bevorzugt. Der naheliegende Schätzer für p ist $\hat{p} = \bar{p} = \frac{S_n}{n}$.

Wir beobachten zunächst

$$\begin{aligned}\bar{p} &= \frac{S_n}{n} = \frac{S_n}{n} + p - p = \frac{S_n - pn}{n} + p \\ &= \frac{S_n - pn}{\sqrt{pqn}} \sqrt{pq} + p = \frac{S_n - pn}{\sqrt{pq}\sqrt{n}\sqrt{n}} \sqrt{pq} + p \\ &= \frac{S_n - pn}{\sqrt{pqn}} \sqrt{\frac{pq}{n}} + p \\ &= S_n^* \sqrt{\frac{pq}{n}} + p\end{aligned}$$

Jetzt können wir den zentralen Grenzwertsatz verwenden, denn $S_n^* := \frac{S_n - pn}{\sqrt{pqn}}$ ist für großes n näherungsweise standard-normal, also

$$\mathbb{P}\left(\bar{p} - 2\sqrt{\frac{pq}{n}} < p < \bar{p} + 2\sqrt{\frac{pq}{n}}\right) \approx 0.954$$

Mit dieser können wir noch nicht *konkret* rechnen, p und q sind nicht bekannt. Wir setzen noch \bar{p} für p und \bar{q} für q ein (das ist eine weitere Approximation) und erhalten:

$$\mathbb{P}\left(\bar{p} - 2\sqrt{\frac{\bar{p}\bar{q}}{n}} < p < \bar{p} + 2\sqrt{\frac{\bar{p}\bar{q}}{n}}\right) \approx 0.954$$

Das **zufällige** Intervall

$$\left(\bar{p} - 2\sqrt{\frac{\bar{p}\bar{q}}{n}}, \bar{p} + 2\sqrt{\frac{\bar{p}\bar{q}}{n}} \right)$$

heißt das 95%-**Konfidenzintervall** für den unbekanntem (aber festen) Parameter p . Mit einer Wahrscheinlichkeit von 0.95 enthält dieses zufällige Intervall den Parameter p . Man darf dabei nicht den Fehler machen und p als zufällig aufzufassen. p liegt zufällig im Intervall, weil das Zufalls-Intervall den fixen Wert p enthält. Die Abbildung in Grinsead und Snell [16, S. 333] ist diesbezüglich erhellend!

Die Länge des oben genannten Konfidenzintervall ist von abhängig. Wenn das Konfidenzintervall zu lang ist, dann ist die Umfrage nicht sehr *informativ*. Das Umfrageinstitut möchte, dass das Konfidenzintervall nicht länger als 0.06 ist. Es soll ein Abschätzung für n ermittelt werden, so dass

$$2\sqrt{\frac{\bar{p}\bar{q}}{n}} \leq 0.03$$

Verwendet man noch die Ungleichung $\bar{p}\bar{q} \leq 1/4$. Dann ist $n \geq 1111$ hinreichend.

9.0.10 Beispiel (vgl. Blitzstein und Hwang [4]): Die ZV X_i seien unabhängig Bernoulli mit $\text{Bild}(X_i) = \{0, 1\}$ und $\mathbb{P}(X_i = 1) = \frac{1}{2}$. Gemäß des GgZ gilt f.s.-lim $\bar{X}_N = \mu$.

Es ist $\mathbb{E}(X_i) = \frac{1}{2}$ und $\mathbb{V}(X_i) = \frac{1}{4}$.

Also gemäß ZGS-Approximation

$$\bar{X}_N \sim \mathcal{N} \left(\frac{1}{2}, \frac{1}{4N} \right)$$

Für $N = 100$ liegt mit Wahrscheinlichkeit 0.95 \bar{X}_N im Intervall $[0.4, 0.6]$.

Auf Basis dieser Beobachtung kann man einen *Schnelltest*² für die Fairness einer Münze konstruieren.

► Die folgenden Bemerkungen sind für viele Anwendungen in der statistischen Analyse von wirtschaftswissenschaftlichen Zusammenhängen von großer Bedeutung. Die Botschaft ist: Für große³ N benötigt man die Normalverteilungsannahme nicht und kann trotzdem bequem Inferenzaussagen anhand der Normalverteilung machen.

9.0.11 Definition (Lineares Regressionsmodell): Wir betrachten ein Lineares Modell

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K + u,$$

$$\mathbb{E}(u | X_1, \dots, X_K) = 0,$$

²Das kann man auch anders besser machen. Deshalb liefert die obige Methode nur einen approximativen *Schnelltest*.

³Aber: Die Frage, ob N in einer konkreten Situation groß genug ist, ist regelmäßig diffizil.

wobei für die zweiten Momente

$$\begin{aligned}\mathbb{E}(Y^2) &< \infty \\ \mathbb{E}(X_k^2) &< \infty, k = 1, \dots, K\end{aligned}$$

gilt. Ferner sei die Matrix der theoretischen zweiten Momente der X -Variablen (der Merkmale)

$$\begin{aligned}\mathbf{Q}_{XX} &= \mathbb{E} \left(\begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_K \end{pmatrix} \begin{pmatrix} X_1, X_2, \dots, X_K \end{pmatrix} \right) = \mathbb{E}(XX^T) \\ &= \mathbb{E} \begin{pmatrix} X_1 \cdot X_1 & X_1 \cdot X_2 & \dots & X_1 \cdot X_K \\ X_2 \cdot X_1 & X_2 \cdot X_2 & \dots & X_2 \cdot X_K \\ \dots & \dots & \dots & \dots \\ X_K \cdot X_1 & X_K \cdot X_2 & \dots & X_K \cdot X_K \end{pmatrix}\end{aligned}$$

positiv definit (also insbesondere invertierbar). Dann sprechen wir von einem **Linearen Regressionsmodell**.

9.0.12 Definitionen und Bemerkungen: Wir betrachten ein Lineares Modell für den bedingten Erwartungswert

$$\begin{aligned}Y &= \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K + u \\ \mathbb{E}(u|X_1, \dots, X_K) &= 0\end{aligned}$$

und eine Zufallsstichprobe

$$\mathcal{Y}_1 = (Y_1, X_{11}, \dots, X_{1K})$$

$$\mathcal{Y}_2 = (Y_2, X_{21}, \dots, X_{2K})$$

...

$$\mathcal{Y}_N = (Y_N, X_{N1}, \dots, X_{NK}).$$

Wir **kennen die Koeffizienten** β_1, \dots, β_K **nicht** und wollen diese auf Basis der Zufallsstichprobe **schätzen**.

Wir betrachten für irgendwelche Koeffizienten $\tilde{\beta}_0, \dots, \tilde{\beta}_K$ die **Fehler**

$$\tilde{u}_i = Y_i - \tilde{\beta}_0 - \tilde{\beta}_1 X_{i1} - \dots - \tilde{\beta}_K X_{iK}, \quad i = 1, \dots, N$$

Wir schätzen die Koeffizienten so, dass der *Gesamtfehler* minimal ist:

$$(\hat{\beta}_0, \dots, \hat{\beta}_K) = \arg \min_{\tilde{\beta}_0, \dots, \tilde{\beta}_K} \sum_{i=1}^N (Y_i - \tilde{\beta}_0 - \tilde{\beta}_1 X_{i1} - \dots - \tilde{\beta}_K X_{iK})^2$$

$\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_K)^T$ heißt **kleinste Quadrate Schätzer** der Koeffizienten β des linearen Modells.

9.0.13 Definition und Satz: Gegeben sei eine Stichprobe

der Größe N . Wir definieren die **Regressormatrix**

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1K} \\ X_{21} & X_{22} & \dots & X_{2K} \\ \dots & & & \\ X_{N1} & X_{N2} & \dots & X_{NK} \end{pmatrix}.$$

Die erste Spalte erfasst in Gleichungen mit Matrizen nahezu immer den Achsenabschnitt. Die erste Zufallsvariable ist dann $\mathbf{X}_{\bullet,1} \equiv 1$ und die erste Spalte hat nur 1'en. Wenn man mit Matrizen arbeitet, dann erfasst der Koeffizient β_1 den Achsenabschnitt (anstatt β_0). Wir werden im Folgenden meistens die Matrixschreibweise verwenden.

Erinnerung X ist ein **Zufallsvektor**. \mathbf{X} ist eine **Matrix** in der **jede Zeile** einer Beobachtung – also für ein i – des **Spaltenvektors** X entspricht.

Wenn diese Matrix $\mathbf{X}^T \mathbf{X}$ invertierbar ist, dann können wir den dann eindeutig bestimmten **kleinste Quadrate Schätzer** kompakt in der Form

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

schreiben, wobei

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}$$

ist.

9.0.14 Satz: Gegeben sei eine **Zufallsstichprobe für ein Lineares Regressionsmodell**. Dann ist der kleinste Quadrate Schätzer **unverzerrt**, d.h.

$$\mathbb{E}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{\beta}.$$

Weiterhin gilt für die Varianz der kleinste Quadrate Schätzung

$$\mathbb{V}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{D} \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1}$$

mit

$$\mathbf{D} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \end{pmatrix},$$

wobei $\sigma_i^2 = \mathbb{V}(u_i | X_{i1}, \dots, X_{iK})$.

Beweis: Hansen [17].

9.0.15 Satz: Wir betrachten für $N \in \mathbb{N}$ eine Zufallsstichprobe für ein lineares Regressionsmodell mit dem kleinsten quadratischen Schätzer $\hat{\boldsymbol{\beta}}^N = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$.

Dann gilt

$$\text{n.W.-lim } \hat{\boldsymbol{\beta}}^N = \boldsymbol{\beta}.$$

Gilt

$$\mathbb{E}(Y^4) < \infty, \mathbb{E}(X_k^4) < \infty$$

Dann gilt

$$\sqrt{n} (\hat{\boldsymbol{\beta}}^N - \boldsymbol{\beta}) \xrightarrow{\text{i.V.}} N(\mathbf{0}, \mathbf{V}_\beta),$$

wobei

$$\mathbf{V}_\beta = \mathbf{Q}_{XX}^{-1} \boldsymbol{\Omega} \mathbf{Q}_{XX}^{-1}$$

und

$$\mathbf{Q}_{XX} = \mathbb{E}(X X^T), \boldsymbol{\Omega} = \mathbb{E}(X X^T u^2).$$

Es seien

$$\hat{\mathbf{Q}}_{XX} = \frac{1}{N} \sum_{i=1}^N X_{i,\bullet} X_{i,\bullet}^T, \hat{\mathbf{Q}}_{XY} = \frac{1}{N} \sum_{i=1}^N X_{i,\bullet} Y_i^T$$

und

$$\hat{\Omega} = \frac{1}{N} \sum_{i=1}^N X_{i,\bullet} X_{i,\bullet}^T \hat{u}_i^2 \quad .$$

Der Schätzer

$$\hat{\mathbf{V}}_{\beta}^{\text{HC0}} = \hat{\mathbf{Q}}_{XX}^{-1} \hat{\Omega} \hat{\mathbf{Q}}_{XX}^{-1}$$

für \mathbf{V}_{β} ist konsistent:

$$\hat{\mathbf{V}}_{\beta}^{\text{HC0}} \xrightarrow{\text{n.W.}} \mathbf{V}_{\beta} \quad .$$

10 Maßtheorie

10.1 Einführung zur Maßtheorie

10.1.1 Bemerkung: i.) DAS großartige Buch zur Maß- und Integrationstheorie ist von **Elstrodt** [8]. **Für dieses Kapitel besonders geeignet ist aber unbedingt Henze** [18, Kapitel 8]. Ich habe auch andere Bücher geprüft, aber meine Favoriten sind Elstrodt und Henze. Eine andere sehr gute Quelle ist **Küchler** [23]. Eine ebenfalls sehr gute Quelle ist das berühmte Buch von Billingsley [2].

ii.) Wir haben schon (in WT1) etwas Maßtheorie betrieben. Ein Wahrscheinlichkeitsmaß ist nämlich ein Maß mit der zusätzlichen Eigenschaft $\mathbb{P}(\Omega) = 1$.

Wir betreiben jetzt Maßtheorie allgemeiner. **Wir wollen systematisch (axiomatisch) erklären, wie man Mengen messen kann.** Es geht dabei nicht mehr nur um Ereignisse und deren Wahrscheinlichkeiten, sondern z.B. auch um

Anzahl, Länge, Fläche und Volumen von passenden Mengen.

Dabei gehen wir so vor. Wir definieren zunächst **möglichst einfach** sogenannte **Inhalte auf möglichst einfachen Strukturen (einfachen Teilmengensystemen)** (sogenannten **Halbringen**). Dann **setzen** wir so einen **Inhalt** zu einem **Maß** auf der erzeugten **σ -Algebra fort**. Wir geben schließlich Bedingungen an, so dass diese Fortsetzung eindeutig ist.

Für uns – mit Blick auf die Anwendung in der WT – am wichtigsten sind die folgenden beiden Maße:

- Das **Lebesgue-Maß** auf $(\mathbb{R}, \mathcal{B})$ bzw. auf $(\mathbb{R}^n, \mathcal{B}^n)$. Wikipedia über Henri Lebesgue.
- **Lebesgue-Stieltjes-Maße** auch auf $(\mathbb{R}, \mathcal{B})$ bzw. auf $(\mathbb{R}^n, \mathcal{B}^n)$.
Lebesgue-Stieltjes-Maße bei Wikipedia. Wikipedia über Thomas Stieltjes

10.2 Mengensysteme

10.2.1 Definition: Es sei $\Omega \neq \emptyset$ eine nicht-leere Menge. Ein Teilmengensystem $\mathcal{H} \subset \mathcal{P}(\Omega)$ heißt **Halbring** (oder **Semi-Ring**) auf Ω , wenn gilt:

- i.) $\emptyset \in \mathcal{H}$,

$$\text{ii.) } A, B \in \mathcal{H} \Rightarrow A \cap B \in \mathcal{H},$$

$$\text{iii.) } A, B \in \mathcal{H} \Rightarrow \text{es gibt disjunkte } C_1, \dots, C_n \in \mathcal{H} \text{ mit} \\ A \setminus B = C_1 \uplus \dots \uplus C_n.$$

10.2.2 Bemerkung: Bei der Bedingung iii.) wird **nicht** gefordert, dass $A \setminus B \in \mathcal{H}$ gilt! Es wird (nur) gefordert, dass sich $A \setminus B$ als disjunkte Vereinigung von Mengen aus \mathcal{H} schreiben lässt.

10.2.3 Beispiel: i.) Die **Loravalle**¹ $\{(a, b] \mid a, b \in \mathbb{R}, a \leq b\}$ bilden einen Halbring.

ii.) Die Menge der Kreise $\{B_\varepsilon(x) \mid x \in \mathbb{R}^2, \varepsilon \geq 0\}$ bilden **keinen Halbring**, denn der Schnitt von zwei Kreisen ist i.A. kein Kreis.

10.2.4 Definition: a.) Es sei Ω eine nicht-leere Menge. Ein Teilmengensystem $\mathcal{R} \subset \mathcal{P}(\Omega)$ heißt **Ring**, wenn gilt:

$$\text{i.) } \emptyset \in \mathcal{R},$$

$$\text{ii.) } A, B \in \mathcal{R} \Rightarrow A \cup B \in \mathcal{R},$$

$$\text{iii.) } A, B \in \mathcal{R} \Rightarrow A \setminus B \in \mathcal{R}.$$

b.) Ist \mathcal{A} ein Ring mit $\Omega \in \mathcal{A}$, dann ist \mathcal{A} eine **Algebra**.

¹Ich glaube außer mir, verwendet niemand den Begriff Loravalle. Mal sehen ...

c.) Eine Algebra \mathcal{A} bzw. ein Ring \mathcal{R} heißt **σ -Algebra** respektive **σ -Ring**, falls die Abgeschlossenheit ii.) bezüglich \cup auch für abzählbare Vereinigungen gilt.

10.3 Inhalt, Prämaß, Maß

10.3.1 Definiton: Es sei $\Omega \neq \emptyset$ und $\mathcal{H} \subset \mathcal{P}(\Omega)$ ein Halbring.

i.) Eine Abbildung $\mu : \mathcal{H} \rightarrow [0, \infty]$ heißt **Inhalt**, falls $\mu(\emptyset) = 0$ und für alle $A_i \in \mathcal{H}, i = 1, \dots, n$ mit

$$\bigcup_{i=1}^n A_i \in \mathcal{H}, A_i \cap A_j = \emptyset, i \neq j$$

gilt:

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

(Beachte: Bei der Eigenschaft der σ -Additivität wird $A_i \in \mathcal{H}$ und $\biguplus A_i \in \mathcal{H}$ vorausgesetzt).

ii.) Ein σ -additiver Inhalt auf \mathcal{H} heißt **Prämaß** auf \mathcal{H} .

iii.) Ein **Maß** ist ein auf einer σ -Algebra definiertes Prämaß.

iv.) Wenn μ ein Maß auf einer σ -Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ ist, dann heißt das Triple $(\Omega, \mathcal{A}, \mu)$ ein **Maßraum**.

10.3.2 Bemerkung: Wenn $(\Omega, \mathcal{A}, \mu)$ ein Maßraum mit $\mu(\Omega) = 1$ ist, dann ist $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum

10.3.3 Bemerkung: Es sei Ω eine Menge. Dann ist $\mathcal{P}(\Omega)$ eine σ -Algebra. Warum suchen wir nicht nach einem Maß auf ganz $\mathcal{P}(\Omega)$? Für den Fall, dass Ω endlich oder abzählbar unendlich ist, haben wir (in WT1) regelmäßig $\mathcal{P}(\Omega)$ verwendet. Warum jetzt so kompliziert? Weil es im Allgemeinen nicht geht; jedenfalls nicht so wie wir es gerne hätten. Die Aussage wollen wir jetzt substantivieren. Drei berühmte Resultate sollen diese Unmöglichkeit vorführen. Wir betrachten $\Omega = \mathbb{R}^n$, da dieser Raum für viele praktische Probleme (Länge, Fläche, Volumen) angemessen ist. Ein berühmtes Buch über die Probleme beim mathematischen Messen ist Tomkowicz und Wagon [32]. Dort werden die unten genannten Aussagen auch mathematisch genau untersucht (und bewiesen).

Bemerkung: Eine Abbildung $\beta : \mathbb{R}^n \rightarrow \mathbb{R}^n$ heißt **Bewegung (Isometrie)**, falls $\|x - y\|_2 = \|\beta(x) - \beta(y)\|_2$ für alle x, y gilt.

Eine Bewegung erhält also alle Abstände. Wenn wir eine Menge $M \subset \mathbb{R}^n$ messen wollen, dann ist es (wohl) vernünftig, dass sich das Maß der Menge nicht ändert, wenn wir die Menge (nur) bewegen.

Satz (Hausdorff): Das Inhaltsproblem ist auf

\mathbb{R}^n , $n \geq 3$ unlösbar. D.h. Es gibt **keine** Abbildung $m : \mathcal{P}(\mathbb{R}^n) \rightarrow \mathbb{R}$ mit

- m ist endlich additiv
- m ist bewegungsinvariant, d.h. für alle Bewegungen β ist $m(\beta(A)) = m(A)$
- m ist normiert, d.h. $m([0, 1]^n) = 1$

Über Felix Hausdorff bei Wikipeada

Satz (Vitali): Das Maßproblem ist unlösbar.²

Über Giuseppe Vitali bei Wikipeada

Satz (Banach, Tarski): Es sei $n \geq 1$ und $A, B \subset \mathbb{R}^n$ seien **beliebige** Mengen mit nicht-leerem Inneren. Dann gibt es abzählbar viele Mengen $C_k \subset \mathbb{R}^n$ und Bewegungen $\beta_k : \mathbb{R}^n \rightarrow \mathbb{R}^n$, so dass

$$A = \bigsqcup C_k$$
$$B = \bigsqcup \beta_k(C_k)$$

Über Stefan Banach bei Wikipeada

Über Alfred Tarski bei Wikipeada

²Für alle n .

10.4 Fortsetzungssätze

10.4.1 Definiton: Es sei $\mathcal{K} \subset \mathcal{P}(\Omega)$ ein Teilmengensystem. Dann heißt

$$\bigcap_{\substack{\mathcal{A} \text{ ist eine } \sigma\text{-Algebra} \\ \mathcal{A} \supset \mathcal{K}}} \mathcal{A} =: \sigma(\mathcal{K})$$

die von \mathcal{K} **erzeugte σ -Algebra**. $\sigma(\mathcal{K})$ ist die kleinste σ -Algebra, die \mathcal{K} enthält.

Wenn also $\tilde{\mathcal{A}}$ eine σ -Algebra mit $\mathcal{K} \subset \tilde{\mathcal{A}}$ ist, dann gilt $\sigma(\mathcal{K}) \subset \tilde{\mathcal{A}}$.

10.4.2 Bemerkung: i.) Das obige Konstruktionsprinzip funktioniert, denn der Schnitt beliebig vieler σ -Algebren ist eine σ -Algebra (siehe Elstrodt [8, Seite 17]). Für Halbringe gilt das nicht. Da wir mit Halbringen anfangen, ist das aber unproblematisch. Denn: Wir suchen nicht nach erzeugten Halbringen, sondern wir erzeugen mit Halbringen.

ii.) Es gibt eine Reihe anderer Mengensysteme, die je nach Fragestellung (und insbesondere in Beweisen) wichtig sind: Monotone Klassen, π -System, λ -System (Dynkin-Systeme). Auf diese Mengensystem sei hier nur aber nur kurz hingewiesen. Für Details vergleiche Elstrodt [8] und Kűchler [23].

10.4.3 Bemerkung: Wir haben den folgenden Plan. Wir starten mit einem Inhalt ν auf einem Halbring \mathcal{H} . Das ist nämlich eine relativ übersichtliche Situation. Wir wollen ν zu einem Maß μ auf der von \mathcal{H} erzeugten σ -Algebra **fortsetzen**.

10.4.4 Satz (Fortsetzungssatz von Caratheodory): Es sei \mathcal{H} ein Halbring auf Ω . ν ein Inhalt auf \mathcal{H} .

Für $A \in \mathcal{P}(\Omega)$ definieren wir das äußere Maß von A

$$\mu(A) = \inf \left\{ \sum_{j=1}^{\infty} \nu(A_j) \mid A_j \in \mathcal{H}, A \subset \bigcup_{j=1}^{\infty} A_j \right\}$$

Ist μ ein Prämaß, so ist μ ein Maß auf der von \mathcal{H} erzeugten σ -Algebra $\mathcal{A} := \sigma(\mathcal{H})$ und es gilt $\mu|_{\mathcal{H}} = \nu$.

Beweis: Henze [18, Seite 313]

Über Caratheodory bei Wikipeda

10.4.5 Bemerkung: Die obige Definition von μ durch die summierten Inhalte der Überdeckungen nennt man die **Approximation von außen**.

10.4.6 Definition: Ein Inhalt ν heißt **σ -endlich** (σ -finit), falls es eine Folge von Mengen $E_j \in \mathcal{H}$ mit $\nu(E_j) < \infty$ und $\Omega = \bigcup E_j$.

10.4.7 Satz (Eindeutigkeit der Fortsetzung): Es sei \mathcal{H} ein Halbring auf Ω und ν ein σ -endlicher Inhalt auf \mathcal{H} , dann ist die Fortsetzung gemäß des Fortsetzungssatz von Caratheodory eindeutig.

10.4.8 Bemerkung: Der Beweis der Eindeutigkeit bedient sich des Konzept der Dynkin-Systeme.

10.4.9 Bemerkung: Das **Programm nach Caratheodory**³ geht also so:

1. Auf einem **Halbring** \mathcal{H} definieren wir einen **Inhalt** ν .
2. Wir weisen nach, dass der Inhalt ν sogar σ -additiv ist; ν ist also ein **Prämaß**.
3. Die Sätze von Caratheodory beschreiben wie wir ein **Maß** auf $\sigma(\mathcal{H})$ erhalten.

Also:

Inhalt auf einem Halbring \leadsto Prämaß \leadsto Maß auf einer σ -Algebra

10.4.10 Satz: Es gilt

- i.) Die Menge der **Loravalle** $\{(a, b] \mid a, b \in \mathbb{R}, a \leq b\}$ bilden einen Halbring.

³Auch diesen Begriff verwendet außer mir wohl keiner.

ii.) Die Abbildung

$$\nu : \mathcal{H} \rightarrow \mathbb{R}, \nu((a, b]) \mapsto b - a$$

ist ein Inhalt.

iii.) ν ist sogar ein Prämaß.

Beweis: Vgl. Billingsley [2, Seite 13ff]

10.4.11 Bemerkung: Der Beweis des vorhergehenden Satzes verwendet den Überdeckungssatz von Heine-Borel: Eine Teilmenge des \mathbb{R}^n ist genau dann kompakt (also beschränkt und abgeschlossen), wenn jede offene Überdeckung eine endliche Teilüberdeckung enthält.

10.4.12 Definition: Die von den Loravallen erzeugte σ -Algebra heißt **Borel'sche σ -Algebra**; sie wird mit $\mathcal{B}(\mathbb{R})$ bezeichnet. Die Mengen in \mathcal{B} heißen **Borelmengen**. Auf den Loravallen definieren wir den Inhalt ν wie im vorhergehenden Satz. Das dann wie im Satz von Caratheodory definierte Maß $\lambda : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{R}$ heißt **Borel-Lebesgue-Maß**.

10.5 Maße mit F definieren

► Für die flexible Definition des Erwartungswertes einer Zufallsvariable X benötigen wir sogenannte **Stieltjes-Integrale** (bzw. eigentlich Lebesgue-Stieltjes-Integrale; dazu später mehr). Die Idee kann man an der folgenden Approximation erfassen:

$$\begin{aligned}\mathbb{E}(X) &= \int X dF \\ &\approx x_1 \cdot (F(b_1) - F(a_1)) + x_2 \cdot (F(b_2) - F(a_2)) + \dots\end{aligned}$$

Wir stellen uns vor, dass der Integrationsbereich in Intervalle unterteilt wird: $(a_1, b_1] \cup (a_2, b_2] \cup \dots$ und Zwischenwerte $x_1 \in (a_1, b_1], \dots$ entnommen werden. Um diese Idee nachzuvollziehen betrachten wir die entsprechende Approximation beim Riemann-Integral:

$$\int x dx = x_1 \cdot (b_1 - a_1) + x_2 \cdot (b_2 - a_1) + \dots$$

Während die Zwischenwerte $x_i \in (a_i, b_i]$ beim Riemann-Integral mit den **Lebesgue-Inhalten** $b_i - a_i$ multipliziert werden, werden die Zwischenwerte beim Stieltjes-Integral mit den **Stieltjes-Inhalten** $F(b_i) - F(a_i)$ multipliziert. Wenn wir uns weiter vorstellen, dass F eine Verteilungsfunktion einer Zufallsvariable X ist, dann erhalten wir

$$F(b_i) - F(a_i) = \mathbb{P}^X((a_i, b_i]).$$

Der Stieltjes Inhalt erfasst also **die Wahrscheinlichkeit** für einen Wert im Intervall $(a_i, b_i]$ anstatt *nur* die Länge des Intervalls $(a_i, b_i]$. Wann wäre die Länge angemessen?

10.5.1 Satz: i.) Es sei $F : \mathbb{R} \rightarrow \mathbb{R}$ nicht-fallend. Dann ist $\nu((a, b]) = F(b) - F(a)$ **ein endlicher Inhalt** auf den Lo-vallen \mathcal{H} .⁴ Für zwei nicht-fallende Funktionen $F, G : \mathbb{R} \rightarrow \mathbb{R}$ gilt $\nu_F = \nu_G$ genau dann, wenn $F - G$ konstant ist.

ii.) Ist $\nu : \mathcal{H} \rightarrow \mathbb{R}$ ein endlicher Inhalt und $F : \mathbb{R} \rightarrow \mathbb{R}$ definiert durch

$$F(x) = \begin{cases} \nu((0, x]) & \text{für } x \geq 0 \\ -\nu((x, 0]) & \text{für } x < 0 \end{cases}.$$

Dann ist F nicht-fallend und $\nu = \nu_F$.

Beweis: Siehe Elstrodt [8, Seite 39]

► Der Lebesgue-Inhalt λ ergibt sich für $F(x) = x$.

10.5.2 Definition: Für nicht-fallendes $F : \mathbb{R} \rightarrow \mathbb{R}$ heißt ν_F der zu F gehörende **Stieltjes Inhalt**.

10.5.3 Satz: Es sei $F : \mathbb{R} \rightarrow \mathbb{R}$ nicht-fallend und ν_F der zu F gehörende **Stieltjes-Inhalt**. ν_F ist genau dann ein **Prä-maß**, wenn F von rechts stetig ist.

⁴Ein Inhalt heißt endlich, falls $\nu(H) < \infty$ für alle $H \in \mathcal{H}$.

Beweis: Siehe Elstrodt [8, Seite 39]

10.5.4 Bemerkung: Die beiden vorhergehenden Sätze charakterisieren alle endlichen Inhalte bzw. alle endlichen Prämaße auf den Loravallen. Ein endlicher Inhalt wird durch eine bis auf Konstante eindeutig bestimmte nicht fallende Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$ charakterisiert. Ein endliches Prämaß wird durch eine bis auf Konstante eindeutig bestimmte nicht fallende von rechts stetig Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$ charakterisiert.

10.5.5 Definition: Es sei F nicht-fallend und von rechts stetig und ν_F der zu F gehörende **Stieltjes Inhalt**. Das dann wie im Satz von Caratheodory definierte Maß $\lambda_F : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ heißt **Borel-Lebesgue-Stieltjes-Maß**.

10.5.6 Bemerkung: Wir haben ein bemerkenswertes Programm abgearbeitet. Es ist ganz offensichtlich, dass die Loravalle für die Anwendung sehr wichtig sind. **Wir wollen regelmäßig mindestens Loravalle messen.**

Wir haben auch gesehen, wie alle Inhalte beschrieben werden können, nämlich durch eine nicht-fallende Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$. Wenn wir noch fordern, dass F von rechts stetig ist, dann erhalten so nach dem Programm nach Caratheodory passende Maße (**Fortsetzungen**) auf \mathcal{B} .

► Agenda: Wir wollen F in einen stetigen und einen diskreten (Sprunganteil) zerlegen.

10.5.7 Definition: Eine Funktion $J : \mathbb{R} \rightarrow \mathbb{R}$ heißt **Sprungfunktion**, wenn es

- eine abzählbare Menge der Sprungstellen $\Delta \subset \mathbb{R}$ gibt und
- eine positive Funktion $p : \Delta \rightarrow (0, \infty)$ der Sprunghöhen mit $\sum_{y \in \Delta \cap [-n, n]} p(y) < \infty$ für alle $n \in \mathbb{N}$ und
- $\alpha \in \mathbb{R}$,

so dass

$$J(x) = \begin{cases} \alpha + \sum_{y \in \Delta \cap (0, x]} p(y) & \text{für } x \geq 0 \\ \alpha - \sum_{y \in \Delta \cap (x, 0]} p(y) & \text{für } x < 0 \end{cases}.$$

10.5.8 Satz: Es sei $F : \mathbb{R} \rightarrow \mathbb{R}$ nicht-fallend und von rechts stetig. Dann ist die Menge Δ der **Unstetigkeitsstellen** von F abzählbar.

Beweis: Vgl. Elstrodt [8, S. 43]

10.5.9 Satz: i.) Es sei $F : \mathbb{R} \rightarrow \mathbb{R}$ nicht-fallend und von rechts stetig. Dann gibt es eine **nicht-fallende und von rechts stetige Sprungfunktion** J und eine **nicht-fallende**

und stetige Funktion $H : \mathbb{R} \rightarrow \mathbb{R}$, so dass $F = J + H$.

ii.) Es gilt $\nu_F = \nu_J + \nu_H$.

iii.) Bis auf additive Konstanten sind die Funktionen J, H eindeutig bestimmt.

Beweis: Elstrodt [8, Seite 43f]

10.5.10 Bemerkung: Wir können gemäß des vorhergehenden Satzes also insbesondere Verteilungsfunktion in einen stetigen und einen Sprungteil zerlegen.

10.6 Vollständigkeit

10.6.1 Definition: Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Dieser Maßraum heißt **vollständig**, wenn jede Teilmenge $A \subset N$ einer Nullmenge N (d.h. $\mu(N) = 0$) auch zu \mathcal{A} gehört: $A \in \mathcal{A}$.

10.6.2 Satz: Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Es sei \mathcal{N} das System aller Nullmengen. Wir betrachten das Mengensystem

$$\tilde{\mathcal{A}} = \mathcal{A} \cup \mathcal{N}$$

und auf $\tilde{\mathcal{A}}$, die durch

$$\tilde{\mu}(\tilde{A}) := \mu(A), \tilde{A} = A \cup N, A \in \mathcal{A}, N \in \mathcal{N}$$

definierte Abbildung $\tilde{\mu}$.

Dann ist $(\Omega, \tilde{\mathcal{A}}, \tilde{\mu})$ ein vollständiger Maßraum.

Beweis: Vgl. Elstrodt [8, Seite 71]

10.6.3 Definition: Das Maß $\tilde{\mu}$ aus dem vorhergehenden Satz heißt **Vervollständigung** von μ .

10.6.4 ...

11 Messbare Abbildungen

Quellen: Henze [18], Elstrodt [8], Küchler [23].

11.1 Grundlegende Begriffe

Für die Integration (im nächsten Kapitel) benötigen wir unbedingt den Begriff der **Messbarkeit von Abbildungen**. Die Messbarkeit steht zudem in einem unmittelbaren Zusammenhang zur Information, die in \mathcal{A} enthalten ist. Salopp formuliert: Eine \mathcal{A} -messbare Abbildung kann nicht mehr Information als \mathcal{A} anzeigen.

11.1.1 Definition: Es sei $f : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ eine Abbildung zwischen messbaren Räumen, dann heißt f **messbar** falls $f^{-1}(A') \in \mathcal{A}$ für alle $A' \in \mathcal{A}'$ gilt [also $f^{-1}(\mathcal{A}') \subset \mathcal{A}$]. Also: Das Urbild jeder im Zielraum messbaren Menge ist im Startraum Raum messbar.

► Wir benötigen den Begriff *messbar* insbesondere, wenn wir auf der Zielmenge Ω' messen wollen, aber *zunächst* nur ein Maß auf dem Definitionsbereich Ω haben. Das ist bei Zufallsvariablen der Fall. Die sind auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definiert und nehmen Werte in \mathbb{R} an. Um das *Verhalten* der **Werte** $f(\omega)$ der Zufallsvariablen zu erfassen, benötigen wir das Bildmaß (die Verteilung). Um das Bildmaß zu definieren, benötigen wir die Messbarkeit.

► Wir erinnern uns $f^{-1}(A')$ ist die Menge aller $\omega \in \Omega$, die von f in die Menge A' abgebildet werden: $f(\omega) \in A'$.

11.1.2 Definition und Satz: Es sei $f : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ eine messbare Abbildung und μ ein Maß auf (Ω, \mathcal{A}) . Dann ist

$$\nu : \begin{cases} \mathcal{A}' \rightarrow [0, \infty] \\ A' \mapsto \nu(A') := \mu(f^{-1}(A')) \end{cases}$$

ein Maß auf (Ω', \mathcal{A}') .

ν heißt das **Bildmaß** von μ bezüglich f .

Wesentlich vereinfacht wird der Nachweis der Messbarkeit durch den folgenden Satz.

11.1.3 Satz: Es sei $f : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ eine Abbildung und \mathcal{E}' ein Erzeuger von \mathcal{A}' ; also $\sigma(\mathcal{E}') = \mathcal{A}'$. Dann gilt: f ist

genau dann messbar, wenn $f^{-1}(\mathcal{E}') \subset \mathcal{A}$.

► Wir müssen also nur $f^{-1}(E') \in \mathcal{A}$ für die $E' \in \mathcal{E}'$ untersuchen!

11.1.4 Definition: Es sei $f : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ eine Abbildung. Die kleinste σ -Algebra auf Ω , so dass f messbar ist, heißt die von f **erzeugte σ -Algebra**. Sie wird mit $\sigma(f)$ bezeichnet.

Kleinste σ -Algebra bedeutet: Wenn $\tilde{\mathcal{A}}$ eine σ -Algebra auf Ω ist, so dass f messbar ist, dann gilt $\sigma(f) \subset \tilde{\mathcal{A}}$.

11.1.5 Satz: Es gilt

$$\sigma(f) = \{f^{-1}(A') \mid A' \in \mathcal{A}'\}.$$

Beweis: Natürlich muss $\{f^{-1}(A') \mid A' \in \mathcal{A}'\} \subset \sigma(f)$ gelten, denn sonst wäre f nicht $\sigma(f)$ -messbar. Außerdem bleiben alle Eigenschaften einer σ -Algebra bei der Bildung der Urbilder erhalten. Also ist $\{f^{-1}(A') \mid A' \in \mathcal{A}'\}$ eine σ -Algebra. Dann kann $\{f^{-1}(A') \mid A' \in \mathcal{A}'\}$ keine echte Teilmenge von $\sigma(f)$ sein. Sonst wäre $\{f^{-1}(A') \mid A' \in \mathcal{A}'\}$ eine kleinere σ -Algebra bezüglich der f messbar ist.

11.1.6 Beispiel: Wir betrachten $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$ und $\Omega' = \mathbb{R}$, $\mathcal{A}' = \mathcal{B}$. Wir betrach-

ten die Abbildung $f : \Omega \rightarrow \mathbb{R}$ mit

$$f(\omega) = \begin{cases} 8, & \text{falls } \omega \in \{1, 2, 3\} \\ 11, & \text{falls } \omega \in \{4, 5, 6\}. \end{cases}$$

a.) f ist nicht messbar.

b.) Es ist $\sigma(f) = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\}$.

Wir beobachten auch, dass $f(1) = 8 \neq 11 = f(5)$ ist. Die Abbildung f ist auf der Menge $\{1, 3, 5\}$ nicht konstant.

► $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$ erfasst das Experiment des abgeklebten Würfels. f zeigt insbesondere an, ob ein Ergebnis aus $\{1, 2, 3\}$ gewürfelt wurde. Das *funktioniert* jedoch nicht, wenn die Information eigentlich $\mathcal{A} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$ ist. Wir können 1 und 5 eigentlich – also bei der Information \mathcal{A} – nicht unterscheiden. Die Abbildung zeigt aber unterschiedliche Werte für $\omega = 1$ und $\omega = 5$ an. Mit der Beobachtung von f kann man $\omega = 1$ versus $\omega = 5$ doch entschieden. Messbarkeit von f würde bedeuten, dass so was **nicht passiert**: An den Werten von f kann man keine über \mathcal{A} hinausgehende Information ablesen.

► Lösung im Unterricht!

11.1.7 Satz: i.) Es sei c eine reelle Zahl. Die Abbildung $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}), \omega \mapsto c$ ist messbar.

ii.) Die Indikatorabbildung $\mathbb{1}_A : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ für eine Menge $A \subset \Omega$ ist genau dann messbar, wenn $A \in \mathcal{A}$. Die Messbarkeit der Indikatorfunktion entspricht also der Messbarkeit der Menge, die angezeigt werden soll.

iii.) Es sei $A \subset \Omega$. Die kleinste σ -Algebra, so dass $\mathbb{1}_A : \Omega \rightarrow (\mathbb{R}, \mathcal{B})$ messbar ist, ist $\{\emptyset, \Omega, A, A^c\}$.

Beweis: i.) Es sei $B \in \mathcal{B}$. Wenn $c \in B$ gilt, dann $f(\omega) = c \in B$ für alle $\omega \in \Omega$. Also $f^{-1}(B) = \Omega$. Wenn $c \notin B$ gilt, dann $f(\omega) = c \notin B$ für alle $\omega \in \Omega$. Also $f^{-1}(B) = \emptyset$. Natürlich gilt $\emptyset, \Omega \in \mathcal{A}$ (das gilt für jede σ -Algebra). Also ist f messbar.

ii.) “ \Rightarrow ”: Die Abbildung $\mathbb{1}_A$ sei messbar. Das Loravall $(0.5, 1.5]$ ist natürlich eine Borelmenge (denn Borelmengen werden mit den Loravallen erzeugt). Wegen der Messbarkeit von $\mathbb{1}_A$, muss $A = \mathbb{1}_A^{-1}((0.5, 1.5]) \in \mathcal{A}$ gelten.

“ \Leftarrow ”: Es sei $A \in \mathcal{A}$; dann ist auch $A^c \in \mathcal{A}$. Natürlich sind auch $\emptyset, \Omega \in \mathcal{A}$. Wir halten schon mal fest, dass unter der Voraussetzung $A \in \mathcal{A}$ die vier Mengen $\emptyset, \Omega, A, A^c \in \mathcal{A}$ in \mathcal{A} liegen.

Es sei $B \in \mathcal{B}$. Wenn $0, 1 \notin B$, dann $\mathbb{1}_A^{-1}(B) = \emptyset$. Wenn $0 \notin B$ und $1 \in B$, dann $\mathbb{1}_A^{-1}(B) = A$. Wenn $0 \in B$ und $1 \notin B$, dann $\mathbb{1}_A^{-1}(B) = A^c$. Wenn $0, 1 \in B$, dann $\mathbb{1}_A^{-1}(B) = \Omega$. Also ist $\mathbb{1}_A^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathcal{B}$.

iii.) Es sei $A \subset \Omega$. $\{\emptyset, \Omega, A, A^c\}$ ist eine σ -Algebra; in der Tat ist $\{\emptyset, \Omega, A, A^c\}$ die kleinste σ -Algebra, die A enthält. Wir haben im Beweis von ii.) gesehen, dass es genau vier Möglichkeiten für $\mathbb{1}_A^{-1}(B)$, $B \in \mathcal{B}$ gibt und zwar genau die Mengen $\{\emptyset, \Omega, A, A^c\}$. Also ist $\mathbb{1}_A^{-1}$ messbar bezüglich $\{\emptyset, \Omega, A, A^c\}$ und eine kleinere σ -Algebra kann es nicht geben.

11.1.8 Satz: Es sei K endlich und $\mathcal{Z} = \{Z_j | j \in K\}$ eine Zerlegung von $\Omega \neq \emptyset$. Dann gilt

$$\sigma(\mathcal{Z}) = \left\{ \bigcup_{k \in J} Z_k \mid J \subset K \right\}.$$

Es folgt insbesondere, dass Mengen $\emptyset \neq A \subsetneq Z_i$ nicht in $\sigma(\mathcal{Z})$ liegen können; $A \notin \sigma(\mathcal{Z})$.

► Feinere Information als Z_i gibt es in $\sigma(\mathcal{Z})$ nicht.

Beweis: Wir beobachten zunächst, dass

$$\mathcal{A} = \left\{ \bigcup_{k \in J, J \subset K} Z_k \right\}$$

eine σ -Algebra ist. Dann beobachten wir: Wenn \mathcal{A}' eine σ -Algebra ist, die \mathcal{Z} enthält. Dann sind auch alle Menge aus \mathcal{A} in \mathcal{A}' ; die Mengen aus \mathcal{A} ergeben sich als Vereinigung aus Mengen von \mathcal{Z} und \mathcal{A}' ist vereinigungsstabil. Also ist – wie behauptet – \mathcal{A} die kleinste σ -Algebra, die \mathcal{Z} enthält.

11.1.9 Satz: Wenn f eine Abbildung mit endlichem $\text{Bild}(f) = \{c_1, \dots, c_n\}$ ist (o.B.d.A. c_i verschieden), dann definieren die Mengen $Z_i = f^{-1}(\{c_i\})$ eine Zerlegung $\mathcal{Z} = \{Z_1, \dots, Z_n\}$ und es gilt

$$\sigma(f) = \sigma(\mathcal{Z}) = \left\{ \bigcup_{k \in J} Z_k \mid J \subset K \right\}.$$

Beweis: Die Mengen $Z_i = f^{-1}(\{c_i\})$ bilden (natürlich) eine Zerlegung von Ω , denn $\text{Bild}(f) = \{c_1, \dots, c_n\}$.

Wir müssen nachweisen, dass alle Mengen $f^{-1}(B)$, $B \in \mathcal{B}$ in $\{\bigcup_{k \in J} Z_k \mid J \subset K\}$ vorkommen und alle Mengen in $\{\bigcup_{k \in J} Z_k \mid J \subset K\}$ auch in $\{f^{-1}(B) \mid B \in \mathcal{B}\}$ sind.

Es sei B eine Borelmeng. Zunächst: Wenn $f^{-1}(B) = \emptyset$, dann ist $c_k \notin B$ alle k . Also ist auch $\bigcup_{j \text{ mit } c_j \in B} Z_j = \emptyset$. Umgekehrt, ist $\bigcup_{j \text{ mit } c_j \in B} Z_j = \emptyset$, dann ist $c_k \notin B$ alle k . Dann gilt auch $f^{-1}(B) = \emptyset$.

Es sei B eine Borelmenge mit $f^{-1}(B) \neq \emptyset$. Wir wollen

$$f^{-1}(B) = \bigcup_{j \text{ mit } c_j \in B} Z_j$$

zeigen.

Es sei $\omega \in f^{-1}(B)$. Es sei $B \ni f(\omega) = c_l$ für ein geeignetes l . Dann ist $\omega \in Z_l$ und folglich auch $\omega \in \bigcup_{j \text{ mit } c_j \in B} Z_j$.

Ist $\omega \in \bigcup_{j \text{ mit } c_j \in B} Z_j$, dann gibt es ein l mit $\omega \in Z_l$ und $c_l \in B$. Also $\omega \in f^{-1}(B)$.

Wir haben nachgewiesen, dass

$$f^{-1}(B) = \bigcup_{j \text{ mit } c_j \in B} Z_j$$

gilt.

Auf der linken Seite stehen alle Mengen aus $\sigma(f)$ und auf der rechten alle Mengen aus $\sigma(\mathcal{Z})$.

11.1.10 Satz: Es sei $\mathcal{Z} = \{Z_1, \dots, Z_n\}$ eine Zerlegung von Ω und $f : \Omega \rightarrow \mathbb{R}$ eine Abbildung. f ist genau dann $\sigma(\mathcal{Z})$ -messbar, wenn f auf den Mengen Z_k der Zerlegung konstant ist.

An den Werten von $f(\omega) = c_j$ kann man nicht ablesen, welches $\omega \in Z_j$ eingetreten ist; nur das irgendein $\omega \in Z_j$ eingetreten ist.

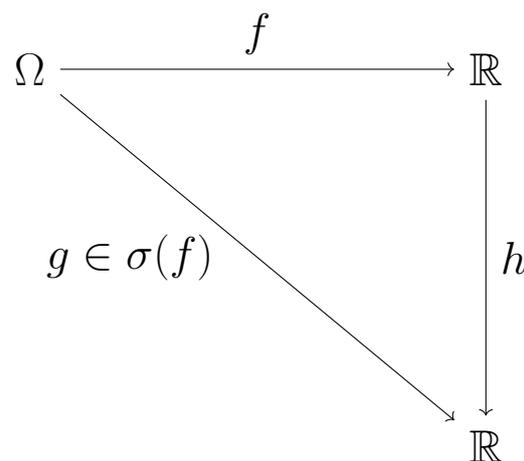
Beweis: Angenommen f ist auf Z_1 nicht konstant, d.h. es gibt $c_1, c_2, c_1 \neq c_2$ und $\omega_1, \omega_2 \in Z_1$ mit $f(\omega_1) = c_1 \neq c_2 = f(\omega_2)$. Wäre f $\sigma(\mathcal{Z})$ -messbar, dann müsste $f^{-1}(\{c_1\}) \in \sigma(\mathcal{Z})$ gelten (denn $\{c_1\}$ ist eine Borelmenge). Also auch $A = f^{-1}(\{c_1\}) \cap Z_1 \in \sigma(\mathcal{Z})$; denn $\sigma(\mathcal{Z})$ ist eine σ -Algebra. Dann ist $\emptyset \neq A \subsetneq Z_1$; die strikte \subsetneq Relation gilt wegen $\omega_2 \in Z_1$.

Ein solches A kann es in $\sigma(\mathcal{Z})$ nicht geben (das haben wir vorne beobachtet). Also kann f nicht $\sigma(\mathcal{Z})$ -messbar sein.

Angenommen f ist auf den Mengen Z_j konstant. Für jede Borelmenge B gilt dann (das zeigt man wie im vorhergehenden Beweis)

$$f^{-1}(B) = \bigcup_{j \text{ mit } c_j \in B} Z_j.$$

Also $f^{-1}(B) \in \sigma(\mathcal{Z})$.



$\exists g \in \sigma(f) \Rightarrow \exists h : \mathbb{R} \rightarrow \mathbb{R}$ mit $g(\omega) = h(f(\omega))$

Abbildung 11.1: Schema zum Faktorisierungslemma

11.1.11 Satz (Faktorisierungslemma): Es sei $f : \Omega \rightarrow \mathbb{R}$ eine Abbildung und $\sigma(f) = f^{-1}(\mathcal{B})$, die von f auf Ω erzeugte σ -Algebra. Ferner sei $g : \Omega \rightarrow \mathbb{R}$ eine $\sigma(f)$ -messbare Abbildung. Genau dann gibt es eine messbare Abbildung $h : \mathbb{R} \rightarrow \mathbb{R}$ mit $g(x) = h(f(x))$.

Anschaulich: $\sigma(f)$ erfasst die Information (denn es ist die kleinste σ -Algebra, so dass f messbar ist), die zu f gehört. Wenn auch g zu dieser Information passt, dann ist g *nur* eine messbare Transformation von f . g kann keine weitere Information *anzeigen*.

Beweis: Vgl. K uchler [23, Seite 115] und  bungsaufgabe oder Shao [29, Seite 37] oder Henze [18, Seite 175].

Beweis: Wir betrachten zun chst eine elementare Funktion g , d.h.

$$g = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}.$$

Gem ss Voraussetzung ist g messbar bezuglich der von f erzeugten σ -Algebra $\sigma(f) = f^{-1}(\mathcal{B})$. Dann muss $A_i \in \sigma(f)$ gelten. Da $\sigma(f) = f^{-1}(\mathcal{B})$ muss es Borelmengen $B_i \in \mathcal{B}$ geben, so dass $A_i = f^{-1}(B_i)$ gilt.

Dann definieren wir

$$h = \sum_{i=1}^n \alpha_i \mathbb{1}_{B_i}$$

und rechnen nach, dass

$$g = h \circ f$$

also

$$g(\omega) = h(f(\omega)) \text{ f ur alle } \omega \in \Omega$$

gilt.

.... weiter in der Mitschrift

11.1.12 Beispiel: Es sei $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $f(1) = f(3) = f(5) = 8$ und $f(2) = f(4) = f(6) = 11$. Dann ist $\sigma(f) = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$.

Ferner sei $g(1) = 1$ und $g(k) = 0, k = 2, \dots, 6$. Die Abbildung g ist auf $\{1, 3, 5\}$ nicht konstant und deshalb auch nicht $\sigma(f)$ -messbar. Gemäß des vorherigen Satzes kann es ein h nicht geben. Angenommen doch. Angenommen es gibt ein h mit

$$g(\omega) = h(f(\omega)) \text{ für alle } \omega \in \Omega.$$

Dann

$$1 = g(1) = h(f(1)) = h(8)$$

$$0 = g(3) = h(f(3)) = h(8)$$

Wir erhalten den Widerspruch $0 = 1$. Es kann so ein h nicht geben.

► Weitere Beispiele im Unterricht

11.2 Treppenfunktionen

11.2.1 Definition: Eine messbare Funktion $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ mit endlicher Bildmenge heißt **Treppenfunktion**. Beachte, dass die Messbarkeit Teil der Definition ist!

11.2.2 Satz: Wenn $f : \Omega \rightarrow \mathbb{R}$ eine Treppenfunktion ist, dann gibt es reelle Zahlen $\alpha_1, \dots, \alpha_m$ und $A_1, \dots, A_m \in \mathcal{A}$ mit

$$f = \sum_{j=1}^m \alpha_j \mathbb{1}_{A_j}.$$

Hinweis: Die Messbarkeit von f und $A_1, \dots, A_m \in \mathcal{A}$ sind Teil der Definition der Treppenfunktion. Eine Abbildung, die exakt so *aussieht* wie eine Treppe, deren A_1, \dots, A_m nicht alle in \mathcal{A} liegen, nennen wir nicht Treppenfunktion.

11.2.3 Definition: f ist eine Treppenfunktion in **kanonischer Form**, wenn die α_j paarweise verschieden sind.

11.2.4 Satz: Man kann eine Treppenfunktion immer in kanonischer Form darstellen.

11.3 $\bar{\mathbb{R}}$ und numerische Funktionen

► Für die Entwicklung des Integrationsbegriff ist es vorteilhaft, auch ∞ und $-\infty$ als Funktionswerte zu zulassen. Dafür benötigen wir eine entsprechende **Erweiterung** der reellen Zahlen/Achse sowie der Borelmengen.

11.3.1 Definition und Behauptung: Wir definieren den Maßraum $(\bar{\mathbb{R}}, \bar{\mathcal{B}})$

$$\begin{aligned}\bar{\mathbb{R}} &= \mathbb{R} \cup \{\infty, -\infty\} \\ \bar{\mathcal{B}} &= \{B \cup E \mid B \in \mathcal{B}, E \subset \{-\infty, \infty\}\}.\end{aligned}$$

Eine Funktion $f : (\Omega, \mathcal{A}) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ heißt **numerische Funktion**.

11.3.2 Bemerkung: Es sei (a_i) eine nicht fallende Folge in \mathbb{R} . Wenn die Folge beschränkt ist, dann konvergiert so eine Folge. Wenn sie unbeschränkt ist dann *konvergiert* sie **nur auf der erweiterten Achse** und zwar gegen ∞ .

► Bezogen auf die Messarbeit wird es immerhin nicht komplizierter:

11.3.3 Satz: Eine numerische Funktion $f : (\Omega, \mathcal{A}) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$

ist genau dann messbar, wenn $f^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathcal{B}$.

Beweis: Vgl. Küchler [23, Seite 92]

Der folgende wichtige Satz deutet an, wie wir später vorgehen wollen: Wir betrachten zunächst Treppenfunktionen und bilden dann einen Grenzübergang. Ob dabei *Eigenschaften* erhalten bleiben, **müssen** wir natürlich überprüfen bzw. durch passende Voraussetzungen sicherstellen.

11.3.4 Satz (Approximationssatz): Eine nicht-negative numerische Funktion $f : (\Omega, \mathcal{A}) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ ist genau dann messbar, wenn es eine Folge (u_n) nicht negativer Treppenfunktionen mit $(u_n) \nearrow f$ gibt.

Beweis: vgl. Küchler [23, Seite 100] oder Tappe [31, Seite 104]

12 Integration

12.0.1 Definition: Für eine nicht-negative Treppenfunktion

$$f = \sum_{j=1}^m \alpha_j \mathbb{1}_{A_j}$$

auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$ definieren wir das **Integral von f bezüglich μ** (über Ω):¹

$$\int f d\mu = \int f(x) d\mu(x) = \sum_{j=1}^m \alpha_j \mu(A_j)$$

12.0.2 Behauptung: Für alle $A \in \mathcal{A}$ gilt

$$\int \mathbb{1}_A d\mu = \mu(A).$$

Wenn $\mu = \mathbb{P}$ ein Wahrscheinlichkeitsmaß ist, dann hat die

¹Das Integralzeichen wurde gemäß Elstrodt [8, S. 135] von Leibniz eingeführt und stellt ein stilisiertes S (das für Summe steht) dar. In Elstrodt [8] findet sich dort auch der Hinweis, dass der Begriff Integral von Johann Bernoulli geprägt wurde und erstmals in einer Arbeit von Jacob Bernoulli erschien.

Behauptung die Form

$$\mathbb{E}(\mathbb{1}_A) = \int \mathbb{1}_A d\mathbb{P} = \mathbb{P}(A).$$

12.0.3 Satz: i.) [**Linearität**] Für alle $\alpha, \beta \in \mathbb{R}_{\geq 0}$ und nicht-negative Treppenfunktionen f, g gilt

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

ii.) [**Monotonie**] Gilt für nicht-negative Treppenfunktionen $f \leq g$, dann gilt

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

12.0.4 Definition: i.) Für eine messbare nicht-negative numerische Funktion $f : (\Omega, \mathcal{A}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ definieren wir das **Integral bezüglich μ** durch

$$\int f d\mu = \sup \left\{ \int u d\mu \mid \text{nicht-negative Treppenfunktion } u \leq f \right\}.$$

ii.) Es sei f eine messbare numerische Funktion und $f^+ = \max\{f, 0\}$, $f^- = \max\{-f, 0\}$. Gilt

$$\int f^+ d\mu < \infty, \int f^- d\mu < \infty,$$

dann heißt f **integrierbar** und wir schreiben $f \in \mathcal{L}^1(\mu)$.

Wir definieren dann das **Integral bezüglich μ**

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

12.0.5 Satz: Es sei $f : (\Omega, \mathcal{A}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ eine nicht-negative numerische Funktion. Dann gilt für jede Folge nicht-negativer **Treppenfunktionen** $(g_i)_{i \in \mathbb{N}}$ mit $g_i \nearrow f$

$$\int g_i d\mu \nearrow \int f d\mu.$$

Mit anderen Worten:

$$\lim \int g_i d\mu = \int f d\mu = \int \lim g_i d\mu.$$

12.0.6 Bemerkung: i.) Wenn wir ein Integral durch Grenzübergang berechnen/definieren wollen, dann können wir irgendeine Folge mit $g_i \nearrow f$ wählen.

ii.) Unter bestimmten Voraussetzungen können wir also \int und \lim vertauschen. Vgl. dazu auch den folgenden Satz sowie später den Satz über die majorisierte Konvergenz.

12.0.7 Satz von der monotonen Konvergenz: Es sei $(f_n)_{n \in \mathbb{N}}$ eine monoton wachsende Folge von messbaren **nicht-negativen** numerischen Funktionen. Dann gilt:

$$\int (\lim f_n) d\mu = \lim \int f_n d\mu.$$

Beweis: Vgl. Elstrodt [8, Seite 139]

12.0.8 Satz von der majorisierten Konvergenz: Es sei $f_n : \Omega \rightarrow \bar{\mathbb{R}}$ eine Folge von messbaren Abbildungen, die fast überall konvergiert. Ferner sei $g \in \mathcal{L}^1$ eine nicht-negative Abbildung mit $|f_n| \leq g$ fast überall. Dann gilt:

i.) Für alle $n \in \mathbb{N}$ ist $f_n \in \mathcal{L}^1$ und es gibt $f \in \mathcal{L}^1$ mit $f_n \rightarrow f$ fast überall.

ii.) Für jedes $f \in \mathcal{L}^1$ mit $f_n \rightarrow f$ fast überall gilt

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu = \int \left(\lim_{n \rightarrow \infty} f_n \right) d\mu$$

und

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

Beweis: Siehe für i.) Bauer [1, 95f].

12.0.9 Satz: Es seien $f, g \in \mathcal{L}^1$ und $\alpha, \beta \in \mathbb{R}$. Dann ist auch $\alpha f + \beta g \in \mathcal{L}^1$ und es gilt die **Linearität der Integration:**

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

12.0.10 Satz (Monotonie der Integration): Für $f, g \in$

\mathcal{L}^1 mit $f \leq g$ gilt

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

12.0.11 Satz (Dreiecksungleichung für die Integration): Für $f \in \mathcal{L}^1$ gilt

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

12.0.12 Satz: Es sei $f : \Omega \rightarrow \bar{\mathbb{R}}$ eine messbare Funktion. Dann sind die folgenden Aussagen äquivalent:

- i.) $f \in \mathcal{L}^1$ (d.h. f ist integrierbar).
- ii.) Es gibt eine nicht-negative integrierbare Funktion $g \in \mathcal{L}^1$ mit $|f| \leq g$. (so ein g heißt **Majorante**)
- iii.) $f^+, f^- \in \mathcal{L}^1$.
- iv.) Es gibt nicht-negative integrierbare Funktionen $g, h \in \mathcal{L}^1$ mit $f = g - h$.
- v.) $|f| \in \mathcal{L}^1$.

12.0.13 Folgerung: Wenn $\mu(\Omega) < \infty$, dann ist jede beschränkte messbare Abbildung integrierbar.

Beweis: Eine integrierbare Majorante ist $g = M$.

12.0.14 Bemerkung: Die Funktion

$$f(x) = \begin{cases} 1 & \text{falls } x \in \mathbb{Q} \\ 0 & \text{falls } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

heißt **Dirichlet Funktion** und ist berühmt (berüchtigt). Für die Einschränkung $g = f|_{[0,1]}$ von f auf das Intervall $[0, 1]$ gilt $f \in \mathcal{L}^1([0, 1], \mathcal{B}_{[0,1]}, \lambda_{[0,1]})$. Diese Funktion ist jedoch nicht Riemann-integrierbar.

12.0.15 Bemerkung: $f \in \mathcal{L}^1$ impliziert, dass $\mu(\{x : |f| = \infty\}) = 0$.

Parameterintegrale

12.0.16 Satz: Es sei $f : (a, b) \times \Omega \rightarrow \mathbb{R}$ eine Abbildung mit:

- i.) Für alle *Parameter* $t \in (a, b)$ ist $f(t, \cdot) \in \mathcal{L}^1$
- ii.) Für fast alle $x \in \Omega$ ist $f(\cdot, x) : (a, b) \rightarrow \mathbb{R}$ stetig in $t_0 \in (a, b)$.
- iii.) Es gibt ein Intervall $(c, d) \ni t_0$ und eine nicht-negative Majorante $g \in \mathcal{L}^1$, so dass für alle $t \in (c, d)$ fast überall $|f(t, \cdot)| \leq g$ gilt.

Dann ist $F : (a, b) \rightarrow \mathbb{R}$ mit

$$F(t) = \int f(t, x) d\mu(x), t \in (a, b)$$

in t_0 stetig. Also

$$\begin{aligned} \lim_{t \rightarrow t_0} F(t) &= \lim_{t \rightarrow t_0} \int f(t, x) d\mu(x) = \int \lim_{t \rightarrow t_0} f(t, x) d\mu(x) \\ &= \int f(t_0, x) d\mu(x) \end{aligned}$$

Beweis: Vgl. Bauer [1, Seite 101f] oder Henze [18, Seite 337]

12.0.17 Satz: Es sei $f : (a, b) \times \Omega \rightarrow \mathbb{R}$ eine Abbildung mit:

i.) Für alle $t \in (a, b)$ ist $f(t, \cdot) \in \mathcal{L}^1$

ii.) Die partielle Ableitung

$$\left(\frac{\partial}{\partial t} f(t, x) \right)_{|t_0}$$

existiert für alle $x \in \Omega$.

iii.) Es gibt ein Intervall $(c, d) \ni t_0$ und eine nicht-negative Majorante $g \in \mathcal{L}^1$, so dass für alle $t \in (c, d) \cap (a, b), t \neq t_0$

$$\left| \frac{f(t, x) - f(t_0, x)}{t - t_0} \right| \leq g(x) \text{ f.ü..}$$

Dann ist $F : (a, b) \rightarrow \mathbb{R}$

$$F(t) = \int f(t, x) d\mu(x)$$

in t_0 differenzierbar und $\frac{\partial}{\partial t} f(t, x)|_{t_0} \in \mathcal{L}^1$ integrierbar und es gilt

$$F'(t_0) = \frac{d}{dt} F(t)|_{t_0} = \frac{d}{dt} \int f(t, x) d\mu(x) = \int \frac{\partial}{\partial t} f(t, x) d\mu(x).$$

Beweis: Vgl. Bauer [1, Seite 102] oder Henze [18, Seite 337]

Riemann-Integral

12.0.18 Satz: i.) Eine beschränkte Funktion $f : [a, b] \rightarrow \mathbb{R}$ ist genau dann **Riemann-integrierbar**, wenn die Menge der Unstetigkeitsstellen eine Lebesgue-Nullmenge bilden.

Eine Menge $N \subset \mathbb{R}$ heißt **Lebesgue-Nullmenge**, falls $\lambda(N) = 0$.

ii.) Wenn eine beschränkte Funktion $f : [a, b] \rightarrow \mathbb{R}$ Riemann-integrierbar ist, dann stimmen Riemann-Integral und Lebesgue-Integral überein.

Beweis: Vgl. Elstrodt [8, S. 166]

12.0.19 Satz: Es sei $I \subset \mathbb{R}$ ein Intervall und $f : I \rightarrow \mathbb{R}$ auf

jedem kompakten Intervall in I Riemann-integrierbar. Dann gilt:

- i.) f ist genau dann Lebesgue-integrierbar, wenn $|f|$ uneigentlich Riemann-integrierbar über I ist.
- ii.) Wenn $|f|$ uneigentlich Riemann-integrierbar über I ist, dann stimmen das Lebesgue-Integral und das uneigentliche Riemann-Integral überein.

Beweis: Vgl. Elstrodt [8, S. 168]

Wir können trotz dieser scheinbaren Überlegenheit nicht auf das Riemann-Integral verzichten.

12.0.20 Bemerkung: Das uneigentliche Riemann-Integral

$$R\text{-}\int_0^{\infty} \frac{\sin(x)}{x} dx$$

existiert (Elstrodt [8, S. 168f]). Aber $|\frac{\sin(x)}{x}|$ ist nicht über $(0, \infty)$ uneigentlich Riemann-integrierbar. Also ist $\frac{\sin(x)}{x}$ auch nicht über $(0, \infty)$ Lebesgue-integrierbar. Ausgerechnet das Integral über $\frac{\sin(x)}{x}$ benötigt man für die Inversionsformel von Levy für charakteristische Funktionen

12.0.21 Bemerkung: Das vorliegende Kapitel liefert nur einen CRASH-Kurs der Integrationstheorie. Details finden Sie in der **Maß- und Integrationsbibel** Elstrodt [8]. Für

unsere Zwecke sind insbesondere noch die Bücher von Küchler [23] und Henze [18] sehr empfehlenswert.

13 Bedingte Erwartung – bedingte Wahrscheinlichkeit

Das Thema dieses Kapitels ist relativ schwer. Meine favorisierten Quellen sind die jeweiligen Kapitel aus Billingsley [2], Küchler [23], Williams [33] und die immer gute Quelle Henze [18].

13.1 Bedingte Erwartungen gegeben B

Bedingte Erwartungen gegeben B

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X \cdot \mathbb{1}_B)}{\mathbb{P}(B)} \quad (13.1)$$

Bedingte Erwartungen gegeben B

Y diskret

$$\mathbb{E}(Y|B) = \sum_y y\mathbb{P}(y|B) \quad (13.2)$$

Y mit Dichte

$$\mathbb{E}(Y|B) = \int yf(y|B)dy \quad (13.3)$$

$$f(y|B) = \frac{\mathbb{P}(B|Y=y)f(y)}{\mathbb{P}(B)} \quad (13.4)$$

Bedingte Erwartungen – Fallunterscheidungsformel

$$\mathbb{E}(Y) = \sum_{i=1}^n \mathbb{E}(Y|B_i)\mathbb{P}(B_i) \quad (13.5)$$

Varianzzerlegung

$$\mathbb{V}(Y) = \mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X)) \quad (13.6)$$

13.2 Radon-Nikodym

Zunächst führen wir den für den späteren Gebrauch (insbesondere in Beweisen) wichtigen Satz von Radon und Nikodym ein.

13.2.1 Definition: Es seien μ, ν zwei σ -finite Maße auf dem Maßraum (Ω, \mathcal{F}) . μ heißt **absolut stetig** bezüglich ν , falls für alle $B \in \mathcal{F}$ mit $\nu(B) = 0$ auch $\mu(B) = 0$ gilt. In diesem Fall schreiben wir $\mu \ll \nu$.

Gilt $\mu \ll \nu$ und $\nu \ll \mu$, dann heißen μ und ν **äquivalent**.

Die folgende Proposition erklärt, warum man von *Stetigkeit* spricht.

13.2.2 Proposition: Es seien μ, ν zwei **endliche** Maße¹ auf dem Maßraum (Ω, \mathcal{F}) . Dann ist μ genau dann absolut stetig bezüglich ν , falls es für alle $\epsilon > 0$ ein $\delta > 0$ gibt, so dass für alle $A \in \mathcal{F}$ mit $\nu(A) < \delta$ ist $\mu(A) < \epsilon$.

13.2.3 Satz (Radon-Nikodym): Es seien μ, ν zwei σ -finite Maße auf dem Maßraum (Ω, \mathcal{F}) .

i.) Wenn $\mu \ll \nu$, dann gibt es eine nicht-negative numerische Funktion $f : \Omega \rightarrow \bar{\mathbb{R}}_{\geq 0}$ mit

a.) f ist messbar.

b.) Für alle $B \in \mathcal{F}$ gilt

$$\mu(B) = \int_B f d\nu = \int \mathbb{1}_B f d\nu.$$

¹Wahrscheinlichkeitsmaße sind endlich!

ii.) f ist ν -fast überall eindeutig.

13.2.4 Definition: Die Funktion f aus dem Satz von Radon-Nikodym heißt **Radon-Nikodym-Ableitung** von μ bezüglich ν . Man schreibt:

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

13.3 Bedingte Erwartung bezüglich \mathcal{F}

Im folgenden benötigen wir regelmäßig mehrere σ -Algebren. Wir beachten dabei folgende Konventionen: \mathcal{A} bezeichnet die *maximale* Information des Experiments $(\Omega, \mathcal{A}, \mathbb{P})$; \mathcal{A} wie Alles.

Für Teilinformationen verwenden wir Sub- σ -Algebren, für die wir die typischerweise die Notation \mathcal{F} verwenden. Was so eine Sub- σ -Algebra erfasst, wollen wir an dem abgebildeten **abgeklebten Würfel** erklären.

Die ungeraden Zahlen sind rot abgeklebt und die geraden Zahlen blau. Man kann die Zahlen unter dem Klebeband erkennen, wir sollen/wollen uns aber vorstellen, dass wir die Zahlen unter dem Klebeband nicht erkennen



können. Wir wissen lediglich welche Zahlen wie abgeklebt

sind.

- i.) Bestimmen Sie den Erwartungswert, wenn Sie rot sehen!
- ii.) Wie hoch ist die Wahrscheinlichkeit für $\{1, 3\}$, wenn Sie rot sehen?

13.3.1 Bemerkung: Wir betrachten eine integrierbare Zufallsvariable $X \in \mathcal{L}^1$. Als Lagemaß enthält der Erwartungswert $\mathbb{E}(X)$ Information über X . Gemäß des GgZ verteilen sich bei unabhängiger identischer Wiederholung die Realisierungen **durchschnittlich** um $\mathbb{E}(X)$. Diese Information über die Verteilung ist aber grob; nur eine einzelne Zahl. Oft haben wir **Zusatzinformation/Teilinformation/Fallunterscheidungen** über die Werte von X .

Wir betrachten beispielsweise die Augenzahl X eines fairen Würfels. Dann ist bekanntlich $\mathbb{E}(X) = 3.5$. Angenommen die geraden Zahlen auf dem Würfel werden blau und die ungeraden Zahlen rot angeklebt. Dann können wir zwei *Werte* ermitteln; je nachdem, welche Farbe wir sehen (je nach Ergebnis ω) :

$$\mathbb{E}(X|\text{Farbinfo})(\omega) = \begin{cases} 3 & \text{falls } \omega \in \{1, 3, 5\} \quad [\text{wir sehen rot}] \\ 4 & \text{falls } \omega \in \{2, 4, 6\} \quad [\text{wir sehen blau}] \end{cases}$$

Wir bemerken, dass **die bedingte Erwartung eine Zufallsvariable** ist, den ihr Wert hängt von der zufälligen Zie-

hung ω ab!

Diese Perspektive wollen wir formalisieren. Es ist $\Omega = \{1, 2, 3, \dots, 6\}$ und $\mathcal{A} = \mathcal{P}(\Omega)$. Wir wissen schon, dass man die Teilinformation (die Farbinformation bei abgeklebten Würfel) durch die folgende **Sub- σ -Algebra** erfassen kann:

$$\mathcal{F} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}.$$

Wenn die Teilinformation \mathcal{F} vorliegt, dann können wir für die Mengen in \mathcal{F} *entscheiden*, ob sie eingetreten sind oder nicht.² Wenn wir beispielsweise *rot* sehen, dann wissen wir, dass $\{1, 3, 5\}$ eingetreten ist. Wir wissen auch, dass $\{2, 4, 6\}$ nicht eingetreten ist. Wir können jedoch nicht entscheiden, ob 3 eingetreten bzw. nicht eingetreten ist. So genau/fein ist die Teilinformation \mathcal{F} nicht.

Angenommen Y ist eine Zufallsvariable, die **nur die Teilinformation** in \mathcal{F} **verwendet**. Mathematisch fordern wir dann: Y ist \mathcal{F} -messbar.

13.3.2 Satz (Kolmogorov): Es sei $X : \Omega \rightarrow \mathbb{R}$ eine integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Es sei \mathcal{F} eine Sub- σ -Algebra von \mathcal{A} . Dann gibt es eine \mathbb{P} -f.s. eindeutig bestimmte integrierbare Zufallsvariable Y mit:

²Wir beachten aber den Hinweis von Billingsley (1995, S. 57 ff), dass wir streng zwischen mathematischen Aussagen und deren Interpretation unterscheiden müssen.

- a.) Y ist \mathcal{F} -messbar.
- b.) Für alle $F \in \mathcal{F}$ gilt

$$\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X).$$

Beweis: Henze [18, Seite 171]

13.3.3 Definition: Eine Zufallsvariable Y wie aus dem vorhergehenden Satz heißt **bedingte Erwartung**³ von X bezüglich \mathcal{F} . Für die bedingte Erwartung schreiben wir $\mathbb{E}(X|\mathcal{F})$.

13.3.4 Bemerkung: Was bedeuten die beiden Bedingungen in der Definition der bedingten Erwartung?

Die Bedingung a.) bedeutet, dass Y sich nur auf die Information in \mathcal{F} bezieht.

Die Bedingung b.) bedeutet, dass sich $Y = \mathbb{E}(X|\mathcal{F})$ auf **allen Ereignissen** $F \in \mathcal{F}$ durchschnittlich wie X verhält:

$$\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X).$$

► Diese Aussage wird deutlicher durch die nächste Proposition.

³Vorsicht: bedingte Erwartung und nicht bedingter Erwartungswert. Die gleichen sich, sind aber nicht dasselbe. Details folgen.

13.3.5 Proposition: Angenommen \mathcal{F} wird von einer Zerlegung $F_i, i = 1, \dots, n, \Omega = \bigsqcup F_i$ erzeugt; wir unterstellen zudem $\mathbb{P}(F_i) > 0$. $\mathbb{E}(X|\mathcal{F})$ ist \mathcal{F} -messbar und deshalb auf den Mengen F_i konstant. Es gilt in der Tat

$$\mathbb{E}(X|\mathcal{F})(\omega) = \frac{\mathbb{E}(\mathbb{1}_{F_i}X)}{\mathbb{P}(F_i)} \text{ für } \omega \in F_i.$$

Wenn X zudem diskret ist, dann gilt

$$\mathbb{E}(X|\mathcal{F})(\omega) = \sum_{x \in \text{Bild}(X)} x \cdot \mathbb{P}(X = x | F_i) \text{ für } \omega \in F_i.$$

Beweis: Da Y \mathcal{F} -messbar ist, ist Y konstant auf F_i ; sagen wir $= \alpha_i$. Gemäß Satz von Kolmogorov gilt

$$\mathbb{E}(\mathbb{1}_{F_i}Y) = \mathbb{E}(\mathbb{1}_{F_i}X).$$

Es ist

$$\mathbb{E}(\mathbb{1}_{F_i}Y) = \alpha_i \mathbb{P}(F_i)$$

$$\mathbb{E}(X|\mathcal{F})(\omega) = \alpha_i = \frac{\mathbb{E}(\mathbb{1}_{F_i}X)}{\mathbb{P}(F_i)}, \omega \in F_i$$

Wenn X diskret ist, dann $\text{Bild}(X) = \{x_1, x_2, \dots\}$ (x_i unterschiedlich).

Wir beobachten

$$X = \sum_{k=1}^{\infty} \mathbb{1}_{\{X=x_k\}} x_k$$

und weiterhin

$$\begin{aligned} \mathbb{E}(\mathbb{1}_{F_i} X) &= \mathbb{E}\left(\mathbb{1}_{F_i} \sum_{k=1}^{\infty} \mathbb{1}_{\{X=x_k\}} x_k\right) \\ &= \mathbb{E}\left(\sum_{k=1}^{\infty} x_k \mathbb{1}_{F_i} \mathbb{1}_{\{X=x_k\}}\right) \\ &= \mathbb{E}\left(\sum_{k=1}^{\infty} x_k \mathbb{1}_{F_i \cap \{X=x_k\}}\right) \\ &= \sum_{k=1}^{\infty} x_k \mathbb{E}(\mathbb{1}_{F_i \cap \{X=x_k\}}) \\ &= \sum_{k=1}^{\infty} x_k \mathbb{P}(F_i \cap \{X = x_k\}) \\ &= \sum_{k=1}^{\infty} x_k \mathbb{P}(F_i \cap \{X = x_k\}) \\ &= \sum_{k=1}^{\infty} x_k \mathbb{P}(\{X = x_k\} | F_i) \mathbb{P}(F_i) \end{aligned}$$

Also

$$\mathbb{E}(X|\mathcal{F})(\omega) = \alpha_i = \sum_{k=1}^{\infty} x_k \mathbb{P}(\{X = x_k\} | F_i), \omega \in F_i$$

13.3.6 Bemerkung: Wir müssen den Unterschied zwischen

$$\mathbb{E}(X|\mathcal{F})(\omega), \omega \in F_i$$

und

$$\mathbb{E}(\mathbb{1}_{F_i}X)$$

beachten. Der Wert $\mathbb{E}(X|\mathcal{F})(\omega), \omega \in F_i$ berücksichtigt die Wahrscheinlichkeit im Nenner, dass $\omega \in F_i$ (also $\mathbb{P}(F_i)$). Angenommen eine 1 wurde gewürfelt, aber die ungeraden Zahlen sind rot abgeklebt. Wenn wir mit diesem Wissen die bedingte Erwartung bestimmen, dann rechnen wir $\frac{1}{3}1 + \frac{1}{3}3 + \frac{1}{3}5$. Die Wahrscheinlichkeit $\frac{1}{3}$ entspricht nicht der Wahrscheinlichkeit für eine 1 [die wäre $\frac{1}{6}$]. Die Wahrscheinlichkeit $\frac{1}{3} = \frac{1/6}{1/2}$ berücksichtigt im Nenner die Wahrscheinlichkeit für eine ungerade Zahl $\mathbb{P}(\{1, 3, 5\}) = \frac{1}{2}$, denn wir wissen, dass $\{1, 3, 5\}$ eingetreten ist, denn wir sehen rot. Der Wert $\mathbb{E}(\mathbb{1}_{F_i}X)$ verwendet die ursprünglichen (a-priori) Wahrscheinlichkeiten \mathbb{P} und *nullt* alle Werte mit Urbildern außerhalb F_i . Wir erhalten also für $\mathbb{E}(\mathbb{1}_{F_i}X)$ **fast** so etwas wie einen *bedingten* Erwartungswert; die ungenullten Wahrscheinlichkeiten geben untereinander die richtigen Wahrscheinlichkeits-Proportionen an, addieren sich aber nicht zu Eins. Wenn wir diese *a-priori* Wahrscheinlichkeiten durch $\mathbb{P}(F_i) = \frac{1}{2}$ teilen, dann erhalten wir die bedingte Erwartung $\mathbb{E}(X|\mathcal{F})(\omega), \omega \in F_i$ berechnet mit den bedingten Wahrscheinlichkeiten.

Der Vorteil von $\mathbb{E}(\mathbb{1}_F X)$ besteht darin, dass wir die Voraussetzung $\mathbb{P}(F) > 0$ nicht benötigen. Für die allgemeine Definition ist das komfortabler.

► Der folgende Satz fasst die wichtigsten Eigenschaften der bedingten Erwartung zusammen. Die Formel sind SEHR wichtig für das Verständnis der bedingten Erwartungen und für das Rechnen mit bedingten Erwartungen. Die Eigenschaft f.) nennt man manchmal das **Gesetz von der iterierten Erwartung** oder die **Turmeigenschaft** der bedingten Erwartung. Die Turmeigenschaft und die Regel bei Unabhängigkeit g.) habe ich beim praktischen Rechnen mehr als 1000 mal angewendet!

Die Eigenschaft i.) zeigt die große Bedeutung der bedingten Erwartung für die Anwendung. Die bedingte Erwartung ist die **beste Prognose**, wenn der Prognostiker über die Information in \mathcal{F} verfügt [und der **mittlere quadratische Fehler** das Qualitätskriterium ist].

13.3.7 Satz: Es seien X, X_1, X_2 integrierbare Zufallsvariablen und $\mathcal{F} \subset \mathcal{A}$ eine Sub- σ -Algebra von \mathcal{A} . Dann gilt

a.) Für alle $\alpha_1, \alpha_2 \in \mathbb{R}$ gilt \mathbb{P} -f.ü.

$$\mathbb{E}(\alpha_1 X_1 + \alpha_2 X_2 | \mathcal{F}) = \alpha_1 \mathbb{E}(X_1 | \mathcal{F}) + \alpha_2 \mathbb{E}(X_2 | \mathcal{F})$$

b.) Ist $X_1 \leq X_2$, so gilt \mathbb{P} -f.ü.

$$\mathbb{E}(X_1 | \mathcal{F}) \leq \mathbb{E}(X_2 | \mathcal{F})$$

c.) Ist $X(\omega) = a$, \mathbb{P} -f.ü., dann ist \mathbb{P} -f.ü.

$$\mathbb{E}(X|\mathcal{F}) = a$$

d.) Ist $\mathcal{F} = \{\emptyset, \Omega\}$, dann ist \mathbb{P} -f.ü.

$$\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X)$$

e.) Ist $\mathcal{F} = \mathcal{A}$, dann ist \mathbb{P} -f.ü.

$$\mathbb{E}(X|\mathcal{F}) = X$$

f.) Ist $\mathcal{G} \subset \mathcal{F}$ eine Sub- σ -Algebra von \mathcal{F} , dann gilt \mathbb{P} -f.ü.

$$\mathbb{E}(\mathbb{E}(X|\mathcal{F})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G}) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{F})$$

g.) Ist V eine \mathcal{F} -messbare integrierbare Zufallsvariable, dann ist \mathbb{P} -f.ü.

$$\mathbb{E}(VX|\mathcal{F}) = V\mathbb{E}(X|\mathcal{F})$$

h.) Sind $\sigma(X)$ und \mathcal{F} unabhängig, dann gilt \mathbb{P} -f.ü.

$$\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X)$$

i.) Ist $X \in \mathcal{L}^2$, dann ist $\mathbb{E}(X|\mathcal{F}) \in \mathcal{L}^2$ und es gilt

$$\mathbb{E}[(X - \mathbb{E}(X|\mathcal{F}))^2] = \min_Y \{\mathbb{E}[(X - Y)^2] \mid Y \in \mathcal{L}^2, Y \in \mathcal{F}\}$$

13.3.8 Definition: Es sei X eine integrierbare Zufallsvariable und $Z : \Omega \rightarrow \mathbb{R}^n$ ein Zufallsvektor. Es sei $\mathcal{F} = Z^{-1}(\mathcal{B}^n)$, d.h. \mathcal{F} ist die von Z erzeugte σ -Algebra. . Dann heißt $\mathbb{E}(X|Z) = \mathbb{E}(X|\mathcal{F})$ die **bedingte Erwartung** von X gegeben Z .

► In der vorhergehenden Definition stammt die Teilinformation also aus einem Zufallsvektor, aber indirekt aus $\sigma(Z)$ und nicht direkt aus dem Wert den Z angenommen hat. Die naheliegende Idee: Die Teilinformation kann direkt **dem angenommenen Wert des Zufallsvektor Z** entnommen werden (und nicht *indirekt* der σ -Algebra $\mathcal{F} = Z^{-1}(\mathcal{B}^n)$). Das stimmt, wie das folgende sogenannte Faktorisierungslemma zeigt. Dieser Zugang ist sogar *natürlicher*.

13.3.9 Satz (Faktorisierungslemma für die bedingte Erwartung): Es sei X eine integrierbare Zufallsvariable und $Z : \Omega \rightarrow \mathbb{R}^n$ ein Zufallsvektor. Es sei $\mathcal{F} = \sigma(Z) = Z^{-1}(\mathcal{B}^n)$ die durch Z erzeugte σ -Algebra. Es gibt eine \mathcal{B}^n - \mathcal{B} messbare Abbildung $h : \mathbb{R}^n \rightarrow \mathbb{R}$ mit

$$\mathbb{E}(X|Z)(\omega) = h(Z(\omega)) \quad \mathbb{P}\text{-f.ü. .}$$

z $\mathcal{F} = \sigma(Z)$ dann $\mathbb{E}(X|\mathcal{F})(\omega) = \mathbb{E}(X|Z = Z(\omega))$

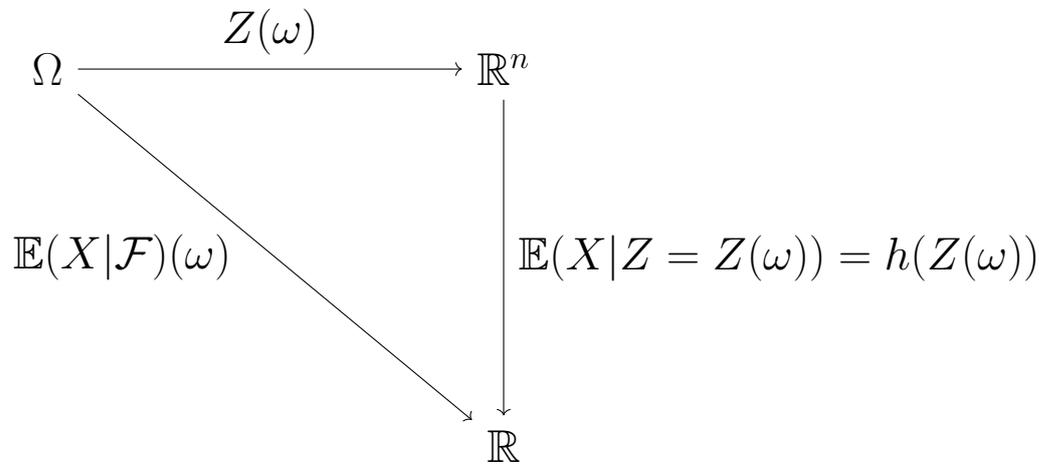


Abbildung 13.1: Faktorisierungslemma für bedingte Erwartungen bzw. bedingter Erwartungswert

► Wir wollen Erwartungen bezüglich X ermitteln. Dazu können wir die Information beachten, die von Z erzeugt wird. Wenn ω eingetreten ist, dann ist die Erwartung $\mathbb{E}(X|Z)(\omega)$ (und damit besser informiert als $\mathbb{E}(X)$). Alternativ können wir den Wert $Z(\omega)$ verwenden und $h(Z(\omega))$ ermitteln. h liefert uns also die bedingte Erwartung als Funktion des Wertes $Z(\omega)$, den die Bedingung annimmt. Es fügt sich also alles schön zusammen. Wir definieren dementsprechend den bedingten Erwartungswert wie folgt.

13.3.10 Definition: Die im vorhergehenden Satz genannte Abbildung $h : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt der **bedingte Erwartungs-**

wert von X gegeben $Z = z$. Wir schreiben dafür

$$h(z) = \mathbb{E}(X|Z = z), z \in \mathbb{R}^n.$$

$z \mapsto \mathbb{E}(X|Z = z)$ ist also eine Abbildung von \mathbb{R}^n nach \mathbb{R} . Das Argument ist z (hier blau markiert).

13.3.11 Bemerkung: i.) Es gilt gemäß Definition \mathbb{P} -f.ü.

$$\mathbb{E}(X|\sigma(Z))(\omega) = \mathbb{E}(X|Z)(\omega) = \mathbb{E}(X|Z = Z(\omega)).$$

ii.) Wir müssen trotzdem⁴ die **bedingten Erwartungen** $\mathbb{E}(X|\mathcal{F})$ von den **bedingten Erwartungswerten** $\mathbb{E}(X|Z = z)$ unterscheiden. In beiden Fällen werden Erwartungen von X gebildet und es kommt das *gleiche* heraus. Aber:

- Der Definitionsbereich von $\mathbb{E}(X|\mathcal{F})$ ist Ω . Was erwarten wir für X , wenn wir den Information $\sigma(Z)$ haben und ω eingetreten ist.
- Der Definitionsbereich von $\mathbb{E}(X|Z = z)$ ist \mathbb{R}^n . Was erwarten wir für X , wenn wir den Information haben, dass z eingetreten ist.
- Die gleiche Antwort erhält man, wenn $z = Z(\omega)$ ist.

► Zwei für die angewandte besonders wichtige Fälle – die

⁴Trotz des Gleichheitszeichens.

man zudem auch elementarer hätte behandeln können – sollen in den anschließenden Propositionen behandelt werden. (1) Das Bild von Z ist endlich und (2) X und Z haben eine gemeinsame Dichte.

13.3.12 Proposition: Es seien X und Z reellwertige Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ und $\text{Bild}(Z) = \{z_1, \dots, z_n\}$ mit $\mathbb{P}(Z = z_i) > 0$. Die von Z erzeugte σ -Algebra wird durch die Zerlegung $\{Z = z_i\}$ erzeugt. Ferner, eine Zufallsvariable Y ist genau dann bezüglich $\sigma(Z)$ messbar, wenn Y auf den Mengen $\{Z = z_i\}$ konstant ist.

Es sei für $z_i \in \text{Bild}(Z)$

$$\mathbb{P}(X \in B | Z = z_i) = \frac{\mathbb{P}(X \in B \wedge Z = z_i)}{\mathbb{P}(Z = z_i)}$$

Dann gilt für $z_i \in \text{Bild}(Z)$

$$\mathbb{E}(X | Z = z_i) = \int X d\mathbb{P}(X \in dx | Z = z_i).$$

Ist auch X diskret, d.h. gilt $\text{Bild}(X) = \{x_1, \dots, x_k\}$, dann ist

$$\begin{aligned} \mathbb{E}(X | Z = z_i) &= \sum x_i \mathbb{P}(X = x_i | Z = z_i) \\ &= \sum x_i \frac{\mathbb{P}(X = x_i \wedge Z = z_i)}{\mathbb{P}(Z = z_i)} \end{aligned}$$

13.3.13 Beispiel: $\Omega = \{\clubsuit, \spadesuit, \diamondsuit, \heartsuit\}$ mit den Wahrscheinlichkeiten (in der entsprechenden Reihenfolge) $\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}$.

Die Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ habe die Werte (wieder in der Reihenfolge) 3, 4, 5, 6 und die die Zufallsvariable $Z : \Omega \rightarrow \mathbb{R}$ habe die Werte (wieder in der Reihenfolge) 10, 10, 11, 11.

Bestimmen Sie:

a.) $\mathbb{E}(X|\sigma(Z))$ und

b.) $\mathbb{E}(X|Z = z)$.

13.3.14 Proposition: Es seien X und Z zwei (reellwertige) Zufallsvariablen mit gemeinsamer Dichte f . Es gilt also für $B_1, B_2 \in \mathcal{B}$:

$$\mathbb{P}((X, Z) \in B_1 \times B_2) = \int_{B_2} \int_{B_1} f(x, z) dx dz.$$

Wir bezeichnen die Randdichten mit f_X bzw. f_Z und wir definieren

$$f(x|z) = \begin{cases} \frac{f(x,z)}{f_Z(z)} & \text{falls } f_Z(z) \neq 0 \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt für den bedingten Erwartungswert von X gegebenen $Z = z$

$$\mathbb{E}(X|Z = z) = \int_{\mathbb{R}} x f(x|z) dx.$$

Beweis: Es sei

$$g(z) = \int_{\mathbb{R}} x f(x|z) dx.$$

Wir müssen beweisen, dass für alle $B \in \mathcal{B}$

$$\mathbb{E}(\mathbb{1}_B(Z)g(Z)) = \mathbb{E}(\mathbb{1}_B(Z)X)$$

gilt. Es gilt in der Tat

$$\begin{aligned} \mathbb{E}(\mathbb{1}_B(Z)g(Z)) &= \int \mathbb{1}_B(z) \left(\int_{\mathbb{R}} x f(x|z) dx \right) f_Z(z) dz \\ &= \int \mathbb{1}_B(z) \left(\int x \frac{f(x, z)}{f_Z(x)} dx \right) f_Z(z) dz \\ &= \int \int \mathbb{1}_B(z) x f(x, z) dx dz \\ &= \mathbb{E}(\mathbb{1}_B(Z)X) \end{aligned}$$

13.3.15 Beispiel: ...

$$f(x, y) = \frac{1}{2} \mathbb{1}_D(x, y)$$

13.4 Bedingte Wahrscheinlichkeit und bedingte Verteilung

13.4.1 Bemerkung: Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Es seien $A, B \in \mathcal{A}$ Ereignisse, wobei $0 < \mathbb{P}(B) < 1$. Wir interessieren uns für die Wahrscheinlichkeit von A .

Ohne *Information* ist die gleich $\mathbb{P}(A)$. Wenn A und B nicht unabhängig sind, dann können wir die Wahrscheinlichkeitseinschätzung für das Eintreten von A anpassen, wenn uns Information über B bzw. B^c vorliegt:

$$f(\omega) = \begin{cases} \mathbb{P}(A|B) & \text{falls } \omega \in B \\ \mathbb{P}(A|B^c) & \text{falls } \omega \in B^c \end{cases}$$

Wir betrachten wieder den abgeklebten Würfel (rot ungerade, blau gerade) und das Ereignis *mindestens eine vier* gewürfelt zu haben; $A = \{4, 5, 6\}$. Ohne Information ist $\mathbb{P}(A) = \frac{1}{2}$. Wenn wir rot sehen, dann ist $\mathbb{P}(A|\text{rot}) = \frac{1}{3}$ und $\mathbb{P}(A|\text{blau}) = \frac{2}{3}$

Beachte: $f(\omega)$ ist eine Zufallsvariable, denn sie hängt vom Ergebnis ab (über das wir aber nicht volle Kenntnis haben). $f(\omega)$ beantwortet die Frage: Welche Erwartung haben wir, wenn ω eingetreten ist, wir aber nur \mathcal{F} (die Farbe) wissen.

Wir wollen die Situation formalisieren. Die Vorinformation wird wieder durch eine σ -Algebra repräsentiert, also $\mathcal{F} = \{\emptyset, \Omega, B, B^c\}$. Wir schreiben $f = \mathbb{P}(A|\mathcal{F})$. f ist eine Zufallsvariable, denn je nach Ergebnis ergibt sich eine andere bedingte Wahrscheinlichkeit

f hat zwei Eigenschaften

- a.) f ist \mathcal{F} -messbar

b.) Für alle $F \in \mathcal{F}$ gilt

$$\int_F \mathbb{P}(A|\mathcal{F})d\mathbb{P} = \mathbb{P}(A \cap F)$$

Es ist für $F = \{1, 3, 5\}$

$$\mathbb{P}(A \cap F) = \frac{1}{6}$$

und mit $x_F = \mathbb{P}(A|\mathcal{F})(\{1\}) = \mathbb{P}(A|\mathcal{F})(\{3\}) = \mathbb{P}(A|\mathcal{F})(\{5\})$
[diese Werte müssen wegen der geforderten \mathcal{F} -Messbarkeit gleich sein]

$$\begin{aligned} \int_F \mathbb{P}(A|\mathcal{F})d\mathbb{P} &= \mathbb{E}(\mathbb{1}_F \mathbb{P}(A|\mathcal{F})) \\ &= \mathbb{1}_F(1)\mathbb{P}(\{1\})x_F + \mathbb{1}_F(2)\mathbb{P}(\{2\})x_f + \dots + \mathbb{1}_F(6)\mathbb{P}(\{6\})x_f \\ &= \mathbb{P}(\{1\})x_F + \mathbb{P}(\{3\})x_F + \mathbb{P}(\{5\})x_F \\ &= (\mathbb{P}(\{1\}) + \mathbb{P}(\{3\}) + \mathbb{P}(\{5\}))x_F = \frac{1}{2}x_F \end{aligned}$$

Also gilt

$$x_F = \frac{1/6}{1/2} = \frac{1}{3} = \mathbb{P}(A|F)$$

13.4.2 Satz: Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{F} \subset \mathcal{A}$ eine Sub- σ -Algebra von \mathcal{A} und $A \in \mathcal{A}$ ein Ereignis. Dann gilt

i.) $\nu(F) = \mathbb{P}(A \cap F)$, $F \in \mathcal{F}$ ist ein endliches WM auf \mathcal{F} .

ii.) $\nu \ll \mathbb{P}$.

iii.) Es gibt eine \mathcal{F} -messbare \mathbb{P} -integrierbare Abbildung f mit

$$\mathbb{P}(A \cap F) = \nu(F) = \int_F f d\mathbb{P}, F \in \mathcal{F}$$

13.4.3 Definition: Die Funktion f aus dem vorhergehenden Satz heißt die **bedingte Wahrscheinlichkeit** von A bezüglich \mathcal{F} . Sie wird mit

$$f = \mathbb{P}(A|\mathcal{F})$$

13.4.4 Satz: Die folgenden beiden Bedingungen charakterisieren die bedingte Wahrscheinlichkeit:

a.) $\mathbb{P}(A|\mathcal{F})$ ist \mathcal{F} -messbar

b.) Für alle $F \in \mathcal{F}$ gilt

$$\int_F \mathbb{P}(A|\mathcal{F}) d\mathbb{P} = \mathbb{P}(A \cap F)$$

13.4.5 Satz:

$$\mathbb{P}(A|\mathcal{F}) = \mathbb{E}(1_A|\mathcal{F})$$

13.4.6 Bemerkung: Manchmal ist die folgende Formel *leich-*

ter

$$\int_F \mathbb{P}(A|\mathcal{F})d\mathbb{P} = \mathbb{E}(\mathbb{1}_F\mathbb{P}(A|\mathcal{F}))$$

13.4.7 Bemerkung: Für die Interpretation der Bedingung b.) beachten wir für den diskreten Fall $\Omega = \{\omega_1, \dots, \omega_k\}$ einerseits

$$\begin{aligned} \int_F \mathbb{P}(A|\mathcal{F})d\mathbb{P} &= \int_{\Omega} \mathbb{1}_F \mathbb{P}(A|\mathcal{F})d\mathbb{P} \\ &= \sum_{i=1}^k \mathbb{1}_F(\omega_k)\mathbb{P}(A|\mathcal{F})(\omega_k)\mathbb{P}(\omega_k) \end{aligned}$$

und andererseits die diskrete Cocktail-Formel

$$\mathbb{P}(A) = \mathbb{P}(A|\{\omega_1\})\mathbb{P}(\{\omega_1\}) + \dots + \mathbb{P}(A|\{\omega_k\})\mathbb{P}(\{\omega_k\})$$

Also lautet die Bedingungen b.)

$$\mathbb{P}(A \cap F) = \int_{\Omega} \mathbb{1}_F \mathbb{P}(A|\mathcal{F})d\mathbb{P}.$$

Bei dieser Bedingung wird also das **Prinzips der Cocktail-Formel** auf alle Ereignisse $A \cap F, F \in \mathcal{F}$ angewendet.

Wir betrachten nochmal $\mathcal{F} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$ und $A = \{4, 5, 6\}$. Dann ist (diesmal) für $\tilde{F} = \{2, 4, 6\}$

$$\mathbb{P}(A \cap \tilde{F}) = \frac{2}{6}$$

und mit $x_{\tilde{F}} = \mathbb{P}(A|\mathcal{F})(\{2\}) = \mathbb{P}(A|\mathcal{F})(\{4\}) = \mathbb{P}(A|\mathcal{F})(\{6\})$
[diese Werte müssen wegen der geforderten \mathcal{F} -Messbarkeit
gleich sein]

$$\begin{aligned}\int_{\Omega} \mathbb{1}_{\tilde{F}} \mathbb{P}(A|\mathcal{F}) d\mathbb{P} &= \mathbb{P}(\{2\}) x_{\tilde{F}} + \mathbb{P}(\{4\}) x_{\tilde{F}} + \mathbb{P}(\{6\}) x_{\tilde{F}} \\ &= (\mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\})) x_{\tilde{F}} = \frac{1}{2} x_{\tilde{F}}.\end{aligned}$$

Also

$$\frac{2}{6} = \frac{1}{2} x_{\tilde{F}}$$

bzw.

$$x_{\tilde{F}} = \frac{2/6}{1/2} = \frac{2}{3} = \mathbb{P}(A|\tilde{F}).$$

13.4.8 Satz: Es gilt

a.) Es ist \mathbb{P} -f.ü.

$$\mathbb{P}(\emptyset|\mathcal{F}) = 0$$

b.) Es ist \mathbb{P} -f.ü.

$$\mathbb{P}(\Omega|\mathcal{F}) = 1$$

c.) Für alle $A \in \mathcal{A}$ ist \mathbb{P} -f.ü.

$$0 \leq \mathbb{P}(A|\mathcal{F}) \leq 1$$

d.) Für $(A_i)_{i \in \mathbb{N}}$ mit $A_i \cap A_j = \emptyset, i \neq j$ gilt \mathbb{P} -f.ü.

$$\mathbb{P} \left(\biguplus_{i=1}^{\infty} A_i | \mathcal{F} \right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i | \mathcal{F})$$

13.4.9 Bemerkung: i.) Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{F} \subset \mathcal{A}$ eine Sub- σ -Algebra von \mathcal{A} , $\omega \in \Omega$. Dann ist

$$A \mapsto \mathbb{P}(A | \mathcal{F})(\omega)$$

i.A. kein Wahrscheinlichkeitsmaß, obwohl die Eigenschaften von $\mathbb{P}(A | \mathcal{F})$ die Vermutung begründen. Die Eigenschaften gelten nur \mathbb{P} -f.ü. und die Ausnahmemengen hängen i.A. von A ab. Man kann in der Tat Beispiele konstruieren, so dass $\mathbb{P}(A | \mathcal{F})$ kein WM ist. Unter bestimmten Voraussetzungen – die zudem in der Anwendung besonders relevant sind – kann man aber doch aus den $\mathbb{P}(A | \mathcal{F}), A \in \mathcal{A}$ ein Wahrscheinlichkeitsmaß konstruieren.

ii.) In diesem Paragraphen haben wir die bedingte Erwartung $\mathbb{E}(X | \mathcal{F})$ definiert. (Unbedingte) Erwartung sind \mathbb{P} -Integrale. **Sind bedingte Erwartungen Integrale bezüglich bedingter Wahrscheinlichkeiten?** In der Tat haben wir gerade bedingte Verteilungen $\mathbb{P}(A | \mathcal{F})$ definiert. Die Vermutung

ist naheliegend, dass

$$\mathbb{E}(X|\mathcal{F}) = \int X d\mathbb{P}(\cdot|\mathcal{F}).$$

gilt. Die Aussage aus i.) zeigt, dass man vorsichtig argumentieren muss, denn $\mathbb{P}(\cdot|\mathcal{F})$ ist im allgemeinen kein WM ist.

13.4.10 Satz: Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, X eine Zufallsvariable, $\mathcal{F} \subset \mathcal{A}$ eine Sub- σ -Algebra von \mathcal{A} . Dann gibt es eine Abbildung (einen sogenannten **Markovkern**)

$$\mu : \mathcal{B} \times \Omega \rightarrow \mathbb{R}$$

mit

a.) Für alle $\omega \in \Omega$ ist

$$\mu(\cdot, \omega) = \begin{cases} \mathcal{B} \rightarrow [0, 1] \\ B \mapsto \mu(B, \omega) \end{cases}$$

ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$.

b.) Für alle $B \in \mathcal{B}$ ist

$$\mu(B, \cdot) = \mathbb{P}(X \in B|\mathcal{F})(\cdot).$$

μ heißt die **bedingte Verteilung** von X bezüglich \mathcal{F} .

Beweis: Siehe Billingsley [2, S. 339]

13.4.11 Satz: Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, X eine reellwertige Zufallsvariable, $\mathcal{F} \subset \mathcal{A}$ eine Sub- σ -Algebra von \mathcal{A} und $\mu(\cdot, \cdot) = \mathbb{P}(X \in \cdot | \mathcal{F})(\cdot)$ eine bedingte Verteilung von X bezüglich \mathcal{F} . Es sei ferner $f : \mathbb{R} \rightarrow \mathbb{R}$ eine \mathcal{B} -messbare Abbildung und $f \circ X$ eine \mathbb{P} -integrierbare Zufallsvariable. Dann gilt

$$\begin{aligned}\mathbb{E}(f(X)|\mathcal{F})(\omega) &= \int_{\mathbb{R}} f(X) \mu(d\lambda, \omega) \\ &= \int_{\mathbb{R}} f(X) d\mathbb{P}(\cdot|\mathcal{F})(\omega) \\ &= \int_{\mathbb{R}} f(X) \mathbb{P}(d\lambda|\mathcal{F})(\omega).\end{aligned}$$

Beweis: Vgl. Mürmann [24, Abschnitt 13.3]

13.4.12 Beispiel: Prototypisches Beispiel ... und *Formel*

13.5 Totale Wahrscheinlichen

13.5.1 Satz (Gesetz(e) von der totalen Wahrscheinlichkeit):

i.) X, Y haben Dichten

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy \quad .$$

ii.) Y hat eine Dichte, X diskret

$$\mathbb{P}(X = x) = \int_{-\infty}^{\infty} \mathbb{P}(X = x|Y = y) f_Y(y) dy \quad .$$

iii.) X hat eine Dichte, Y diskret

$$f_X(x) = \sum_y f_X(x|Y = y) \mathbb{P}(Y = y) \quad .$$

Beweis: Blitzstein und Hwang [4].

13.5.2 Satz und Definition: Faltung ...

14 Erzeugende Funktionen

15 Charakteristische Funktionen

15.0.1 Bemerkung: i.) **Komplexe Zahlen** geben wir regelmäßig in der Form

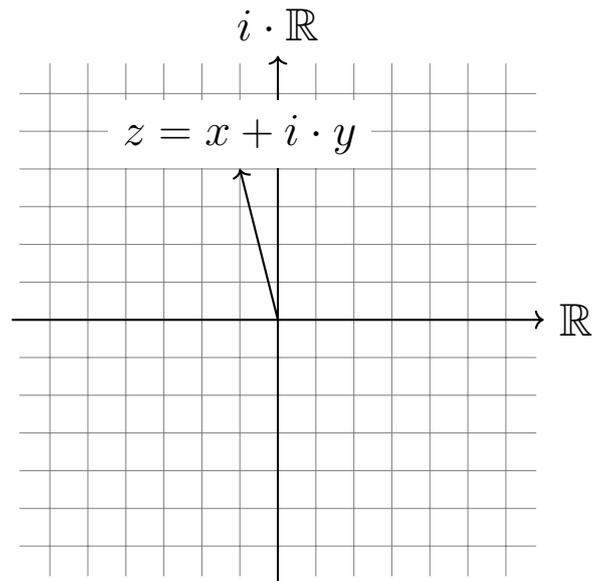
$$z = x + iy$$

an. Dabei ist $i = \sqrt{-1}$. x heißt **Realteil** von z und y der **Imaginärteil** von z . Die komplexe Zahl $\bar{z} = x - iy$ heißt die zu z **konjugiert komplexe Zahl**.

ii.) Unter einer komplexen Zahl kann man sich auch einen Punkt $(x, y)^T$ der komplexen Ebene vorstellen:

$$z = x + iy = \begin{pmatrix} x \\ y \end{pmatrix}$$

Die Abbildung zeigt die komplexe Ebene mit der komplexen Zahl $z = -\frac{1}{2} + 2i$.

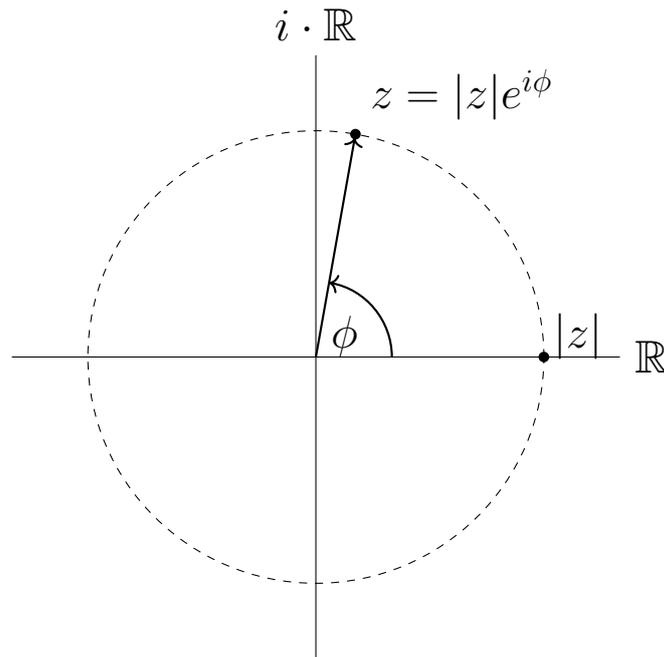


iii.) Es gilt: $\|z\|^2 = z\bar{z}$, $\|z\| = \sqrt{x^2 + y^2}$ und $z = \bar{z} \Leftrightarrow z \in \mathbb{R}$.

iv.) Es gilt

$$e^{i\phi} = \cos(\phi) + i \sin(\phi)$$
$$e^{-i\phi} = \overline{e^{i\phi}}$$
$$\|e^{i\phi}\| = 1$$

Bekanntlich kann man dann komplexe Zahlen auch mittels Betrag und Winkeln darstellen.



15.0.2 Definition: Es sei X eine Zufallsvariable. Die Abbildung

$$\psi_X : \mathbb{R} \rightarrow \mathbb{C}, t \mapsto \mathbb{E}(e^{itX})$$

heißt **charakteristische Funktion** von X .

15.0.3 Bemerkung: i.) Eigentlich wäre es besser von der charakteristische Funktion einer Verteilung zu sprechen, denn die Verteilungsfunktion legt die charakteristische Funktion fest.

ii.) Die Abbildung $t \mapsto \mathbb{E}(e^{itX})$ kann man sich als Graph in \mathbb{R}^3 veranschaulichen. Für jedes t erhält man eine komplexe Zahl $\mathbb{E}(e^{itX}) = \mathbb{E} \cos(tX) + i \cdot \mathbb{E} \sin(tX)$ (die man sich als 2-dimensionalen Vektor vorstellen kann). Insgesamt erhält man

Kurven wie sie in beiden Panels der Abbildung 15.1 dargestellt sind. Dabei ist t auf der vertikalen Achse abgetragen. Die blaue Kurve zeigen die CF der **Standardnormalverteilung** mit $\mu = 0$ und $\sigma = 1$; die Kurve verläuft komplett in der reellen Ebene. Die rote Kurve zeigt die CF der **Normalverteilung** mit $\mu = 5$ und $\sigma = 1$. Die beiden Panels zeigen die gleichen CFen; der Blickwinkel ist variiert.

Die Abbildung 15.2 zeigt die CFen der Gleichverteilungen auf $[-1, 1]$ (blau) bzw. auf $[0, 1]$ (rot). Die CF der Gleichverteilung auf $[-1, 1]$ ist reellwertig. Es gilt $\Psi_{\text{Unif}[-1,1]}(t) = \frac{\sin(t)}{t}$.

iii.) Die für die charakteristische Funktion relevante Information über die Zufallsvariable X ist bekanntlich komplett durch die Verteilungsfunktion, also in der Abbildung

$$F^X : \mathbb{R} \rightarrow [0, 1], x \mapsto \mathbb{P}(X \leq x)$$

enthalten.

Nach der **Transformation zur charakteristische Funktion** erhalten wir die Abbildung

$$\psi_X : \mathbb{R} \rightarrow \mathbb{C}, t \mapsto \mathbb{E}(e^{itX}).$$

Später werden wir sehen, dass wir die Transformation ein-

deutig umkehren können. Schematisch:

$$F_X \xrightarrow{\text{cf-Transformation}} \psi_X$$
$$F_X \xleftarrow{\text{Inverse-cf-transformation}} \psi_X$$

Fazit: Man kann die Verteilung einer Zufallsvariablen wahlweise durch die **Verteilungsfunktion** oder die **charakteristische Funktion** repräsentieren.

15.0.4 Beispiel: Es sei $X \sim \text{Bernoulli}(p)$. Dann gilt

$$\psi_X(t) = \mathbb{E}(e^{itX}) = (1-p)e^{it \cdot 0} + pe^{it \cdot 1} = (1-p) + pe^{it}.$$

$e^{it \cdot 0}$ und $e^{it \cdot 1}$ sind Punkte auf dem Einheitskreis. Von diesen Punkten bilden wir die Konvexkombination Erwartungswertes.

15.0.5 Satz. Es sei X eine Zufallsvariable mit charakteristischer Funktion ψ_X . Dann gilt

- i.) $\psi_X(0) = 1$ und $|\psi_X(t)| \leq 1$ für alle $t \in \mathbb{R}$.
- ii.) ψ_X ist gleichmäßig stetig.
- iii.) $\psi_X(-t) = \overline{\psi_X(t)}$ für alle $t \in \mathbb{R}$.
- iv.) $\psi_{a+bX}(t) = e^{itb} \cdot \psi_X(at)$ für alle $a, b, t \in \mathbb{R}$.

15.0.6 Satz (Inversionsatz von Levy): Es sei $\psi_X(t)$ die charakteristische Funktion der Zufallsvariable X mit der Verteilung \mathbb{P}^X . Dann

$$\lim_{T \nearrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \psi_X(t) dt = \frac{1}{2} \mathbb{P}^X(\{a\}) + \mathbb{P}^X((a, b)) + \frac{1}{2} \mathbb{P}^X(\{b\})$$

Wenn zudem $\int_{\mathbb{R}} |\Psi_X(t)| dt < \infty$, dann hat X eine stetige Dichte und es gilt

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \psi_X(t) dt,$$

so dass wir eine schöne *Dualität*

$$\psi_X(t) = \int_{\mathbb{R}} e^{-itx} f(x) dt$$

gilt.

15.0.7 Satz (Stetigkeitssatz von Levy-Cramer): Es sei X_j eine Folge von Zufallsvariablen mit Verteilungsfunktionen $X_j \sim F_j$ und charakteristischen Funktionen ψ_j . Dann ist äquivalent

- i.) Es gibt eine Verteilungsfunktion F mit $F_j \xrightarrow{i.V.} F$.
- ii.) Für jedes $t \in \mathbb{R}$ existiert $\psi(t) = \lim_{j \rightarrow \infty} \psi_j(t)$ und die Funktion $\psi(t)$ ist stetig im Nullpunkt.

Falls i.) oder ii.) gilt, dann ist ψ die charakteristische Funk-

tion von F :

$$\psi(t) = \int e^{itx} dF(x).$$

15.0.8 Satz: Wenn die Zufallsvariablen X_1, \dots, X_n **unabhängig** sind, dann gilt

$$\psi_{X_1+\dots+X_n} = \prod_{j=1}^n \psi_{X_j}$$

15.0.9 Bemerkung:

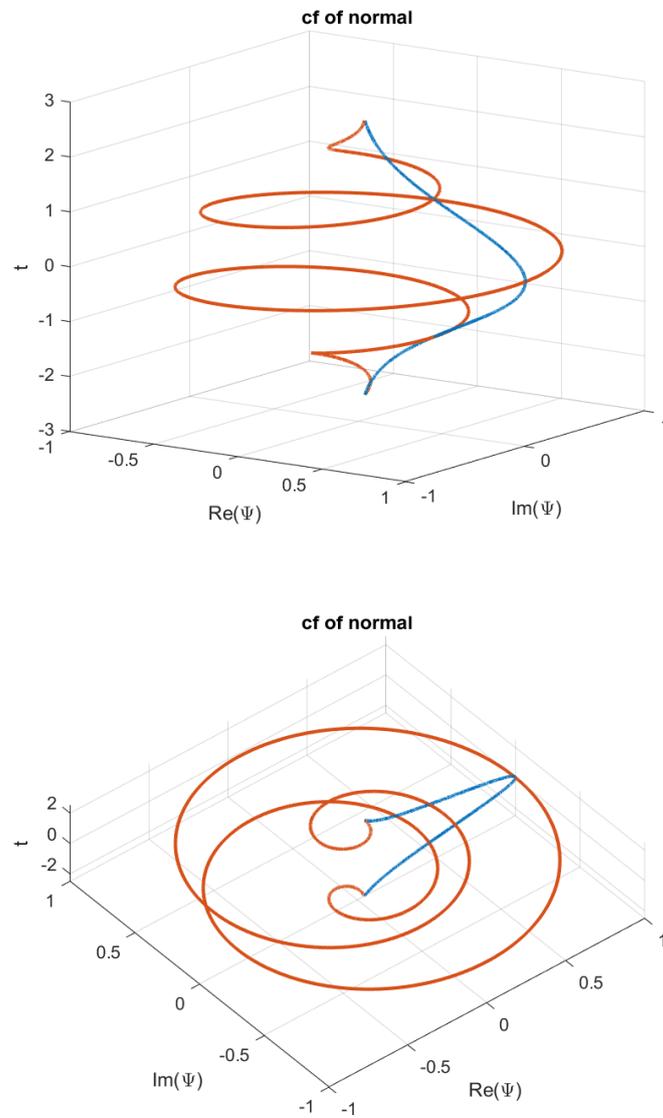


Abbildung 15.1: CF der Standardnormalverteilung (blau) und der Normalverteilung mit $\mu = 5$ und $\sigma = 1$. Quelle: Eigene Darstellung mit dem Matlab Code `plot3dSomeCFs.m`.

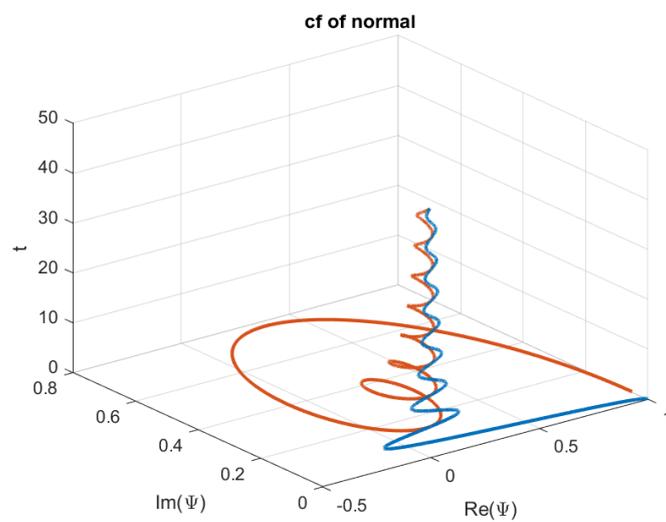


Abbildung 15.2: CFen der Gleichverteilungen auf $[-1, 1]$ (blau) bzw. auf $[0, 1]$ (rot). Quelle: Eigene Darstellung mit dem Matlab Code `plot3dSomeCFs.m`.

Literaturverzeichnis

- [1] Heinz **Bauer**, 1992, Maß- und Integrationstheorie, 2. Auflage, De Gruyter
- [2] Patrick **Billingsley**, 1995, Probability and Measure, 3th. Auflage; Wiley
- [3] Norman **Biggs**, 2002, Discrete Mathematics, 2. ed.: Oxford University Press
- [4] Joseph **Blitzstein**, Jessica Hwang, Introduction to Probability, CRC Press
- [5] Kai Chung **Chung**, 2000, A Course in Probability Theory, Wiley
- [6] Rick **Durrett**, 2019, Probability: Theory and Examples, Cambridge University Press
- [7] Alexander McNeil, Rüdiger Frey, Paul Embrechts, 2015, Quantitative Risk Management, 2. Auflage, Princeton

University Press

- [8] Jürgen **Elstrodt**, 2018, Maß- und Integrationstheorie, 8. Auflage, Springer
- [9] Nasrollah **Etemadi**; 1981; An elementary proof of the strong law of large numbers; Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete; Volume 55, Seiten 119-122.
- [10] William **Feller**; 1968; An Introduction to Probability – Theory and its Application; 3ed; Wiley.
- [11] Ludwig **Fahrmeir**, Christian **Heumann**, Rita **Künstler**, Iris **Pigeot**, Gerhard **Tutz**, 2016, Statistik – Der Weg zur Datenanalyse, 8. Auflage, Springer
- [12] Hans **Föllmer** und Alexander **Schied**, 2016, Stochastic Finance – An introduction in discrete time, 4. Auflage, De Gruyter
- [13] Hans-Otto **Georgii**; 2015, Stochastik, Fünfte Auflage, De Gruyter
- [14] Geoffrey **Grimmett** und David **Strikzaker**, 2020, Probability and Random Processes, Fourth Edition, Oxford University Press
- [15] Geoffrey **Grimmett** und David **Strikzaker**, 2020, One

- tousand exercises in probability, Third Edition, Oxford University Press
- [16] Charles **Grinstead** und Laurie **Snell**, 2006, Introduction to Probability, https://chance.dartmouth.edu/teaching_aids/books_articles/probability_book/book.html [abgerufen am 20210612]
- [17] Bruce **Hansen**; 2022; Econometrics; Princeton University Press
- [18] Norbert **Henze**, 2019, Stochastik – Einführung mit Grundzügen der Maßtheorie, Springer
- [19] Jean **Jacod** und Philip **Protter**, 2004, Probability Essentials, Springer
- [20] Samuel **Karlin**, Howard Taylor; 1975; A first course in Stochastic Processes; 2ed.; Academic Press
- [21] Achim **Klenke**, 2020, Wahrscheinlichkeitstheorie, 4. Auflage, Springer Spektrum, Berlin
- [22] Ulrich **Krengel**, 2005, Einführung in die Wahrscheinlichkeitstheorie und Statistik, 8. Auflage, Springer
- [23] Uwe **Küchler**, 2016, Maßtheorie für Statistiker, 8. Auflage, Springer

- [24] Michael **Mürmann**, 2014, Wahrscheinlichkeitstheorie und Stochastische Prozesse, Springer
- [25] David **Mazur**, 2010, Combinatorics, MAA Press
- [26] Ralph Tyrrell **Rockafellar**, Stanislav **Uryasev**; 2002; Conditional value-at-risk for general loss distributions; Journal of Banking & Finance; Volume 26, Seiten 1443-1471.
- [27] Walter **Rudin**, 1976, Principle of Real Analysis, Springer
- [28] Rene **Schilling**; 2021; Measure, Integral, Probability & Processes; Rene-Schilling-Dresden
- [29] Jun **Shao**, 2003, Mathematical Statistics, Springer
- [30] **Shreve, Steven**, 2004, Stochastic Calculus for Finance 2, Springer, Heidelberg
- [31] Stefan **Tappe**, 2013, Einführung in die Wahrscheinlichkeitstheorie, Springer
- [32] Grzegorz **Tomkowicz** und Stan **Wagon**, 2016, The Banach-Tarski Paradox, 2. Auflage, Cambridge University Press
- [33] David **Williams**, 1991, Probability with Martingales,

Cambridge University Press

- [34] Amy **Wagaman** und Robert **Dobrow**, 2021, Probability – With Applications and ξ , 2. Auflage, Wiley