



---

# Elementare Wahrscheinlichkeitsrechnung und Statistik

---

Universität Ulm  
Institut für Stochastik

Vorlesungsskript  
Prof. Dr. Volker Schmidt  
Stand: Wintersemester 2008/09

ULM, IM FEBRUAR 2009

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>7</b>
1.1	Was ist Stochastik? . . . . .	7
1.2	Typische Fragestellungen und Ergebnisse . . . . .	7
1.3	Beispiel . . . . .	8
<b>2</b>	<b>Ereignisse und Wahrscheinlichkeiten</b>	<b>9</b>
2.1	Ereignisse als Mengen . . . . .	9
2.2	Ereignissysteme . . . . .	10
2.3	Wahrscheinlichkeitsmaße . . . . .	12
2.3.1	Definition und elementare Eigenschaften . . . . .	12
2.3.2	Weitere Eigenschaften von Wahrscheinlichkeitsmaßen . . . . .	15
2.4	Endliche Wahrscheinlichkeitsräume . . . . .	17
2.4.1	Laplacescher Wahrscheinlichkeitsraum . . . . .	17
2.4.2	Einfache Urnenmodelle . . . . .	18
2.5	Geometrische Wahrscheinlichkeiten . . . . .	20
2.6	Bedingte Wahrscheinlichkeiten . . . . .	23
2.6.1	Definition und Multiplikationssatz . . . . .	23
2.6.2	Formel der totalen Wahrscheinlichkeit; Bayessche Formel . . . . .	24
2.7	Stochastische Unabhängigkeit . . . . .	26
<b>3</b>	<b>Zufallsvariablen und Zufallsvektoren</b>	<b>29</b>
3.1	Definition von Zufallsvariablen . . . . .	29
3.2	Verteilung und Verteilungsfunktion . . . . .	30
3.2.1	Diskrete Zufallsvariablen; Wahrscheinlichkeitsfunktion . . . . .	31
3.2.2	Grundlegende Klassen diskreter Verteilungen (Zusammenfassung) . . . . .	32
3.2.3	Verteilungsfunktion; absolutstetige Zufallsvariablen . . . . .	34
3.3	Zufallsvektoren . . . . .	40
3.3.1	Definition, Verteilung und Verteilungsfunktion . . . . .	40
3.3.2	Eigenschaften multivariater Verteilungsfunktionen . . . . .	42
3.3.3	Weitere Beispiele von Zufallsvektoren . . . . .	44
3.3.4	Bedingte Wahrscheinlichkeitsfunktion; bedingte Verteilung; bedingte Dichte . . . . .	45
3.3.5	Unabhängige Zufallsvariablen . . . . .	46
3.3.6	Beispiele: Zufallsvektoren mit unabhängigen Komponenten . . . . .	48
3.4	Funktionen von Zufallsvektoren . . . . .	50
3.4.1	Zusammengesetzte Abbildungen . . . . .	50
3.4.2	Lineare Transformation . . . . .	52
3.4.3	Quadrierung . . . . .	53
3.4.4	Summe, Produkt und Quotient von unabhängigen Zufallsvariablen . . . . .	54
3.4.5	Unabhängigkeit zusammengesetzter Abbildungen . . . . .	57

<b>4</b>	<b>Weitere Charakteristiken von Zufallsvariablen</b>	<b>58</b>
4.1	Erwartungswert . . . . .	58
4.1.1	Definition und Berechnungsformeln . . . . .	58
4.1.2	Alternative Integral-Darstellungen . . . . .	61
4.1.3	Weitere Eigenschaften des Erwartungswertes . . . . .	63
4.1.4	Integral-Darstellung mittels Quantilfunktion . . . . .	65
4.2	Varianz und höhere Momente . . . . .	67
4.2.1	Definition und elementare Eigenschaften . . . . .	67
4.2.2	Transformationssatz und Berechnungsformeln . . . . .	69
4.3	Gemischte Momente . . . . .	72
4.3.1	Transformationssatz für Zufallsvektoren . . . . .	72
4.3.2	Multiplikationsformel und Kovarianz . . . . .	73
4.3.3	Linearer Zusammenhang von Zufallsvariablen . . . . .	77
4.3.4	Erwartungswertvektor und Kovarianzmatrix . . . . .	77
4.4	Ungleichungen für Momente und Wahrscheinlichkeiten . . . . .	80
4.4.1	Ungleichungen vom $L^p$ -Typ . . . . .	80
4.4.2	Jensen-Ungleichung . . . . .	83
4.4.3	Tschebyschew-Ungleichung; Markow-Ungleichung . . . . .	84
<b>5</b>	<b>Konvergenzarten und Grenzwertsätze</b>	<b>87</b>
5.1	Konvergenzarten . . . . .	87
5.1.1	Definitionen und elementare Eigenschaften . . . . .	87
5.1.2	Charakterisierung der Verteilungskonvergenz . . . . .	91
5.1.3	Konvergenz zusammengesetzter Abbildungen; Satz von Slutsky . . . . .	93
5.2	Gesetz der großen Zahlen . . . . .	97
5.2.1	Schwaches Gesetz der großen Zahlen . . . . .	98
5.2.2	Starkes Gesetz der großen Zahlen . . . . .	99
5.2.3	Anwendungsbeispiele . . . . .	105
5.3	Zentraler Grenzwertsatz . . . . .	110
5.3.1	Zentraler Grenzwertsatz für Summen von unabhängigen Zufallsvariablen . . . . .	111
5.3.2	Anwendungsbeispiele . . . . .	117
5.3.3	Charakteristische Funktionen . . . . .	121
5.3.4	Bedingungen von Lindeberg und Ljapunow . . . . .	128

<b>6</b>	<b>Stichproben und Stichprobenfunktionen</b>	<b>135</b>
6.1	Zufallsstichprobe . . . . .	135
6.2	Stichprobenfunktionen . . . . .	136
6.2.1	Stichprobenmittel . . . . .	136
6.2.2	Stichprobenvarianz . . . . .	139
6.3	Beispiel: Normalverteilte Stichprobenvariablen . . . . .	144
6.3.1	Gammaverteilung und $\chi^2$ -Verteilung . . . . .	144
6.3.2	Unabhängigkeit und Transformation von Zufallsvektoren . . . . .	148
6.3.3	Verteilung von Stichprobenmittel und Stichprobenvarianz . . . . .	149
6.3.4	t-Verteilung . . . . .	152
<b>7</b>	<b>Tabellen für Verteilungsfunktionen und Quantile</b>	<b>155</b>

# Literatur

H. Dehling, B. Haupt  
Einführung in die Wahrscheinlichkeitstheorie und Statistik  
Springer-Verlag, Berlin 2003

L. Dümbgen  
Stochastik für Informatiker  
Springer-Verlag, Berlin 2003

C. Hesse  
Angewandte Wahrscheinlichkeitstheorie  
Vieweg-Verlag, Braunschweig 2003

H.-O. Georgii  
Stochastik  
Walter de Gruyter, Berlin, New York 2002

A. Irle  
Wahrscheinlichkeitstheorie und Statistik  
Teubner, Stuttgart, Leipzig, Wiesbaden 2001

J. Jacod und P. Protter  
Probability essentials  
Springer-Verlag, Berlin 2003

A.F. Karr  
Probability  
Springer-Verlag, New York 1993

U. Krengel  
Einführung in die Wahrscheinlichkeitstheorie und Statistik  
Vieweg-Verlag, Braunschweig 2002

R. Meester  
Introduction to Probability Theory  
Birkhäuser-Verlag, Basel, Cambridge 2003

S. Resnick  
A probability path  
Birkhäuser-Verlag, Basel 1999

A.N. Shiryaev  
Probability  
Springer-Verlag, New York 1996  
(deutsche Übersetzung: Wahrscheinlichkeit  
Deutscher Verlag der Wissenschaften, Berlin, 1988.)

## Klassiker und weitere Bücher zur Vertiefung des Stoffes

H. Bauer  
Wahrscheinlichkeitstheorie  
Verlag De Gruyter, Berlin 1991

P. Billingsley  
Probability and Measure  
J. Wiley & Sons, New York 1995

L. Breiman  
Probability  
SIAM, Philadelphia, 1993

W. Feller  
An introduction to probability theory and its applications. Vol I/II  
J. Wiley & Sons, New York 1970/71

P. Gänssler und W. Stute  
Wahrscheinlichkeitstheorie  
Springer-Verlag, Berlin 1977

O. Kallenberg  
Foundations of modern probability  
Springer-Verlag, New York 2001

# 1 Einleitung

## 1.1 Was ist Stochastik?

Der Begriff *Stochastik* stammt ursprünglich aus dem Griechischen und bedeutet dort: die Kunst des geschickten Vermutens. Die mathematische Stochastik befasst sich mit der Beschreibung und Untersuchung von Ereignissen, zeitlichen Entwicklungen bzw. räumlichen Strukturen, die *vom Zufall beeinflusst* werden. Solche Ereignisse, Entwicklungen bzw. Strukturen werden oft durch *Daten* dokumentiert, für deren Analyse die *Statistik* – ein Teilgebiet der Stochastik – geeignete Methoden bereitstellt.

## 1.2 Typische Fragestellungen und Ergebnisse

Zu den Aufgaben der Stochastik gehört die Bewertung von Ereignissen, Entwicklungen bzw. Strukturen durch die Bestimmung ihrer *Wahrscheinlichkeit*, die eine Maßzahl für die Chance ihres Eintretens ist.

Das Phänomen „Zufall“ kommt in vielfältiger Weise in zahlreichen Bereichen des täglichen Lebens vor, z.B. bei der Vorhersage von zukünftigen Aktienkursen bzw. Zinssätzen, bei der Wettervorhersage, bei Würfel- bzw. Kartenspielen, beim Zahlenlotto, usw.

**Stochastische Modelle** sind in vielen Disziplinen der Wissenschaft ein wichtiges Hilfsmittel, so in

- Informatik und Ingenieurwissenschaften (z.B. bei der Dimensionierung und Leistungsanalyse von Kommunikations- und Rechnersystemen)
- Physik, Chemie und Materialwissenschaften (z.B. bei der Strukturanalyse von Werkstoffen, Untersuchung der Rauheit von technischen Oberflächen)
- Wirtschaftswissenschaften (z.B. beim Risikomanagement von Versicherungen und Banken, Analyse von Finanzmärkten)
- Biologie und Medizin (z.B. bei der Bildanalyse zur Untersuchung mikroskopischer Gewebestrukturen bzw. intrazellulärer Netzwerke)

Es ist üblich, die Stochastik in die folgenden **Teilgebiete** zu unterteilen:

- Wahrscheinlichkeitsrechnung
- Statistik
- stochastische Prozesse und Felder (z.B. Markov-Modelle)
- stochastische Simulation (z.B. Markov-Chain-Monte-Carlo)

**Typische Fragestellungen und Ergebnisse** der Stochastik sind

- geschlossene Formeln für die Wahrscheinlichkeit von Ereignissen, Entwicklungen bzw. Strukturen (oftmals nur unter restriktiven Modellannahmen möglich)
- Grenzwertsätze (Näherungslösungen, z.B. Gesetz der großen Zahlen, Zentraler Grenzwertsatz)
- Methoden zur Schätzung unbekannter Modellparameter; Tests hypothetischer Modellannahmen (Signifikanztests)
- Kopplung von stochastischer Modellierung, statistischer Datenanalyse und Computer-Simulation

### 1.3 Beispiel

Betrachten Roulette-Spiel mit 38 möglichen Ausgängen, nämlich 18 rote Felder, 18 schwarze Felder und 2 grüne Felder. Betrachten Spieler, der auf „Rot“ setzt. Er gewinnt 1 Euro mit der Wahrscheinlichkeit  $\frac{18}{38}$  ( $= \frac{9}{19}$ ) und verliert 1 Euro mit der Wahrscheinlichkeit  $\frac{20}{38}$  ( $= \frac{10}{19}$ ). Sei nun  $X_n$  der zufällige „Gewinn“ beim  $n$ -ten Spiel. Dann gilt:

$$P(X_n = 1) = \frac{9}{19}, \quad P(X_n = -1) = \frac{10}{19}.$$

Die Zufallsgrößen  $X_1, X_2, \dots$  sind *unabhängig* und *identisch verteilt*. Betrachten Gesamtgewinn  $S_n = X_1 + \dots + X_n$  aus den ersten  $n$  Spielen. Die Folge  $S_1, S_2, \dots$  wird *Random Walk* (bzw. *zufällige Irrfahrt*) genannt.

Wie groß ist der erwartete Gewinn  $\mathbb{E} X_n$  (Erwartungswert) beim  $n$ -ten Spiel? Wie groß ist der erwartete Gesamtgewinn  $\mathbb{E} S_n$  aus den ersten  $n$  Spielen? Es gilt:

$$\begin{aligned} \mathbb{E} X_n &= 1 \cdot \frac{9}{19} + (-1) \cdot \frac{10}{19} = -0.05263 \left( = -\frac{1}{19} \right) \\ \mathbb{E} S_n &= \mathbb{E} X_1 + \dots + \mathbb{E} X_n = -n \cdot 0.05263 \left( = -\frac{n}{19} \right) \end{aligned}$$

**(Schwaches) Gesetz der großen Zahlen** „Für große  $n$  ist  $\frac{S_n}{n}$  nahe bei  $\mathbb{E} X_1$  mit hoher Wahrscheinlichkeit.“

**Zentraler Grenzwertsatz** „Für große  $n$  lässt sich die Wahrscheinlichkeit  $P\left(a \leq \frac{S_n - \mathbb{E} S_n}{\sqrt{n}} \leq b\right)$  durch die Normalverteilung approximieren, wobei  $a, b \in \mathbb{R}$  mit  $a < b$  beliebige, jedoch fest vorgegebene Toleranzgrenzen sind.“

**Beachte** Den Begriffsbildungen „Zufallsgröße“, „unabhängig“, „identisch verteilt“, „für große  $n$ “, „nahe bei  $\mathbb{E} X_1$ “, „mit hoher Wahrscheinlichkeit“ bzw. „Normalverteilung“ liegen mathematische Definitionen zugrunde. Sie gehören zu den Grundbegriffen der Stochastik, die in den folgenden Abschnitten detailliert erläutert werden.



## 2 Ereignisse und Wahrscheinlichkeiten

### 2.1 Ereignisse als Mengen

Wir modellieren Ereignisse als Mengen. Dabei ist eine *Menge* eine Zusammenfassung von wohldefinierten und unterscheidbaren Dingen (Elemente) zu einem Ganzen.

**Schreibweise**  $\Omega$  Grundmenge,  $\omega$  Element

$\omega \in \Omega$ :  $\omega$  ist Element von  $\Omega$

$\omega \notin \Omega$ :  $\omega$  ist nicht Element von  $\Omega$

$A = \{a, b, c, \dots\}$ : Die Menge  $A$  besteht aus den Elementen  $a, b, c, \dots$

$A = \{\omega : \omega \in \Omega, \omega \text{ hat Eigenschaft } E\}$ :  $A$  besteht aus denjenigen Elementen  $\omega$  von  $\Omega$ , die die Eigenschaft  $E$  haben.

**Beispiel**  $\Omega = \mathbb{N}$ ,  $A = \{2, 4, 6, \dots\} = \{n : n \in \mathbb{N}, n \text{ ist durch } 2 \text{ teilbar}\}$

Der Vergleich von Ereignissen erfolgt durch den Vergleich der Mengen, durch die die Ereignisse modelliert werden.

#### Definition

1.  $A_1 \subset A_2$  bedeutet,  $A_1$  ist Teilmenge von  $A_2$ , d.h., aus  $\omega \in A_1$  folgt  $\omega \in A_2$
2.  $A_1 = A_2$ , falls  $A_1 \subset A_2$  und  $A_2 \subset A_1$ .

Betrachten Ereignisse, die bei einem Zufallsexperiment (z.B. Münzwurf, Werfen eines Würfels, Roulette-Spiel, Erzeugen einer Pseudozufallszahl mit einem Zufallszahlengenerator) eintreten können. Dann ist

- $\Omega$  = Menge aller möglichen Versuchsergebnisse (Grundmenge, Grundgesamtheit, Merkmalraum, Stichprobenraum); Ereignis  $\Omega$  = sicheres Ereignis (tritt immer ein)
- $A_1, A_2 \subset \Omega$ : Ereignisse = Teilmengen von Versuchsergebnissen mit bestimmten Eigenschaften;
- $\{\omega\} \subset \Omega$ : Elementarereignis = ein (einzelnes) Versuchsergebnis;
- Angenommen: bei einem Versuch wird das Ergebnis  $\omega$  erzielt. Dann sagen wir: Das Ereignis  $A$  tritt ein, falls  $\omega \in A$ .
- Für  $A_1 \subset A_2$  gilt: Wenn  $A_1$  eintritt, dann tritt auch  $A_2$  ein.

**Beispiel** (einmaliges Würfeln):  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ; Elementarereignisse  $\{1\}, \{2\}, \dots, \{6\}$ .

Das Ereignis  $A_1 = \{2\}$  tritt genau dann ein, wenn die Zahl 2 gewürfelt wird.

Das Ereignis  $A_2 = \{2, 4, 6\}$  tritt genau dann ein, wenn eine gerade Zahl gewürfelt wird.

Also gilt:  $A_1 \subset A_2$ , d.h., wenn  $A_1$  eintritt, dann tritt auch  $A_2$  ein.

**Definition** Diejenige Teilmenge von  $\Omega$ , die *kein* Element enthält, heißt *leere Menge* und wird mit  $\emptyset$  bezeichnet.

**Beachte** Das Ereignis  $\emptyset$  tritt niemals ein und wird deshalb *unmögliches Ereignis* genannt.

## 2.2 Ereignissysteme

Aus gegebenen Ereignissen  $A_1, A_2, \dots$  kann man durch deren „Verknüpfung“ weitere Ereignisse bilden. Dies wird durch die folgenden Mengenoperationen modelliert.

### Mengenoperationen und ihre probabilistische Bedeutung

1. *Vereinigungsmenge*  $A_1 \cup A_2$ : Menge aller Elemente, die zu  $A_1$  oder  $A_2$  gehören.  
(Ereignis  $A_1 \cup A_2$  = mindestens eines der Ereignisse  $A_1$  oder  $A_2$  tritt ein)  
 $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \dots$  = Menge aller Elemente, die zu *mindestens einer* der Mengen  $A_i$  gehören.  
(Ereignis  $\bigcup_{i=1}^{\infty} A_i$  = mindestens eines der Ereignisse  $A_1, A_2, \dots$  tritt ein)
2. *Schnittmenge*  $A_1 \cap A_2$ : Menge aller Elemente, die zu  $A_1$  und  $A_2$  gehören  
(Ereignis  $A_1 \cap A_2$  = beide Ereignisse  $A_1$  und  $A_2$  treten ein).  
*Beachte*: Zwei Ereignisse  $A_1, A_2 \subset \Omega$  mit  $A_1 \cap A_2 = \emptyset$  heißen *unvereinbar*, d.h., sie können nicht gleichzeitig eintreten.  
 $\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \dots$ : Menge aller Elemente, die zu *jeder* der Mengen  $A_i$  gehören.  
(Ereignis  $\bigcap_{i=1}^{\infty} A_i$  = sämtliche Ereignisse  $A_1, A_2, \dots$  treten ein)
3. *Differenzmenge*  $A_1 \setminus A_2$ : Menge aller Elemente von  $A_1$ , die *nicht* zu  $A_2$  gehören.  
Spezialfall:  $A^c = \Omega \setminus A$  (*Komplement*) (Ereignis  $A^c$  = Ereignis  $A$  tritt nicht ein)
4. *Symmetrische Mengendifferenz*  $A_1 \triangle A_2$ : Menge aller Elemente, die zu  $A_1$  oder  $A_2$ , jedoch *nicht* zu beiden gehören.

**Beispiel** Sei  $\Omega = \{a, b, c, d\}$ ,  $A_1 = \{a, b, c\}$ ,  $A_2 = \{b, d\}$ .

Dann gilt  $A_1 \cup A_2 = \{a, b, c, d\}$ ;  $A_1 \cap A_2 = \{b\}$ ;  $A_1 \setminus A_2 = \{a, c\}$ ;  $A_1^c = \{d\}$ ;  $A_1 \triangle A_2 = \{a, c, d\}$ .

### Beachte

- Das Ereignis  $A_1 \cup A_2$  tritt genau dann ein, wenn  $A_1$  oder  $A_2$  oder beide eintreten.
- Das Ereignis  $A_1 \cap A_2$  tritt genau dann ein, wenn  $A_1$  und  $A_2$  eintreten.
- Das Ereignis  $A_1 \setminus A_2$  tritt genau dann ein, wenn  $A_1$  eintritt und  $A_2$  nicht eintritt.
- Das Ereignis  $A_1 \triangle A_2$  tritt genau dann ein, wenn  $A_1$  oder  $A_2$  eintreten und nicht beide eintreten.
- Die Mengen  $A_1, A_2, \dots \subset \Omega$  heißen *paarweise disjunkt*, wenn  $A_i \cap A_j = \emptyset$  für beliebige  $i \neq j$ .  
(Die Ereignisse  $A_1, A_2, \dots \subset \Omega$  heißen *paarweise unvereinbar*, wenn  $A_i \cap A_j = \emptyset$  für beliebige  $i \neq j$ )

**Lemma 2.1** Für beliebige Mengen  $A_1, A_2 \subset \Omega$  gilt:

$$A_1 \setminus A_2 = A_1 \cap A_2^c,$$

$$A_1 \triangle A_2 = (A_1 \cup A_2) \setminus (A_1 \cap A_2) = (A_1 \setminus A_2) \cup (A_2 \setminus A_1).$$

**Beweis** klar

**Beachte** Jede beliebige Folge von Mengen  $A_1, A_2, \dots \subset \Omega$  kann man in eine Folge von paarweise disjunkten Mengen  $A'_1, A'_2, \dots$  überführen: Sei  $A'_1 = A_1, A'_2 = A_2 \setminus A'_1, A'_3 = A_3 \setminus (A'_1 \cup A'_2), \dots$

Dann gilt:  $A'_i \cap A'_j = \emptyset$  für  $i \neq j$ , und  $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} A'_n$

**Weitere Eigenschaften** Seien  $A, B, C \subset \Omega$  beliebige Teilmengen. Dann gelten

- *Eindeutigkeitsgesetze*:  $A \cup \emptyset = A, A \cap \emptyset = \emptyset, A \cup \Omega = \Omega, A \cap \Omega = A$   
(allgemein: falls  $A \subset B$ , dann gilt  $A \cap B = A, A \cup B = B$ )
- *de Morgansche Gesetze*:  $(A \cup B)^c = A^c \cap B^c, (A \cap B)^c = A^c \cup B^c$
- *Assoziativ-Gesetze*:  $A \cup (B \cup C) = (A \cup B) \cup C, A \cap (B \cap C) = (A \cap B) \cap C$
- *Distributiv-Gesetze*:  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Es ist oft nicht zweckmäßig, *alle* möglichen Teilmengen von  $\Omega$  in die Modellierung einzubeziehen, sondern man betrachtet nur die Familie derjenigen Teilmengen von  $\Omega$ , deren Wahrscheinlichkeiten tatsächlich von Interesse sind. Diese Mengenfamilie soll jedoch abgeschlossen sein bezüglich der Operationen  $\cup, \cap, \setminus$ , was durch die folgende Begriffsbildung erreicht wird.

**Definition** Eine nichtleere Familie  $\mathcal{F}$  von Teilmengen von  $\Omega$  heißt *Algebra*, falls

- (A1)  $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$   
 (A2)  $A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cup A_2 \in \mathcal{F}$

**Beispiel**  $\Omega = \{a, b, c, d\}$ ,  $\mathcal{F}_1 = \{\emptyset, \{a\}, \{b, c, d\}, \Omega\}$  ist eine Algebra,  $\mathcal{F}_2 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$  ist dagegen keine Algebra.

**Lemma 2.2** Sei  $\mathcal{F}$  eine Algebra und  $A_1, A_2, \dots, A_n \in \mathcal{F}$ . Dann gilt

1.  $\emptyset, \Omega \in \mathcal{F}$
2.  $A_1 \cap A_2 \in \mathcal{F}$
3.  $A_1 \setminus A_2 \in \mathcal{F}$
4.  $\bigcup_{i=1}^n A_i \in \mathcal{F}, \bigcap_{i=1}^n A_i \in \mathcal{F}$

**Beweis**

1. Weil  $\mathcal{F}$  nicht leer ist, gibt es ein  $A \in \mathcal{F}, A \subset \Omega$ . Also gilt  $A^c \in \mathcal{F}$  wegen (A1) bzw.  $\underbrace{A \cup A^c}_{=\Omega} \in \mathcal{F}$  wegen (A2) bzw.  $\underbrace{\Omega^c}_{=\emptyset} \in \mathcal{F}$  wegen (A1).
2. Außerdem ergibt sich aus den Gesetzen von de Morgan, dass  $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c \in \mathcal{F}$ .
3. Auf ähnliche Weise ergibt sich, dass  $A_1 \setminus A_2 = A_1 \cap A_2^c = (A_1^c \cup A_2)^c \in \mathcal{F}$
4. Dies ergibt sich mit vollständiger Induktion aus (A2) bzw. aus der zweiten Teilaussage. □

Um Grenzwerte bilden zu können, ist es erforderlich, dass das Mengensystem  $\mathcal{F}$  nicht nur abgeschlossen ist bezüglich Vereinigung bzw. Durchschnitt von *endlich* vielen Mengen, sondern auch bezüglich Vereinigung bzw. Durchschnitt von *abzählbar unendlich* vielen Mengen. Dies wird durch die Hinzunahme der folgenden Bedingung erreicht.

**Definition** Eine Algebra  $\mathcal{F}$  heißt  $\sigma$ -Algebra, wenn zusätzlich gilt:

- (A3)  $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ .

**Beispiel einer Algebra, die keine  $\sigma$ -Algebra ist**

Sei  $\Omega = \mathbb{N}$ , und sei  $\mathcal{F}$  die Familie derjenigen Teilmengen  $A$  von  $\mathbb{N}$ , so dass entweder  $A$  oder  $A^c$  nur endlich viele Elemente hat. Das Mengensystem  $\mathcal{F}$  ist eine Algebra, jedoch keine  $\sigma$ -Algebra.

**Beachte**

- Das Paar  $(\Omega, \mathcal{F})$  heißt *Messraum*, falls  $\mathcal{F}$  eine  $\sigma$ -Algebra ist.
- Für jedes  $\Omega$  ist die Potenzmenge  $\mathcal{P}$ , d.h. die Familie *aller* Teilmengen von  $\Omega$ , stets eine  $\sigma$ -Algebra.
- Wenn  $\Omega$  endlich oder abzählbar unendlich ist, dann kann  $\mathcal{F} = \mathcal{P}$  gewählt werden. Bei nicht abzählbarem  $\Omega$  (z.B.  $\Omega = \mathbb{R}$  oder  $\Omega = [0, 1]$ ) muss eine kleinere  $\sigma$ -Algebra betrachtet werden (*nicht*  $\mathcal{P}$ ).

Außerdem benötigen wir die folgende Begriffe zur Bildung von Grenzwerten.

**Definition** Seien  $A, A_1, A_2, \dots \subset \Omega$  beliebige Teilmengen von  $\Omega$ .

- Dann heißt die Menge

$$\limsup_n A_n := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \quad \left( = \{ \omega \in \Omega : \forall k \geq 1 \exists n \geq k \text{ mit } \omega \in A_n \} \right) \quad (1)$$

der *Limes Superior* der Folge  $\{A_n\}$ ,

- und

$$\liminf_n A_n := \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n \quad \left( = \{ \omega \in \Omega : \exists k \geq 1 \forall n \geq k \text{ mit } \omega \in A_n \} \right) \quad (2)$$

heißt der *Limes Inferior* der Folge  $\{A_n\}$ .

- Außerdem sagt man, dass die Folge  $\{A_n\}$  gegen die Menge  $A$  *konvergiert*, falls

$$\liminf_n A_n = \limsup_n A_n \quad \left( \stackrel{\text{Def}}{=} A \right). \quad (3)$$

**Schreibweise** Wenn die Folge  $\{A_n\}$  gegen die Menge  $A$  konvergiert, d.h., wenn (3) gilt, dann schreiben wir  $A = \lim_n A_n$  bzw. einfach  $A_n \rightarrow A$ .

**Lemma 2.3** Seien  $A, A_1, A_2, \dots \subset \Omega$  beliebige Teilmengen von  $\Omega$ .

1. Wenn  $A_1 \subset A_2 \subset \dots$ , dann gilt  $A_n \rightarrow A := \bigcup_{n=1}^{\infty} A_n$ . (Schreibweise:  $A_n \uparrow A$ )
2. Wenn  $A_1 \supset A_2 \supset \dots$ , dann gilt  $A_n \rightarrow A := \bigcap_{n=1}^{\infty} A_n$ . (Schreibweise:  $A_n \downarrow A$ )

**Beweis** Wir zeigen nur die erste Teilaussage. Der Beweis der zweiten Teilaussage ist analog. Sei  $A = \bigcup_{n=1}^{\infty} A_n$ . Für jedes  $k$  gilt dann  $\bigcup_{n=k}^{\infty} A_n = A$ . Also ist  $\limsup_n A_n = A$ . Andererseits gilt  $\bigcap_{n=k}^{\infty} A_n = A_k$ , d.h.,  $\liminf_n A_n = \bigcup_{k=1}^{\infty} A_k = A$ .  $\square$

## 2.3 Wahrscheinlichkeitsmaße

### 2.3.1 Definition und elementare Eigenschaften

Gegeben sei ein Messraum  $(\Omega, \mathcal{F})$ . Betrachten eine Mengenfunktion, d.h. eine Abbildung  $P : \mathcal{F} \rightarrow [0, 1]$ , die jeder Menge  $A \in \mathcal{F}$  eine Zahl  $P(A) \in [0, 1]$  zuordnet. Dann heißt  $P(A)$  Wahrscheinlichkeit des Ereignisses  $A \in \mathcal{F}$ .

**Definition (Axiome von Kolmogorow)**

- Die Mengenfunktion  $P : \mathcal{F} \rightarrow [0, 1]$  heißt *Wahrscheinlichkeitsmaß* auf  $\mathcal{F}$ , falls
  - (P1)  $P(\Omega) = 1$  („Normiertheit“)
  - (P2)  $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$  für paarweise disjunkte  $A_1, A_2, \dots \in \mathcal{F}$  („ $\sigma$ -Additivität“)
- Wenn  $(\Omega, \mathcal{F})$  ein Messraum und  $P$  ein Wahrscheinlichkeitsmaß auf  $\mathcal{F}$  ist, dann heißt das Tripel  $(\Omega, \mathcal{F}, P)$  *Wahrscheinlichkeitsraum*.

**Theorem 2.1** Sei  $(\Omega, \mathcal{F}, P)$  ein Wahrscheinlichkeitsraum und  $A, A_1, A_2, \dots \in \mathcal{F}$ . Dann gilt

1.  $P(A^c) = 1 - P(A)$
2.  $A_1 \subset A_2 \Rightarrow P(A_1) \leq P(A_2)$
3.  $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
4.  $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$

**Beweis**

1. Es gilt

$$1 = P(\Omega) = P(A \cup A^c) = P(A \cup A^c \cup \emptyset \cup \emptyset \cup \dots) = P(A) + P(A^c) + P(\emptyset) + P(\emptyset) + \dots$$

Hieraus folgt, dass  $P(\emptyset) = 0$ , d.h.  $1 = P(\Omega) = P(A) + P(A^c)$ .

2. Außerdem gilt

$$\begin{aligned}
 P(A_2) &= P(A_1 \cup (A_2 \setminus A_1)) \\
 &= P(A_1 \cup (A_2 \setminus A_1) \cup \emptyset \cup \emptyset \dots) \\
 &= P(A_1) + \underbrace{P(A_2 \setminus A_1)}_{\geq 0} + \underbrace{P(\emptyset) + P(\emptyset) + \dots}_{=0} \\
 &\Rightarrow P(A_2) \geq P(A_1)
 \end{aligned}$$

3. Weil  $P(A \setminus B) = P(A) - P(A \cap B)$  für  $A, B \in \mathcal{F}$  mit  $A \supset B$  (vgl. den Beweis der Teilaussage 2) und weil

$$A_1 \cup A_2 = (A_1 \setminus (A_1 \cap A_2)) \cup (A_1 \cap A_2) \cup (A_2 \setminus (A_1 \cap A_2)),$$

gilt

$$\begin{aligned}
 P(A_1 \cup A_2) &= P((A_1 \setminus (A_1 \cap A_2)) \cup (A_1 \cap A_2) \cup (A_2 \setminus (A_1 \cap A_2))) \\
 &= P(A_1 \setminus (A_1 \cap A_2)) + P(A_1 \cap A_2) + P(A_2 \setminus (A_1 \cap A_2)) \\
 &= P(A_1) - P(A_1 \cap A_2) + P(A_1 \cap A_2) + P(A_2) - P(A_1 \cap A_2) \\
 &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \\
 &\leq P(A_1) + P(A_2).
 \end{aligned}$$

4. folgt unmittelbar aus 3. □

**Beachte** Aus dem Beweis bzw. aus den Aussagen von Theorem 2.1 ergibt sich sofort, dass

- $P(\emptyset) = 0$ ,

- $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$  für jede *endliche* Folge  $A_1, \dots, A_n \in \mathcal{F}$  von paarweise disjunkten Mengen,
- $P(A_2 \setminus A_1) = P(A_2) - P(A_1)$ , falls  $A_1 \subset A_2$ ,
- $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$  für jede beliebige Folge  $A_1, \dots, A_n \in \mathcal{F}$ .

In Verallgemeinerung der 3. Teilaussage von Theorem 2.1 ergibt sich außerdem die folgende *Siebformel*.

**Korollar 2.1** Für jedes  $n = 1, 2, \dots$  und jede Folge  $A_1, \dots, A_n \in \mathcal{F}$  gilt

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < \dots < k_i \leq n} P(A_{k_1} \cap \dots \cap A_{k_i}). \quad (4)$$

**Beweis** (Induktion)

- Induktionsanfang:  $n = 1$  ist klar; für  $n = 2$  ist (4) identisch mit der 3. Teilaussage von Theorem 2.1.
- Induktionsannahme: (4) gelte für ein  $n \geq 2$ .
- Induktionsschritt: Wir zeigen nun, dass dann (4) auch für  $n + 1$  gilt. Und zwar ist

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(A_{n+1} \setminus \left(A_{n+1} \cap \bigcup_{i=1}^n A_i\right) \cup \bigcup_{i=1}^n A_i\right) \\ &= P(A_{n+1}) - P\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right) + P\left(\bigcup_{i=1}^n A_i\right) \\ \text{Ind.-annahme} \quad &\stackrel{=}{=} P(A_{n+1}) - \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < \dots < k_i \leq n} P(A_{k_1} \cap \dots \cap A_{k_i} \cap A_{n+1}) \\ &\quad + \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < \dots < k_i \leq n} P(A_{k_1} \cap \dots \cap A_{k_i}) \\ &= \sum_{i=1}^{n+1} (-1)^{i-1} \sum_{1 \leq k_1 < \dots < k_i \leq n+1} P(A_{k_1} \cap \dots \cap A_{k_i}). \quad \square \end{aligned}$$

Darüber hinaus kann man mit Hilfe von Theorem 2.1 zeigen, dass Wahrscheinlichkeitsmaße stetig sind bezüglich der monotonen Konvergenz von Mengen.

**Korollar 2.2** Seien  $A_1, A_2, \dots \in \mathcal{F}$ . Dann gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i), \quad \text{falls } A_1 \subset A_2 \subset \dots, \quad (5)$$

und

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i), \quad \text{falls } A_1 \supset A_2 \supset \dots \quad (6)$$

**Beweis** Sei  $A_1 \subset A_2 \subset \dots$ . Mit der zusätzlichen Notation  $A_0 = \emptyset$  gilt dann

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} (A_i \setminus A_{i-1})\right) = \sum_{i=1}^{\infty} P(A_i \setminus A_{i-1}) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i \setminus A_{i-1}) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

Damit ist (5) bewiesen. Der Beweis von (6) ist analog. Dabei kann die bereits gezeigte Formel (5) genutzt werden, wenn zu den Komplementen übergegangen wird. Und zwar gilt

$$\begin{aligned} P\left(\bigcap_{i=1}^{\infty} A_i\right) &= 1 - P\left(\bigcup_{i=1}^{\infty} A_i^c\right) \stackrel{(5)}{=} 1 - \lim_{i \rightarrow \infty} P(A_i^c) \\ &= \lim_{i \rightarrow \infty} (1 - P(A_i^c)) = \lim_{i \rightarrow \infty} P(A_i). \end{aligned}$$

□

### 2.3.2 Weitere Eigenschaften von Wahrscheinlichkeitsmaßen

Die *Subadditivität* von Wahrscheinlichkeitsmaßen, die in Teilaussage 4 von Theorem 2.1 betrachtet wurde, gilt nicht nur für zwei bzw. endlich viele Ereignisse, sondern auch für Folgen von *unendlich* vielen Ereignissen.

**Theorem 2.2** Sei  $(\Omega, \mathcal{F}, P)$  ein Wahrscheinlichkeitsraum, und  $A_1, A_2, \dots \in \mathcal{F}$  seien beliebige Ereignisse. Dann gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i). \quad (7)$$

**Beweis** Anstelle der Folge  $\{A_n\}$  betrachten wir die Folge  $\{A'_n\}$  von paarweise disjunkten Mengen, wobei

$$A'_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i.$$

Dann gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A'_i\right) = \sum_{i=1}^{\infty} P(A'_i) \leq \sum_{i=1}^{\infty} P(A_i),$$

wobei sich die letzte Ungleichung aus der Teilaussage 2 von Theorem 2.1, d.h., aus der Monotonie von Wahrscheinlichkeitsmaßen ergibt, denn offenbar gilt  $A'_n \subset A_n$ . □

Das folgende Korollar wird in der Literatur das *Lemma von Borel–Cantelli* genannt, vgl. auch Theorem 2.7.

**Korollar 2.3** Sei  $(\Omega, \mathcal{F}, P)$  ein Wahrscheinlichkeitsraum und  $A_1, A_2, \dots \in \mathcal{F}$  eine beliebige Folge von Ereignissen. Dann gilt

$$P(\limsup_n A_n) = 0, \quad (8)$$

falls

$$\sum_{i=1}^{\infty} P(A_i) < \infty. \quad (9)$$

**Beweis** Aus Korollar 2.2 und Theorem 2.2 ergibt sich, dass

$$\begin{aligned} P(\limsup_n A_n) &= P\left(\bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} A_i\right) \stackrel{(6)}{=} \lim_{k \rightarrow \infty} P\left(\bigcup_{i=k}^{\infty} A_i\right) \\ &\stackrel{(7)}{\leq} \lim_{k \rightarrow \infty} \sum_{i=k}^{\infty} P(A_i) = 0, \end{aligned}$$

wobei sich die letzte Gleichheit aus der Summierbarkeitsbedingung (9) ergibt.  $\square$

Wir diskutieren nun den Zusammenhang zwischen der  $\sigma$ -Additivität und gewissen Stetigkeitseigenschaften von Mengenfunktionen.

**Theorem 2.3** Sei  $(\Omega, \mathcal{F})$  ein beliebiger Messraum, und sei  $P : \mathcal{F} \rightarrow [0, 1]$  eine beliebige additive Mengenfunktion, d.h.  $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$  gelte für jede endliche Folge von paarweise disjunkten Mengen  $A_1, \dots, A_n \in \mathcal{F}$ . Außerdem gelte  $P(\Omega) = 1$ . Dann sind die folgenden Aussagen äquivalent:

1.  $P$  ist  $\sigma$ -additiv (und damit ein Wahrscheinlichkeitsmaß),
2.  $P(A_n) \uparrow P(A)$ , falls  $A_n \uparrow A \in \mathcal{F}$ ,
3.  $P(A_n) \downarrow P(A)$ , falls  $A_n \downarrow A \in \mathcal{F}$ ,
4.  $P(A_n) \downarrow 0$ , falls  $A_n \downarrow \emptyset$ .

**Beweis** Wir führen einen zyklischen Beweis, d.h., wir zeigen, dass die folgenden Implikationen richtig sind:

$$1. \implies 2. \implies 3. \implies 4. \implies 1.$$

1.  $\implies$  2.  $P$  sei  $\sigma$ -additiv, und es gelte  $A_n \uparrow A \in \mathcal{F}$ . Die Behauptung ergibt sich dann genauso wie im Beweis von Korollar 2.2.
2.  $\implies$  3. Es gelte  $A_n \downarrow A \in \mathcal{F}$ , d.h. insbesondere, dass  $A = \bigcap_{n=1}^{\infty} A_n$ . Hieraus folgt, dass  $A_n^c \uparrow A^c \in \mathcal{F}$ , wobei  $A^c = \bigcup_{n=1}^{\infty} A_n^c$ . Also ergibt sich aus Teilaussage 2, dass

$$P(A_n) = 1 - P(A_n^c) \downarrow 1 - P(A^c) = P(A).$$

3.  $\implies$  4. Diese Implikation gilt offensichtlich, weil die Teilaussage 4 ein Spezialfall von Teilaussage 3 ist.
4.  $\implies$  1. Seien  $A_1, A_2, \dots \in \mathcal{F}$  paarweise disjunkte Mengen. Dann gilt

$$\bigcup_{i=n+1}^{\infty} A_i \downarrow \emptyset \quad \text{für } n \rightarrow \infty.$$

Also ist

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^n A_i \cup \bigcup_{i=n+1}^{\infty} A_i\right) \\ &\stackrel{\text{Additivität von } P}{=} \sum_{i=1}^n P(A_i) + P\left(\bigcup_{i=n+1}^{\infty} A_i\right) \\ &\stackrel{\text{Nullstetigkeit von } P}{\longrightarrow} \sum_{i=1}^{\infty} P(A_i), \end{aligned}$$

wobei sich die zweite Gleichheit aus der (endlichen) Additivität von  $P$  ergibt, die im Theorem vorausgesetzt wird.  $\square$



Für Wahrscheinlichkeitsmaße, d.h. für  $\sigma$ -additive Mengenfunktionen lassen sich die in Theorem 2.3 betrachteten drei Stetigkeitseigenschaften wie folgt zu *einer* Stetigkeitseigenschaft zusammenfassen.

**Theorem 2.4** *Sei  $(\Omega, \mathcal{F}, P)$  ein Wahrscheinlichkeitsraum, und  $A, A_1, A_2, \dots \in \mathcal{F}$  seien beliebige Ereignisse. Dann gilt*

$$P(A_n) \rightarrow P(A), \quad \text{falls } A_n \rightarrow A. \quad (10)$$

### Beweis

- Um die Behauptung zu zeigen, nutzen wir die folgenden beiden Ungleichungen:

$$P(\liminf_n A_n) \leq \liminf_{n \rightarrow \infty} P(A_n) \quad (11)$$

und

$$\limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_n A_n). \quad (12)$$

- Dabei ergibt sich die Ungleichung (11) aus Korollar 2.2 und Theorem 2.2:

$$\begin{aligned} P(\liminf_n A_n) &= P\left(\bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} A_i\right) \\ &\stackrel{(5)}{=} \lim_{k \rightarrow \infty} P\left(\bigcap_{i=k}^{\infty} A_i\right) \\ &\stackrel{\text{Teilaussage 2 von Theorem 2.2}}{\leq} \lim_{k \rightarrow \infty} \inf_{i \geq k} P(A_i) \\ &= \liminf_{k \rightarrow \infty} P(A_k). \end{aligned}$$

- Die Ungleichung (12) lässt sich auf analoge Weise zeigen. Weil

$$A = \liminf_n A_n = \limsup_n A_n,$$

ergibt sich nun aus (11) und (12), dass

$$\limsup_{n \rightarrow \infty} P(A_n) \leq P(A) \leq \liminf_{n \rightarrow \infty} P(A_n),$$

und damit, dass  $P(A) = \lim_n P(A_n)$ . □

## 2.4 Endliche Wahrscheinlichkeitsräume

### 2.4.1 Laplacescher Wahrscheinlichkeitsraum

Für jedes  $A \subset \Omega$  bezeichne  $|A|$  die Anzahl der Elemente, die zu  $A$  gehören. Es gelte  $|\Omega| < \infty$  (und damit auch  $|A| < \infty$  für jedes  $A \subset \Omega$ ). Ein Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  mit  $|\Omega| < \infty$  heißt *endlicher Wahrscheinlichkeitsraum*.

**Definition** Ein endlicher Wahrscheinlichkeitsraum  $(\Omega, \mathcal{P}(\Omega), P)$ , bei dem alle Elementarereignisse die *gleiche* Wahrscheinlichkeit haben, d.h.,  $P(\{\omega\}) = \frac{1}{|\Omega|}, \forall \omega \in \Omega$ , heißt *Laplacescher Wahrscheinlichkeitsraum*.

**Beachte** Sei  $(\Omega, \mathcal{P}(\Omega), P)$  ein Laplacescher Wahrscheinlichkeitsraum. Wegen der  $\sigma$ -Additivität von Wahrscheinlichkeitsmaßen gilt dann für jedes  $A \subset \Omega$

$$P(A) = \frac{|A|}{|\Omega|},$$

denn offenbar gilt  $A = \bigcup_{\omega \in A} \{\omega\}$  mit  $\{\omega\} \cap \{\omega'\} = \emptyset$  für  $\omega \neq \omega'$  und somit

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}.$$

Die so gegebene Wahrscheinlichkeit  $P(A) = |A|/|\Omega|$  heißt *Laplacesche Wahrscheinlichkeit* von  $A$ .

**Beispiel** (zweimaliges Würfeln)

$\omega = (i, j)$  ( $i$  = Augenzahl beim 1. Wurf;  $j$  = Augenzahl beim 2. Wurf)

$\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$ ;  $|\Omega| = 36$ ;  $\mathcal{F} = \mathcal{P}(\Omega)$ .

Sei beispielsweise  $A = \{\text{Gesamtaugenzahl} \geq 10\} = \{(6, 6), (6, 5), (5, 6), (6, 4), (4, 6), (5, 5)\}$ .

Dann gilt  $|A| = 6$  und somit  $P(A) = \frac{6}{36} = \frac{1}{6}$ .

### 2.4.2 Einfache Urnenmodelle

Gegeben sei eine Urne mit  $N$  Objekten (z.B. Kugeln), die mit den Zahlen  $1, 2, \dots, N$  numeriert werden. Aus dieser Urne werden  $n$  Objekte „zufällig“ entnommen. Ergebnis des gesamten Losvorganges ist ein  $n$ -Tupel  $(i_1, \dots, i_n)$ . Dabei gibt  $i_j$  die Nummer des Objektes an, das bei der  $j$ -ten Ziehung entnommen wird. Wir betrachten vier verschiedene Arten von Losvorgängen, die sich durch die folgenden Auswahlarten ergeben:

- *mit* Zurücklegen (d.h., Mehrfachziehungen sind möglich)
- *ohne* Zurücklegen (d.h., jedes Objekt kann maximal einmal gezogen werden)
- *mit* Reihenfolge (d.h.  $(1, 1, 4, 2) \neq (1, 2, 4, 1)$ )
- *ohne* Reihenfolge (d.h.  $(1, 1, 4, 2) = (1, 2, 4, 1)$ )

Sei  $D = \{1, 2, \dots, N\}$  die Menge der Objekte, die sich zu Beginn des Zufallsexperimentes in der Urne befinden. Die Grundmengen  $\Omega_I - \Omega_{IV}$ , die die vier verschiedenen Arten von Losvorgängen modellieren, haben die folgende Gestalt:

1. *Auswahl mit Reihenfolge und mit Zurücklegen*

$$\Omega_I = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in D \text{ für } i = 1, \dots, n\} = D^n$$

und somit  $|\Omega_I| = N^n$

2. *Auswahl mit Reihenfolge und ohne Zurücklegen*

$$\Omega_{II} = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in D, \omega_i \neq \omega_j \text{ für } i \neq j\}$$

1. Ziehung:  $N$  Möglichkeiten

2. Ziehung:  $N - 1$  Möglichkeiten

$\vdots$

$n$ -te Ziehung:  $(N - n + 1)$  Möglichkeiten

Also:  $|\Omega_{II}| = N \cdot (N - 1) \cdot \dots \cdot (N - n + 1) = \frac{N!}{(N-n)!}$

Wichtiger Spezialfall:  $n = N$  (*Permutationen*)  $\Rightarrow |\Omega_{II}| = N!$

## 3. Auswahl ohne Reihenfolge und ohne Zurücklegen

$$\Omega_{III} = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in D, \omega_1 < \omega_2 < \dots < \omega_n\}$$

Also:  $|\Omega_{III}| = \frac{\Omega_{II}}{n!} = \frac{N!}{(N-n)! \cdot n!} = \binom{N}{n}$  (Binomialkoeffizient)

## 4. Auswahl ohne Reihenfolge und mit Zurücklegen

$$\Omega_{IV} = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in D, \omega_1 \leq \omega_2 \leq \dots \leq \omega_n\}$$

und somit  $|\Omega_{IV}| = \binom{N+n-1}{n} = \frac{(N+n-1)!}{n! \cdot (N-1)!}$

## Zusammenfassung

Auswahl vom Umfang $n$ aus $\{1, 2, \dots, N\}$	mit Zurücklegen	ohne Zurücklegen	
mit Reihenfolge	$ \Omega_I  = N^n$	$ \Omega_{II}  = \frac{N!}{(N-n)!}$	unterscheidbare Marken
ohne Reihenfolge	$ \Omega_{IV}  = \binom{N+n-1}{n}$	$ \Omega_{III}  = \binom{N}{n}$	nicht unterscheidbare Marken
	mit Mehrfachbelegung	ohne Mehrfachbelegung	Verteilung von $n$ Marken auf $N$ Zellen

## Beispiele

1. Wir werfen 4 identische Würfel nacheinander und achten dabei auf die Reihenfolge der erzielten Augenzahlen.

*Frage:* Wie groß ist die Wahrscheinlichkeit, dass die vier Augenzahlen voneinander verschieden sind?

*Lösung:* Laplacescher Wahrscheinlichkeitsraum (Auswahlen mit Reihenfolge und Zurücklegen)

$$\Rightarrow |\Omega_I| = 6^4 \text{ (Anzahl der möglichen Fälle)}$$

$$A = \{\text{Auswahlen, bestehend aus 4 unterschiedlichen Augenzahlen}\} \subset \Omega_I$$

$$|A| = 6 \cdot 5 \cdot 4 \cdot 3 \text{ (Anzahl der günstigen Fälle)}$$

$$P(A) = \frac{|A|}{|\Omega_I|} = \frac{6 \cdot 5 \cdot 4 \cdot 3}{6^4} = \frac{5}{18}$$

2. Zahlenlotto:  $n = 6$  aus  $N = 49$  (ohne Reihenfolge und ohne Zurücklegen)

*Frage:* Wie groß ist die Wahrscheinlichkeit, mindestens 4 Richtige zu haben?

$$\text{Lösung: } D = \{1, \dots, 49\}, |\Omega_{III}| = \binom{49}{6}, \Omega_{III} = \{\omega = \{\omega_1, \dots, \omega_6\}, \omega_i \in D, \omega_1 < \dots < \omega_6\}$$

$$A_i = \{\text{genau } i \text{ Richtige}\}, P(A_4 \cup A_5 \cup A_6) = ?$$

$$\text{Weil } A_i \cap A_j = \emptyset \text{ für } i \neq j, \text{ gilt } P(A_4 \cup A_5 \cup A_6) = P(A_4) + P(A_5) + P(A_6).$$

$$\text{Dabei ist } |A_4| = \binom{6}{4} \binom{43}{2}, |A_5| = \binom{6}{5} \binom{43}{1}, |A_6| = \binom{6}{6} \binom{43}{0} = 1$$

$$\Rightarrow P(A) = \frac{\binom{6}{4} \binom{43}{2} + \binom{6}{5} \binom{43}{1} + 1}{\binom{49}{6}} = 0.000987$$

**Beachte** Dieses Beispiel ist ein Spezialfall der *hypergeometrischen Verteilung*.

**Hypergeometrische Verteilung**

Betrachten Urne mit  $N$  Objekten, wobei zwei Typen von Objekten vorhanden seien, mit den Teilanzahlen  $S$  und  $R$  (z.B.  $S$  schwarze Kugeln,  $R$  rote Kugeln);  $N = S + R$

Sei  $n$  = Anzahl der insgesamt entnommenen Kugeln;  $s$  = Anzahl der entnommenen schwarzen Kugeln

Sei  $P(s, n; S, N)$  die Wahrscheinlichkeit,  $s$  schwarze Kugeln bei der Entnahme von  $n$  Kugeln zu ziehen.

Dann gilt

$$P(s, n; S, N) = \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}}.$$

**2.5 Geometrische Wahrscheinlichkeiten**

Während bei der Definition der Laplace'schen Wahrscheinlichkeiten, vgl. Abschnitt 2.4.1,

*Quotienten von Anzahlen*

gebildet werden, betrachtet man bei geometrischen Wahrscheinlichkeiten

*Quotienten von Flächeninhalten bzw. Volumina.*

**Beispiel**

- Sei beispielsweise  $\Omega = [0, 1]^2$  das Einheitsquadrat in der euklidischen Ebene  $\mathbb{R}^2$ ,
  - und sei  $\mathcal{F} = \mathcal{B}([0, 1]^2)$  die *Borel- $\sigma$ -Algebra* von Teilmengen des Einheitsquadrates  $[0, 1]^2$ ,
  - d.h.,  $\mathcal{B}([0, 1]^2)$  ist die kleinste  $\sigma$ -Algebra von Teilmengen von  $[0, 1]^2$ , die alle Rechtecke  $(a, b) \times (a', b')$  enthält;  $0 < a < b < 1$ ,  $0 < a' < b' < 1$ . Schreibweise:

$$\mathcal{B}([0, 1]^2) = \sigma\left(\underbrace{\{(a, b) \times (a', b'), 0 < a < b < 1, 0 < a' < b' < 1\}}_{\text{Erzeugersystem}}\right).$$

- Die (geometrische) Wahrscheinlichkeit  $P(A)$  von  $A \in \mathcal{B}([0, 1]^2)$  ist dann gegeben durch den Ansatz

$$P(A) = \frac{|A|}{|[0, 1]^2|} = |A|,$$

wobei  $|A|$  den "Flächeninhalt" (genauer: den Wert des 2-dimensionalen Lebesgue-Maßes) von  $A$  bezeichnet.

Eine allgemeinere Variante der Definition der geometrischen Wahrscheinlichkeit lautet wie folgt.

**Definition**

- Sei  $d \geq 1$  eine (beliebige, jedoch fest vorgegebene) natürliche Zahl, und sei  $\mathbb{R}^d$  der  $d$ -dimensionale euklidische Raum.
  - Weiter sei  $\mathcal{B}(\mathbb{R}^d)$  die Borel- $\sigma$ -Algebra auf  $\mathbb{R}^d$ , d.h.,

$$\mathcal{B}(\mathbb{R}^d) = \sigma\left(\{(a_1, b_1) \times \dots \times (a_d, b_d), a_i, b_i \in \mathbb{R}^d, a_i < b_i\}\right).$$

- Sei  $\Omega \in \mathcal{B}(\mathbb{R}^d)$  eine beliebige (Borelsche) Teilmenge in  $\mathbb{R}^d$  mit

$$0 < |\Omega| < \infty,$$

wobei  $|\Omega|$  das  $d$ -dimensionale „Volumen“ (genauer: den Wert des  $d$ -dimensionalen Lebesgue-Maßes) von  $\Omega$  bezeichnet.

- Die Wahrscheinlichkeit von  $A \in \mathcal{B}(\mathbb{R}^d) \cap \Omega$  ist dann gegeben durch

$$P(A) = \frac{|A|}{|\Omega|}, \quad (13)$$

wobei  $\mathcal{B}(\mathbb{R}^d) \cap \Omega$  die Spur- $\sigma$ -Algebra der Borel-Mengen in  $\Omega$  bezeichnet.

- Das Tripel  $(\Omega, \mathcal{B}(\mathbb{R}^d) \cap \Omega, P)$  heißt dann *geometrischer Wahrscheinlichkeitsraum*.

### Beispiel (Buffonsches Nadelexperiment)

- Das Buffonsche Nadelexperiment ist eine Methode zur (numerischen) Bestimmung der Zahl  $\pi$ , die auf stochastischen Gesetzmäßigkeiten beruht.
  - Der „Erfinder“ ist Georges Louis Leclerc Comte de Buffon (1707–1788).
  - Algorithmische Versionen solcher Verfahren sind unter der Bezeichnung „Monte-Carlo-Simulation“ bekannt.
- Betrachten das System

$$K = \{(x, y) : (x, y) \in \{\dots, -1, 0, 1, \dots\} \times \mathbb{R}\} \subset \mathbb{R}^2$$

von parallelen und äquidistanten (vertikalen) Geraden in der euklidischen Ebene  $\mathbb{R}^2$ .

- Werfen eine Nadel (d.h. eine Strecke) mit der Länge 1 „willkürlich“ in die Ebene  $\mathbb{R}^2$ , wobei mit „willkürlich“ das folgende stochastische Modell gemeint ist.
- Betrachten die Größen  $s$  und  $t$ , die die relative Lage der Nadel bezüglich des Geradensystems  $K$  beschreiben, wobei
  - $s$  der (orthogonale) Abstand des Nadelmittelpunktes zur nächsten linksliegenden Nachbargeraden von  $K$  ist,
  - $t$  der Winkel ist, den die Nadel zum Lot auf die Geraden von  $K$  bildet.
- Betrachten den geometrischen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{B}(\mathbb{R}^2) \cap \Omega, P)$  mit  $\Omega = [0, 1] \times [-\pi/2, \pi/2]$ .
- Bestimmen die Wahrscheinlichkeit des Ereignisses  $A \subset \Omega$  mit

$$A = \left\{ (s, t) \in \Omega : 0 < s < \frac{1}{2} \cos t \right\} \cup \left\{ (s, t) \in \Omega : 1 - \frac{1}{2} \cos t < s < 1 \right\},$$

dass die Nadel eine der Geraden von  $K$  schneidet.

- Es gilt

$$\begin{aligned} P(A) &= P\left((s, t) \in \Omega : 0 < s < \frac{1}{2} \cos t\right) + P\left((s, t) \in \Omega : 1 - \frac{1}{2} \cos t < s < 1\right) \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t \, dt + \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t \, dt \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos t \, dt = \frac{2}{\pi}, \end{aligned}$$

d.h., für die (geometrische) Wahrscheinlichkeit des Ereignisses  $A$ , dass die Nadel eine der Geraden von  $K$  schneidet, gilt

$$P(A) = \frac{2}{\pi}. \quad (14)$$

**Beachte**

- Aus der Gleichung (14) ergibt sich eine Methode zur experimentellen Bestimmung der Zahl  $\pi$ , die auf dem sogenannten *Gesetz der großen Zahlen* beruht, vgl. Abschnitt 5.2.2.
- Und zwar werfen wir die Nadel  $n$ -mal, wobei  $n$  eine hinreichend große natürliche Zahl sein sollte.
- Seien  $(s_1, t_1), \dots, (s_n, t_n)$  die Ergebnisse der  $n$  durchgeführten Experimente.
- Betrachten die Funktionswerte  $x_1 = x(s_1, t_1), \dots, x_n = x(s_n, t_n)$  mit

$$x(s, t) = \begin{cases} 1, & \text{falls } s < \frac{1}{2} \cos t \text{ oder } 1 - \frac{1}{2} \cos t < s, \\ 0 & \text{sonst,} \end{cases}$$

d.h., die Indikatoren der Ereignisse, ob die Nadel beim jeweiligen Wurf eine der Geraden von  $K$  schneidet oder nicht.

- Aus (14) und aus dem starken Gesetz der großen Zahlen folgt dann, dass das arithmetische Mittel  $y_n = n^{-1} \sum_{i=1}^n x_i$  mit „großer Wahrscheinlichkeit“ eine gute Näherung der Zahl  $2/\pi$  ist.
- Mit anderen Worten: Für große  $n$  ist  $2/y_n$  mit großer Wahrscheinlichkeit eine gute Näherung der Zahl  $\pi$ .

Das folgende Beispiel soll deutlich machen, dass es (ähnlich wie bei der Laplace'schen Wahrscheinlichkeit) auch bei der geometrischen Wahrscheinlichkeit sehr wichtig ist, die Grundmenge  $\Omega$  geeignet zu wählen.

**Beispiel** (Bertrandsches Paradoxon)

- In den Kreis  $B(o, 1) \subset \mathbb{R}^2$  mit Mittelpunkt im Nullpunkt und Radius Eins werde „auf gut Glück“ eine Sehne gelegt.
- Wie groß ist die Wahrscheinlichkeit des Ereignisses  $A$ , dass die Sehne länger als  $\sqrt{3}$  ist (wobei  $\sqrt{3}$  die Seitenlänge des einbeschriebenen gleichseitigen Dreiecks ist).
- **Beachte:** Das Problem ist „inkorrekt“ gestellt und erfordert zunächst eine Präzisierung, was genau mit der Sprechweise „auf gut Glück“ gemeint ist.
- **Modell 1:** Der Mittelpunkt der Sehne werde „auf gut Glück“ in den Kreis  $B(o, 1)$  gelegt. Mit  $\Omega = B(o, 1)$  und  $A = \{\omega \in \Omega : 0 \leq |\omega| < 0.5\}$ , wobei  $|\omega|$  die Länge des Vektors  $\omega$  bezeichnet, ergibt sich dann die Wahrscheinlichkeit

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\pi/4}{\pi} = \frac{1}{4}.$$

- **Modell 2:** Ein Endpunkt der Sehne sei fest vorgegeben, und der andere Endpunkt werde „auf gut Glück“ auf die Kreislinie  $\partial B(o, 1)$  gelegt. Mit  $\Omega = (-\pi/2, \pi/2)$  und  $A = \{\omega \in \Omega : -\pi/6 < \omega < \pi/6\}$  ergibt sich dann die Wahrscheinlichkeit

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\pi/3}{\pi} = \frac{1}{3}.$$

- **Modell 3:** Die Richtung der Sehne sei fest vorgegeben (o.B.d.A. vertikal), und der Mittelpunkt der Sehne werde „auf gut Glück“ in das Intervall  $(-1, 1)$  gelegt. Mit  $\Omega = (-1, 1)$  und  $A = \{\omega \in \Omega : -0.5 < \omega < 0.5\}$ , ergibt sich dann die Wahrscheinlichkeit

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1}{2}.$$

## 2.6 Bedingte Wahrscheinlichkeiten

Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum.

### 2.6.1 Definition und Multiplikationssatz

Häufig verfügen wir bei der Durchführung von Experimenten über Vorinformationen, die bei der Berechnung von Wahrscheinlichkeiten interessierender Ereignisse berücksichtigt werden sollen.

Bei manchen Untersuchungen wird jedoch lediglich (hypothetisch) angenommen, dass eine bestimmte Vorinformation vorliegt, wobei dann unter dieser hypothetischen Annahme gerechnet wird. Diese sogenannte *Bayessche Methodik* wird im weiteren Verlauf der Vorlesung noch genauer diskutiert.

### Beispiele

#### 1. Skatenspiel

- Die Kenntnis der eigenen 10 Karten soll als Vorinformation über die Verteilung der übrigen 22 Karten genutzt werden.
  - Markieren die 32 Karten mit den Zahlen  $1, 2, \dots, 32$ .
  - Betrachten Laplaceschen Wahrscheinlichkeitsraum, wobei  $\Omega$  die Menge aller Permutationen von 32 Elementen ist ( $N = n = 32$ ; mit Reihenfolge und ohne Zurücklegen)
  - Gesucht sei die Wahrscheinlichkeit des Ereignisses  $A_2 \cap A_3$ , wobei  $A_2 = \{\text{Spieler 2 hat } x \text{ Asse}\}$ ,  $A_3 = \{\text{Spieler 3 hat } y \text{ Asse}\}$ , unter der Bedingung, dass das Ereignis  $A_1 = \{\text{Spieler 1 hat die Karten mit den Nummern } k_1, \dots, k_{10}\}$  eintritt.
- *Lösungsansatz:* Beziehen die Anzahl der Permutationen, bei denen  $A_2 \cap A_3$  eintritt, nicht auf die Gesamtanzahl  $32!$  aller möglichen Permutationen, sondern lediglich auf diejenigen Permutationen, bei denen das Ereignis  $A_1$  eintritt.
  - D.h., die gesuchte Wahrscheinlichkeit ist die (bedingte) relative Häufigkeit  $|(A_2 \cap A_3) \cap A_1|/|A_1|$
  - Dabei benutzen wir die Schreibweise:

$$P(A_2 \cap A_3 \mid A_1) = \frac{|(A_2 \cap A_3) \cap A_1|}{|A_1|}$$

und nennen diese Größe *bedingte Wahrscheinlichkeit* des Ereignisses  $A_2 \cap A_3$  unter der Bedingung, dass das Ereignis  $A_1$  eintritt.

#### 2. Urnenmodell

- Betrachten Urne mit  $N$  Objekten ( $S$  schwarze,  $R$  rote Kugeln), d.h.  $N = S + R$ , vgl. Abschnitt 3.2.2;
- 2 Objekte,  $2 \leq N$ , sollen insgesamt ausgewählt werden (ohne Zurücklegen);
- Sei  $A$  das Ereignis, beim zweiten Versuch „schwarz“ zu ziehen, und sei  $B$  das Ereignis, beim ersten Versuch „rot“ zu ziehen.
- Gesucht ist die bedingte Wahrscheinlichkeit  $P(A \mid B)$ , beim zweiten Versuch „schwarz“ zu ziehen, falls beim ersten Versuch „rot“ gezogen wird.
- Es gilt

$$P(A \mid B) = \frac{|A \cap B|}{|B|} = \frac{R \cdot S}{R \cdot S + R(R-1)} = \frac{S}{N-1}.$$

Dies führt zu der folgenden (allgemeineren) Begriffsbildung.

**Definition** Seien  $A, B \in \mathcal{F}$  beliebige Ereignisse mit  $P(B) > 0$ . Dann heißt

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (15)$$

die *bedingte Wahrscheinlichkeit* von  $A$  unter der Bedingung  $B$ .

**Beachte** Die Definitionsgleichung (15) kann in der Form  $P(A \cap B) = P(B)P(A | B)$  geschrieben werden. Durch Iteration dieser Überlegung ergibt sich der folgende *Multiplikationssatz*.

**Theorem 2.5** Seien  $A_1, A_2, \dots, A_n \in \mathcal{F}$  Ereignisse mit  $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$ . Dann gilt:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \dots P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}). \quad (16)$$

**Beweis** klar

**Beispiel** (Skatspiel)

- Betrachten das Ereignis  $A_i = \{\text{Spieler } i \text{ erhält genau ein As}\}; i = 1, 2, 3$ .
- Gesucht ist die Wahrscheinlichkeit  $P(A_1 \cap A_2 \cap A_3)$ , dass jeder der drei Spieler genau ein As erhält?
- *Lösung:* Es gilt

$$P(A_1) = \frac{\binom{4}{1} \binom{28}{9}}{\binom{32}{10}} = 0.428, \quad P(A_2 | A_1) = \frac{\binom{3}{1} \binom{19}{9}}{\binom{22}{10}} = 0.42857, \quad P(A_3 | A_1 \cap A_2) = \frac{\binom{2}{1} \binom{10}{9}}{\binom{12}{10}} = 0.303.$$

- Hieraus und aus (16) ergibt sich

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) = 0.0556.$$

## 2.6.2 Formel der totalen Wahrscheinlichkeit; Bayessche Formel

Bei der Berechnung der Wahrscheinlichkeit  $P(A)$  eines Ereignisses  $A \in \mathcal{F}$  ist es manchmal nützlich, die (unbedingte) Wahrscheinlichkeit  $P(A)$  als gewichtete Summe von bedingten Wahrscheinlichkeiten darzustellen.

Hierfür ist es erforderlich, den Grundraum  $\Omega$  wie folgt in (messbare) Teilmengen zu zerlegen.

**Definition** Sei  $n \in \mathbb{N}$  eine beliebige natürliche Zahl, und sei  $B_1, B_2, \dots, B_n \in \mathcal{F}$  eine (endliche) Folge von Ereignissen mit den Eigenschaften

- (Z1)  $B_i \cap B_j = \emptyset$  für  $i \neq j$ ,
- (Z2)  $\bigcup_{i=1}^n B_i = \Omega$ ,
- (Z3)  $P(B_i) > 0$  für alle  $i = 1, \dots, n$ .

Dann heißt  $B_1, B_2, \dots, B_n$  *messbare Zerlegung* von  $\Omega$ .

**Theorem 2.6** Sei  $A \in \mathcal{F}$  ein beliebiges Ereignis und  $B_1, B_2, \dots, B_n$  eine messbare Zerlegung von  $\Omega$ . Dann gilt

- Formel der totalen Wahrscheinlichkeit

$$P(A) = \sum_{j=1}^n P(B_j)P(A | B_j), \quad (17)$$



- Bayessche Formel

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)} \quad (18)$$

für jedes  $i = 1, \dots, n$ , wobei in (18) vorausgesetzt wird, dass  $P(A) > 0$ .

**Beweis** Aus (Z1)–(Z3) und aus der Additivität des Wahrscheinlichkeitsmaßes  $P$  ergibt sich, dass

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P\left(A \cap \left(\bigcup_{i=1}^n B_i\right)\right) \\ &= P\left(\bigcup_{i=1}^n (A \cap B_i)\right) = \sum_{i=1}^n P(A \cap B_i) \\ &= \sum_{i=1}^n P(B_i) \frac{P(A \cap B_i)}{P(B_i)} = \sum_{i=1}^n P(B_i)P(A | B_i), \end{aligned}$$

wobei im letzten Schritt die Definitionsgleichung (15) benutzt wird. Damit ist (17) bewiesen. Aus (15) und (17) ergibt sich nun

$$P(B_i | A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)}. \quad \square$$

**Beachte** Die Aussagen von Theorem 2.6 bleiben gültig, wenn anstelle einer Zerlegung von  $\Omega$  in endlich viele Teilmengen eine *unendliche* Folge  $B_1, B_2, \dots \in \mathcal{F}$  von Ereignissen mit den Eigenschaften

$$(Z'1) \quad B_i \cap B_j = \emptyset \text{ für } i \neq j,$$

$$(Z'2) \quad \bigcup_{i=1}^{\infty} B_i = \Omega,$$

$$(Z'3) \quad P(B_i) > 0 \text{ für alle } i = 1, 2, \dots$$

betrachtet wird. Die Formeln (17) und (18) sind dann lediglich wie folgt zu modifizieren:

$$P(A) = \sum_{j=1}^{\infty} P(B_j)P(A | B_j), \quad (19)$$

bzw.

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^{\infty} P(B_j)P(A | B_j)} \quad (20)$$

für jedes  $i = 1, 2, \dots$ , wobei in (20) erneut vorausgesetzt wird, dass  $P(A) > 0$ .

### Beispiel

- Betrachten eine Fußballmannschaft, deren Siegeschance je Bundesliga-Spiel bei 75% liegt, falls ihr Kapitän in guter Form ist.
  - Wenn ihr Kapitän jedoch nicht in guter Form ist, dann betrage ihre Siegeschance nur 40%.
  - Bei 70% aller Bundesliga-Spiele seiner Mannschaft sei der Kapitän in guter Form.
  - Gesucht ist die Wahrscheinlichkeit, dass
    1. die Mannschaft ein Bundesliga-Spiel gewinnt,
    2. der Kapitän bei einem Bundesliga-Spiel in guter Form ist, obwohl die Mannschaft das Spiel nicht gewinnt.
- *Lösung:* Zerlegen den Grundraum  $\Omega$  auf zwei verschiedene Weisen in zwei Komponenten.

- Sei  $A = \{\text{Mannschaft gewinnt Bundesliga-Spiel}\}$ ,  $A^c = \{\text{Mannschaft gewinnt Bundesliga-Spiel nicht}\}$   
bzw.  
 $B = \{\text{Kapitän ist in guter Form}\}$ ,  $B^c = \{\text{Kapitän ist nicht in guter Form}\}$
- Dann gilt  $P(A | B) = 0.75$ ,  $P(A | B^c) = 0.40$ ,  $P(B) = 0.70$
- Aus (17) bzw. (18) ergibt sich nun

$$\begin{aligned} P(A) &= P(A | B)P(B) + P(A | B^c)P(B^c) \\ &= 0.75 \cdot 0.70 + 0.40 \cdot 0.30 = 0.645 \end{aligned}$$

bzw.

$$\begin{aligned} P(B | A^c) &= \frac{P(A^c | B)P(B)}{P(A^c | B)P(B) + P(A^c | B^c)P(B^c)} \\ &= \frac{0.25 \cdot 0.70}{0.25 \cdot 0.70 + 0.60 \cdot 0.30} = 0.493. \end{aligned}$$

## 2.7 Stochastische Unabhängigkeit

Der Begriff der stochastischen Unabhängigkeit zweier Ereignisse  $A, B \in \mathcal{F}$  ist mit der intuitiven Vorstellung verbunden, dass die bedingte Wahrscheinlichkeit  $P(A | B)$  des Ereignisses  $A$  unter der Bedingung  $B$  mit der „unbedingten“ Wahrscheinlichkeit  $P(A)$  von  $A$  übereinstimmt, d.h.  $P(A | B) = P(A)$ , wobei  $P(B) > 0$  vorausgesetzt wird.

Es ist jedoch zweckmäßiger, die folgende (äquivalente) Gleichung  $P(A \cap B) = P(A)P(B)$  zu betrachten, weil durch sie auch der Fall  $P(B) = 0$  erfasst wird.

### Definition

1. Die Ereignisse  $A, B \in \mathcal{F}$  heißen *unabhängig*, falls  $P(A \cap B) = P(A)P(B)$ .
2. Sei  $A_1, A_2, \dots, A_n \in \mathcal{F}$  eine beliebige Familie von Ereignissen. Dann sagt man, dass  $A_1, A_2, \dots, A_n$  *unabhängige Ereignisse* sind, falls für jede Teilmenge  $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$  gilt:

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k P(A_{i_j}). \quad (21)$$

### Beachte

1. Der Begriff der Unabhängigkeit wird auch für *unendliche* Familien von Ereignissen benötigt. Man sagt, dass  $A_1, A_2, \dots \in \mathcal{F}$  *unabhängige Ereignisse* sind, falls für jede *endliche* Teilmenge  $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots\}$  die Bedingung (21) erfüllt ist.
2. Das folgende Beispiel zeigt, dass die Unabhängigkeit von Ereignis-Paaren  $A_{i_1}, A_{i_2}$  im allgemeinen *nicht* die (vollständige) Unabhängigkeit der gesamten Folge  $A_1, A_2, \dots, A_n$  impliziert.

### Beispiel

- Sei  $(\Omega, \mathcal{F}, P)$  gegeben durch  $\Omega = \{1, 2, 3, 4\}$ ,  $\mathcal{F} = \mathcal{P}(\Omega)$ ,  $P(\{k\}) = \frac{1}{4}$  für jedes  $k \in \Omega$ .
- Sei  $A_k = \{k, 4\}$  für  $k = 1, 2, 3$ . Dann sieht man leicht, dass
  1. die Paare  $A_1, A_2$  bzw.  $A_2, A_3$  bzw.  $A_1, A_3$  jeweils unabhängig sind,
  2. die Ereignisse  $A_1, A_2, A_3$  jedoch *nicht* unabhängig sind.

- Denn es gilt

$$P(A_i \cap A_{i+1}) = P(\{4\}) = \frac{1}{4} = P(A_i) \cdot P(A_{i+1}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad \text{für } i = 1, 2$$

bzw.

$$P(A_1 \cap A_3) = P(\{4\}) = \frac{1}{4} = P(A_1) \cdot P(A_3) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

- Jedoch

$$P(A_1 \cap A_2 \cap A_3) = P(\{4\}) = \frac{1}{4} \neq P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{1}{8}.$$

Die folgende Invarianzeigenschaft unabhängiger Ereignisse bezüglich der Komplementbildung wird im weiteren Verlauf der Vorlesung mehrfach benötigt.

**Lemma 2.4** Sei  $A_1, \dots, A_n \in \mathcal{F}$  eine beliebige Folge von unabhängigen Ereignissen. Dann sind auch die Ereignisse  $A_1^c, \dots, A_n^c$  unabhängig.

### Beweis

- Wir betrachten nur den Fall  $n = 2$ . Dann ergibt sich aus den allgemeinen Rechenregeln für Wahrscheinlichkeitsmaße (vgl. Theorem 2.1) und aus der Unabhängigkeit der Ereignisse  $A_1$  und  $A_2$ , dass

$$\begin{aligned} P(A_1^c)P(A_2^c) &= (1 - P(A_1))(1 - P(A_2)) = 1 - P(A_1) - P(A_2) + P(A_1)P(A_2) \\ &= 1 - P(A_1) - P(A_2) + P(A_1 \cap A_2) = 1 - P(A_1 \cup A_2) \\ &= P((A_1 \cup A_2)^c) = P(A_1^c \cap A_2^c). \end{aligned}$$

- Für beliebiges  $n \geq 2$  ergibt sich die Behauptung auf ähnliche Weise mit vollständiger Induktion.  $\square$

In Ergänzung von Korollar 2.3 können wir nun den zweiten Teil des *Lemmas von Borel–Cantelli* formulieren und beweisen.

**Theorem 2.7** Sei  $A_1, A_2, \dots \in \mathcal{F}$  eine beliebige Folge von unabhängigen Ereignissen, so dass

$$\sum_{i=1}^{\infty} P(A_i) = \infty. \quad (22)$$

Dann gilt

$$P(\limsup_n A_n) = 1. \quad (23)$$

### Beweis

- Für reelle Zahlen  $\alpha_k, \dots, \alpha_{k+l}$  mit  $k, l = 1, 2, \dots$  gilt bekanntlich  $\log(1 - \alpha_i) \leq -\alpha_i$ , falls  $0 \leq \alpha_i \leq 1$ .  
 $\implies$  Hieraus folgt, dass

$$\log\left(\prod_{n=k}^{k+l} (1 - \alpha_n)\right) \leq -\sum_{n=k}^{k+l} \alpha_n \quad \text{bzw.} \quad \prod_{n=k}^{k+l} (1 - \alpha_n) \leq \exp\left(-\sum_{n=k}^{k+l} \alpha_n\right).$$

- Weil mit den Ereignissen  $A_1, A_2, \dots$  auch die Ereignisse  $A_1^c, A_2^c, \dots$  unabhängig sind (vgl. Lemma 2.4), ergibt sich somit

$$\begin{aligned} P\left(\bigcap_{n=k}^{k+l} A_n^c\right) &= \prod_{n=k}^{k+l} (1 - P(A_n)) \\ &\leq \exp\left(-\sum_{n=k}^{k+l} P(A_n)\right). \end{aligned}$$

- Aus (22) folgt, dass bei festem  $k$  die rechte Seite für  $l \rightarrow \infty$  gegen 0 strebt.  
 $\Rightarrow$  Hieraus ergibt sich mit Hilfe von Korollar 2.2, dass

$$P\left(\bigcap_{n=k}^{\infty} A_n^c\right) \stackrel{\text{Korollar 2.2}}{=} \lim_{l \rightarrow \infty} P\left(\bigcap_{n=k}^{k+l} A_n^c\right) \leq \lim_{l \rightarrow \infty} \exp\left(-\sum_{n=k}^{k+l} P(A_n)\right) = 0.$$

$\Rightarrow$  Hieraus ergibt sich, dass

$$1 - P(\limsup_n A_n) = P\left(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n^c\right) \leq \sum_{k=1}^{\infty} P\left(\bigcap_{n=k}^{\infty} A_n^c\right) = 0.$$

□

### 3 Zufallsvariablen und Zufallsvektoren

#### 3.1 Definition von Zufallsvariablen

Betrachten einen beliebigen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  und ein beliebiges Element  $\omega \in \Omega$ , wobei wir so wie bisher  $\{\omega\}$  als Elementarereignis bzw. Versuchsergebnis interpretieren.

**Beachte** Häufig interessiert nicht  $\omega$  selbst, sondern eine (quantitative oder qualitative) Kennzahl  $X(\omega)$  von  $\omega$ , d.h., wir betrachten die Abbildung  $\omega \rightarrow X(\omega)$ .

#### Beispiele

1.  $\Omega$  = Menge von Eintragungen in einem Telefonbuch  
 $\omega$  = Familienname,  $X(\omega)$  = Anzahl der Buchstaben von  $\omega$   
oder  
 $\omega$  = Telefonnummer,  $X(\omega)$  = Anzahl der Ziffer „1“ in  $\omega$
2. zweimaliges Würfeln  $\Omega = \{\omega = (\omega_1; \omega_2), \omega_i \in \{1, \dots, 6\}\}$   
Augensumme  $\omega \rightarrow X(\omega) = \omega_1 + \omega_2$   
Sei  $A = \{\omega : X(\omega) = 10\} = \{(6, 4), (5, 5), (4, 6)\}$  bzw. allgemeiner  $A = \{\omega : X(\omega) = k\}$ , wobei  $k \in \{2, \dots, 12\}$ .  
Gesucht ist die Wahrscheinlichkeit  $P(A)$ . Hierfür ist es erforderlich, dass  $A \in \mathcal{F}$ .  
Allgemein muss also  $\{\omega : \omega \in \Omega, X(\omega) = k\} \in \mathcal{F}$  für jedes  $k = 2, \dots, 12$  gelten.  
Bei diesem Beispiel ist das gleichbedeutend mit  $\{\omega : \omega \in \Omega, X(\omega) \leq x\} \in \mathcal{F}$  für jedes  $x \in \mathbb{R}$ .

Das führt zu der folgenden Begriffsbildung.

**Definition** Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum. Die Abbildung  $X : \Omega \rightarrow \mathbb{R}$  heißt *Zufallsvariable* (bzw. *Zufallsgröße*), falls

$$\{\omega : \omega \in \Omega, X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}. \quad (1)$$

#### Beachte

1. Die Regularitätsbedingung (1) wird *Messbarkeit* der Abbildung  $X$  bezüglich der  $\sigma$ -Algebra  $\mathcal{F}$  genannt.
2. In vielen Fällen interessiert nicht nur die Wahrscheinlichkeit, dass die Werte  $X(\omega)$  der Zufallsvariablen  $X$  einen vorgegebenen *Schwellenwert*  $x$  nicht überschreiten, d.h., dass  $X$  Werte im Intervall  $B = (-\infty, x]$  annimmt.
3. Oftmals interessiert auch die Wahrscheinlichkeit, dass  $X$  Werte in einer allgemeineren Teilmenge  $B \subset \mathbb{R}$  annimmt, wobei  $B$  beispielsweise die Vereinigung von mehreren disjunkten Intervallen sein kann.
4. Deshalb wird nicht nur im Grundraum  $\Omega$ , sondern auch im *Bildraum*  $\mathbb{R}$  ein System von Teilmengen von  $\mathbb{R}$  betrachtet, das abgeschlossen bezüglich der Mengenoperationen  $\cup, \cap, \setminus$  ist.
5. Dabei wird oft die *Borel- $\sigma$ -Algebra*  $\mathcal{B}(\mathbb{R})$  betrachtet, die definiert ist als die kleinste  $\sigma$ -Algebra von Teilmengen von  $\mathbb{R}$ , die alle offenen Intervalle  $(a, b)$  enthält;  $-\infty < a < b < \infty$ . D.h.

$$\mathcal{B}(\mathbb{R}) = \sigma \left( \underbrace{\{(a, b), -\infty < a < b < \infty\}}_{\text{Erzeugersystem}} \right)$$

6. Insbesondere enthält  $\mathcal{B}(\mathbb{R})$  auch alle halboffenen bzw. abgeschlossenen Intervalle, denn es gilt

$$(a, b] = \bigcap_{n=1}^{\infty} (a, b + n^{-1}) \in \mathcal{B}(\mathbb{R}), \quad [a, b) = \bigcap_{n=1}^{\infty} (a - n^{-1}, b) \in \mathcal{B}(\mathbb{R}), \quad [a, b] = \bigcap_{n=1}^{\infty} (a - n^{-1}, b + n^{-1}) \in \mathcal{B}(\mathbb{R}).$$

7. Für jede abzählbare Teilmenge  $C = \{x_1, x_2, \dots\}$  von  $\mathbb{R}$  gilt  $C \in \mathcal{B}(\mathbb{R})$ , denn für jedes  $x \in \mathbb{R}$  gilt  $\{x\} = \bigcap_{n=1}^{\infty} (x - n^{-1}, x + n^{-1}) \in \mathcal{B}(\mathbb{R})$  und damit auch  $C = \bigcup_{i=1}^{\infty} \{x_i\} \in \mathcal{B}(\mathbb{R})$ .

### 3.2 Verteilung und Verteilungsfunktion

Die Regularitätsbedingung (1) kann durch die folgende (scheinbar schärfere, in Wirklichkeit jedoch äquivalente) Bedingung ersetzt werden.

**Theorem 3.1** Die Abbildung  $X : \Omega \rightarrow \mathbb{R}$  ist genau dann eine Zufallsvariable, wenn

$$\{\omega : \omega \in \Omega, X(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{B}(\mathbb{R}). \quad (2)$$

**Beweis** Offenbar folgt (1) aus (2). Es genügt also zu zeigen, dass auch umgekehrt (2) aus (1) folgt.

- Es gelte also die Zugehörigkeitsrelation (1).
- Wir zeigen zuerst, dass das Mengensystem

$$\mathcal{G} = \{B : B \in \mathcal{B}(\mathbb{R}), X^{-1}(B) \in \mathcal{F}\} \quad (3)$$

eine  $\sigma$ -Algebra ist, wobei  $X^{-1}(B) = \{\omega : \omega \in \Omega, X(\omega) \in B\}$  das Urbild von  $B$  bezüglich der Abbildung  $X$  ist.

- Es ist klar, dass  $\mathbb{R} \in \mathcal{G}$ , weil  $X^{-1}(\mathbb{R}) = \Omega \in \mathcal{F}$ .
- Außerdem gilt  $B^c \in \mathcal{G}$  für jedes  $B \in \mathcal{G}$ , weil  $X^{-1}(B^c) = (X^{-1}(B))^c \in \mathcal{F}$ .
- Analog ergibt sich, dass  $B_1 \cup B_2 \in \mathcal{G}$  für beliebige  $B_1, B_2 \in \mathcal{G}$ , weil
 
$$X^{-1}(B_1 \cup B_2) = (X^{-1}(B_1) \cup X^{-1}(B_2)) \in \mathcal{F},$$
 bzw., dass  $B_1 \cup B_2 \cup \dots \in \mathcal{G}$  für beliebige  $B_1, B_2, \dots \in \mathcal{G}$ .
- Also ist das in (3) gegebene Mengensystem eine  $\sigma$ -Algebra.
- Darüber hinaus bedeutet die Bedingung (1), dass  $(-\infty, x] \in \mathcal{G}$  für jedes  $x \in \mathbb{R}$ .
  - Hieraus folgt, dass  $(x, \infty) = (-\infty, x]^c \in \mathcal{G}$  und  $(-\infty, x) = \bigcup_{n=1}^{\infty} (-\infty, x - n^{-1}] \in \mathcal{G}$ .
  - Deshalb gilt  $(a, b) = (-\infty, b) \cap (a, \infty) \in \mathcal{G}$  für  $-\infty < a \leq b < \infty$ .
  - Also gehört das Erzeugersystem  $\{(a, b), -\infty < a < b < \infty\}$  der Borel- $\sigma$ -Algebra  $\mathcal{B}(\mathbb{R})$  zu  $\mathcal{G}$ .
  - Dies bedeutet, dass  $\mathcal{B}(\mathbb{R}) \subset \mathcal{G}$ .
- Damit ist gezeigt, dass (2) aus (1) folgt. □

Dies führt zu der folgenden Begriffsbildung.

#### Definition

- Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum, und  $X : \Omega \rightarrow \mathbb{R}$  sei eine beliebige Zufallsvariable.
- Die Verteilung der Zufallsvariablen  $X$  ist die Mengenfunktion  $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  mit

$$P_X(B) = P(\{\omega : \omega \in \Omega, X(\omega) \in B\}) \quad \forall B \in \mathcal{B}(\mathbb{R}). \quad (4)$$

#### Beachte

1. Die in (4) definierte Mengenfunktion  $P_X$  ist ein *Wahrscheinlichkeitsmaß* auf dem Messraum  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , denn  $P_X$  ist
  - normiert, weil  $P_X(\mathbb{R}) = P(\Omega) = 1$ , und
  - $\sigma$ -additiv, weil für paarweise disjunkte  $B_1, B_2, \dots \in \mathcal{B}(\mathbb{R})$

$$P_X\left(\bigcup_{i=1}^{\infty} B_i\right) = P\left(X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right)\right) = P\left(\bigcup_{i=1}^{\infty} X^{-1}(B_i)\right) = \sum_{i=1}^{\infty} P(X^{-1}(B_i)) = \sum_{i=1}^{\infty} P_X(B_i).$$

2. Die Abbildung  $P \rightarrow P_X$  nennt man *Maßtransport* vom Messraum  $(\Omega, \mathcal{F})$  in den Messraum  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

Die folgende Kurzschreibweise ist üblich:  $P(X \in B) = P(\{\omega : \omega \in \Omega, X(\omega) \in B\}) \quad \forall B \in \mathcal{B}(\mathbb{R})$

Speziell:  $P(X \leq x) = P(\{\omega : \omega \in \Omega, X(\omega) \leq x\}) \quad \forall x \in \mathbb{R}$

### 3.2.1 Diskrete Zufallsvariablen; Wahrscheinlichkeitsfunktion

Wir betrachten in dieser Vorlesung zwei (Grund-) Typen von Zufallsvariablen: *diskrete* Zufallsvariablen und *absolutstetige* Zufallsvariablen.

**Definition** Die Zufallsvariable  $X$  (bzw. ihre Verteilung) heißt *diskret*, falls es eine abzählbare Teilmenge  $C \subset \mathbb{R}$  gibt, so dass  $P(X \in C) = 1$ .

**Beachte** Der Begriff der absolutstetigen Zufallsvariablen wird später in Abschnitt 3.2.3 eingeführt. Wir erwähnen jedoch schon jetzt ein wichtiges Unterscheidungsmerkmal:

1. diskrete Zufallsvariablen haben einen abzählbaren Wertebereich, z.B. wenn  $\Omega = X^{-1}(C)$  mit  $C \subset \mathbb{N}, \mathbb{Z}, \mathbb{Q} \subset \mathbb{R}$ . Sei beispielsweise  $X =$  Augensumme bei zweimaligem Würfeln  $\Rightarrow X : \Omega \rightarrow \{2, 3, \dots, 12\}$
2. absolutstetige Zufallsvariablen haben einen überabzählbaren Wertebereich, z.B.  $[a, b], [a, \infty), (-\infty, b], \mathbb{R}$  z.B. Roulette mit drehbarem Zeiger und „kontinuierlicher“ Skala, wobei  $X =$  Wert des Spiels = Winkel des Zeigers, d.h.  $X : \Omega \rightarrow [0, 2\pi)$

**Definition** Sei  $X$  eine diskrete Zufallsvariable, d.h., es gebe eine abzählbare Menge  $C = \{x_1, x_2, \dots\}$ , so dass  $P(X \in C) = 1$ . Dann heißt die Folge  $p_1, p_2, \dots$  mit  $p_k = P(X = x_k)$  *Wahrscheinlichkeitsfunktion* (bzw. *Zähldichte*) von  $X$ .

**Beachte**

1. Für jede Wahrscheinlichkeitsfunktion  $\{p_k\}$  gilt offenbar  $p_k \in [0, 1]$  für jedes  $k = 1, 2, \dots$  und  $\sum_{k=1}^{\infty} p_k = 1$ .
2. Die Verteilung einer diskreten Zufallsvariablen  $X$  wird eindeutig durch die Wahrscheinlichkeitsfunktion  $\{p_k\}$  bestimmt, denn es gilt dann für jedes  $B \in \mathcal{B}(\mathbb{R})$

$$P_X(B) = P_X(B \cap C) = P_X\left(\bigcup_{i: x_i \in B} \{x_i\}\right) = \sum_{i: x_i \in B} P_X(\{x_i\}) = \sum_{i: x_i \in B} p_i.$$

3. Für jedes  $x_k \in C$  heißt die Zahl  $p_k = P(X = x_k)$  *Einzelwahrscheinlichkeit* von  $X$ .

**Beispiele**

#### 1. zweimaliges Würfeln

- Sei  $X =$  Summe der Augenzahlen beim zweimaligen Würfeln.
- Dann gilt  $P(X \in C) = 1$  mit  $C = \{2, 3, \dots, 12\}$ , d.h.  $x_k = k + 1$ .
- Die Wahrscheinlichkeitsfunktion von  $X$  ist gegeben durch:

$x_k$	2	3	4	5	6	7	8	9	10	11	12
$p_k$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

#### 2. Bernoulli-Schema

- Einmaliger Münzwurf:  $\Omega_1 = \{0, 1\}$ , „0“ = Wappen, „1“ = Zahl
- $n$ -maliger Münzwurf: Für  $i = 1, \dots, n$  setzen wir  $\Omega_i = \{0, 1\}$ , wobei  $P_i(\{\omega_i\}) = \frac{1}{2}$  im Fall einer fairen Münze bzw. allgemein  $P_i(\{1\}) = a_i$  und  $P_i(\{0\}) = 1 - a_i$ .
- Bei *identischen Versuchsbedingungen* nehmen wir an, dass  $a_1 = a_2 = \dots = a_n = p$ ,  $p \in [0, 1]$ .
- *Unabhängige Versuche* modellieren wir durch den Ansatz  $(\Omega, \mathcal{F}, P)$  mit

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in \{0, 1\}\}, \quad \mathcal{F} = \mathcal{P}(\Omega),$$

wobei  $P$  das *Produktmaß* auf  $\mathcal{F}$  ist, für das gilt:  $P(\{\omega\}) = \prod_{i=1}^n P_i(\{\omega_i\})$ .

- Sei  $\omega = (\omega_1, \dots, \omega_n) \in \Omega$  ein beliebiges Elementarereignis. Dann gilt

$$P(\{\omega\}) = \prod_{i:\omega_i=1} a_i \prod_{j:\omega_j=0} (1 - a_j)$$

- Für  $\omega \in \Omega$  mit  $|\{i : \omega_i = 1\}| = k$  und  $a_1 = a_2 = \dots = a_n = p$  gilt dann insbesondere

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

- Deuten „1“ als Erfolg und „0“ als Misserfolg.
- Sei  $X = \text{Anzahl der Erfolge bei } n \text{ Versuchen}$ . Wenn  $a_1 = a_2 = \dots = a_n = p$ , dann gilt

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \forall k = 0, 1, \dots, n.$$

- Sprechweise:  $X$  ist *binomialverteilt* mit den Parametern  $n$  und  $p$ .
- Spezialfall: Wenn  $n = 1$ , dann sagen wir, dass  $X$  *Bernoulli-verteilt* ist.

### 3.2.2 Grundlegende Klassen diskreter Verteilungen (Zusammenfassung)

#### 1. Hypergeometrische Verteilung (Urnenmodell)

- Betrachten Urne mit  $N$  Objekten ( $S$  schwarze,  $R$  rote Kugeln), d.h.  $N = S + R$ ;
- $n$  Objekte,  $n \leq N$ , sollen insgesamt ausgewählt werden;
- $X = \text{Anzahl von schwarzen Kugeln bei } n \text{ Entnahmen}$

$$X : \Omega \rightarrow \{0, 1, \dots, \min\{n, S\}\}, \quad p_k = P(X = k) = \frac{\binom{S}{k} \binom{N-S}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \dots, \min\{S, n\}$$

- Die hypergeometrische Verteilung hat 3 Parameter:  $N, S \leq N, n \leq N$ ;
- Berechnung der Einzelwahrscheinlichkeiten  $p_k$  ist *schwierig*, falls  $N$  und  $S$  groß sind;
- *Ausweg*: Binomialverteilung liefert asymptotische Näherungsformel, falls  $N, S \rightarrow \infty, \frac{S}{N} \xrightarrow{S, N \rightarrow \infty} p$

#### 2. Binomialverteilung (Bernoulli-Schema)

- $n$ -maliger Münzwurf;
- identische Erfolgswahrscheinlichkeiten  $a_1 = \dots = a_n = p$ ;
- $X = \text{Anzahl der Erfolge bei } n \text{ Versuchen}$

$$X : \Omega \rightarrow \{0, 1, \dots, n\}, \quad p_k = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n$$

- Für jedes  $k = 0, 1, \dots, n$  gilt

$$\frac{\binom{S}{k} \binom{N-S}{n-k}}{\binom{N}{n}} \rightarrow \binom{n}{k} p^k (1 - p)^{n-k},$$



falls  $N, S \rightarrow \infty$ ,  $\frac{S}{N} \xrightarrow{S, N \rightarrow \infty} p$ , denn

$$\begin{aligned} \frac{\binom{S}{k} \binom{N-S}{n-k}}{\binom{N}{n}} &= \frac{\frac{S!}{k!(S-k)!} \frac{(N-S)!}{(n-k)!(N-S-n+k)!}}{\frac{N!}{n!(N-n)!}} \\ &= \frac{n!}{k!(n-k)!} \frac{S(S-1)\dots(S-k+1)(N-S)(N-S-1)\dots(N-S-n+k+1)}{N(N-1)\dots(N-n+1)} \\ &= \binom{n}{k} \underbrace{\frac{S}{N} \dots \frac{S-k+1}{N-k+1}}_{\rightarrow p^k} \underbrace{\frac{N-S}{N-k} \dots \frac{N-S-(n-k)+1}{N-n+1}}_{(1-p)^{n-k}} \\ &\rightarrow \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

- Die Binomialverteilung hat 2 Parameter:  $n$  und  $p$  (Schreibweise:  $\text{Bin}(n, p)$ );
- Berechnung der Einzelwahrscheinlichkeiten  $p_k$  ist *schwierig*, falls  $n$  groß und  $p$  klein (oder nahe bei 1) ist;
- *Ausweg*: Poisson-Verteilung liefert asymptotische Näherungsformel, falls  $n \rightarrow \infty$  und  $p \rightarrow 0$  mit  $n \cdot p \rightarrow \lambda$ , so dass  $0 < \lambda < \infty$ .

### 3. Poisson-Verteilung (Gesetz der seltenen Ereignisse)

- Betrachten diskrete Zufallsvariable  $X : \Omega \rightarrow \{0, 1, \dots\}$ ;
- die Einzelwahrscheinlichkeiten  $p_k = P(X = k)$  seien gegeben durch

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots; 0 < \lambda < \infty$$

- Die Poisson-Verteilung hat 1 Parameter:  $\lambda$  (Schreibweise:  $\text{Poi}(\lambda)$ );
- *Gesetz der seltenen Ereignisse*: Für jedes  $k = 0, 1, \dots$  gilt

$$\binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda},$$

falls  $n \rightarrow \infty$  und  $p \rightarrow 0$  mit  $n \cdot p \rightarrow \lambda$ , denn

$$\begin{aligned} \binom{n}{k} p^k (1-p)^{n-k} &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{1}{k!} \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\rightarrow 1} \underbrace{(np)^k}_{\rightarrow \lambda^k} \underbrace{\frac{(1-p)^n}{(1-p)^k}}_{\approx \frac{(1-\frac{\lambda}{n})^n}{1} \rightarrow e^{-\lambda}} \\ &\rightarrow \frac{1}{k!} \lambda^k e^{-\lambda}. \end{aligned}$$

### 4. geometrische Verteilung

- Sei  $X$  = die Anzahl der Versuche bis zum ersten Erfolg im Bernoulli-Schema (mit  $n = \infty$ ), d.h.  $X : \Omega \rightarrow \{1, 2, \dots\}$ ;
- die Einzelwahrscheinlichkeiten  $p_k = P(X = k)$  sind dann gegeben durch

$$p_k = p(1-p)^{k-1}, \quad k = 1, 2, \dots;$$

- Die geometrische Verteilung hat 1 Parameter:  $p \in (0, 1)$ .

### 3.2.3 Verteilungsfunktion; absolutstetige Zufallsvariablen

Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum und  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable.

**Definition** Die Funktion  $F_X : \mathbb{R} \rightarrow [0, 1]$  mit

$$F_X(x) = P(X \leq x) \quad \forall x \in \mathbb{R} \quad (5)$$

heißt *Verteilungsfunktion* von  $X$ , wobei  $F_X(x) = P(X \leq x) = P(X \in (-\infty, x]) = P_X((-\infty, x])$ .

Wir diskutieren nun zunächst einige Eigenschaften von Verteilungsfunktionen.

**Theorem 3.2** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable und  $F_X : \mathbb{R} \rightarrow [0, 1]$  ihre Verteilungsfunktion. Dann gilt

1. Asymptotisches Verhalten im Unendlichen:

$$F_X(-\infty) := \lim_{x \rightarrow -\infty} F_X(x) = 0, \quad F_X(\infty) := \lim_{x \rightarrow \infty} F_X(x) = 1, \quad (6)$$

2. Monotonie:

$$F_X(x) \leq F_X(x+h) \quad \forall x \in \mathbb{R} \text{ und } h \geq 0, \quad (7)$$

3. Rechtsstetigkeit:  $F_X(x)$  ist rechtsseitig stetig, d.h. für jede Folge  $\{h_n\}$  mit  $h_n \geq 0$  und  $\lim_{n \rightarrow \infty} h_n = 0$  gilt

$$\lim_{n \rightarrow \infty} F_X(x+h_n) = F_X(x) \quad \forall x \in \mathbb{R}. \quad (8)$$

#### Beweis

Zu 2. Weil  $(-\infty, x] \subset (-\infty, x+h]$ , ergibt sich aus Teilaussage 2 von Theorem 2.1, dass

$$F_X(x) = P_X((-\infty, x]) \leq P_X((-\infty, x+h]) = F_X(x+h).$$

Zu 1. Wir zeigen nur die erste Teilaussage von (6). Wegen der Monotonie von  $F_X$  können wir o.B.d.A. annehmen, dass  $x$  monoton gegen  $-\infty$  konvergiert. Aus Korollar 2.2 ergibt sich dann, dass

$$\begin{aligned} \lim_{x \rightarrow -\infty} F_X(x) &= \lim_{x \rightarrow -\infty} P_X((-\infty, x]) \\ &= P_X\left(\bigcap_{x \leq 0} (-\infty, x]\right) \\ &= P_X(\emptyset) = 0. \end{aligned}$$

Der Beweis der zweiten Teilaussage von (6) verläuft analog.

Zu 3. Ähnlich wie im Beweis von Teilaussage 1 ergibt sich aus Korollar 2.2, dass

$$\begin{aligned} \lim_{n \rightarrow \infty} F_X(x+h_n) &= \lim_{n \rightarrow \infty} P_X((-\infty, x+h_n]) \\ &= P_X\left(\bigcap_{n \geq 1} (-\infty, x+h_n]\right) \\ &= P_X((-\infty, x]) \\ &= F_X(x). \end{aligned}$$

□

**Beachte**

1. Mit Hilfe der Verteilungsfunktion  $F_X$  lassen sich auch die folgenden Wahrscheinlichkeiten ausdrücken

$$P(a \leq X \leq b), \quad P(a < X \leq b), \quad P(a < X < b), \quad P(a \leq X < b),$$

denn es gilt beispielsweise

$$\begin{aligned} P(a \leq X \leq b) &= P(\{X \leq b\} \setminus \{X < a\}) \\ &= P(X \leq b) - P(X < a) \\ &= F_X(b) - \lim_{h \downarrow 0} F_X(a - h). \end{aligned}$$

2. Im allgemeinen gilt jedoch nicht  $F_X(a) = \lim_{h \downarrow 0} F_X(a - h)$ , sondern

$$F_X(a) = \lim_{h \downarrow 0} F_X(a - h) + P(X = a), \quad (9)$$

denn

$$\begin{aligned} P(X = a) &= P\left(\bigcap_{n=1}^{\infty} \{a - n^{-1} < X \leq a\}\right) \\ &= \lim_{n \rightarrow \infty} P(a - n^{-1} < X \leq a) \\ &= \lim_{n \rightarrow \infty} (P(X \leq a) - P(X \leq a - n^{-1})) \\ &= F_X(a) - \lim_{n \rightarrow \infty} F_X(a - n^{-1}). \end{aligned}$$

3. In Theorem 3.2 wurde gezeigt, dass

- Verteilungsfunktionen monotone und beschränkte Funktionen sind.
- Hieraus folgt, dass Verteilungsfunktionen für jedes  $\varepsilon > 0$  nur endlich viele Sprungstellen besitzen können, deren Sprunghöhen größer als  $\varepsilon$  sind.
- Insgesamt können Verteilungsfunktionen also höchstens abzählbar viele Sprungstellen besitzen.

4. Für die Verteilungsfunktion  $F_X$  einer diskreten Zufallsvariablen  $X$  gilt für jedes  $x \in \mathbb{R}$ :

$$F_X(x) = P(X \leq x) = P\left(\bigcup_{k: x_k \leq x} \{X = x_k\}\right) = \sum_{k: x_k \leq x} P(X = x_k) = \sum_{k: x_k \leq x} p_k,$$

wobei  $p_k = P(X = x_k)$ .

5. Die Verteilungsfunktion  $F_X$  einer diskreten Zufallsvariablen  $X$  ist eine sogenannte *Treppenfunktion*, d.h. eine stückweise konstante Funktion mit der Sprunghöhe  $p_k$  im Punkt  $x_k$ ; vgl. Abbildung 1. In Abbildung 2 wird als ein Beispiel dieser Art die Verteilungsfunktion der Poisson-Verteilung dargestellt.

Um zu zeigen, dass der Zusammenhang zwischen Verteilung und Verteilungsfunktion von Zufallsvariablen eindeutig ist, benötigen wir einen Satz über monotone Mengensysteme, der ein wichtiges Hilfsmittel in der Wahrscheinlichkeitsrechnung ist.

**Definition** Sei  $\Omega$  eine beliebige Menge.

1. Eine Familie  $\mathcal{G}$  von Teilmengen von  $\Omega$  heißt *d-System* (Dynkin-System) auf  $\Omega$ , falls

- $\Omega \in \mathcal{G}$ ,
- $A \setminus B \in \mathcal{G}$  für beliebige  $A, B \in \mathcal{G}$  mit  $A \supset B$ ,
- $\bigcup_{i=1}^{\infty} A_i \in \mathcal{G}$  für beliebige  $A_1, A_2, \dots \in \mathcal{G}$  mit  $A_1 \subset A_2 \subset \dots$

2. Eine nichtleere Familie  $\mathcal{G}$  von Teilmengen von  $\Omega$  heißt  $\pi$ -System auf  $\Omega$ , falls  $\bigcap_{i=1}^n A_i \in \mathcal{G}$  für jedes  $n = 1, 2, \dots$  und für beliebige  $A_1, \dots, A_n \in \mathcal{G}$ .

### Beachte

- Jede  $\sigma$ -Algebra auf  $\Omega$  ist gleichzeitig ein  $d$ -System und auch ein  $\pi$ -System auf  $\Omega$ .
- Umgekehrt ist jedes System von Teilmengen von  $\Omega$ , dass sowohl ein  $d$ -System als auch ein  $\pi$ -System auf  $\Omega$  ist, eine  $\sigma$ -Algebra auf  $\Omega$ .
- *Schreibweise:* Sei  $d(\mathcal{G})$  das kleinste  $d$ -System, das das erzeugende System  $\mathcal{G}$  von Teilmengen von  $\Omega$  umfasst. Analog bezeichne  $\sigma(\mathcal{G})$  die kleinste  $\sigma$ -Algebra, die das erzeugende System  $\mathcal{G}$  von Teilmengen von  $\Omega$  umfasst.

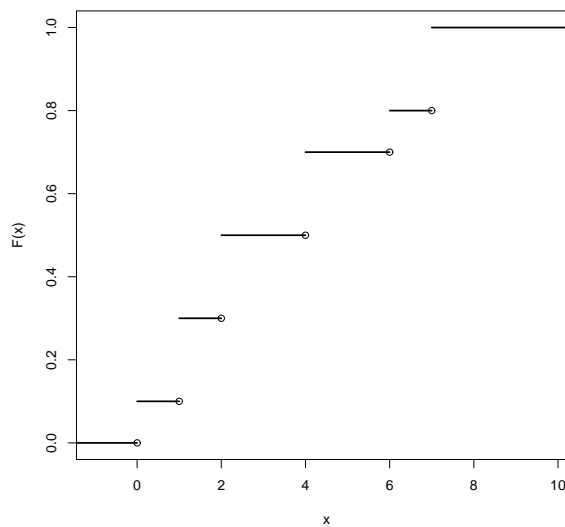


Abbildung 1: Verteilungsfunktion einer diskreten Zufallsvariablen

**Lemma 3.1** (Monotone class theorem) *Sei  $\mathcal{G}$  ein  $\pi$ -System auf  $\Omega$ . Dann gilt*

$$\sigma(\mathcal{G}) = d(\mathcal{G}). \quad (10)$$

Der *Beweis* von Lemma 3.1 geht über den Rahmen dieser einführenden Vorlesung hinaus und wird deshalb weggelassen.

**Theorem 3.3** *Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable. Dann wird die Verteilung  $P_X$  von  $X$  durch die Verteilungsfunktion  $F_X$  von  $X$  eindeutig bestimmt.*

### Beweis

- Betrachten zwei beliebige Zufallsvariablen  $X, Y : \Omega \rightarrow \mathbb{R}$  und
- nehmen an, dass ihre Verteilungsfunktionen übereinstimmen, d.h.,

$$F_X(x) = F_Y(x) \quad \forall x \in \mathbb{R}. \quad (11)$$

- Zu zeigen ist, dass dann

$$P_X(B) = P_Y(B) \quad \forall B \in \mathcal{B}(\mathbb{R}). \quad (12)$$

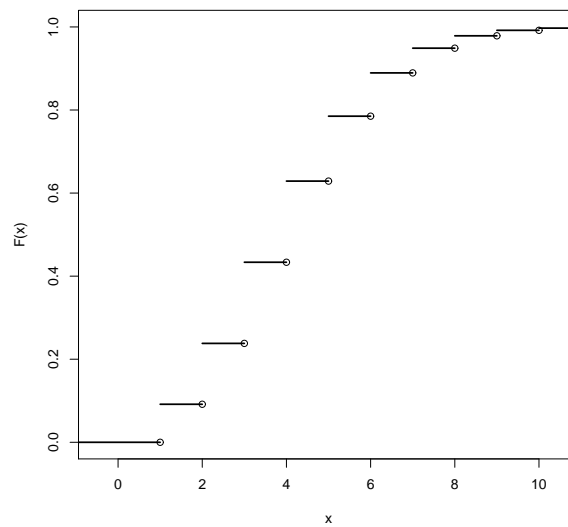


Abbildung 2: Verteilungsfunktion der Poisson-Verteilung

- Sei  $\mathcal{G} = \{(a, b] : -\infty < a \leq b < \infty\}$  das System aller links-offenen und rechts-abgeschlossenen Intervalle in  $\mathbb{R}$ .
- Es ist klar, dass  $\mathcal{G}$  ein  $\pi$ -System ist.
- Aus (11) und

$$P_X((a, b]) = F_X(b) - F_X(a) = F_Y(b) - F_Y(a) = P_Y((a, b])$$

folgt, dass (12) für jedes  $B \in \mathcal{G}$  gilt.

- Außerdem kann man sich leicht überlegen, dass (12) auch für jedes  $B \in d(\mathcal{G})$  gilt.
- Aus Lemma 3.1 folgt nun, dass (12) für jedes  $B \in \sigma(\mathcal{G}) = \mathcal{B}(\mathbb{R})$  gilt.  $\square$

**Definition** Die Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  (bzw. ihre Verteilung) heißt *absolutstetig*, falls die Verteilungsfunktion  $F_X$  von  $X$  die folgende *Integraldarstellung*

$$F_X(x) = \int_{(-\infty, x]} f_X(y) dy \quad \forall x \in \mathbb{R} \quad (13)$$

besitzt, wobei  $f_X : \mathbb{R} \rightarrow [0, \infty)$  eine (Lebesgue-integrierbare) Funktion mit nichtnegativen Werten ist, die *Dichte* (bzw. *Wahrscheinlichkeitsdichte*) von  $X$  genannt wird. Das Integral in (13) wird im allgemeinen als Lebesgue-Integral aufgefasst.

Die Verteilungsfunktion  $F_X$  (und damit auch die Verteilung  $P_X$ ) einer absolutstetigen Zufallsvariablen  $X$  wird in dem folgenden Sinne *eindeutig* durch die Dichte  $f_X$  bestimmt.

### Theorem 3.4

1. Die Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  ist genau dann absolutstetig, wenn sich die Verteilung  $P_X$  von  $X$  darstellen lässt in der Form:

$$P_X(B) = \int_B f_X(y) dy \quad \forall B \in \mathcal{B}(\mathbb{R}). \quad (14)$$

2. Die Zufallsvariablen  $X, Y : \Omega \rightarrow \mathbb{R}$  seien absolutstetig. Es gilt  $P_X = P_Y$  genau dann, wenn

$$f_X(x) = f_Y(x) \quad (15)$$

für fast alle  $x \in \mathbb{R}$ , d.h., (15) gilt für alle  $x \in \mathbb{R} \setminus B$ , wobei die (Borelsche) „Ausnahmemenge“  $B \subset \mathbb{R}$  das Lebesgue-Maß 0 hat.

### Beweis

- Die Hinlänglichkeit der Bedingung (14) ist offensichtlich, denn es genügt, in (14) die spezielle Borel-Menge  $B = (-\infty, x]$  einzusetzen, um (13) zu erhalten.
- Die Notwendigkeit der Bedingung (14) ergibt sich aus (13) und aus dem eindeutigen Zusammenhang zwischen Verteilung und Verteilungsfunktion, vgl. Theorem 3.3.
- Die Notwendigkeit und Hinlänglichkeit von Bedingung (15) ergibt sich aus den allgemeinen (Eindeutigkeits-) Eigenschaften von Lebesgue-Integralen, vgl. die Vorlesung Analysis III.  $\square$

### Beachte

1. Bei vielen Anwendungen ist die Dichte  $f_X$  eine (zumindest stückweise) stetige Funktion. Das Integral in der Definitionsgleichung (13) kann dann auch als ein uneigentliches Riemann-Integral aufgefasst werden.
2. Wenn  $X$  absolutstetig ist, dann hat die Verteilungsfunktion  $F_X$  keine Sprünge, d.h.,  $F_X$  ist eine (im üblichen Sinne) stetige Funktion. Hieraus und aus (9) folgt insbesondere, dass

$$P(X = x) = 0 \quad \forall x \in \mathbb{R}. \quad (16)$$

3. Die Verteilungsfunktion  $F_X$  einer absolutstetigen Zufallsvariablen  $X$  ist jedoch im allgemeinen *nicht* überall differenzierbar. Und zwar ist  $F_X$  dort nicht differenzierbar, wo die Dichte  $f_X$  Sprungstellen hat, falls  $f_X$  stückweise stetig ist.

**Beispiele** Um die Verteilung einer absolutstetigen Zufallsvariablen  $X$  zu beschreiben, genügt es, die Dichte  $f_X$  zu betrachten, weil durch  $f_X$  die Verteilungsfunktion  $F_X$  und damit auch die Verteilung  $P_X$  von  $X$  eindeutig bestimmt wird.

1. *Normalverteilung*  $N(\mu, \sigma^2)$  mit den Parametern  $\mu \in \mathbb{R}$  und  $\sigma^2 > 0$ :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \forall x \in \mathbb{R} \quad (17)$$

Spezialfall: *Standardnormalverteilung*  $N(0, 1)$ . Dann nimmt die Dichte  $f_X(x)$  in (17) die folgende Form an:

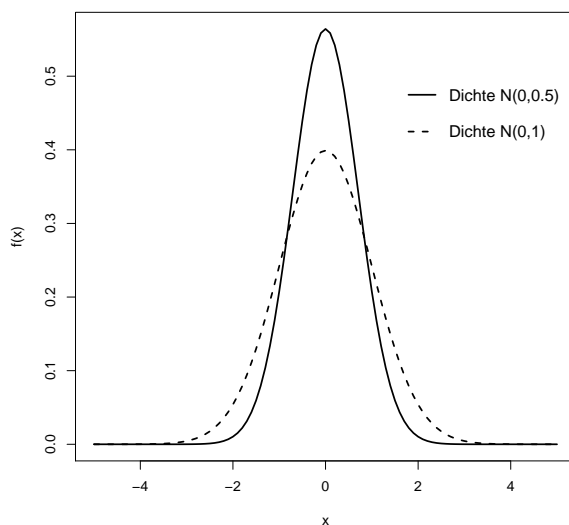
$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad \forall x \in \mathbb{R} \quad (18)$$

In Abbildung 3 wird die Dichte der Normalverteilung für großes bzw. kleines  $\sigma^2$  dargestellt.

2. *Exponentialverteilung*  $\text{Exp}(\lambda)$  mit Parameter  $\lambda > 0$ :

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

Die Dichte der Exponentialverteilung ist in Abbildung 4 dargestellt. Für die zugehörige Verteilungsfunktion  $F_X : \mathbb{R} \rightarrow [0, 1]$  gilt  $F_X(x) = 1 - e^{-\lambda x}$ , falls  $x \geq 0$ , und  $F_X(x) = 0$ , falls  $x < 0$ .

Abbildung 3: Dichte der Normalverteilung ( $\sigma^2$  groß bzw. klein)

3. Gleichverteilung  $U(a, b)$  mit den Parametern  $a, b \in \mathbb{R}$ , wobei  $a < b$ :

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

Die Dichte der Gleichverteilung ist in Abbildung 5 dargestellt. Für die zugehörige Verteilungsfunktion  $F_X : \mathbb{R} \rightarrow [0, 1]$  gilt

$$F_X(x) = \begin{cases} 0, & \text{falls } x < a, \\ \frac{x-a}{b-a}, & \text{falls } a \leq x \leq b \\ 1, & \text{falls } x > b. \end{cases}$$

### Beachte

- Absolutstetige Verteilungen treten oft als (asymptotische) Näherungslösungen auf.
- So lässt sich beispielsweise die Binomialverteilung mit den Parametern  $n$  und  $p$  durch die Standardnormalverteilung approximieren, falls  $n$  groß ist. Dies ist der folgende *zentrale Grenzwertsatz von DeMoivre-Laplace*.

**Theorem 3.5** Für jedes  $n \geq 1$  sei die Zufallsvariable  $X_n$  binomialverteilt mit den Parametern  $n$  und  $p$ . Dann gilt für beliebige  $p \in (0, 1)$  und  $a, b \in \mathbb{R}$  mit  $a < b$

$$\lim_{n \rightarrow \infty} P\left(a < \frac{X_n - np}{\sqrt{np(1-p)}} \leq b\right) = P(a < X \leq b) \quad \left(= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx\right), \quad (19)$$

wobei die Zufallsvariable  $X$  standardnormalverteilt ist, d.h.  $X \sim N(0, 1)$ .

Der *Beweis* von Theorem 3.5 wird zunächst weggelassen und später, in Abschnitt 5.3 der Vorlesung, in einem allgemeineren Zusammenhang nachgeholt.

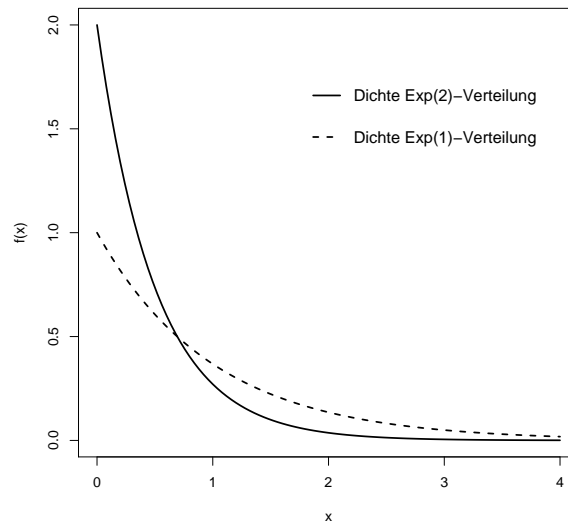


Abbildung 4: Dichte der Exponentialverteilung

**Beachte** Wenn für die Schranken  $a$  und  $b$  in (19)

$$a = \frac{k-1-np}{\sqrt{np(1-p)}} \quad \text{und} \quad b = \frac{k-np}{\sqrt{np(1-p)}}$$

eingesetzt wird, dann ergibt sich aus (19), dass für große  $n$

$$\begin{aligned} P(X_n = k) &= P(k-1 < X_n \leq k) \\ &= P\left(\frac{k-1-np}{\sqrt{np(1-p)}} < \frac{X_n - np}{\sqrt{np(1-p)}} \leq \frac{k-np}{\sqrt{np(1-p)}}\right) \\ &\approx P\left(\frac{k-1-np}{\sqrt{np(1-p)}} < X \leq \frac{k-np}{\sqrt{np(1-p)}}\right) \\ &\approx \frac{1}{\sqrt{np(1-p)}} f_X\left(\frac{k-np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

für jedes  $k = 0, 1, \dots, n$ , wobei  $f_X$  die in (18) gegebene Dichte der Standardnormalverteilung ist.

### 3.3 Zufallsvektoren

#### 3.3.1 Definition, Verteilung und Verteilungsfunktion

Bei Anwendungen besteht oft die Notwendigkeit, nicht nur eine Kennzahl  $X(\omega)$ , sondern gleichzeitig mehrere Kennzahlen  $X_1(\omega), \dots, X_n(\omega)$  von  $\omega \in \Omega$  zu betrachten.

#### Beispiele

##### 1. zweimaliges Würfeln

- Als Grundraum wählen wir so wie bisher  $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$ , vgl. Abschnitt 2.4.1.
- Sei  $X : \Omega \rightarrow \{0, 1, 2\}$  bzw.  $Y : \Omega \rightarrow \{0, 1, 2\}$  die (zufällige) Anzahl, mit der die Augenzahl „6“ bzw. „1“ beim zweimaligen Würfeln erzielt wird.



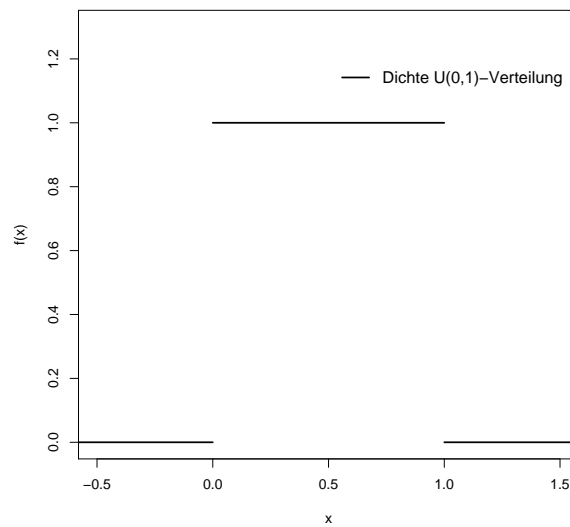


Abbildung 5: Dichte der Gleichverteilung

- Dann gilt für die Wahrscheinlichkeiten  $P(X = x, Y = y)$  bzw. für die Einzelwahrscheinlichkeiten  $P(X = x)$  und  $P(Y = y)$  von  $X$  und  $Y$ :

$P(X = x, Y = y)$		$y$			$P(X = x)$
		0	1	2	
$x$	0	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$	$\frac{25}{36}$
	1	$\frac{8}{36}$	$\frac{2}{36}$	0	$\frac{10}{36}$
	2	$\frac{1}{36}$	0	0	$\frac{1}{36}$
$P(Y = y)$		$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	

- Aus der Tabelle kann man auch die Einzelwahrscheinlichkeiten der Summe  $X + Y$  erhalten. Beispielsweise gilt

$$P(X + Y = 1) = P(X = 1, Y = 0) + P(X = 0, Y = 1) = \frac{8}{36} + \frac{8}{36} = \frac{16}{36}.$$

- Analog ergibt sich

$$P(X + Y = 0) = \frac{16}{36}, \quad P(X + Y = 2) = \frac{4}{36}$$

bzw.

$$P(X \cdot Y = 1) = \frac{2}{36}, \quad P(X \cdot Y = 0) = \frac{34}{36}.$$

## 2. Analyse von Kommunikationsnetzen

- Betrachten ein Kommunikationsnetz mit  $n$  Komponenten.
- Sei  $\Omega = \Omega_1 \times \dots \times \Omega_n$ , wobei  $\Omega$  die Menge aller möglichen Momentanzustände  $\omega = (\omega_1, \dots, \omega_n)$  des Netzes und  $\Omega_i$  die Menge aller möglichen Momentanzustände  $\omega_i$  der  $i$ -ten Komponente bezeichnet;  $i = 1, \dots, n$ .
- Dann kann beispielsweise durch die Abbildung  $\omega \rightarrow X_i(\omega)$  die Belastung  $X_i(\omega)$  der  $i$ -ten Komponente in Abhängigkeit vom Momentanzustand  $\omega$  des Netzes modelliert werden.

- Die (globale) Belastung des gesamten Netzes kann dann durch den Vektor  $(X_1(\omega), \dots, X_n(\omega))$  beschrieben werden.

Die gleichzeitige Betrachtung mehrerer Kennzahlen  $X_1(\omega), \dots, X_n(\omega)$  von  $\omega \in \Omega$  führt zum Begriff des Zufallsvektors.

**Definition** Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum, und sei  $X_1, \dots, X_n$  eine beliebige Folge von Zufallsvariablen  $X_i : \Omega \rightarrow \mathbb{R}; i = 1, \dots, n$ .

- Die Abbildung  $X = (X_1, \dots, X_n)$  von  $\Omega$  nach  $\mathbb{R}^n$  heißt dann  $n$ -dimensionaler *Zufallsvektor* mit den Komponenten  $X_1, \dots, X_n$ .
- Die *Verteilung* des Zufallsvektors  $X$  ist die Mengenfunktion  $P_X : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$  mit

$$P_X(B) = P(\omega : \omega \in \Omega, X(\omega) \in B) \quad \forall B \in \mathcal{B}(\mathbb{R}^n). \quad (20)$$

- Die Funktion  $F_X : \mathbb{R}^n \rightarrow [0, 1]$  mit

$$F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) \quad (21)$$

heißt (gemeinsame bzw. multivariate) *Verteilungsfunktion* des Zufallsvektors  $X = (X_1, \dots, X_n)$ .

**Theorem 3.6** Die Abbildung  $X : \Omega \rightarrow \mathbb{R}^n$  ist genau dann ein Zufallsvektor, wenn

$$\{\omega : \omega \in \Omega, X(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{B}(\mathbb{R}^n). \quad (22)$$

Die Verteilung  $P_X$  von  $X$  wird eindeutig durch die Verteilungsfunktion  $F_X$  von  $X$  bestimmt.

Der *Beweis* ist analog zum Beweis der Theoreme 3.1 und 3.3. Er wird deshalb weggelassen.

### 3.3.2 Eigenschaften multivariater Verteilungsfunktionen

**Theorem 3.7** Sei  $X : \Omega \rightarrow \mathbb{R}^n$  ein beliebiger Zufallsvektor und  $F_X : \mathbb{R}^n \rightarrow [0, 1]$  seine Verteilungsfunktion. Dann gilt

1. Asymptotisches Verhalten im Unendlichen: Für beliebige  $i \in \{1, \dots, n\}$  und  $x_1, \dots, x_n \in \mathbb{R}$  gilt
  - (i)  $F_X(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = 0$ , wobei

$$F_X(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = \lim_{x_i \rightarrow -\infty} F_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n);$$

- (ii)  $F_X(\infty, \dots, \infty) = 1$ , wobei  $F_X(\infty, \dots, \infty) = \lim_{x_1, \dots, x_n \rightarrow \infty} F_X(x_1, \dots, x_n)$ ;
- (iii)  $F_X(\infty, \dots, \infty, x_i, \infty, \dots, \infty) = F_{X_i}(x_i)$ , wobei  $F_X(\infty, \dots, \infty, x_i, \infty, \dots, \infty)$  analog zu den in (i)–(ii) betrachteten Grenzwerten definiert wird und  $F_{X_i}$  Randverteilungsfunktion von  $X$  genannt wird.

2. Monotonie:  $F_X(x_1 + h_1, \dots, x_n + h_n) \geq F_X(x_1, \dots, x_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}, h_1, \dots, h_n \geq 0$
3. Rechtsstetigkeit:  $F_X(x_1, \dots, x_n) = \lim_{h_1, \dots, h_n \downarrow 0} F_X(x_1 + h_1, \dots, x_n + h_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}$

Der *Beweis* ist analog zum Beweis von Theorem 3.2 und wird deshalb weggelassen.

#### Definition

1. Der Zufallsvektor  $X = (X_1, \dots, X_n)$  heißt *diskret*, wenn es eine abzählbare Menge  $C \subset \mathbb{R}^n$  gibt, so dass  $P(X \in C) = 1$ .

2. Sei  $X = (X_1, \dots, X_n)$  ein diskreter Zufallsvektor. Dann heißt  $\{P(X = x), x \in C\}$  *Wahrscheinlichkeitsfunktion* von  $X$ .

### Beachte

- Wenn  $X = (X_1, \dots, X_n)$  ein diskreter Zufallsvektor ist, dann sind auch seine Komponenten  $X_1, \dots, X_n$  diskrete Zufallsvariablen, d.h., es gibt abzählbare Mengen  $C_1, \dots, C_n \subset \mathbb{R}^n$ , so dass

$$P(X_1 \in C_1) = \dots = P(X_n \in C_n) = 1.$$

- Dabei gilt  $C = \{x = (x_1, \dots, x_n) : x_1 \in C_1, \dots, x_n \in C_n\}$ .
- Durch die Wahrscheinlichkeitsfunktion  $\{P(X = x), x \in C\}$  des Zufallsvektors  $X = (X_1, \dots, X_n)$  lassen sich die Wahrscheinlichkeitsfunktionen

$$\{P(X_1 = x_1), x_1 \in C_1\}, \dots, \{P(X_n = x_n), x_n \in C_n\}$$

seiner Komponenten  $X_1, \dots, X_n$  ausdrücken, denn es gilt

$$\begin{aligned} P(X_i = x_i) &= \sum_{y_1 \in C_1} \dots \sum_{y_{i-1} \in C_{i-1}} \sum_{y_{i+1} \in C_{i+1}} \dots \sum_{y_n \in C_n} P(X_1 = y_1, \dots, X_{i-1} = y_{i-1}, X_i = x_i, X_{i+1} = y_{i+1}, \\ &\quad \dots, X_n = y_n). \end{aligned}$$

**Definition** Der Zufallsvektor  $X = (X_1, \dots, X_n)$  heißt *absolutstetig*, wenn es eine (Borel-messbare) Funktion  $f_X : \mathbb{R}^n \rightarrow [0, \infty)$  gibt, so dass

$$F_X(x_1, \dots, x_n) = \int_{(-\infty, x_n] \times \dots \times (-\infty, x_1]} f_X(y_1, \dots, y_n) d(y_1, \dots, y_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}. \quad (23)$$

Die Funktion  $f_X$  heißt die (gemeinsame) *Dichte* von  $X$ .

**Theorem 3.8** Sei  $X = (X_1, \dots, X_n)$  ein absolutstetiger Zufallsvektor mit der Dichte  $f_X$ . Dann gilt:

1.

$$P_X(B) = \int_B f_X(y) dy \quad \forall B \in \mathcal{B}(\mathbb{R}^n). \quad (24)$$

2. Die Komponenten  $X_1, \dots, X_n$  von  $X$  sind absolutstetige Zufallsvariablen, wobei die Dichte  $f_{X_i} : \mathbb{R} \rightarrow [0, \infty)$  von  $X_i$  gegeben ist durch

$$f_{X_i}(x_i) = \underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}}}_{(n-1)\text{-mal}} f_X(y_1, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_n) dy_1 \dots dy_{i-1} dy_{i+1} \dots dy_n. \quad (25)$$

### Beweis

- Aus der Definitionsgleichung (23) und aus dem eindeutigen Zusammenhang zwischen der Verteilung  $P_X$  und der Verteilungsfunktion  $F_X$  von  $X$  ergibt sich unmittelbar die Gültigkeit von (24).
- Außerdem ergibt sich aus Teilaussage 1.3 von Theorem 3.7 und aus der Definitionsgleichung (23), dass

$$\begin{aligned} F_{X_i}(x_i) &= F_X(\infty, \dots, \infty, x_i, \infty, \dots, \infty) \\ &= \int_{(-\infty, x_i]} \underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}}}_{(n-1)\text{-mal}} f_X(y_1, \dots, y_{i-1}, y_i, y_{i+1}, \dots, y_n) dy_1 \dots dy_{i-1} dy_{i+1} \dots dy_n dy_i. \end{aligned}$$

- Aus dieser Darstellungsformel für die Verteilungsfunktion  $F_{X_i}$  folgt unmittelbar, dass  $X_i$  absolutstetig ist und die in (25) angegebene Dichte besitzt.  $\square$

### Beachte

- Die in (25) betrachtete Funktion  $f_{X_i}$  heißt *Randdichte* von  $X$ .
- Die Umkehrung der zweiten Teilaussage von Theorem 3.8 gilt im allgemeinen nicht.
- Beispielsweise ist der Zufallsvektor  $Y = (X, \dots, X)$  mit  $X \sim N(0, 1)$  *nicht* absolutstetig, obwohl seine Komponenten absolutstetig sind.

### 3.3.3 Weitere Beispiele von Zufallsvektoren

#### 1. $n$ -maliger Münzwurf

- Betrachten den in Abschnitt 3.2.1 eingeführten Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  mit der Grundmenge

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in \{0, 1\}\}$$

- bei identischen Erfolgswahrscheinlichkeiten  $a_1 = \dots = a_n = p$ ;
- Sei  $X$  = Anzahl der Erfolge bei  $n$  Versuchen. Dann ist  $X$  binomialverteilt, d.h.

$$p_k = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

- Außerdem betrachten wir die (zufällige) Nummer  $Y$  desjenigen Versuches, bei dem zum ersten Mal ein Erfolg eintritt, d.h.

$$Y(\omega) = \begin{cases} \inf\{i \geq 1 : \omega_i = 1\}, & \text{falls } X(\omega) \geq 1 \\ n+1, & \text{falls } X(\omega) = 0 \end{cases}$$

- Gesucht sind die Einzelwahrscheinlichkeiten des Zufallsvektors  $(X, Y)$ :

$$p(x, y) = P(X = x, Y = y) \quad \text{für } 0 \leq x \leq n, 1 \leq y \leq n+1$$

- Offenbar gilt  $p(0, n+1) = (1-p)^n$  und  $p(0, y) = 0$  für  $1 \leq y \leq n$ .
- Sei jetzt  $X(\omega) = x \geq 1$  und  $Y(\omega) = y$ , d.h.,  $\omega_i = 0$  für  $i < y$  und  $\omega_y = 1$ , und es müssen genau  $x-1$  Einsen unter  $\omega_{y+1}, \dots, \omega_n$  sein.
- Dafür gibt es  $\binom{n-y}{x-1}$  Möglichkeiten, falls  $x-1 \leq n-y$ , und 0 Möglichkeiten, falls  $x-1 > n-y$ .
- Also ist

$$p(x, y) = \begin{cases} \binom{n-y}{x-1} p^x (1-p)^{n-x}, & \text{falls } x-1 \leq n-y, \\ 0, & \text{falls } x-1 > n-y. \end{cases}$$

- Spezialfall. Sei jetzt  $n = 3$  und  $p = \frac{1}{2}$ . Dann sind die Wahrscheinlichkeiten  $P(X = x, Y = y)$  bzw. die (Rand-) Wahrscheinlichkeiten  $P(X = x)$  und  $P(Y = y)$  in der folgenden Tabelle gegeben:

$P(X = x, Y = y)$		$y$				$P(X = x)$
		1	2	3	4	
$x$	0	0	0	0	$\frac{1}{8}$	$\frac{1}{8}$
	1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{3}{8}$
	2	$\frac{2}{8}$	$\frac{1}{8}$	0	0	$\frac{3}{8}$
	3	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$P(Y = y)$		$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	

## 2. zweidimensionale integrierte Gleichverteilung

- Sei  $X = (X_1, X_2)$  ein absolutstetiger Zufallsvektor mit der (gemeinsamen) Dichte

$$f_X(x_1, x_2) = \begin{cases} 4x_1x_2, & \text{falls } 0 \leq x_1, x_2 \leq 1 \\ 0 & \text{sonst} \end{cases}$$

- Für die Randdichten folgt dann aus (25), dass

$$f_{X_1}(x_1) = \int_0^1 f_X(x_1, x_2) dx_2 = 4x_1 \cdot \frac{1}{2} = 2x_1 \quad \forall x_1 \in [0, 1]$$

und analog

$$f_{X_2}(x_2) = \int_0^1 f_X(x_1, x_2) dx_1 = 2x_2 \quad \forall x_2 \in [0, 1]$$

- Wir wollen nun noch die Wahrscheinlichkeit  $P(X_1 \leq 2X_2)$  berechnen. Wegen (24) gilt

$$\begin{aligned} P(X_1 \leq 2X_2) &= P_X((x_1, x_2) : 0 \leq x_1, x_2 \leq 1, x_1 \leq 2x_2) \\ &= \int_0^{\frac{1}{2}} \int_0^{2x_2} 4x_1x_2 dx_1 dx_2 + \int_{\frac{1}{2}}^1 \int_0^1 4x_1x_2 dx_1 dx_2 \\ &= \int_0^{\frac{1}{2}} 4x_2 \left[ \frac{x_1^2}{2} \right]_0^{2x_2} dx_2 + \int_{\frac{1}{2}}^1 4x_2 \left[ \frac{x_1^2}{2} \right]_0^1 dx_2 \\ &= 8 \left[ \frac{x_2^4}{4} \right]_0^{\frac{1}{2}} + 2 \left[ \frac{x_2^2}{2} \right]_{\frac{1}{2}}^1 = \frac{7}{8}. \end{aligned}$$

## 3.3.4 Bedingte Wahrscheinlichkeitsfunktion; bedingte Verteilung; bedingte Dichte

Analog zu dem in Abschnitt 2.6.1 eingeführten Begriff der bedingten Wahrscheinlichkeit definieren wir nun die Begriffe der bedingten Verteilung bzw. der bedingten Dichte.

**Definition** (bedingte Wahrscheinlichkeitsfunktion und bedingte Verteilung)

- Sei  $(X, Y)$  ein diskreter Zufallsvektor mit  $P((X, Y) \in C) = 1$  für eine abzählbare Menge  $C = \{(x_i, y_j) : i, j \in \mathbb{N}\}$ .
- Für jedes  $j \in \mathbb{N}$  mit  $P(Y = y_j) > 0$  heißt dann  $\{P(X = x_i | Y = y_j), i \in \mathbb{N}\}$  die *bedingte Wahrscheinlichkeitsfunktion* von  $X$  unter der Bedingung  $\{Y = y_j\}$ .
- Sie bestimmt die *bedingte Verteilung*  $\{P(X \in B | Y = y_j), B \in \mathcal{B}(\mathbb{R})\}$  von  $X$  unter der Bedingung  $\{Y = y_j\}$  eindeutig.
- Analog heißt  $\{P(Y = y_j | X = x_i), j \in \mathbb{N}\}$  bzw.  $\{P(Y \in B | X = x_i), B \in \mathcal{B}(\mathbb{R})\}$  für jedes  $i \in \mathbb{N}$  mit  $P(X = x_i) > 0$  die bedingte Wahrscheinlichkeitsfunktion bzw. bedingte Verteilung von  $Y$  unter der Bedingung  $\{X = x_i\}$ .

**Definition** (bedingte Dichte)

- Sei  $(X, Y)$  ein absolutstetiger Zufallsvektor mit der gemeinsamen Dichte  $f_{(X,Y)}(x, y)$ .
- Dann heißt die Funktion  $f_{X|Y=y} : \mathbb{R} \rightarrow [0, \infty)$  mit

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}$$

die *bedingte Dichte* von  $X$  unter der Bedingung  $\{Y = y\}$ , wobei  $f_Y(y) > 0$  vorausgesetzt wird.

- Analog heißt die Funktion  $f_{Y|X=x} : \mathbb{R} \rightarrow [0, \infty)$  mit

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$$

die bedingte Dichte von  $Y$  unter der Bedingung  $\{X = x\}$ , wobei  $f_X(x) > 0$  vorausgesetzt wird.

### Beispiele

#### 1. zweimaliges Würfeln

Sei  $X : \Omega \rightarrow \{0, 1, 2\}$  bzw.  $Y : \Omega \rightarrow \{0, 1, 2\}$  die (zufällige) Anzahl, mit der die Augenzahl „6“ bzw. „1“ beim zweimaligen Würfeln erzielt wird; vgl. Abschnitt 3.3.1. Dann ergeben sich die folgenden bedingten Wahrscheinlichkeitsfunktionen von  $X$  unter der Bedingung  $\{Y = j\}$ :

$P(X = i   Y = j)$		$i$		
		0	1	2
$j$	0	$\frac{16}{25}$	$\frac{8}{25}$	$\frac{1}{25}$
	1	$\frac{8}{10}$	$\frac{2}{10}$	0
	2	1	0	0

#### 2. integrierte Gleichverteilung

Sei  $(X, Y)$  ein absolutstetiger Zufallsvektor mit der gemeinsamen Dichte

$$f_{(X,Y)}(x,y) = \begin{cases} 4xy, & \text{falls } 0 \leq x, y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Für  $y \in (0, 1]$  gilt dann für die bedingte Dichte von  $X$  unter der Bedingung  $\{Y = y\}$ :

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} = \begin{cases} 2x, & \text{falls } x \in [0, 1] \\ 0 & \text{sonst} \end{cases}$$

Die bedingte Dichte  $f_{X|Y=y}(x)$  stimmt also bei diesem Beispiel mit der (unbedingten Rand-) Dichte  $f_X(x)$  von  $X$  überein; vgl. Abschnitt 3.3.3.

### 3.3.5 Unabhängige Zufallsvariablen

Die Unabhängigkeit von Zufallsvariablen wird durch den in Abschnitt 2.7 eingeführten Begriff der Unabhängigkeit von Ereignissen ausgedrückt.

So heißen zwei Zufallsvariablen  $X_1, X_2 : \Omega \rightarrow \mathbb{R}$  unabhängig, wenn die Ereignisse  $\{X_1 \leq x_1\}$  und  $\{X_2 \leq x_2\}$  für beliebige  $x_1, x_2 \in \mathbb{R}$  unabhängig sind.

Für Folgen von Zufallsvariablen wird der Begriff der Unabhängigkeit folgendermaßen gebildet.

**Definition** Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum.

1. Die Zufallsvariablen  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  heißen *unabhängig*, falls

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n. \quad (26)$$

2. Sei  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  eine beliebige (unendliche) Folge von Zufallsvariablen. Dann sagt man, dass  $X_1, X_2, \dots$  *unabhängige Zufallsvariablen* sind, falls jede endliche Teilfolge  $X_{i_1}, \dots, X_{i_k}$  von  $X_1, X_2, \dots$  aus unabhängigen Zufallsvariablen besteht.

**Beachte**

- Aus den Definitionen der Verteilungsfunktionen  $F_{(X_1, \dots, X_n)}$  und  $F_{X_1}, \dots, F_{X_n}$  ergibt sich sofort, dass die Definitionsgleichung (26) äquivalent ist mit

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \dots P(X_n \leq x_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}. \quad (27)$$

- Darüber hinaus kann man zeigen, dass (27) und damit auch (26) äquivalent ist mit der folgenden (scheinbar schärferen) Bedingung.

**Theorem 3.9** *Die Zufallsvariablen  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  sind genau dann unabhängig, wenn*

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n) \quad \forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}). \quad (28)$$

**Beweis**

- Es ist klar, dass (27) aus (28) folgt. Hierfür genügt es,  $B_i = (-\infty, x_i]$  zu setzen.
- Umgekehrt ergibt sich (28) aus (27) aus dem Monotone Class Theorem (vgl. Theorem 3.1), wobei ähnlich wie im Beweis von Theorem 3.3 vorgegangen werden kann.  $\square$

Hieraus und aus den Definitionsgleichungen (13) und (23) von  $f_{(X_1, \dots, X_n)}(x_1, \dots, x_n)$  und  $f_{X_1}(x_1), \dots, f_{X_n}(x_n)$  ergibt sich unmittelbar die folgende Charakterisierung der Unabhängigkeit von diskreten bzw. absolutstetigen Zufallsvariablen.

**Theorem 3.10**

1. Sei  $X = (X_1, \dots, X_n)$  ein diskreter Zufallsvektor mit

$$P(X \in C) = 1$$

für eine abzählbare Menge  $C \subset \mathbb{R}^n$ . Seine Komponenten  $X_1, \dots, X_n$  sind genau dann unabhängige Zufallsvariablen, wenn

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n) \quad \forall (x_1, \dots, x_n) \in C. \quad (29)$$

2. Sei  $X = (X_1, \dots, X_n)$  ein absolutstetiger Zufallsvektor. Seine Komponenten  $X_1, \dots, X_n$  sind genau dann unabhängige Zufallsvariablen, wenn

$$f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n) \quad (30)$$

für fast alle  $x_1, \dots, x_n \in \mathbb{R}$  gilt, d.h., für alle  $x = (x_1, \dots, x_n) \in \mathbb{R}^n \setminus B$ , wobei die (Borelsche) „Ausnahmemenge“  $B \subset \mathbb{R}^n$  das ( $n$ -dimensionale) Lebesgue-Maß 0 hat.

**Beweis**

- Sei  $X = (X_1, \dots, X_n)$  ein diskreter Zufallsvektor.
  - Offenbar folgt dann (29) aus (28). Hierfür genügt es,  $B_i = \{x_i\}$  zu setzen.

– Umgekehrt ergibt sich (28) sofort aus (29), denn es gilt

$$\begin{aligned}
 P(X_1 \in B_1, \dots, X_n \in B_n) &= \sum_{x_1 \in B_1 \cap C_1} \dots \sum_{x_n \in B_n \cap C_n} P(X_1 = x_1, \dots, X_n = x_n) \\
 &\stackrel{(29)}{=} \sum_{x_1 \in B_1 \cap C_1} \dots \sum_{x_n \in B_n \cap C_n} P(X_1 = x_1) \dots P(X_n = x_n) \\
 &= \sum_{x_1 \in B_1 \cap C_1} P(X_1 = x_1) \dots \sum_{x_n \in B_n \cap C_n} P(X_n = x_n) \\
 &= P(X_1 \in B_1) \dots P(X_n \in B_n),
 \end{aligned}$$

wobei  $C_i \subset \mathbb{R}$  eine abzählbare Menge ist mit  $P(X_i \in C_i) = 1$ ;  $i = 1, \dots, n$ .

– Damit ist die erste Teilaussage bewiesen.

- Sei nun  $X = (X_1, \dots, X_n)$  ein absolutstetiger Zufallsvektor mit unabhängigen Komponenten.

– Dann ergibt sich aus (23) und (26), dass für beliebige  $x_1, \dots, x_n \in \mathbb{R}$

$$\begin{aligned}
 &\int_{(-\infty, x_n]} \dots \int_{(-\infty, x_1]} f_X(y_1, \dots, y_n) dy_1 \dots dy_n \\
 &= F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n) = \int_{(-\infty, x_1]} f_{X_1}(y_1) dy_1 \dots \int_{(-\infty, x_n]} f_{X_n}(y_n) dy_n \\
 &= \int_{(-\infty, x_n]} \dots \int_{(-\infty, x_1]} f_{X_1}(y_1) \dots f_{X_n}(y_n) dy_1 \dots dy_n.
 \end{aligned}$$

– Hieraus und aus den allgemeinen (Eindeutigkeits-) Eigenschaften von Lebesgue-Integralen folgt die Gültigkeit von (30) für fast alle  $x_1, \dots, x_n \in \mathbb{R}$ .

– Umgekehrt ergibt sich aus (23) und (30) sofort die Gültigkeit von (26) und damit die Unabhängigkeit der Komponenten  $X_1, \dots, X_n$  von  $X$ .  $\square$

### 3.3.6 Beispiele: Zufallsvektoren mit unabhängigen Komponenten

#### 1. $n$ -maliger Münzwurf

- Wir betrachten den in Abschnitt 3.2.1 eingeführten Wahrscheinlichkeitsraum  $(\Omega, \mathcal{P}(\Omega), P)$  mit der Grundmenge

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \{\omega : \omega = (\omega_1, \dots, \omega_n), \omega_i \in \{0, 1\}\}$$

und dem Wahrscheinlichkeitsmaß  $P$ , das durch

$$P(\{\omega\}) = \prod_{i:\omega_i=1} a_i \prod_{j:\omega_j=0} (1 - a_j) \quad (31)$$

gegeben ist, wobei  $0 \leq a_1, \dots, a_n \leq 1$ .

- Außerdem betrachten wir die Zufallsvariablen  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ , die durch die Projektion  $X_i(\omega) = \omega_i$  für  $i = 1, \dots, n$  gegeben seien.
- Aus (29) und (31) folgt unmittelbar, dass  $X_1, \dots, X_n$  unabhängige Zufallsvariablen sind.
- Man kann jedoch auch in diesem Modell Zufallsvariablen konstruieren, die nicht unabhängig sind. Sei nämlich  $a_1 = 0.5$ , und sei  $Y_1 : \Omega \rightarrow \mathbb{R}$  gegeben durch

$$Y_1(\omega) = \begin{cases} 1 & \text{falls } \omega_1 = 0 \\ 0 & \text{falls } \omega_1 = 1 \end{cases}$$



- Dann sind  $X_1$  und  $Y_1$  nicht unabhängig, denn es gilt

$$P(X_1 = 1, Y_1 = 1) = 0 \quad \neq \quad P(X_1 = 1)P(Y_1 = 1) = \frac{1}{4}.$$

## 2. zweidimensionale Normalverteilung

- Betrachten zwei unabhängige Zufallsvariablen  $X_1, X_2$ , die standardnormalverteilt sind. D.h.,

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

- Für die (gemeinsame) Dichte  $f_X(x_1, x_2)$  des Zufallsvektors  $X = (X_1, X_2)$  gilt

$$f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right).$$

- Man sagt dann, dass auch der Zufallsvektor  $X = (X_1, X_2)$  *standardnormalverteilt* ist.
- *Verallgemeinerung*: Sei  $\rho \in [0, 1)$ , und sei  $X = (X_1, X_2)$  ein absolutstetiger Zufallsvektor mit der Dichte

$$f_X(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2}\right) \quad \forall x_1, x_2 \in \mathbb{R}.$$

- Dann gilt für die (Rand-) Dichte  $f_{X_1}(x_1)$  von  $X_1$

$$\begin{aligned} f_{X_1}(x_1) &= \int_{\mathbb{R}} f_X(x_1, x_2) dx_2 \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2} \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2}\right) dx_2 \\ &\stackrel{\text{quadratische Ergänzung}}{=} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \frac{x_1^2(1-\rho^2)}{1-\rho^2}\right) \int_{\mathbb{R}} \exp\left(-\frac{1}{2} \frac{(x_2 - \rho x_1)^2}{1-\rho^2}\right) dx_2. \end{aligned}$$

- Durch die Substitution

$$u = \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}} \quad \text{bzw.} \quad du = \frac{1}{\sqrt{1-\rho^2}} dx_2$$

ergibt sich also

$$f_{X_1}(x_1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}x_1^2\right) \underbrace{\int_{\mathbb{R}} \exp\left(-\frac{1}{2}u^2\right) du}_{=\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right).$$

- Analog gilt

$$f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_2^2\right).$$

- Die Komponenten  $X_1, X_2$  des Zufallsvektors  $X = (X_1, X_2)$  sind also für *jedes*  $\rho \in [0, 1)$  standardnormalverteilt.
- Beachte jedoch, dass  $X_1$  und  $X_2$  nur dann unabhängig sind, wenn  $\rho = 0$ . Denn für  $\rho > 0$  gilt

$$f_X(x_1, x_2) \neq f_{X_1}(x_1) \cdot f_{X_2}(x_2).$$

### 3.4 Funktionen von Zufallsvektoren

#### 3.4.1 Zusammengesetzte Abbildungen

**Beispiel** (Klassifikation)

- Manchmal ist es zweckmäßig, die Werte von Zufallsvariablen  $X : \Omega \rightarrow \mathbb{R}$  zu klassifizieren. Dabei wird der Wertebereich  $\mathbb{R}$  von  $X$  in  $m$  Klassen zerlegt, die wir mit der Menge der ersten  $m$  natürlichen Zahlen  $\{1, 2, \dots, m\}$  identifizieren.
- Mit anderen Worten: Außer der Zufallsvariablen  $X$  betrachten wir noch eine weitere Abbildung  $\varphi : \mathbb{R} \rightarrow \{1, 2, \dots, m\}$ .
- Durch Nacheinanderausführung der Abbildungen  $X$  und  $\varphi$  ergibt sich dann die Abbildung  $\varphi(X) : \Omega \rightarrow \{1, 2, \dots, m\}$  mit  $\varphi(X)(\omega) = \varphi(X(\omega))$ , die jedem  $\omega \in \Omega$  die Klasse  $\varphi(X)(\omega)$  zuordnet.
- Um die Wahrscheinlichkeit bestimmen zu können, dass die Zufallsvariable  $X$  Werte in Klasse  $i$  annimmt, muss gewährleistet sein, dass

$$\{\omega : \omega \in \Omega, \varphi(X)(\omega) = i\} \in \mathcal{F}. \quad (32)$$

- Die Abbildung  $\varphi(X)$  muss also die Regularitätseigenschaft einer Zufallsvariablen besitzen, d.h. der Messbarkeitsbedingung (1) genügen.
- Um dies zu erreichen, wird über die Abbildung  $\varphi$  vorausgesetzt, dass  $\{x : x \in \mathbb{R}, \varphi(x) = i\}$  eine Teilmenge von  $\mathbb{R}$  aus  $\mathcal{B}(\mathbb{R})$ , d.h. eine Borel-Menge ist, für jedes  $i = 1, 2, \dots, m$ .
- Man kann zeigen, dass dann (32) für jedes  $i \in \{1, 2, \dots, m\}$  bzw.  $\{\omega : \omega \in \Omega, \varphi(X)(\omega) \leq x\} \in \mathcal{F}$  für jedes  $x \in \mathbb{R}$  gilt, d.h.,  $\varphi(X)$  ist eine Zufallsvariable.

**Beachte**

- Es interessieren auch Funktionen  $\varphi : \mathbb{R}^n \rightarrow \{1, 2, \dots, m\}$  von Zufallsvektoren  $X : \Omega \rightarrow \mathbb{R}^n$ .
- Damit auch in diesem Fall eine Bedingung an  $\varphi$  formuliert werden kann, so dass die zusammengesetzte Abbildung  $\varphi(X) : \Omega \rightarrow \{1, 2, \dots, m\}$  eine Zufallsvariable ist, wird die *Borel- $\sigma$ -Algebra*  $\mathcal{B}(\mathbb{R}^n)$  von Teilmengen des  $\mathbb{R}^n$  betrachtet.
- Sie ist definiert als die kleinste  $\sigma$ -Algebra von Teilmengen des  $\mathbb{R}^n$ , die alle offenen Quader  $\times_{i=1}^n (a_i, b_i)$  enthält;  $-\infty < a_i < b_i < \infty$ . D.h.

$$\mathcal{B}(\mathbb{R}^n) = \sigma \left( \underbrace{\left\{ \times_{i=1}^n (a_i, b_i), -\infty < a_i < b_i < \infty \right\}}_{\text{Erzeugersystem}} \right).$$

- Die Abbildung  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  heißt *Borel-messbar*, wenn

$$\{x : x \in \mathbb{R}^n, \varphi(x) \leq y\} \in \mathcal{B}(\mathbb{R}^n) \quad \forall y \in \mathbb{R}. \quad (33)$$

- Zwei einfache Beispiele solcher (Borel-messbarer) Abbildungen  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  sind gegeben durch  $\varphi(x_1, \dots, x_n) = x_1 + \dots + x_n$  bzw.  $\varphi(x_1, \dots, x_n) = x_1 \cdot \dots \cdot x_n$ , vgl. auch Abschnitt 3.4.4.

Allgemein gilt für zusammengesetzte Abbildungen

**Theorem 3.11** *Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum, und sei  $X : \Omega \rightarrow \mathbb{R}^n$  ein beliebiger Zufallsvektor. Außerdem sei  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  eine Borel-messbare Abbildung, d.h., die Bedingung (33) sei erfüllt. Dann ist die zusammengesetzte Abbildung  $\varphi(X) : \Omega \rightarrow \mathbb{R}$  mit  $\varphi(X)(\omega) = \varphi(X(\omega))$  eine (reellwertige) Zufallsvariable, d.h., es gilt*

$$\{\omega : \omega \in \Omega, \varphi(X)(\omega) \leq z\} \in \mathcal{F} \quad \forall z \in \mathbb{R}. \quad (34)$$

**Beweis** Um die Gültigkeit von (34) zu zeigen, genügt es zu beachten, dass für das Urbild  $\varphi^{-1}(-\infty, z]$  wegen (33) gilt

$$\varphi^{-1}(-\infty, z] = \{x : x \in \mathbb{R}^n, \varphi(x) \leq z\} \in \mathcal{B}(\mathbb{R}^n) \quad \forall z \in \mathbb{R}.$$

Hieraus und aus der Borel-Messbarkeit von  $X : \Omega \rightarrow \mathbb{R}^n$  folgt, dass

$$\{\omega : \omega \in \Omega, \varphi(X)(\omega) \leq z\} = \{\omega : \omega \in \Omega, X(\omega) \in \varphi^{-1}(-\infty, z]\} \in \mathcal{F}$$

für jedes  $z \in \mathbb{R}$ . □

### Beispiel

- Sei  $X = (X_1, \dots, X_n)$  ein beliebiger Zufallsvektor;  $n \geq 2$ .
- Betrachten die folgende Abbildung  $\varphi : \mathbb{R}^n \rightarrow \{0, 1, \dots, m\}$ , die als Klassifikation der Werte von  $X$  aufgefasst werden kann.
- Sei  $\varepsilon > 0$  eine beliebige, jedoch fest vorgegebene Toleranzgrenze.
- Für jedes  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  sei

$$\varphi(x) = |\{(i, j) : 1 \leq i < j \leq n, -\varepsilon \leq x_i - x_j \leq \varepsilon\}|, \quad (35)$$

d.h.,  $\varphi(x)$  ist die Anzahl derjenigen Paare von Komponenten  $x_i, x_j$  von  $x$ , die sich um nicht mehr als  $\varepsilon$  voneinander unterscheiden.

- Der Wertebereich  $\mathbb{R}^n$  von  $X$  wird dabei in  $m + 1 = \binom{n}{2} + 1$  Klassen zerlegt.
- Es ist nicht schwierig zu zeigen, dass die in (35) gegebene Abbildung die Regularitätsbedingung (33) erfüllt.

Eine weitere wichtige Klasse von Borel-messbaren Funktionen ist durch das folgende Lemma gegeben.

**Lemma 3.2** Jede stetige Funktion  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  genügt der Bedingung (33), d.h., jede stetige Funktion ist Borel-messbar.

**Beweis** Sei  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  eine stetige Abbildung. Die Begründung der Behauptung lässt sich nun in folgende Schritte zerlegen:

- Weil  $\varphi^{-1}(-\infty, y] = \{x : x \in \mathbb{R}^n, \varphi(x) \leq y\}$  und weil für stetige Abbildungen die Urbilder abgeschlossener Mengen erneut abgeschlossene Mengen sind, ist  $\{x : x \in \mathbb{R}^n, \varphi(x) \leq y\}$  eine abgeschlossene Teilmenge des  $\mathbb{R}^n$ .
- Außerdem kann man sich leicht überlegen, dass sich jede offene Teilmenge des  $\mathbb{R}^n$  als abzählbare Vereinigung von Quadern darstellen lässt.
- Also ist jede offene Teilmenge des  $\mathbb{R}^n$  eine Borel-Menge.
- Deshalb ist auch jede abgeschlossene Teilmenge des  $\mathbb{R}^n$  (als das Komplement einer offenen Menge) eine Borel-Menge.
- Es gilt also  $\{x : x \in \mathbb{R}^n, \varphi(x) \leq y\} \in \mathcal{B}(\mathbb{R}^n)$ . □

**Korollar 3.1** Sei  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  ein Polynom, d.h., es gelte

$$\varphi(x_1, \dots, x_n) = \sum_{i=0}^k a_i x_1^{m_{i1}} \dots x_n^{m_{in}} \quad (36)$$

für ein  $k \in \mathbb{N}$  und für gewisse Konstanten  $a_0, a_1, \dots, a_k \in \mathbb{R}$ ,  $m_{i1}, \dots, m_{in} \geq 0$ . Dann ist  $\varphi$  Borel-messbar.

**Beweis** Die Behauptung ergibt sich unmittelbar aus Lemma 3.2 und aus der Tatsache, dass jedes Polynom eine stetige Abbildung ist. □

In den folgenden Abschnitten 3.4.2 – 3.4.4 wird eine Reihe von Spezialfällen diskutiert, bei denen  $\varphi$  die in (36) gegebene Form hat.

### 3.4.2 Lineare Transformation

Ein wichtiger Spezialfall einer zusammengesetzten Abbildung ist die lineare Transformation von Zufallsvariablen, wobei  $n = 1$  und  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  mit  $\varphi(x) = ax + b$ ;  $a, b \in \mathbb{R}$ .

**Theorem 3.12** *Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable und  $a, b \in \mathbb{R}$  beliebige Zahlen mit  $a \neq 0$ . Dann ist  $aX + b$  eine Zufallsvariable, und*

1. die Verteilungsfunktion von  $aX + b$  ist gegeben durch

$$F_{aX+b}(x) = \begin{cases} F_X\left(\frac{x-b}{a}\right), & \text{falls } a > 0, \\ 1 - F_X\left(\frac{x-b}{a}\right) + P\left(X = \frac{x-b}{a}\right), & \text{falls } a < 0. \end{cases} \quad (37)$$

2. Wenn  $X$  absolutstetig ist mit der Dichte  $f_X$ , dann ist auch  $aX + b$  absolutstetig mit der Dichte

$$f_{aX+b}(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right). \quad (38)$$

#### Beweis

- Wenn  $a > 0$ , dann gilt

$$F_{aX+b}(x) = P(aX + b \leq x) = P\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right).$$

- Analog ergibt sich für  $a < 0$

$$\begin{aligned} F_{aX+b}(x) &= P(aX + b \leq x) \\ &= P\left(X \geq \frac{x-b}{a}\right) \\ &= 1 - P\left(X < \frac{x-b}{a}\right) \\ &= 1 - F_X\left(\frac{x-b}{a}\right) + P\left(X = \frac{x-b}{a}\right). \end{aligned}$$

- Damit ist (37) bewiesen.
- Sei nun  $X$  absolutstetig. Weil  $\varphi(x) = ax + b$ , gilt dann  $\varphi^{-1}(x) = a^{-1}(x - b)$  und somit  $(\varphi^{-1})'(x) = a^{-1}$ . Die Behauptung (38) ergibt sich nun aus der folgenden allgemeinen Transformationsformel (40) für Dichten absolutstetiger Zufallsvariablen.  $\square$

**Theorem 3.13** *Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum, und sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable.*

1. Wenn die Abbildung  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  stetig und streng monoton wachsend ist, dann gilt für die Verteilungsfunktion  $F_{\varphi(X)}$  von  $\varphi(X)$

$$F_{\varphi(X)}(x) = F_X(\varphi^{-1}(x)), \quad (39)$$

wobei  $\varphi^{-1}$  die Umkehrfunktion von  $\varphi$  bezeichnet.

2. Sei nun  $X$  absolutstetig mit der Dichte  $f_X$ , sei  $B \subset \mathbb{R}$  eine zusammenhängende offene Menge, so dass  $P(X \in B) = 1$ , und sei  $\varphi$  stetig differenzierbar auf  $B$  mit  $\varphi'(x) \neq 0$  für alle  $x \in B$ . Dann ist  $\varphi(X)$  absolutstetig, und es gilt

$$f_{\varphi(X)}(y) = f_X(\varphi^{-1}(y)) |(\varphi^{-1})'(y)| \quad (40)$$

für alle  $y \in \varphi(B) := \{\varphi(x) : x \in B\}$ .

**Beweis**

- Für jedes  $x \in \mathbb{R}$  gilt

$$F_{\varphi(X)}(x) = P(\varphi(X) \leq x) = P(X \leq \varphi^{-1}(x)) = F_X(\varphi^{-1}(x)).$$

- Damit ist (39) bewiesen.
- Wir zeigen (40) nur für den Spezialfall, dass  $B = \mathbb{R}$ . Dann können wir ohne Einschränkung der Allgemeinheit voraussetzen, dass  $\varphi'(x) > 0$  für alle  $x \in \mathbb{R}$ . (Wenn  $\varphi'(x) < 0$  für alle  $x \in \mathbb{R}$ , dann verläuft der Beweis analog.)
- Sei also  $\varphi'(x) > 0$  für alle  $x \in \mathbb{R}$ .
- Aus Teilaussage 1 ergibt sich dann, dass

$$F_{\varphi(X)}(x) = F_X(\varphi^{-1}(x)) = \int_{(-\infty, \varphi^{-1}(x)]} f_X(u) du = \int_{(-\infty, x]} f_X(\varphi^{-1}(v)) |(\varphi^{-1})'(v)| dv,$$

wobei sich die letzte Gleichheit mit  $u = \varphi^{-1}(v)$  aus den allgemeinen Substitutionsregeln für Lebesgue-Integrale ergibt.

- Hieraus und aus dem eindeutigen Zusammenhang zwischen Verteilungsfunktion und Dichte von absolutstetigen Zufallsvariablen ergibt sich (40), vgl. die Theoreme 3.3 und 3.4.  $\square$

**Beispiel**

- Sei  $X$  standardnormalverteilt, und  $\sigma > 0$ ,  $\mu \in \mathbb{R}$  seien beliebige Konstanten.
- Gemäß Theorem 3.12 ist dann die Zufallsvariable  $Y = \sigma X + \mu$  absolutstetig, und es gilt

$$f_Y(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \quad (41)$$

- Eine Zufallsvariable  $Y$ , deren Dichte durch (41) gegeben ist, heißt *normalverteilt* mit den Parametern  $\mu$  und  $\sigma^2$ ; vgl. auch Abschnitt 3.2.3. Dabei verwenden wir die Schreibweise  $Y \sim N(\mu, \sigma^2)$ .
- Ausgehend von einer normalverteilten Zufallsvariablen  $Y \sim N(\mu, \sigma^2)$  ergibt sich umgekehrt durch die lineare Transformation  $X = (Y - \mu)/\sigma$  eine standardnormalverteilte Zufallsvariable  $X \sim N(0, 1)$ .

**3.4.3 Quadrierung**

**Theorem 3.14** *Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable. Dann gilt*

1. *für die Verteilungsfunktion von  $X^2$*

$$F_{X^2}(x) = \begin{cases} F_X(\sqrt{x}) - F_X(-\sqrt{x}) + P(X = -\sqrt{x}), & \text{falls } x \geq 0, \\ 0, & \text{falls } x < 0. \end{cases}$$

2. *Wenn  $X$  absolutstetig ist mit der Dichte  $f_X$ , dann ist auch  $X^2$  absolutstetig, und es gilt*

$$f_{X^2}(x) = \begin{cases} \frac{1}{2\sqrt{x}}(f_X(\sqrt{x}) + f_X(-\sqrt{x})), & \text{falls } x > 0, \\ 0, & \text{falls } x \leq 0. \end{cases} \quad (42)$$

**Beweis** Wir können ähnlich wie im Beweis von Theorem 3.12 vorgehen:

- Für  $x < 0$  gilt offenbar  $F_{X^2}(x) = 0$ .
- Für  $x \geq 0$  gilt

$$\begin{aligned}
 F_{X^2}(x) &= P(X^2 \leq x) \\
 &= P(-\sqrt{x} \leq X \leq \sqrt{x}) \\
 &= P(X \leq \sqrt{x}) - P(X < -\sqrt{x}) \\
 &= P(X \leq \sqrt{x}) - P(X \leq -\sqrt{x}) + P(X = -\sqrt{x}) \\
 &= F_X(\sqrt{x}) - F_X(-\sqrt{x}) + P(X = -\sqrt{x}).
 \end{aligned}$$

- Damit ist die erste Teilaussage bewiesen.
- Im absolutstetigen Fall gilt für  $x \geq 0$

$$\begin{aligned}
 F_{X^2}(x) &= P(-\sqrt{x} \leq X \leq \sqrt{x}) \\
 &= \int_{(-\sqrt{x}, \sqrt{x})} f_X(u) du \\
 &= \int_{(0, \sqrt{x})} f_X(u) du + \int_{(-\sqrt{x}, 0)} f_X(u) du \\
 &= \int_{(0, x)} \frac{1}{2\sqrt{v}} f_X(\sqrt{v}) dv + \int_{(0, x)} \frac{1}{2\sqrt{v}} f_X(-\sqrt{v}) dv \\
 &= \int_{(0, x)} \frac{1}{2\sqrt{v}} (f_X(\sqrt{v}) + f_X(-\sqrt{v})) dv,
 \end{aligned}$$

wobei in der vorletzten Gleichheit die Substitution  $u = \sqrt{v}$  im ersten Integral und die Substitution  $u = -\sqrt{v}$  im zweiten Integral verwendet wurde.

- Damit ist gezeigt, dass die Verteilungsfunktion  $F_{X^2}$  von  $X^2$  eine Integraldarstellung mit der in (42) gegebenen Dichte besitzt.  $\square$

**Beispiel** Wenn  $X \sim N(0, 1)$ , dann ergibt sich aus (42):

$$f_{X^2}(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} \exp(-\frac{x}{2}), & \text{falls } x > 0, \\ 0, & \text{falls } x \leq 0. \end{cases} \quad (43)$$

**Beachte** Die Summe von  $n$  unabhängigen (und identisch verteilten) Zufallsvariablen, deren Dichte durch (43) gegeben ist, heißt  $\chi^2$ -verteilt mit  $n$  sogenannten Freiheitsgraden. Die  $\chi^2$ -Verteilungen werden in der Vorlesung Statistik I genauer diskutiert. Sie sind eine Familie von sogenannten *statistischen Prüfverteilungen*.

### 3.4.4 Summe, Produkt und Quotient von unabhängigen Zufallsvariablen

**Theorem 3.15** Sei  $X = (X_1, X_2) : \Omega \rightarrow \mathbb{R}^2$  ein absolutstetiger Zufallsvektor mit der (gemeinsamen) Dichte  $f_X$ . Dann ist auch die Zufallsvariable  $X_1 + X_2$  absolutstetig, und ihre Dichte ist gegeben durch

$$f_{X_1+X_2}(z) = \int_{\mathbb{R}} f_X(t, z-t) dt \quad \forall z \in \mathbb{R}. \quad (44)$$

Wenn die Zufallsvariablen  $X_1, X_2$  unabhängig sind, dann gilt insbesondere die sogenannte Faltungsformel

$$f_{X_1+X_2}(z) = \int_{\mathbb{R}} f_{X_1}(t) f_{X_2}(z-t) dt \quad \forall z \in \mathbb{R}. \quad (45)$$

**Beweis**

- Wir nutzen die Tatsache, dass zwischen Verteilungsfunktion und Dichte einer Zufallsvariablen eine eindeutige Zuordnung besteht, und zeigen,
- dass das Integral der Funktion in (44) die Verteilungsfunktion von  $X_1 + X_2$  ergibt.
- Und zwar gilt für jedes  $z \in \mathbb{R}$

$$\begin{aligned}
\int_{-\infty}^z \int_{\mathbb{R}} f_X(t, v-t) dt dv &= \int_{\mathbb{R}} \int_{-\infty}^z f_X(t, \underbrace{v-t}_{=u}) dv dt \\
&= \int_{\mathbb{R}} \int_{-\infty}^{z-t} f_X(t, u) du dt \\
&= \int_{\{(t, u): t+u \leq z\}} f_X(t, u) d(u, t) \\
&= P_X((t, u) : t+u \leq z) \\
&= P(X_1 + X_2 \leq z).
\end{aligned}$$

- Die Faltungsformel (45) ergibt sich unmittelbar aus (44), weil  $f_{(X_1, X_2)}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$  für fast alle  $(x_1, x_2) \in \mathbb{R}^2$ , falls  $X_1$  und  $X_2$  unabhängig sind.  $\square$

**Korollar 3.2** Wenn die Zufallsvariablen  $X_1, X_2$  unabhängig sind mit  $X_1 \sim N(\mu_1, \sigma_1^2)$  bzw.  $X_2 \sim N(\mu_2, \sigma_2^2)$ , dann ist auch die Summe  $X_1 + X_2$  normalverteilt mit

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2). \quad (46)$$

**Beweis**

- Aus (45) und aus Formel (17) für die Dichte der Normalverteilung ergibt sich

$$\begin{aligned}
f_{X_1+X_2}(z) &= \int_{\mathbb{R}} f_{X_1}(t) f_{X_2}(z-t) dt \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{t-\mu_1}{\sigma_1}\right)^2\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{z-t-\mu_2}{\sigma_2}\right)^2\right) dt \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{t}{\sigma_1}\right)^2\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{z-(\mu_1+\mu_2)-t}{\sigma_2}\right)^2\right) dt \\
&= \int_{\mathbb{R}} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{t}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{z-(\mu_1+\mu_2)-t}{\sigma_2}\right)^2\right) dt \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\frac{(z-(\mu_1+\mu_2))^2}{\sigma_1^2+\sigma_2^2}\right) g(z),
\end{aligned}$$

wobei

$$\begin{aligned}
g(z) &= \int_{\mathbb{R}} \exp\left(-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(t - \frac{\sigma_1^2(z-(\mu_1+\mu_2))}{\sigma_1^2+\sigma_2^2}\right)^2\right) dt \\
&= \int_{\mathbb{R}} \exp\left(-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2} t^2\right) dt \\
&= \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}t^2\right) dt = \frac{\sqrt{2\pi}\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}.
\end{aligned}$$

Also gilt

$$f_{X_1+X_2}(z) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\frac{z-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)^2\right).$$

- Damit ist gezeigt, dass  $f_{X_1+X_2}$  die Dichte der Normalverteilung  $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$  ist.
- Wegen des eindeutigen Zusammenhanges zwischen Dichte und Verteilung gilt also (46).  $\square$

**Beachte**

- Das in Korollar 3.2 gegebene Additionstheorem für unabhängige und normalverteilte Zufallsvariablen wird auch *Faltungsstabilität* der Normalverteilung genannt.
- Für eine beliebige Anzahl  $n \geq 2$  von unabhängigen Zufallsvariablen  $X_1, \dots, X_n$  mit  $X_i \sim N(\mu_i, \sigma_i^2)$  für alle  $i \in \{1, \dots, n\}$  ergibt sich nun durch Iteration, dass

$$X_1 + \dots + X_n \sim N(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2).$$

Völlig analog zu Theorem 3.15 ergibt sich

**Theorem 3.16** Die Zufallsvariablen  $X_1$  und  $X_2$  seien unabhängig und absolutstetig mit den Dichten  $f_{X_1}$  und  $f_{X_2}$ . Dann sind die Zufallsvariablen  $X_1 \cdot X_2$  und  $X_1/X_2$  absolutstetig, und ihre Dichten sind gegeben durch

$$f_{X_1 \cdot X_2}(z) = \int_{\mathbb{R}} \frac{1}{|t|} f_{X_1}(t) f_{X_2}\left(\frac{z}{t}\right) dt \quad \forall z \in \mathbb{R} \quad (47)$$

bzw.

$$f_{X_1/X_2}(z) = \int_{\mathbb{R}} |t| f_{X_1}(z \cdot t) f_{X_2}(t) dt \quad \forall z \in \mathbb{R}. \quad (48)$$

**Beachte**

- Der Fall  $X_2(\omega) = 0$ , der bei der Bildung des Quotienten  $X_1/X_2$  zur Division durch Null führen würde, tritt nur mit Wahrscheinlichkeit Null auf (weil  $X_2$  absolutstetig ist).
- Deshalb kann  $X_1(\omega)/X_2(\omega)$  für solche  $\omega \in \Omega$  gesondert definiert werden (z.B. können wir dann  $X_1(\omega)/X_2(\omega) = 0$  setzen).

**Beispiel** Wenn  $X_1$  und  $X_2$  unabhängig sind mit  $X_1 \sim N(0,1)$  und  $X_2 \sim N(0,1)$ , dann gilt

$$f_{X_1/X_2}(z) = \frac{1}{\pi(z^2 + 1)} \quad \forall z \in \mathbb{R}. \quad (49)$$

denn aus (48) ergibt sich, dass

$$\begin{aligned} f_{X_1/X_2}(z) &= \int_{\mathbb{R}} |t| f_{X_1}(z \cdot t) f_{X_2}(t) dt \\ &= \int_{\mathbb{R}} |t| \frac{1}{2\pi} \exp\left(-\frac{1}{2}((z \cdot t)^2 + t^2)\right) dt \\ &\stackrel{\text{Symmetrie}}{=} \frac{1}{\pi} \int_0^\infty t \exp\left(-\underbrace{\frac{t^2}{2}(z^2 + 1)}_{=v}\right) dt \\ &= \frac{1}{\pi(z^2 + 1)} \underbrace{\int_0^\infty e^{-v} dv}_{=1} = \frac{1}{\pi(z^2 + 1)}. \end{aligned}$$

**Beachte** Eine absolutstetige Zufallsvariable mit der in (49) gegebenen Dichte heißt *Cauchy-verteilt*.



### 3.4.5 Unabhängigkeit zusammengesetzter Abbildungen

Das folgende Resultat ist eine sehr nützliche Eigenschaft von unabhängigen Zufallsvariablen.

**Theorem 3.17** *Seien  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  unabhängige Zufallsvariablen, und  $\{i_{11}, \dots, i_{1n_1}\}, \dots, \{i_{k1}, \dots, i_{kn_k}\}$  sei eine beliebige Zerlegung der Menge  $\{1, \dots, n\}$  in  $k$  nichtleere, paarweise disjunkte Teilmengen. Für beliebige Borel-messbare Funktionen  $\varphi_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R}, \dots, \varphi_k : \mathbb{R}^{n_k} \rightarrow \mathbb{R}$  sind dann auch*

$$\varphi_1(X_{i_{11}}, \dots, X_{i_{1n_1}}), \dots, \varphi_k(X_{i_{k1}}, \dots, X_{i_{kn_k}})$$

*unabhängige Zufallsvariablen.*

#### Beweis

- Wegen der Unabhängigkeit der Zufallsvariablen  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  ergibt sich aus Formel (28) in Theorem 3.9, dass

$$\begin{aligned} & P((X_{i_{11}}, \dots, X_{i_{1n_1}}) \in B_{i_{11}} \times \dots \times B_{i_{1n_1}}, \dots, (X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in B_{i_{k1}} \times \dots \times B_{i_{kn_k}}) \\ &= P((X_{i_{11}}, \dots, X_{i_{1n_1}}) \in B_{i_{11}} \times \dots \times B_{i_{1n_1}}) \dots P((X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in B_{i_{k1}} \times \dots \times B_{i_{kn_k}}) \end{aligned}$$

für beliebige  $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ .

- Sei  $\mathcal{G}$  das  $\pi$ -System  $\mathcal{G} = \{B_1 \times \dots \times B_{n_1}, B_1, \dots, B_{n_1} \in \mathcal{B}(\mathbb{R})\}$ , und sei  $\mathcal{D}$  das  $d$ -System derjenigen Borel-Mengen  $C_1 \in \mathcal{B}(\mathbb{R}^{n_1})$ , für die gilt

$$\begin{aligned} & P((X_{i_{11}}, \dots, X_{i_{1n_1}}) \in C_1, (X_{i_{21}}, \dots, X_{i_{2n_2}}) \in B_{i_{21}} \times \dots \times B_{i_{2n_2}}, \\ & \quad \dots, (X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in B_{i_{k1}} \times \dots \times B_{i_{kn_k}}) \\ &= P((X_{i_{11}}, \dots, X_{i_{1n_1}}) \in C_1) P((X_{i_{21}}, \dots, X_{i_{2n_2}}) \in B_{i_{21}} \times \dots \times B_{i_{2n_2}}) \\ & \quad \dots P((X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in B_{i_{k1}} \times \dots \times B_{i_{kn_k}}). \end{aligned}$$

- Dann ergibt sich mit Hilfe des Satzes über monotone Klassen (vgl. Theorem 3.1), dass

$$\mathcal{D} \supset d(\mathcal{G}) = \sigma(\mathcal{G}) = \mathcal{B}(\mathbb{R}^{n_1}).$$

- Hieraus und durch wiederholte Anwendung des Satzes über monotone Klassen ergibt sich also, dass

$$\begin{aligned} & P((X_{i_{11}}, \dots, X_{i_{1n_1}}) \in C_1, \dots, (X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in C_k) \\ &= P((X_{i_{11}}, \dots, X_{i_{1n_1}}) \in C_1) \dots P((X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in C_k) \end{aligned} \quad (50)$$

für beliebige  $C_1 \in \mathcal{B}(\mathbb{R}^{n_1}), \dots, C_k \in \mathcal{B}(\mathbb{R}^{n_k})$ .

- Somit gilt für beliebige Borel-Mengen  $B_1, \dots, B_k \in \mathcal{B}(\mathbb{R})$ , dass

$$\begin{aligned} & P(\varphi_1(X_{i_{11}}, \dots, X_{i_{1n_1}}) \in B_1, \dots, \varphi_k(X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in B_k) \\ &= P((X_{i_{11}}, \dots, X_{i_{1n_1}}) \in \varphi_1^{-1}(B_1), \dots, (X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in \varphi_k^{-1}(B_k)) \\ &= P((X_{i_{11}}, \dots, X_{i_{1n_1}}) \in \varphi_1^{-1}(B_1)) \dots P((X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in \varphi_k^{-1}(B_k)) \\ &= P(\varphi_1(X_{i_{11}}, \dots, X_{i_{1n_1}}) \in B_1) \dots P(\varphi_k(X_{i_{k1}}, \dots, X_{i_{kn_k}}) \in B_k), \end{aligned}$$

- wobei sich die vorletzte Gleichheit aus (50) ergibt mit  $C_i = \varphi_i^{-1}(B_i)$ .
- Durch die erneute Anwendung von Theorem 3.9 erhalten wir nun die Behauptung. □

## 4 Weitere Charakteristiken von Zufallsvariablen

Zu den wichtigsten Charakteristiken von Zufallsvariablen gehören deren Momente, insbesondere der Erwartungswert und die Varianz.

### 4.1 Erwartungswert

#### 4.1.1 Definition und Berechnungsformeln

Bevor wir zur allgemeinen Definition des Erwartungswertes kommen, wollen wir die intuitive Bedeutung dieses Begriffes anhand des folgenden Beispiels erläutern.

**Beispiel** (wiederholtes Würfeln)

- Betrachten den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  mit der Grundmenge

$$\Omega = \Omega_1 \times \Omega_2 \times \dots = \{\omega : \omega = (\omega_1, \omega_2, \dots), \omega_i \in \{1, 2, \dots, 6\}\},$$

der Produkt- $\sigma$ -Algebra  $\mathcal{F} = \mathcal{P}(\Omega_1) \otimes \mathcal{P}(\Omega_2) \otimes \dots$  und dem Wahrscheinlichkeitsmaß  $P$ , das durch

$$P(\{\omega : \omega \in \Omega, \omega_{i_1} = j_1, \dots, \omega_{i_k} = j_k\}) = \frac{1}{6^k}$$

gegeben ist;  $k \in \mathbb{N}$ ;  $1 \leq i_1 < \dots < i_k$ ;  $j_1, \dots, j_k \in \{1, 2, \dots, 6\}$ .

- Betrachten die Zufallsvariablen  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ , die gegeben seien durch die Projektion  $X_i(\omega) = \omega_i$  für  $i = 1, 2, \dots$ . D.h.,  $X_i$  ist die (zufällige) Augenzahl, die beim  $i$ -ten Wurf erzielt wird.
- Es ist nicht schwierig zu zeigen, dass  $X_1, X_2, \dots$  unabhängige (und identisch verteilte) Zufallsvariablen sind.
- Betrachten die Zufallsvariable

$$Y_n = \frac{1}{n}(X_1 + \dots + X_n),$$

d.h. die *mittlere Augenzahl* bei  $n$ -maligem Würfeln.

- Man kann zeigen, dass es eine „nichtzufällige“ Zahl  $c \in \mathbb{R}$  gibt, so dass

$$P(\{\omega : \lim_{n \rightarrow \infty} Y_n(\omega) = c\}) = 1, \quad (1)$$

wobei

$$c = \sum_{i=1}^6 i P(X_1 = i) = \frac{1}{6} \sum_{i=1}^6 i = 3.5. \quad (2)$$

- Die Formeln (1) und (2) bedeuten: Wenn die Anzahl  $n$  der durchgeführten Versuche immer größer wird, dann
  - werden die Werte  $Y_n(\omega)$  der mittleren Augenzahl  $Y_n$  immer weniger von der jeweiligen Ausprägung  $\omega$  des Zufalls beeinflusst,
  - strebt das „Zeitmittel“  $Y_n$  bei  $n$  Versuchen gegen das „Scharmittel“  $c$  jedes (einzelnen) Versuches.
- Die Formeln (1) und (2) sind ein Spezialfall des sogenannten *Gesetzes der großen Zahlen*, das im weiteren Verlauf der Vorlesung noch genauer diskutiert wird.
- Das Scharmittel  $c$  in (2) wird *Erwartungswert* der Zufallsvariablen  $X_1$  genannt und mit  $\mathbb{E} X_1$  bezeichnet.

Auf analoge Weise wird der Begriff des Erwartungswertes für beliebige Zufallsvariablen eingeführt.

**Definition** Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum, und sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable mit

$$\int_{\mathbb{R}} |x| P_X(dx) < \infty. \quad (3)$$

Dann heißt die Zufallsvariable  $X$  *integrierbar*, und der *Erwartungswert*  $\mathbb{E} X$  von  $X$  wird durch das folgende (Lebesgue-) Integral definiert:

$$\mathbb{E} X = \int_{\mathbb{R}} x P_X(dx). \quad (4)$$

**Beachte** Der in (4) definierte Erwartungswert  $\mathbb{E} X$  von integrierbaren Zufallsvariablen  $X$  kann auch als *Lebesgue-Stieltjes-Integral* bezüglich der Verteilungsfunktion  $F_X$  von  $X$  eingeführt werden. Und zwar gilt:

$$\mathbb{E} X = \int_{\mathbb{R}} x dF_X(x). \quad (5)$$

Aus der Definitionsgleichung (4) des Erwartungswertes ergibt sich ohne weiteres, wie diese Definitionsgleichung für diskrete bzw. absolutstetige Zufallsvariablen spezifiziert werden kann.

**Theorem 4.1** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable.

1. Wenn  $X$  diskret ist mit  $P(X \in C) = 1$  für eine abzählbare Menge  $C \subset \mathbb{R}$ , dann ist der Erwartungswert  $\mathbb{E} X$  von  $X$  durch das gewichtete Mittel

$$\mathbb{E} X = \sum_{x \in C} x P(X = x) \quad (6)$$

gegeben, wobei vorausgesetzt wird, dass

$$\sum_{x \in C} |x| P(X = x) < \infty. \quad (7)$$

2. Wenn  $X$  absolutstetig ist mit der Dichte  $f_X(x)$ , dann ist der Erwartungswert  $\mathbb{E} X$  von  $X$  durch das Integral

$$\mathbb{E} X = \int_{-\infty}^{\infty} x f_X(x) dx \quad (8)$$

gegeben, wobei vorausgesetzt wird, dass

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty. \quad (9)$$

**Beweis**

- Sei  $X$  diskret mit  $P(X \in C) = 1$  für eine abzählbare Menge  $C \subset \mathbb{R}$ .  
– Dann gilt

$$P_X(B) = \sum_{x \in B \cap C} P(X = x) \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

- Hieraus und aus (4) folgt (6).

- Sei nun  $X$  absolutstetig mit der Dichte  $f_X(x)$ .  
– Aus (3.14) ergibt sich dann, dass

$$P_X(B) = \int_B f_X(x) dx \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

– Hieraus und aus (4) folgt dann (8). □

### Beachte

- Die Summe in (7) bzw. das Integral in (9) kann man als Erwartungswert  $\mathbb{E}|X|$  der Zufallsvariablen  $|X|$  auffassen, vgl. auch den Transformationssatz für Erwartungswerte in Abschnitt 4.2.2.
- Man kann sich leicht Beispiele überlegen, bei denen die Summierbarkeitsbedingung (7) bzw. die Integrierbarkeitsbedingung (9) verletzt ist.
- Wenn die Zufallsgröße  $X$  nur nichtnegative Werte annimmt, d.h.  $P(X \geq 0) = 1$ , dann kann man den Begriff des Erwartungswertes auch einführen, wenn die Bedingung (7) bzw. (9) *nicht* erfüllt ist.
- In diesem Fall wird  $\mathbb{E}X = \infty$  gesetzt.
- Sei  $X$  beispielsweise eine absolutstetige Zufallsvariable mit der Dichte

$$f_X(x) = \frac{1}{\pi(x^2 + 1)} \quad \forall x \in \mathbb{R},$$

d.h.  $X$  ist Cauchy-verteilt, vgl. Abschnitt 3.4.4.

- Dann gilt zwar  $P(0 < |X| < \infty) = 1$ , jedoch

$$\begin{aligned} \mathbb{E}|X| &= \int_{\mathbb{R}} |x| \frac{1}{\pi(1+x^2)} dx \\ &> \int_0^\infty \frac{x}{\pi(1+x^2)} dx \\ &= \frac{1}{\pi} \lim_{t \rightarrow \infty} \left[ \frac{1}{2} \log(1+x^2) \right]_0^t \\ &= \frac{1}{2\pi} \lim_{t \rightarrow \infty} \log(1+t^2) = \infty. \end{aligned}$$

**Beispiele** Wir zeigen nun anhand zweier Beispiele, wie die Formeln (6) und (8) bei der praktischen Bestimmung des Erwartungswertes genutzt werden können.

#### 1. Binomialverteilung

Sei  $X$  binomialverteilt mit den Parametern  $n \in \mathbb{N}$  und  $p \in [0, 1]$ . Dann ergibt sich aus (6), dass

$$\begin{aligned} \mathbb{E}X &= \sum_{i=1}^n i \binom{n}{i} p^i (1-p)^{n-i} \\ &= np \sum_{i=1}^n \binom{n-1}{i-1} p^{i-1} (1-p)^{(n-1)-(i-1)} \\ &= np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \\ &= np (p + (1-p))^{n-1} = np. \end{aligned}$$

#### 2. Normalverteilung

Sei  $X$  normalverteilt mit den Parametern  $\mu \in \mathbb{R}$  und  $\sigma > 0$ . Dann ist  $X$  integrierbar, d.h.,

$$\int_{\mathbb{R}} |x| f_X(x) dx < \infty,$$

denn es gilt

$$\begin{aligned}
& \int_{\mathbb{R}} |x| \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\
&= \int_0^\infty x \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx - \int_{-\infty}^0 x \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\
&= \sigma \int_{-\mu/\sigma}^\infty (\sigma v + \mu) \exp\left(-\frac{1}{2} v^2\right) dv - \sigma \int_{-\infty}^{-\mu/\sigma} (\sigma v + \mu) \exp\left(-\frac{1}{2} v^2\right) dv \\
&\leq 2\sigma^2 \int_0^\infty v \exp\left(-\frac{1}{2} v^2\right) dv + \sigma\mu \int_{-\infty}^\infty \exp\left(-\frac{1}{2} v^2\right) dv \\
&= 2\sigma^2 + \sigma\mu\sqrt{2\pi},
\end{aligned}$$

wobei in der letzten Gleichheit genutzt wurde, dass

$$\begin{aligned}
\left(\int_{-\infty}^\infty \exp\left(-\frac{1}{2} v^2\right) dv\right)^2 &= \int_{-\infty}^\infty \int_{-\infty}^\infty \exp\left(-\frac{1}{2} (x^2 + y^2)\right) dx dy \\
&= \int_0^{2\pi} \int_0^\infty r \exp\left(-\frac{1}{2} r^2\right) dr d\varphi = 2\pi.
\end{aligned}$$

Aus (8) ergibt sich nun, dass

$$\begin{aligned}
\mathbb{E} X &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} x \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\sigma v + \mu) \exp\left(-\frac{1}{2} v^2\right) dv \\
&= \underbrace{\frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} v \exp\left(-\frac{1}{2} v^2\right) dv}_{=0} + \underbrace{\frac{\mu}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2} v^2\right) dv}_{=\sqrt{2\pi}} = \mu.
\end{aligned}$$

#### 4.1.2 Alternative Integral-Darstellungen

Wir leiten zunächst eine Darstellungsformel des Erwartungswertes von nichtnegativen Zufallsvariablen her.

**Theorem 4.2** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable mit  $P(X \geq 0) = 1$ . Dann gilt

$$\mathbb{E} X = \int_0^\infty (1 - F_X(y)) dy. \quad (10)$$

**Beweis** Aus der Definitionsgleichung (4) von  $\mathbb{E} X$  und aus dem Satz von Fubini ergibt sich, dass

$$\begin{aligned}
\mathbb{E} X &= \int_{\mathbb{R}} x P_X(dx) = \int_0^\infty x P_X(dx) = \int_0^\infty \left( \int_0^\infty \mathbb{1}_{(0,x)}(y) dy \right) P_X(dx) \\
&= \int_0^\infty \left( \int_0^\infty \mathbb{1}_{(0,x)}(y) P_X(dx) \right) dy = \int_0^\infty P_X(\{x : x \in \mathbb{R}, x > y\}) dy \\
&= \int_0^\infty P(X > y) dy = \int_0^\infty (1 - F_X(y)) dy.
\end{aligned}$$

□

**Korollar 4.1** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine integrierbare Zufallsvariable. Dann gilt

$$\mathbb{E} X = \int_0^\infty (1 - F_X(y)) dy - \int_{-\infty}^0 F_X(y) dy. \quad (11)$$

**Beweis** Ähnlich wie im Beweis von Theorem 4.2 ergibt sich, dass

$$\begin{aligned}\mathbb{E} X &= \int_{\mathbb{R}} x P_X(dx) = \int_0^\infty x P_X(dx) + \int_{-\infty}^0 x P_X(dx) \\ &= \int_0^\infty x P_X(dx) - \int_{-\infty}^0 (-x) P_X(dx) = \int_0^\infty (1 - F_X(y)) dy - \int_{-\infty}^0 F_X(y) dy,\end{aligned}$$

wobei in der letzten Gleichheit die Formel (10) und die Tatsache genutzt wird, dass

$$\begin{aligned}\int_{-\infty}^0 (-x) P_X(dx) &= \int_{-\infty}^0 \left( \int_0^\infty \mathbb{I}_{(0, -x]}(y) dy \right) P_X(dx) = \int_0^\infty \left( \int_{-\infty}^0 \mathbb{I}_{(0, -x]}(y) P_X(dx) \right) dy \\ &= \int_0^\infty P_X(\{x : x \in \mathbb{R}, -x \geq y\}) dy = \int_0^\infty P_X(\{x : x \in \mathbb{R}, x \leq -y\}) dy \\ &= \int_0^\infty P(X \leq -y) dy = \int_0^\infty F_X(-y) dy = \int_{-\infty}^0 F_X(y) dy.\end{aligned}$$

□

Wir zeigen nun noch, dass der in (4) definierte Erwartungswert  $\mathbb{E} X$  auch in der folgenden Form dargestellt werden kann.

**Theorem 4.3** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable, deren Verteilung der Bedingung (3) genügt. Dann gilt

$$\mathbb{E} X = \int_{\Omega} X(\omega) P(d\omega). \quad (12)$$

**Beweis**

- Wir zeigen die Gültigkeit von (12) zunächst für den Fall, dass  $X = \mathbb{I}_A$  für ein  $A \in \mathcal{F}$ .  
– Dann gilt

$$\mathbb{E} X = \int_{\Omega} \mathbb{I}_A(\omega) P(d\omega), \quad (13)$$

- denn

$$\begin{aligned}\mathbb{E} X &= \mathbb{E} \mathbb{I}_A = 0 \cdot P_X(\{0\}) + 1 \cdot P_X(\{1\}) \\ &= 0 \cdot P(\mathbb{I}_A = 0) + 1 \cdot P(\mathbb{I}_A = 1) \\ &= 0 \cdot P(A^c) + 1 \cdot P(A) = \int_{\Omega} \mathbb{I}_A(\omega) P(d\omega).\end{aligned}$$

- Auf analoge Weise lässt sich die Gültigkeit von (12) für Linearkombinationen von Indikatorvariablen zeigen.  
– Und zwar seien  $A_1, \dots, A_n \in \mathcal{F}$  beliebige Ereignisse und  $x_1, \dots, x_n \in \mathbb{R}$  beliebige reelle Zahlen.  
– Ohne Einschränkung der Allgemeinheit können (und werden) wir annehmen, dass  $A_1, \dots, A_n$  paarweise disjunkte Mengen sind.  
– Für die Zufallsvariable  $X = \sum_{i=1}^n x_i \mathbb{I}_{A_i}$  gilt dann wegen (6) und (13), dass

$$\begin{aligned}\mathbb{E} X &= \mathbb{E} \sum_{i=1}^n x_i \mathbb{I}_{A_i} = \sum_{i=1}^n x_i P(A_i) = \sum_{i=1}^n x_i \mathbb{E} \mathbb{I}_{A_i} \\ &= \sum_{i=1}^n x_i \int_{\Omega} \mathbb{I}_{A_i}(\omega) P(d\omega) = \int_{\Omega} \sum_{i=1}^n x_i \mathbb{I}_{A_i}(\omega) P(d\omega) = \int_{\Omega} X(\omega) P(d\omega).\end{aligned}$$

- Mit Hilfe des Satzes von Beppo Levi (über die monotone Konvergenz von Lebesgue-Integralen) wird schließlich gezeigt, dass (12) für jede Zufallsvariable  $X$  gilt, deren Verteilung der Bedingung (3) genügt.
  - Dabei nutzen wir die Tatsache, dass jede Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  in den positiven Teil  $X^+ = \max\{0, X\}$  bzw. den negativen Teil  $X^- = -\min\{0, X\}$  zerlegt werden kann.
  - Dann gilt  $X = X^+ - X^-$ , und es gibt zwei monoton wachsende Folgen  $\{X_n^+\}$  bzw.  $\{X_n^-\}$  von Linearkombinationen von Indikatorvariablen, so dass  $X_n^+ \uparrow X^+$  und  $X_n^- \uparrow X^-$ .
  - Aus Korollar 4.1 und aus dem Satz über die monotone Konvergenz ergibt sich nun, dass

$$\begin{aligned}
 \mathbb{E} X &= \int_0^\infty (1 - F_X(y)) dy - \int_{-\infty}^0 F_X(y) dy \\
 &= \int_0^\infty P(X^+ > y) dy - \int_0^\infty P(X^- \geq y) dy \\
 &= \int_0^\infty P(\lim_n X_n^+ > y) dy - \int_0^\infty P(\lim_n X_n^- \geq y) dy \\
 &= \int_0^\infty \lim_n P(X_n^+ > y) dy - \int_0^\infty \lim_n P(X_n^- \geq y) dy \\
 &= \lim_n \int_0^\infty P(X_n^+ > y) dy - \lim_n \underbrace{\int_0^\infty P(X_n^- \geq y) dy}_{= \int_0^\infty P(X_n^- > y) dy} .
 \end{aligned}$$

- Durch erneute Anwendung von Korollar 4.1 und des Satzes über die monotone Konvergenz ergibt sich somit, dass

$$\begin{aligned}
 \mathbb{E} X &= \lim_n \mathbb{E} X_n^+ - \lim_n \mathbb{E} X_n^- \\
 &= \lim_n \int_\Omega X_n^+(\omega) P(d\omega) - \lim_n \int_\Omega X_n^-(\omega) P(d\omega) \\
 &= \int_\Omega X^+(\omega) P(d\omega) - \int_\Omega X^-(\omega) P(d\omega) \\
 &= \int_\Omega (X^+(\omega) - X^-(\omega)) P(d\omega) = \int_\Omega X(\omega) P(d\omega) .
 \end{aligned}$$

□

### Beachte

- Die im Beweis von Theorem 4.3 verwendete Methode wird *algebraische Induktion* genannt.
- Sie beruht auf dem Prinzip, die betreffende Aussage zunächst
  - für Indikatoren von Ereignissen,
  - danach für Linearkombinationen von Indikatoren
  - und schließlich (durch monotone Approximation und Zerlegung in Positiv- bzw. Negativteil) für beliebige Zufallsvariablen zu beweisen.
- Wir werden im folgenden noch weitere Aussagen mit dieser Beweismethode herleiten.

#### 4.1.3 Weitere Eigenschaften des Erwartungswertes

Mit Hilfe der Darstellungsformeln des Erwartungswertes, die in Abschnitt 4.1.2 diskutiert wurden, lassen sich weitere Eigenschaften des Erwartungswertes von Zufallsvariablen herleiten.

**Theorem 4.4** Seien  $X, Y, X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen über einem beliebigen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ . Dann gilt:

1. Monotonie: Wenn  $X$  und  $Y$  integrierbar sind mit  $X \leq Y$  f.s., dann gilt

$$\mathbb{E} X \leq \mathbb{E} Y. \quad (14)$$

Wenn  $Y$  integrierbar ist und wenn  $0 \leq X \leq Y$  f.s., dann ist auch  $X$  integrierbar, und es gilt (14).

2. Wenn  $X$  integrierbar ist, dann ist

$$|\mathbb{E} X| \leq \mathbb{E} |X|. \quad (15)$$

3. Linearität: Wenn  $X$  und  $Y$  integrierbar sind, dann ist auch  $aX + bY$  integrierbar für beliebige  $a, b \in \mathbb{R}$ , und es gilt

$$\mathbb{E}(aX + bY) = a\mathbb{E} X + b\mathbb{E} Y. \quad (16)$$

4. monotone Konvergenz: Wenn  $X_n \geq 0$  f.s. für alle  $n = 1, 2, \dots$  und wenn  $X_n \uparrow X$  f.s., dann gilt

$$\mathbb{E} X_n \uparrow \mathbb{E} X. \quad (17)$$

5. majorisierte Konvergenz und  $L_1$ -Konvergenz: Wenn  $Y$  integrierbar ist, wenn  $|X_n| \leq Y$  f.s. für alle  $n = 1, 2, \dots$  und wenn  $X_n \rightarrow X$  f.s., dann ist auch  $X$  integrierbar, und es gilt

$$\lim_{n \rightarrow \infty} \mathbb{E} X_n = \mathbb{E} X \quad (18)$$

und

$$\lim_{n \rightarrow \infty} \mathbb{E} |X_n - X| = 0. \quad (19)$$

6. Wenn  $X = \mathbb{1}_A$  für ein  $A \in \mathcal{F}$ , dann gilt

$$\mathbb{E} X = P(A). \quad (20)$$

7. Wenn  $X$  integrierbar ist mit  $X \geq 0$  f.s. und  $\mathbb{E} X = 0$ , dann gilt

$$X = 0 \quad \text{f.s.} \quad (21)$$

## Beweis

Zu 1) Die Ungleichung (14) ergibt sich unmittelbar aus der Integral-Darstellung (12) des Erwartungswertes und aus der entsprechenden Monotonie-Eigenschaft des Lebesgue-Integrals in (12). Die zweite Teilaussage von 1. ergibt sich auf die gleiche Weise.

Zu 2) Weil mit  $X$  offenbar auch  $|X|$  bzw.  $-|X|$  integrierbar sind und weil  $X \leq |X|$  bzw.  $-|X| \leq X$ , ergibt sich aus (14), dass

$$\mathbb{E} X \leq \mathbb{E} |X| \quad \text{bzw.} \quad -\mathbb{E} |X| = \mathbb{E} (-|X|) \leq \mathbb{E} X,$$

wobei sich die Gleichheit aus der Linearität des Lebesgue-Integrals ergibt.

Zu 3) Die Integrierbarkeit von  $aX + bY$  ergibt sich unmittelbar aus der Ungleichung

$$|aX + bY| \leq |a||X| + |b||Y|,$$

aus (14) und aus der Linearität des Lebesgue-Integrals, denn es gilt

$$\mathbb{E} |aX + bY| \leq \mathbb{E} (|a||X| + |b||Y|) = |a|\mathbb{E} |X| + |b|\mathbb{E} |Y| < \infty.$$

Die Gültigkeit von (16) folgt dann ebenfalls aus der Linearität des Lebesgue-Integrals.

Zu 4/5) Diese Teilaussagen ergeben sich unmittelbar aus dem Sätzen über die monotone bzw. majorisierte Konvergenz von Lebesgue-Integralen.



Zu 6) Wenn  $X = \mathbb{1}_A$  für ein  $A \in \mathcal{F}$ , dann ergibt sich aus der Darstellungsformel (6) für den Erwartungswert diskreter Verteilungen, dass

$$\mathbb{E} X = \mathbb{E} \mathbb{1}_A = 0 \cdot P(\mathbb{1}_A = 0) + 1 \cdot P(\mathbb{1}_A = 1) = P(A),$$

vgl. auch den ersten Teil des Beweises von Theorem 4.3.

Zu 7) Sei  $X$  integrierbar, und es gelte  $X \geq 0$  f.s. und  $\mathbb{E} X = 0$ . Wir führen einen indirekten Beweis und nehmen an, dass  $P(X > 0) > 0$ . Wegen der Stetigkeitseigenschaften von Wahrscheinlichkeitsmaßen (vgl. Theorem 2.3) gilt dann auch  $P(X > \varepsilon) > 0$  für ein  $\varepsilon > 0$ . Hieraus und aus (14) folgt, dass

$$\mathbb{E} X \geq \mathbb{E}(X \mathbb{1}_{\{X > \varepsilon\}}) \geq \mathbb{E}(\varepsilon \mathbb{1}_{\{X > \varepsilon\}}) = \varepsilon \mathbb{E} \mathbb{1}_{\{X > \varepsilon\}} = \varepsilon P(X > \varepsilon) > 0.$$

Dies ist aber im Widerspruch zu  $\mathbb{E} X = 0$ . □

**Korollar 4.2** Sei  $n \in \mathbb{N}$  eine beliebige, jedoch fest vorgegebene natürliche Zahl, und seien  $X_1, \dots, X_n$  beliebige Zufallsvariablen mit  $\mathbb{E} |X_1| < \infty, \dots, \mathbb{E} |X_n| < \infty$ . Dann gilt für beliebige Konstanten  $a_1, \dots, a_n \in \mathbb{R}$

$$\mathbb{E}(a_1 X_1 + \dots + a_n X_n) = a_1 \mathbb{E} X_1 + \dots + a_n \mathbb{E} X_n. \quad (22)$$

**Beweis** Die Behauptung ergibt sich aus (16) mittels vollständiger Induktion.

**Beispiel** (wiederholtes Würfeln)

- Mit Hilfe von Korollar 4.2 können wir das in Formel (2) betrachtete Scharmittel  $c$  als Erwartungswert der mittleren Augenzahl bei  $n$ -maligem Würfeln darstellen.
  - Zur Erinnerung: Wir betrachten den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{P}(\Omega), P)$ , der in Abschnitt 4.1.1 eingeführt worden ist,
  - und die Zufallsvariablen  $X_1, X_2, \dots : \Omega \rightarrow \{1, 2, \dots, 6\}$ , wobei  $X_i$  die zufällige Augenzahl ist, die beim  $i$ -ten Würfeln erzielt wird.
- Dann gilt  $\mathbb{E} X_i = \sum_{j=1}^6 j P(X_i = j) = \frac{1}{6} \sum_{j=1}^6 j = 3.5$ .
- Aus Korollar 4.2 ergibt sich dann für den Erwartungswert  $\mathbb{E} Y_n$ , wobei  $Y_n = n^{-1}(X_1 + \dots + X_n)$  die mittlere Augenzahl bei  $n$ -maligem Würfeln ist, dass

$$\mathbb{E} Y_n = \mathbb{E} \left( \frac{1}{n} (X_1 + \dots + X_n) \right) \stackrel{\text{Korollar 4.2}}{=} \frac{1}{n} (\mathbb{E} X_1 + \dots + \mathbb{E} X_n) = \frac{1}{n} n 3.5 = 3.5.$$

#### 4.1.4 Integral-Darstellung mittels Quantilfunktion

Die folgenden beiden Eigenschaften von verallgemeinerten inversen Funktionen führen zu einer weiteren nützlichen Integral-Darstellung des Erwartungswertes.

**Lemma 4.1** Sei  $F : \mathbb{R} \rightarrow \mathbb{R}$  eine nichtfallende und rechtsseitig stetige Funktion, und sei  $F^{-1} : \mathbb{R} \rightarrow \mathbb{R}$  die verallgemeinerte inverse Funktion mit

$$F^{-1}(y) = \inf\{x : F(x) \geq y\}, \quad (23)$$

wobei  $\inf \emptyset = \infty$  gesetzt wird. Dann gilt:

1.  $F^{-1}$  ist nichtfallend.
2. Es gilt  $y \leq F(x)$  genau dann, wenn  $F^{-1}(y) \leq x$ .

**Beweis**

- Die erste Teilaussage ergibt sich unmittelbar aus der Definitionsgleichung (23).
- Außerdem ergibt sich sofort aus (23), dass  $F^{-1}(y) \leq x$ , falls  $y \leq F(x)$ .
- Sei nun  $F^{-1}(y) \leq x$ . Wegen der Monotonie von  $F$  gibt es dann eine Folge  $\{x_n\}$  derart, dass  $x_n \downarrow x$  und  $F(x_n) \geq y$  für jedes  $n$ .
- Hieraus und aus der rechtsseitigen Stetigkeit von  $F$  folgt, dass  $F(x) \geq y$ .  $\square$

**Definition** Die verallgemeinerte inverse Funktion  $F_X^{-1}$  der Verteilungsfunktion  $F_X$  einer beliebigen Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  heißt *Quantilfunktion* von  $X$ .

**Lemma 4.2** Seien  $v, w : \mathbb{R} \rightarrow \mathbb{R}$  zwei nichtfallende und rechtsseitig stetige Funktionen. Für jede Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  gilt dann für fast jedes  $y \in [0, 1]$

$$F_{v(X)+w(X)}^{-1}(y) = F_{v(X)}^{-1}(y) + F_{w(X)}^{-1}(y). \quad (24)$$

**Beweis**

- Für beliebige Funktionen  $g(t)$ ,  $g'(t)$  bezeichne  $g \circ g'(t)$  die zusammengesetzte Abbildung  $g(g'(t))$ .
- Dann gilt  $F_{v(X)}^{-1}(y) = v \circ F_X^{-1}(y)$  für fast jedes  $y \in [0, 1]$ , denn aus der zweiten Teilaussage von Lemma 4.1 ergibt sich, dass für fast jedes  $y \in [0, 1]$  und  $x \in \mathbb{R}$

$$\begin{aligned} F_{v(X)}^{-1}(y) \leq x &\iff F_{v(X)}(x) \geq y \iff P(v(X) \leq x) \geq y \\ &\iff P(X \leq v^{-1}(x)) \geq y \iff F_X(v^{-1}(x)) \geq y \\ &\iff F_X^{-1}(y) \leq v^{-1}(x) \iff v(F_X^{-1}(y)) \leq x. \end{aligned}$$

- Auf die gleiche Weise ergibt sich, dass  $F_{w(X)}^{-1}(y) = w \circ F_X^{-1}(y)$  und  $F_{(v+w)(X)}^{-1}(y) = (v+w) \circ F_X^{-1}(y)$  für fast jedes  $y \in [0, 1]$ .
- Hieraus folgt, dass für fast jedes  $y \in [0, 1]$

$$F_{v(X)+w(X)}^{-1}(y) = (v+w) \circ F_X^{-1}(y) = v \circ F_X^{-1}(y) + w \circ F_X^{-1}(y) = (F_{v(X)}^{-1} + F_{w(X)}^{-1})(y).$$

- Damit ist (24) bewiesen.  $\square$

**Theorem 4.5** Sei  $X$  integrierbar, d.h.,  $\mathbb{E}|X| < \infty$ . Dann gilt

$$\mathbb{E} X = \int_0^1 F_X^{-1}(y) dy. \quad (25)$$

**Beweis**

- Sei zunächst  $X \geq 0$ . Dann ist auch  $F_X^{-1}(y) \geq 0$  für alle  $y \in (0, 1)$ ,
- und es gilt

$$\begin{aligned} \int_0^1 F_X^{-1}(x) dx &= \int_0^1 \int_0^\infty \mathbb{1}_{(0, F_X^{-1}(x))}(y) dy dx = \int_0^\infty \int_0^1 \mathbb{1}_{(0, F_X^{-1}(x))}(y) dx dy \\ &\stackrel{\text{Lemma 4.1}}{=} \int_0^\infty \int_0^1 \mathbb{1}_{(F_X(y), 1)}(x) dx dy = \int_0^\infty (1 - F_X(y)) dy = \mathbb{E} X, \end{aligned}$$

wobei sich die letzte Gleichheit aus (10) ergibt.

- Wenn  $X \leq 0$ , dann ergibt sich auf analoge Weise, dass

$$\begin{aligned} \int_0^1 F_X^{-1}(x) dx &= - \int_0^1 \int_{-\infty}^0 \mathbb{1}_{(F_X^{-1}(x), 0)}(y) dy dx = - \int_{-\infty}^0 \int_0^1 \mathbb{1}_{(F_X^{-1}(x), 0)}(y) dx dy \\ &\stackrel{\text{Lemma 4.1}}{=} - \int_{-\infty}^0 \int_0^1 \mathbb{1}_{(0, F_X(y))}(x) dx dy = - \int_{-\infty}^0 F_X(y) dy = \mathbb{E} X. \end{aligned}$$

- Für beliebige integrierbare Zufallsvariablen  $X$  erhalten wir (25) nun mittels der Zerlegung  $X = X^+ - X^-$ , denn aus Lemma 4.2 ergibt sich, dass

$$\begin{aligned} \int_0^1 F_X^{-1}(x) dx &= \int_0^1 F_{X^+ - X^-}^{-1}(x) dx = \int_0^1 (F_{X^+}^{-1}(x) + F_{-X^-}^{-1}(x)) dx \\ &= \int_0^1 F_{X^+}^{-1}(x) dx + \int_0^1 F_{-X^-}^{-1}(x) dx \\ &= \mathbb{E} X^+ + \mathbb{E} (-X^-) = \mathbb{E} (X^+ - X^-) = \mathbb{E} X, \end{aligned}$$

wobei sich die vorletzte Gleichheit aus der Linearitätseigenschaft (16) des Erwartungswertes ergibt.  $\square$

## 4.2 Varianz und höhere Momente

### 4.2.1 Definition und elementare Eigenschaften

#### Beachte

- Es ist klar, dass der in (4) definierte Erwartungswert  $\mathbb{E} X$  eindeutig durch die Verteilung  $P_X$  der Zufallsvariablen  $X$  bestimmt wird.
- In Theorem 4.1 hatten wir darüber hinaus gezeigt, dass der Erwartungswert  $\mathbb{E} X$  einer diskreten bzw. absolutstetigen Zufallsvariablen  $X$  eindeutig durch die Wahrscheinlichkeitsfunktion bzw. die Dichte von  $X$  bestimmt wird.
- Umgekehrt ist jedoch im allgemeinen die Verteilung  $P_X$  *nicht* eindeutig durch den Erwartungswert  $\mathbb{E} X$  von  $X$  festgelegt.
- Insbesondere ist die Wahrscheinlichkeitsfunktion bzw. die Dichte einer diskreten bzw. absolutstetigen Zufallsvariablen  $X$  nicht eindeutig durch den Erwartungswert  $\mathbb{E} X$  von  $X$  festgelegt.

#### Beispiel (symmetrische diskrete Gleichverteilung)

- Sei  $n \in \mathbb{N}$  eine beliebige natürliche Zahl und  $X : \Omega \rightarrow \{-n, \dots, -1, 0, 1, \dots, n\}$  eine diskrete Zufallsvariable mit  $P(X = i) = (2n + 1)^{-1}$  für jedes  $i \in \{-n, \dots, n\}$ .
- Dann gilt

$$\mathbb{E} X = \frac{1}{2n + 1} \sum_{i=-n}^n i = \frac{1}{2n + 1} \cdot 0 = 0.$$

- Während der Erwartungswert  $\mathbb{E} X$  also nicht von  $n$  abhängt, sind die Werte von  $X$  mit wachsendem  $n$  immer breiter um  $\mathbb{E} X$  gestreut.
- Eine Charakteristik, die den Streuungsgrad der Werte von  $X$  um den Erwartungswert  $\mathbb{E} X$  misst, ist die *erwartete quadratische Abweichung*  $\mathbb{E}((X - \mathbb{E} X)^2)$  vom Erwartungswert  $\mathbb{E} X$ , genannt die *Varianz* von  $X$ , die bei diesem Beispiel gegeben ist durch

$$\mathbb{E}((X - \mathbb{E} X)^2) = \sum_{i=-n}^n i^2 \frac{1}{2n + 1} = 2 \frac{n(n + 1)(2n + 1)}{6} \frac{1}{2n + 1} = \frac{n(n + 1)}{3}.$$

**Beachte**

- Der Erwartungswert  $\mathbb{E} X$  der Zufallsvariablen  $X$  wird manchmal auch das *erste Moment* von  $X$  genannt.
- Völlig analog lassen sich die Begriffe der Varianz bzw. der höheren Momente einer beliebigen Zufallsvariablen  $X$  einführen,
- und zwar durch die Betrachtung des Erwartungswertes  $\mathbb{E} \varphi(X)$  entsprechend gewählter Funktionen  $\varphi(X)$  von  $X$ .

**Definition**

- Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable mit  $\mathbb{E}(X^2) < \infty$ .
- Betrachten die Abbildung  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  mit  $\varphi(x) = (x - \mathbb{E} X)^2$ .
- Dann heißt der Erwartungswert  $\mathbb{E} \varphi(X)$  der Zufallsvariablen  $\varphi(X)$  die *Varianz* von  $X$  (Schreibweise:  $\text{Var } X$ ).
- Für die Varianz von  $X$  gilt also

$$\text{Var } X = \mathbb{E} \left( (X - \mathbb{E} X)^2 \right). \quad (26)$$

**Beachte**

- Die höheren Momente von Zufallsvariablen werden völlig analog definiert.
- Und zwar sei  $k \in \mathbb{N}$  und  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable mit  $\mathbb{E}(|X|^k) < \infty$ .
- Betrachten die Abbildung  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  mit  $\varphi(x) = x^k$ . Dann heißt der Erwartungswert  $\mathbb{E} \varphi(X) = \mathbb{E}(X^k)$  von  $\varphi(X)$  das *k-te Moment* von  $X$ .
- Betrachten die Abbildung  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  mit  $\varphi(x) = (x - \mathbb{E} X)^k$ . Dann heißt der Erwartungswert  $\mathbb{E} \varphi(X) = \mathbb{E}((X - \mathbb{E} X)^k)$  von  $\varphi(X)$  das *k-te zentrale Moment* von  $X$ .
- Die Varianz  $\text{Var } X$  ist also das zweite zentrale Moment von  $X$ .
- Die Größe  $\sqrt{\text{Var } X}$  heißt die *Standardabweichung* von  $X$ .
- Der Erwartungswert ist eine Lagekenngröße und die Varianz ist eine Variabilitätskenngröße.

Die folgenden elementaren Eigenschaften ergeben sich unmittelbar aus der Definitionsgleichung (26) der Varianz und aus der Linearität des Erwartungswertes, vgl. Formel (16) in Theorem 4.4.

**Theorem 4.6** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable mit  $\mathbb{E}(X^2) < \infty$ . Dann gilt

$$\text{Var } X = \mathbb{E}(X^2) - (\mathbb{E} X)^2, \quad (27)$$

und für beliebige  $a, b \in \mathbb{R}$

$$\text{Var}(aX + b) = a^2 \text{Var } X. \quad (28)$$

**Beweis**

- Aus der Definitionsgleichung (26) der Varianz und aus der Linearität des Erwartungswertes ergibt sich, dass

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

- Damit ist (27) bewiesen.

- Auf analoge Weise ergibt sich aus (27), dass

$$\begin{aligned}
 \text{Var}(aX + b) &= \mathbb{E}((aX + b)^2) - (\mathbb{E}(aX + b))^2 \\
 &= \mathbb{E}((aX)^2 + 2abX + b^2) - (a\mathbb{E}X + b)^2 \\
 &= a^2\mathbb{E}(X^2) + 2ab\mathbb{E}X + b^2 - a^2(\mathbb{E}X)^2 - 2ab\mathbb{E}X - b^2 \\
 &= a^2(\mathbb{E}(X^2) - (\mathbb{E}X)^2) \\
 &= a^2\text{Var}X.
 \end{aligned}$$

- Damit ist auch (28) bewiesen. □

### Beachte

- Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable mit  $\mathbb{E}(X^2) < \infty$ . Dann gilt  $\text{Var} X = 0$  genau dann, wenn

$$P(X = \mathbb{E}X) = 1. \quad (29)$$

- Dies kann man sich folgendermaßen überlegen.
  - Sei  $\text{Var} X = 0$ , d.h., der Erwartungswert der nichtnegativen Zufallsvariablen  $(X - \mathbb{E}X)^2$  ist Null.
  - Aus Teilaussage 7 von Theorem 4.4 ergibt sich dann, dass  $(X - \mathbb{E}X)^2 = 0$  mit Wahrscheinlichkeit 1, d.h., es gilt  $X = \mathbb{E}X$  mit Wahrscheinlichkeit 1.
  - Es sei nun die Bedingung (29) erfüllt. Dann gilt

$$1 = P(X = \mathbb{E}X) = P(|X - \mathbb{E}X| = 0) = P((X - \mathbb{E}X)^2 = 0).$$

- Hieraus folgt, dass  $\mathbb{E}((X - \mathbb{E}X)^2) = 0$ , d.h.  $\text{Var} X = 0$ .

### 4.2.2 Transformationssatz und Berechnungsformeln

Bei der praktischen Berechnung der Varianz ist der folgende *Transformationssatz für Erwartungswerte* nützlich.

**Theorem 4.7** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable, und sei  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  eine Borel-messbare Abbildung, so dass

$$\int_{\mathbb{R}} |\varphi(x)| P_X(dx) < \infty. \quad (30)$$

Für den Erwartungswert  $\mathbb{E}\varphi(X)$  von  $\varphi(X) : \Omega \rightarrow \mathbb{R}$  gilt dann

$$\mathbb{E}\varphi(X) = \int_{\mathbb{R}} \varphi(x) P_X(dx). \quad (31)$$

### Beweis

- Wir führen den Beweis mittels algebraischer Induktion, vgl. auch den Beweis von Theorem 4.3, wo die gleiche Beweismethode verwendet wurde.
- Zunächst nehmen wir an, dass  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  eine Linearkombination von Indikatoren ist, d.h., es gelte

$$\varphi(x) = \sum_{i=1}^n a_i \mathbb{1}_{B_i}(x) \quad \forall x \in \mathbb{R},$$

wobei  $n \in \mathbb{N}$ ,  $a_1, \dots, a_n \in \mathbb{R}$  und  $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ .

- Dann ist  $\varphi(X)$  eine diskrete Zufallsvariable mit

$$\varphi(X) = \sum_{i=1}^n a_i \mathbb{1}_{\{X \in B_i\}}.$$

– Aus der Linearität des Erwartungswertes und aus der Berechnungsformel (6) ergibt sich nun, dass

$$\begin{aligned}\mathbb{E} \varphi(X) &= \sum_{i=1}^n a_i \mathbb{E} \mathbb{1}_{\{X \in B_i\}} = \sum_{i=1}^n a_i P(X \in B_i) \\ &= \sum_{i=1}^n a_i \int_{\mathbb{R}} \mathbb{1}_{B_i}(x) P_X(dx) = \int_{\mathbb{R}} \left( \sum_{i=1}^n a_i \mathbb{1}_{B_i}(x) \right) P_X(dx) \\ &= \int_{\mathbb{R}} \varphi(x) P_X(dx).\end{aligned}$$

- Sei jetzt  $\varphi$  eine beliebige Borel-messbare Abbildung mit  $\varphi(x) \geq 0$  für jedes  $x \in \mathbb{R}$ .
  - Dann gibt es eine monotone Folge  $\varphi_1, \varphi_2, \dots$  von Linearkombinationen von Indikatoren, so dass  $\varphi_n(x) \uparrow \varphi(x)$  für jedes  $x \in \mathbb{R}$
  - und damit auch  $\varphi_n(X)(\omega) \uparrow \varphi(X)(\omega)$  für jedes  $\omega \in \Omega$ .
  - Durch (zweimalige) Anwendung des Satzes über die monotone Konvergenz ergibt sich somit, dass

$$\begin{aligned}\mathbb{E} \varphi(X) &= \mathbb{E} \lim_n \varphi_n(X) = \lim_n \mathbb{E} \varphi_n(X) \\ &= \lim_n \int_{\mathbb{R}} \varphi_n(x) P_X(dx) \\ &= \int_{\mathbb{R}} \lim_n \varphi_n(x) P_X(dx) \\ &= \int_{\mathbb{R}} \varphi(x) P_X(dx).\end{aligned}$$

- Sei schließlich  $\varphi(X)$  eine beliebige integrierbare Zufallsvariable.
  - Dann betrachten wir die Zerlegung  $\varphi(X) = \varphi^+(X) - \varphi^-(X)$  von  $\varphi(X)$  in den positiven Teil  $\varphi^+(X)$  bzw. in den negativen Teil  $\varphi^-(X)$ .
  - Aus der Linearität des Erwartungswertes ergibt sich dann

$$\begin{aligned}\mathbb{E} \varphi(X) &= \mathbb{E} (\varphi^+(X) - \varphi^-(X)) = \mathbb{E} \varphi^+(X) - \mathbb{E} \varphi^-(X) \\ &= \int_{\mathbb{R}} \varphi^+(x) P_X(dx) - \int_{\mathbb{R}} \varphi^-(x) P_X(dx) \\ &= \int_{\mathbb{R}} (\varphi^+(x) - \varphi^-(x)) P_X(dx) \\ &= \int_{\mathbb{R}} \varphi(x) P_X(dx).\end{aligned}$$

□

**Korollar 4.3** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable, und sei  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  eine beliebige Borel-messbare Abbildung.

1. Wenn  $X$  diskret ist mit  $P(X \in C) = 1$  für eine abzählbare Menge  $C \subset \mathbb{R}$ , dann gilt

$$\mathbb{E} \varphi(X) = \sum_{x \in C} \varphi(x) P(X = x), \quad (32)$$

wobei vorausgesetzt wird, dass  $\sum_{x \in C} |\varphi(x)| P(X = x) < \infty$ .

2. Wenn  $X$  absolutstetig ist mit der Dichte  $f_X$ , dann gilt

$$\mathbb{E} \varphi(X) = \int_{-\infty}^{\infty} \varphi(x) f_X(x) dx, \quad (33)$$

wobei vorausgesetzt wird, dass  $\int_{-\infty}^{\infty} |\varphi(x)| f_X(x) dx < \infty$ .

**Beweis**

- Sei  $X$  diskret. Dann ergibt sich (32) unmittelbar aus der Transformationsformel (31).
- Sei nun  $X$  absolutstetig. Dann ergibt sich (33) aus (31) und aus (3.14). □

**Beispiele**1. *Binomialverteilung*

- Sei  $X$  binomialverteilt mit den Parametern  $n \in \mathbb{N}$  und  $p \in [0, 1]$ .
- In Abschnitt 4.1.1 wurde gezeigt, dass  $\mathbb{E} X = np$ .
- Analog ergibt sich aus Korollar 4.3, dass

$$\begin{aligned}
 \mathbb{E}(X^2) &= \sum_{i=1}^n i^2 P(X=i) \\
 &= \sum_{i=1}^n (i(i-1) + i) P(X=i) \\
 &= \sum_{i=1}^n i(i-1) P(X=i) + \sum_{i=1}^n i P(X=i) \\
 &= n(n-1)p^2 + np.
 \end{aligned}$$

- Also ergibt sich aus (27), dass

$$\begin{aligned}
 \text{Var } X &= \mathbb{E}(X^2) - (\mathbb{E} X)^2 \\
 &= n(n-1)p^2 + np - (np)^2 = np(1-p).
 \end{aligned}$$

2. *Normalverteilung*

- Sei  $X$  normalverteilt mit den Parametern  $\mu \in \mathbb{R}$  und  $\sigma > 0$ .
- In Abschnitt 4.1.1 wurde gezeigt, dass  $\mathbb{E} X = \mu$ .
- Somit ergibt sich aus (26) und (33), dass

$$\begin{aligned}
 \text{Var } X &= \mathbb{E}((X - \mu)^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} (x - \mu)^2 \exp\left(-\frac{1}{2}\underbrace{\left(\frac{x - \mu}{\sigma}\right)^2}_{=t}\right) dx \\
 &= \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t^2 \exp\left(-\frac{1}{2}\underbrace{t^2}_{=u}\right) dt \\
 &= \sigma^2 \frac{2}{\sqrt{2\pi}} \int_0^\infty \sqrt{2u} \exp(-u) du \\
 &= \sigma^2 \frac{2}{\sqrt{\pi}} \underbrace{\int_0^\infty \sqrt{u} \exp(-u) du}_{= \Gamma(3/2)} = \sigma^2, \\
 &= \Gamma(3/2) = \frac{1}{2}\Gamma(1/2) = \sqrt{\pi}/2
 \end{aligned}$$

wobei die Gamma-Funktion gegeben ist durch  $\Gamma(p) = \int_0^\infty u^{p-1} e^{-u} du$  für  $p > 0$ .

**Beachte**

- Die Parameter (und damit die Wahrscheinlichkeitsfunktion bzw. die Dichte) von Binomial- bzw. Normalverteilung sind jeweils eindeutig durch den Erwartungswert und die Varianz dieser Verteilungen festgelegt.

- Für die Normalverteilung ist dies offensichtlich. Denn, wie soeben gezeigt, gilt  $\mathbb{E} X = \mu$  und  $\text{Var } X = \sigma^2$ , falls  $X \sim N(\mu, \sigma^2)$ .
- Sei nun  $Y \sim \text{Bin}(n, p)$ . Dann kann man sich leicht überlegen, dass

$$n = \frac{\mu^2}{\mu - \sigma^2}, \quad p = \frac{\mu - \sigma^2}{\mu},$$

wobei

$$\mu = \mathbb{E} Y = np \quad \text{und} \quad \sigma^2 = \text{Var } Y = np(1 - p). \quad (34)$$

- Im allgemeinen wird die Verteilung einer Zufallsvariablen jedoch *nicht eindeutig* durch den Erwartungswert und die Varianz bestimmt.
- *Beispiel.* Der Erwartungswert und die Varianz von  $X \sim N(\mu, \sigma^2)$  stimmen mit dem Erwartungswert und der Varianz von  $Y \sim \text{Bin}(n, p)$  überein, falls  $\mu$  und  $\sigma^2$  durch (34) gegeben sind.

### 4.3 Gemischte Momente

- Neben Erwartungswert, Varianz bzw. höheren Momenten einer (einzelnen) Zufallsvariablen werden außerdem sogenannte *gemischte Momente* für Familien von (endlich vielen) Zufallsvariablen betrachtet.
- Sie sind ein wichtiges Hilfsmittel, um Zusammenhänge zwischen zwei oder mehreren Zufallsvariablen zu quantifizieren.
- Bei der Herleitung dieser Eigenschaften wird die folgende Verallgemeinerung von Theorem 4.7 für Zufallsvektoren benötigt.

#### 4.3.1 Transformationssatz für Zufallsvektoren

Auf die gleiche Weise wie Theorem 4.7, d.h. mittels algebraischer Induktion, ergibt sich der folgende Transformationssatz für Zufallsvektoren.

**Theorem 4.8** *Sei  $X : \Omega \rightarrow \mathbb{R}^n$  ein beliebiger Zufallsvektor, und sei  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  eine Borel-messbare Abbildung, so dass*

$$\int_{\mathbb{R}^n} |\varphi(x)| P_X(dx) < \infty. \quad (35)$$

*Für den Erwartungswert  $\mathbb{E} \varphi(X)$  von  $\varphi(X) : \Omega \rightarrow \mathbb{R}$  gilt dann*

$$\mathbb{E} \varphi(X) = \int_{\mathbb{R}^n} \varphi(x) P_X(dx). \quad (36)$$

Die Aussage von Theorem 4.8 lässt sich wie folgt für diskrete bzw. absolutstetige Zufallsvektoren spezifizieren (genauso wie dies im eindimensionalen Fall in Korollar 4.3 getan wurde).

**Korollar 4.4** *Sei  $X : \Omega \rightarrow \mathbb{R}^n$  ein beliebiger Zufallsvektor, und sei  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  eine beliebige Borel-messbare Abbildung.*

1. *Wenn  $X$  diskret ist mit  $P(X \in C) = 1$  für eine abzählbare Menge  $C \subset \mathbb{R}^n$ , dann gilt*

$$\mathbb{E} \varphi(X) = \sum_{x \in C} \varphi(x) P(X = x), \quad (37)$$

*wobei vorausgesetzt wird, dass*

$$\sum_{x \in C} |\varphi(x)| P(X = x) < \infty.$$



2. Wenn  $X$  absolutstetig ist mit der (gemeinsamen) Dichte  $f_X(x)$ , dann gilt

$$\mathbb{E} \varphi(X) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n, \quad (38)$$

wobei vorausgesetzt wird, dass

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |\varphi(x_1, \dots, x_n)| f_X(x_1, \dots, x_n) dx_1 \dots dx_n < \infty.$$

Der Beweis von Korollar 4.4 verläuft analog zum Beweis von Korollar 4.3. Er wird deshalb weggelassen.

### 4.3.2 Multiplikationsformel und Kovarianz

**Definition** Seien  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen, so dass

$$\mathbb{E} |X_1 \dots X_n| < \infty. \quad (39)$$

Der Erwartungswert  $\mathbb{E}(X_1 \dots X_n)$  des Produktes  $X_1 \dots X_n$  heißt dann *gemischtes Moment* der Zufallsvariablen  $X_1, \dots, X_n$ .

**Beachte** Man kann zeigen, dass die Integrierbarkeitsbedingung (39) erfüllt ist, falls  $\mathbb{E}(|X_i|^n) < \infty$  für jedes  $i \in \{1, \dots, n\}$ . Dies ergibt sich aus der Abschätzung

$$|X_1 \dots X_n| \leq \sum_{i=1}^n |X_i|^n \mathbb{1}_{\{X_i \geq \max\{X_1, \dots, X_n\}\}} \leq \sum_{i=1}^n |X_i|^n.$$

Mit Hilfe des Transformationssatzes für Zufallsvektoren, der in Theorem 4.8 diskutiert wurde, lässt sich nun die folgende *Multiplikationsformel* für den Erwartungswert des Produktes von  $n$  unabhängigen Zufallsvariablen herleiten.

**Theorem 4.9** Seien  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen mit  $\mathbb{E}(|X_i|^n) < \infty$  für jedes  $i \in \{1, \dots, n\}$ . Wenn  $X_1, \dots, X_n$  unabhängig sind, dann gilt

$$\mathbb{E} \left( \prod_{i=1}^n X_i \right) = \prod_{i=1}^n \mathbb{E} X_i. \quad (40)$$

**Beweis**

- Die Komponenten des Zufallsvektors  $X = (X_1, \dots, X_n)$  seien unabhängige Zufallsvariablen.
- Aus Theorem 4.8 ergibt sich dann mit  $\varphi(x) = x_1 \dots x_n$  für  $x = (x_1, \dots, x_n)$ , dass

$$\begin{aligned} \mathbb{E} \left( \prod_{i=1}^n X_i \right) &= \int_{\mathbb{R}^n} x_1 \dots x_n P_X(dx) \\ &= \int_{\mathbb{R}} \dots \int_{\mathbb{R}} x_1 \dots x_n P_{X_1}(dx_1) \dots P_{X_n}(dx_n) \\ &= \int_{\mathbb{R}} x_1 P_{X_1}(dx_1) \dots \int_{\mathbb{R}} x_n P_{X_n}(dx_n) \\ &= \prod_{i=1}^n \mathbb{E} X_i. \end{aligned}$$

□

**Beachte**

- Die Linearitätseigenschaft des Erwartungswertes, die wir in Teilaussage 3 von Theorem 4.4 gezeigt haben, ist so für die Varianz *nicht* zutreffend.
- Für Summen von unabhängigen Zufallsvariablen gilt jedoch das folgende *Additionstheorem* für die Varianz.

**Theorem 4.10** Seien  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  unabhängige Zufallsvariablen mit  $\mathbb{E}(X_i^2) < \infty$  für jedes  $i \in \{1, \dots, n\}$ . Dann gilt

$$\text{Var}(X_1 + \dots + X_n) = \text{Var } X_1 + \dots + \text{Var } X_n. \quad (41)$$

**Beweis**

- Wir zeigen die Gültigkeit von (41) zunächst für den Fall  $n = 2$ .
- Aus (27) und (40) ergibt sich, dass

$$\begin{aligned} \text{Var}(X_1 + X_2) &\stackrel{(27)}{=} \mathbb{E}((X_1 + X_2)^2) - (\mathbb{E}(X_1 + X_2))^2 \\ &= \left( \mathbb{E}(X_1^2) + 2\mathbb{E}(X_1 X_2) + \mathbb{E}(X_2^2) \right) - \left( (\mathbb{E} X_1)^2 + 2\mathbb{E} X_1 \mathbb{E} X_2 + (\mathbb{E} X_2)^2 \right) \\ &\stackrel{(40)}{=} \mathbb{E}(X_1^2) - (\mathbb{E} X_1)^2 + \mathbb{E}(X_2^2) - (\mathbb{E} X_2)^2 \\ &\stackrel{(27)}{=} \text{Var } X_1 + \text{Var } X_2. \end{aligned}$$

- Für beliebiges  $n \in \mathbb{N}$  ergibt sich die Gültigkeit von (41) mittels vollständiger Induktion.
- Dabei verwenden wir die Tatsache, dass wegen Theorem 3.17 aus der Unabhängigkeit von  $X_1, \dots, X_n$  folgt, dass auch die Zufallsvariablen  $X_1 + \dots + X_{n-1}$  und  $X_n$  unabhängig sind.
- Deshalb gilt

$$\begin{aligned} \text{Var}(X_1 + \dots + X_n) &= \text{Var}((X_1 + \dots + X_{n-1}) + X_n) \\ &= \text{Var}(X_1 + \dots + X_{n-1}) + \text{Var}(X_n) \\ &= \text{Var}(X_1) + \dots + \text{Var}(X_{n-1}) + \text{Var}(X_n), \end{aligned}$$

wobei sich die letzten beiden Gleichheiten aus der Induktionsannahme ergeben.  $\square$

Wir diskutieren nun Eigenschaften des gemischten Momentes  $\mathbb{E}(X_1 X_2)$  von zwei beliebigen (nicht notwendig unabhängigen) Zufallsvariablen  $X_1, X_2$ .

In diesem Zusammenhang führen wir zunächst die Begriffe der Kovarianz und des Korrelationskoeffizienten ein.

**Definition** Seien  $X_1, X_2$  beliebige Zufallsvariablen mit  $\mathbb{E}(X_i^2) < \infty$  für  $i = 1, 2$ .

- Der Erwartungswert  $\mathbb{E}((X_1 - \mathbb{E} X_1)(X_2 - \mathbb{E} X_2))$  heißt die *Kovarianz* von  $X_1$  und  $X_2$ , wobei wir die Schreibweise

$$\text{Cov}(X_1, X_2) = \mathbb{E}((X_1 - \mathbb{E} X_1)(X_2 - \mathbb{E} X_2)) \quad (42)$$

verwenden.

- Die Zufallsvariablen  $X_1, X_2$  heißen *unkorreliert*, falls  $\text{Cov}(X_1, X_2) = 0$ .
- Wenn  $\text{Var } X_1 > 0$  und  $\text{Var } X_2 > 0$ , dann heißt die Größe

$$\varrho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var } X_1 \cdot \text{Var } X_2}} \quad (43)$$

der *Korrelationskoeffizient* von  $X_1$  und  $X_2$ .

**Beachte**

- Es ist klar, dass Kovarianz und Korrelationskoeffizient die folgende *Symmetrieeigenschaft* besitzt:

$$\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1), \quad \varrho(X_1, X_2) = \varrho(X_2, X_1). \quad (44)$$

- Außerdem gilt

$$\text{Cov}(X, X) = \text{Var } X, \quad \varrho(X, X) = 1. \quad (45)$$

Darüber hinaus gelten weitere nützliche Rechenregeln und Abschätzungen für Kovarianz bzw. Korrelationskoeffizient.

**Theorem 4.11** Seien  $X_1, X_2$  beliebige Zufallsvariablen mit  $\mathbb{E}(X_i^2) < \infty$  für  $i = 1, 2$ .

1. Dann gilt

$$\text{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E} X_1 \mathbb{E} X_2 \quad (46)$$

2. und für beliebige Zahlen  $a, b, c, d \in \mathbb{R}$

$$\text{Cov}(aX_1 + b, cX_2 + d) = a c \text{Cov}(X_1, X_2). \quad (47)$$

3. Außerdem gilt die Ungleichung von Cauchy–Schwarz

$$\mathbb{E}|X_1 X_2| \leq \sqrt{\mathbb{E}(X_1^2) \mathbb{E}(X_2^2)}, \quad (48)$$

bzw.

$$|\text{Cov}(X_1, X_2)| \leq \sqrt{\text{Var } X_1 \text{Var } X_2}. \quad (49)$$

**Beweis**

Zu 1) Die Formel (46) ergibt sich unmittelbar aus (16) und (42), denn es gilt

$$\begin{aligned} \text{Cov}(X_1, X_2) &\stackrel{(42)}{=} \mathbb{E}((X_1 - \mathbb{E} X_1)(X_2 - \mathbb{E} X_2)) \\ &= \mathbb{E}(X_1 X_2 - X_1 \mathbb{E} X_2 - X_2 \mathbb{E} X_1 + \mathbb{E} X_1 \mathbb{E} X_2) \\ &\stackrel{(16)}{=} \mathbb{E}(X_1 X_2) - \mathbb{E} X_1 \mathbb{E} X_2. \end{aligned}$$

Zu 2) Die Formel (47) ergibt sich durch eine ähnliche einfache Rechnung aus (16) und (46), und zwar gilt

$$\begin{aligned} \text{Cov}(aX_1 + b, cX_2 + d) &\stackrel{(46)}{=} \mathbb{E}((aX_1 + b)(cX_2 + d)) - \mathbb{E}(aX_1 + b) \mathbb{E}(cX_2 + d) \\ &\stackrel{(16)}{=} a c (\mathbb{E}(X_1 X_2) - \mathbb{E} X_1 \mathbb{E} X_2) \\ &\stackrel{(46)}{=} a c \text{Cov}(X_1, X_2). \end{aligned}$$

Zu 3) Wir zeigen nun die Gültigkeit der Ungleichung (48).

- Wenn  $\mathbb{E}(X_1^2) = 0$ , dann gilt  $P(X_1 = 0) = 1$  und somit auch  $\mathbb{E}(X_1 X_2) = 0 \leq \sqrt{\mathbb{E}(X_1^2) \mathbb{E}(X_2^2)}$ .
- Sei jetzt  $\mathbb{E}(X_1^2) > 0$ . Dann gilt für jede Zahl  $a \in \mathbb{R}$

$$\begin{aligned} 0 &\leq \mathbb{E}((aX_1 + X_2)^2) = \mathbb{E}(a^2 X_1^2 + 2a X_1 X_2 + X_2^2) \\ &= a^2 \mathbb{E}(X_1^2) + 2a \mathbb{E}(X_1 X_2) + \mathbb{E}(X_2^2). \end{aligned}$$

- Durch beidseitige Multiplikation mit  $\mathbb{E}(X_1^2)$  bzw. quadratische Ergänzung ergibt sich hieraus, dass

$$\begin{aligned} 0 &\leq a^2 (\mathbb{E}(X_1^2))^2 + 2a \mathbb{E}(X_1^2) \mathbb{E}(X_1 X_2) + \mathbb{E}(X_1^2) \mathbb{E}(X_2^2) \\ &= (a \mathbb{E}(X_1^2) + \mathbb{E}(X_1 X_2))^2 + \mathbb{E}(X_1^2) \mathbb{E}(X_2^2) - (\mathbb{E}(X_1 X_2))^2. \end{aligned}$$

- Hieraus ergibt sich die Gültigkeit von (48) für  $a = -\mathbb{E}(X_1 X_2)/\mathbb{E}(X_1^2)$ , wenn dabei gleichzeitig  $X_1$  bzw.  $X_2$  durch  $|X_1|$  bzw.  $|X_2|$  ersetzt wird.
- Die Gültigkeit von (49) ergibt sich unmittelbar aus (48), wenn in (48) die Zufallsvariablen  $X_1$  bzw.  $X_2$  durch  $X_1 - \mathbb{E} X_1$  bzw.  $X_2 - \mathbb{E} X_2$  ersetzt werden.  $\square$

#### Korollar 4.5

1. Wenn  $X_1$  und  $X_2$  unabhängig sind, dann gilt

$$\text{Cov}(X_1, X_2) = 0. \quad (50)$$

d.h.,  $X_1$  und  $X_2$  sind unkorreliert.

2. Wenn  $\text{Var } X_1 > 0$  und  $\text{Var } X_2 > 0$ , dann gilt

$$-1 \leq \varrho(X_1, X_2) \leq 1. \quad (51)$$

#### Beweis

- Aus der Multiplikationsformel (40) in Theorem 4.9 und aus (46) ergibt sich, dass

$$\text{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E} X_1 \mathbb{E} X_2 = \mathbb{E} X_1 \mathbb{E} X_2 - \mathbb{E} X_1 \mathbb{E} X_2 = 0.$$

Damit ist (50) bewiesen.

- Die Ungleichungen in (51) ergeben sich unmittelbar aus der Ungleichung (49) von Cauchy-Schwarz und aus der Definitionsgleichung (43) des Korrelationskoeffizienten.  $\square$

**Beachte** Die Aussage 1 in Korollar 4.5 lässt sich im allgemeinen *nicht* umkehren, denn aus der Unkorreliertheit zweier Zufallsvariablen  $X_1$  und  $X_2$  folgt im allgemeinen nicht, dass  $X_1$  und  $X_2$  unabhängig sind.

**Beispiel** (zweimaliger Münzwurf)

- Seien  $Y_1, Y_2 : \Omega \rightarrow \{0, 1\}$  zwei unabhängige (und identisch verteilte) Zufallsvariablen, die nur die beiden Werte 0 oder 1 annehmen können, mit

$$P(Y_i = j) = \begin{cases} \frac{1}{2}, & \text{falls } j = 1, \\ \frac{1}{2}, & \text{falls } j = 0, \end{cases}$$

für  $i = 1, 2$ .

- Man kann sich leicht überlegen, dass dann die Zufallsvariablen  $X_1 = Y_1 + Y_2$  und  $X_2 = Y_1 - Y_2$  zwar unkorreliert, jedoch nicht unabhängig sind.
- Denn es gilt  $\mathbb{E} X_1 = 1$ ,  $\mathbb{E} X_2 = 0$  und  $\mathbb{E}(X_1 X_2) = 0$ . Andererseits gilt

$$P(X_1 = 0, X_2 = 0) = \frac{1}{4} \quad \neq \quad \frac{1}{4} \frac{1}{2} = P(X_1 = 0)P(X_2 = 0).$$

### 4.3.3 Linearer Zusammenhang von Zufallsvariablen

Die Korrelation  $\varrho(X_1, X_2)$  zweier Zufallsvariablen  $X_1$  und  $X_2$  mit  $0 < \text{Var } X_1, \text{Var } X_2 < \infty$  kann man als Grad ihres linearen (stochastischen) „Zusammenhanges“ auffassen.

**Theorem 4.12** Seien  $a, b \in \mathbb{R}$  beliebige Zahlen, und  $X_1, X_2$  seien Zufallsvariablen mit  $0 < \text{Var } X_1, \text{Var } X_2 < \infty$ .

1. Die erwartete quadratische Abweichung  $\mathbb{E}((X_2 - (aX_1 + b))^2)$  zwischen den Zufallsvariablen  $X_2$  und  $aX_1 + b$  ist minimal, wenn  $a$  und  $b$  wie folgt gewählt werden:

$$a = \varrho(X_1, X_2) \frac{\sqrt{\text{Var } X_2}}{\sqrt{\text{Var } X_1}}, \quad b = \mathbb{E} X_2 - a \mathbb{E} X_1. \quad (52)$$

2. Insbesondere gilt  $\mathbb{E}((X_2 - (aX_1 + b))^2) = 0$ , d.h.  $P(X_2 = aX_1 + b) = 1$  genau dann, wenn

$$|\varrho(X_1, X_2)| = 1.$$

#### Beweis

- Betrachten die Zufallsvariable  $Y = X_2 - (aX_1 + b)$ .
- Wegen (28) hängt dann  $\text{Var } Y$  nicht von  $b$  ab.
- Deshalb kann man bei der Minimierung von

$$\mathbb{E}(Y^2) = (\mathbb{E} Y)^2 + \text{Var } Y$$

zunächst  $(\mathbb{E} Y)^2$  für jedes feste  $a$  durch die entsprechende Wahl von  $b$  minimieren.

- Es ist klar, dass  $(\mathbb{E} Y)^2$  für  $b = \mathbb{E} X_2 - a \mathbb{E} X_1$  minimal ist, weil dann  $\mathbb{E} Y = 0$  gilt.
- Es ist nun noch  $a$  so zu bestimmen, dass

$$\begin{aligned} \text{Var } Y &= \mathbb{E} \left( ((X_2 - \mathbb{E} X_2) - a(X_1 - \mathbb{E} X_1))^2 \right) \\ &= \text{Var } X_2 - 2a \text{Cov}(X_1, X_2) + a^2 \text{Var } X_1 \\ &= \text{Var } X_1 \left( a - \frac{\text{Cov}(X_1, X_2)}{\text{Var } X_1} \right)^2 + \text{Var } X_2 \left( 1 - \frac{(\text{Cov}(X_1, X_2))^2}{\text{Var } X_1 \text{Var } X_2} \right) \end{aligned}$$

minimal wird, wobei sich die letzte Gleichung durch quadratische Ergänzung ergibt.

- Hieraus folgt, dass  $\text{Var } Y$  minimal ist, falls

$$a = \frac{\text{Cov}(X_1, X_2)}{\text{Var } X_1} = \varrho(X_1, X_2) \frac{\sqrt{\text{Var } X_2}}{\sqrt{\text{Var } X_1}}.$$

- Außerdem folgt hieraus, dass  $\mathbb{E}(Y^2) = 0$  genau dann, wenn  $|\varrho(X_1, X_2)| = 1$ . □

### 4.3.4 Erwartungswertvektor und Kovarianzmatrix

Wir zeigen zunächst, wie der Begriff der Kovarianz genutzt werden kann, um die in Theorem 4.10 angegebene Additionsformel (41) für die Varianz zu verallgemeinern.

**Theorem 4.13** Seien  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen mit  $\mathbb{E}(X_i^2) < \infty$  für jedes  $i \in \{1, \dots, n\}$ .

1. Dann gilt

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var } X_i + 2 \sum_{1 \leq j < k \leq n} \text{Cov}(X_j, X_k). \quad (53)$$

2. Wenn die Zufallsvariablen  $X_1, \dots, X_n$  paarweise unkorreliert sind, dann gilt insbesondere

$$\text{Var}(X_1 + \dots + X_n) = \text{Var} X_1 + \dots + \text{Var} X_n. \quad (54)$$

**Beweis**

- Wir betrachten zunächst den Fall  $n = 2$ .
- Dann ergibt sich sofort aus dem Beweis von Theorem 4.10, dass

$$\text{Var}(X_1 + X_2) = \text{Var} X_1 + \text{Var} X_2 + 2 \text{Cov}(X_1, X_2). \quad (55)$$

- Für beliebiges  $n \in \mathbb{N}$  ergibt sich die Gültigkeit von (53) nun mittels vollständiger Induktion.
- Und zwar erhalten wir aus (55) und aus der Induktionsannahme, dass

$$\begin{aligned} \text{Var}(X_1 + \dots + X_n) &= \text{Var}((X_1 + \dots + X_{n-1}) + X_n) \\ &= \text{Var}(X_1 + \dots + X_{n-1}) + \text{Var}(X_n) \\ &\quad + 2 \text{Cov}(X_1 + \dots + X_{n-1}, X_n) \\ &= \text{Var}(X_1 + \dots + X_{n-1}) + \text{Var}(X_n) + 2 \sum_{1 \leq j \leq n-1} \text{Cov}(X_j, X_n) \\ &\stackrel{\text{Ind.-annahme}}{=} \sum_{i=1}^{n-1} \text{Var} X_i + 2 \sum_{1 \leq j < k \leq n-1} \text{Cov}(X_j, X_k) \\ &\quad + \text{Var}(X_n) + 2 \sum_{1 \leq j \leq n-1} \text{Cov}(X_j, X_n) \\ &= \sum_{i=1}^n \text{Var} X_i + 2 \sum_{1 \leq j < k \leq n} \text{Cov}(X_j, X_k). \end{aligned}$$

- Die Gleichung (54) ergibt sich unmittelbar aus (50) und (53). □

**Beachte** Neben den Erwartungswerten  $\mathbb{E} X_1, \dots, \mathbb{E} X_n$  und den Varianzen  $\text{Var} X_1, \dots, \text{Var} X_n$  sind die in (53) auftretenden Kovarianzen  $\{\text{Cov}(X_i, X_j), 1 \leq i < j \leq n\}$  wichtige Charakteristiken des Zufallsvektors  $X = (X_1, \dots, X_n)$ .

Dies führt zu den folgenden Begriffsbildungen.

**Definition** Seien  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen mit  $\mathbb{E}(X_i^2) < \infty$  für jedes  $i \in \{1, \dots, n\}$ .

1. Der Vektor  $(\mathbb{E} X_1, \dots, \mathbb{E} X_n)$  heißt der *Erwartungswertvektor* des Zufallsvektors  $X = (X_1, \dots, X_n)$ .  
Schreibweise:  $\mathbb{E} X = (\mathbb{E} X_1, \dots, \mathbb{E} X_n)$
2. Die  $n \times n$ -Matrix

$$\text{Cov} X = \begin{pmatrix} \text{Cov}(X_1, X_1) & \cdots & \text{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \cdots & \text{Cov}(X_n, X_n) \end{pmatrix} = (\text{Cov}(X_i, X_j))_{ij}$$

heißt die *Kovarianzmatrix* von  $X = (X_1, \dots, X_n)$ .

**Theorem 4.14** Die Kovarianzmatrix  $\text{Cov } X$  ist

1. symmetrisch, d.h., für beliebige  $i, j \in \{1, \dots, n\}$  gilt

$$\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i). \quad (56)$$

2. nichtnegativ definit, d.h., für jedes  $x^\top = (x_1, \dots, x_n) \in \mathbb{R}^n$  gilt

$$x^\top \text{Cov } X x \geq 0, \quad (57)$$

wobei  $x^\top$  der zu  $x$  transponierte Vektor ist.

**Beweis**

- Die Symmetrieeigenschaft (56) ergibt sich unmittelbar aus der Definitionsgleichung (42) der Kovarianz.
- Außerdem gilt

$$\begin{aligned} x^\top \text{Cov } X x &= \sum_{i,j=1}^n x_i \text{Cov}(X_i, X_j) x_j \\ &= \sum_{i=1}^n x_i \text{Cov}\left(X_i, \sum_{j=1}^n x_j X_j\right) \\ &= \text{Cov}\left(\sum_{i=1}^n x_i X_i, \sum_{j=1}^n x_j X_j\right) \\ &= \text{Var}\left(\sum_{i=1}^n x_i X_i\right) \geq 0. \end{aligned}$$

- Damit ist auch die Ungleichung (57) bewiesen. □

**Beachte** Die Matrix  $\text{Cov } X$  heißt *positiv definit*, falls

$$x^\top \text{Cov } X x > 0 \quad (58)$$

für jedes  $x \in \mathbb{R}^n$  mit  $x \neq 0$ .

**Beispiel** (zweidimensionale Normalverteilung)

- Betrachten nun eine (weitere) Verallgemeinerung der zweidimensionalen Normalverteilung, die in Abschnitt 3.3.6 eingeführt wurde.
- Seien  $\mu_1, \mu_2 \in \mathbb{R}$ ,  $\sigma_1, \sigma_2 > 0$  und  $\varrho \in [0, 1)$  beliebige, jedoch fest vorgegebene Zahlen.
- Sei  $X = (X_1, X_2)$  ein absolutstetiger Zufallsvektor mit der gemeinsamen Dichte

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\varrho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right). \quad (59)$$

- Dann heißt  $X$  *normalverteilt* mit dem Erwartungswertvektor  $\mathbb{E} X = (\mu_1, \mu_2)$  und der Kovarianzmatrix

$$\text{Cov } X = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho \\ \sigma_1\sigma_2\varrho & \sigma_2^2 \end{pmatrix}. \quad (60)$$

**Beachte**

- Die Verteilung eines normalverteilten Zufallsvektors  $X = (X_1, X_2)$  wird *eindeutig* durch den Erwartungswertvektor  $\mathbb{E} X = (\mu_1, \mu_2)$  und die Kovarianzmatrix  $\text{Cov } X$  bestimmt. Für beliebige (nicht normalverteilte) Zufallsvektoren gilt diese Eindeutigkeitsaussage jedoch im allgemeinen nicht.
- Es gilt  $\varrho = \varrho(X_1, X_2)$ . Außerdem ergibt sich aus (59), dass  $f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$  genau dann, wenn  $\varrho = 0$ . Die Komponenten  $X_1, X_2$  eines normalverteilten Zufallsvektors  $X = (X_1, X_2)$  sind also genau dann unabhängig, wenn sie unkorreliert sind.
- Weil  $\varrho < 1$  vorausgesetzt wird, ist die Determinante der Kovarianzmatrix  $\text{Cov } X$  nicht Null, d.h., die Matrix  $\text{Cov } X$  ist positiv definit und invertierbar. Man kann sich deshalb leicht überlegen, dass die in (59) gegebenen Dichte  $f_X$  von  $X$  auch wie folgt dargestellt werden kann: Es gilt

$$f_X(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sqrt{\det K}} \exp\left(-\frac{1}{2}(x - \mu)^\top K^{-1}(x - \mu)\right) \quad (61)$$

für jedes  $x^\top = (x_1, x_2) \in \mathbb{R}^2$ , wobei  $\mu^\top = (\mu_1, \mu_2)$  und  $K^{-1}$  die inverse Matrix zu der in (60) gegebenen Kovarianzmatrix  $K = \text{Cov } X$  bezeichnet.

Schreibweise:  $X \sim N(\mu, K)$ .

- Man kann sich leicht überlegen, dass die Randverteilungen von  $X \sim N(\mu, K)$  (eindimensionale) Normalverteilungen sind mit  $X_i \sim N(\mu_i, \sigma_i^2)$  für  $i = 1, 2$ .
- Manchmal betrachtet man auch Zufallsvektoren  $X = (X_1, X_2)$ , deren Komponenten  $X_1, X_2$  normalverteilt sind mit  $|\varrho(X_1, X_2)| = 1$ . Aus Theorem 4.12 folgt dann, dass  $P(X_2 = aX_1 + b) = 1$  für ein Zahlenpaar  $a, b \in \mathbb{R}$ . In diesem Fall ist der Zufallsvektor  $X = (X_1, X_2)$  *nicht absolutstetig*, obwohl seine Komponenten diese Eigenschaft besitzen.

Für Zufallsvektoren mit einer beliebigen Dimension  $n \in \mathbb{N}$  kann man den Begriff der  $n$ -dimensionalen Normalverteilung einführen, indem man eine zu (61) analoge Dichte-Formel betrachtet.

**Definition**

- Sei  $\mu^\top = (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$  ein beliebiger Vektor, und sei  $K$  eine symmetrische und positiv definite  $n \times n$ -Matrix.
- Sei  $X = (X_1, \dots, X_n)$  ein absolutstetiger Zufallsvektor mit der gemeinsamen Dichte

$$f_X(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{\det K}} \exp\left(-\frac{1}{2}(x - \mu)^\top K^{-1}(x - \mu)\right) \quad (62)$$

für jedes  $x^\top = (x_1, \dots, x_n) \in \mathbb{R}^n$ .

- Man sagt dann, dass der Zufallsvektor  $X = (X_1, \dots, X_n)$  *normalverteilt* ist mit dem Erwartungswertvektor  $\mu = (\mu_1, \dots, \mu_n)^\top$  und der Kovarianzmatrix  $K$ .

**4.4 Ungleichungen für Momente und Wahrscheinlichkeiten****4.4.1 Ungleichungen vom  $L^p$ -Typ**

- In diesem Abschnitt verallgemeinern wir die Ungleichung (48) von Cauchy-Schwarz und leiten weitere Ungleichungen dieses Typs her, die wir *Ungleichungen vom  $L^p$ -Typ* nennen.
- Dabei ist das folgende Hilfsergebnis nützlich, das manchmal die *Ungleichung von Young* genannt wird.



**Lemma 4.3** Sei  $\varphi : [0, \infty) \rightarrow [0, \infty)$  eine stetige und streng monoton wachsende Funktion mit  $\varphi(0) = 0$  und  $\lim_{x \rightarrow \infty} \varphi(x) = \infty$ . Für die Funktionen  $\psi_1 : [0, \infty) \rightarrow [0, \infty)$  und  $\psi_2 : [0, \infty) \rightarrow [0, \infty)$  mit

$$\psi_1(x) = \int_0^x \varphi(y) dy, \quad \psi_2(x) = \int_0^x \varphi^{-1}(y) dy$$

gilt dann

$$ab \leq \psi_1(a) + \psi_2(b) \quad (63)$$

für beliebige  $a, b \geq 0$ , wobei  $\varphi^{-1}$  die zu  $\varphi$  inverse Funktion ist.

**Beweis**

- Wir unterscheiden zwei Fälle:  $\varphi(a) \geq b$  bzw.  $\varphi(a) \leq b$ .
- Sei zunächst  $\varphi(a) \geq b$ .
- Dann gibt es ein  $a' \leq a$ , so dass  $\varphi(a') = b$ , d.h.

$$\varphi(y) \leq b \quad \forall y \in (0, a') \quad \text{und} \quad \varphi(y) \geq b \quad \forall y \in (a', a).$$

- Hieraus folgt, dass

$$a'b = \psi_1(a') + \psi_2(b) \quad \text{und} \quad (a - a')b \leq \psi_1(a) - \psi_1(a'),$$

wobei sich die erste Gleichheit aus der Tatsache ergibt, dass die Werte der Integrale  $\psi_1(a')$  und  $\psi_2(b)$  als die Inhalte von Flächen unter bzw. über der Kurve  $\varphi : [0, a'] \rightarrow [0, b]$  aufgefasst werden können und dass deshalb die Summe  $\psi_1(a') + \psi_2(b)$  mit dem Flächeninhalt des Rechtecks  $[0, a'] \times [0, b]$  übereinstimmt.

- Es gilt somit

$$\begin{aligned} ab &= a'b + (a - a')b \\ &\leq \psi_1(a') + \psi_2(b) + \psi_1(a) - \psi_1(a') \\ &= \psi_1(a) + \psi_2(b). \end{aligned}$$

- Für den Fall  $\varphi(a) \leq b$  verläuft der Beweis analog. □

**Definition**

- Sei  $(\Omega, \mathcal{F}, P)$  ein beliebiger Wahrscheinlichkeitsraum, und sei  $p \in [1, \infty)$  eine beliebige Zahl.
- Mit  $L^p = L^p(\Omega, \mathcal{F}, P)$  bezeichnen wir dann die Familie aller Zufallsvariablen  $X : \Omega \rightarrow \mathbb{R}$ , für die  $\mathbb{E}(|X|^p) < \infty$ .
- Für jedes  $X \in L^p$  heißt  $\mathbb{E}(|X|^p)$  das  $p$ -te absolute Moment von  $X$ .

Wir leiten nun eine Abschätzung für das (erste) absolute Moment des Produktes zweier Zufallsvariablen her.

**Theorem 4.15** (Hölder–Ungleichung) Sei  $p, q > 1$ , so dass

$$\frac{1}{p} + \frac{1}{q} = 1, \quad (64)$$

und seien  $X, Y : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen mit  $X \in L^p$  und  $Y \in L^q$ . Dann gilt  $XY \in L^1$  und

$$\mathbb{E}|XY| \leq (\mathbb{E}(|X|^p))^{1/p} (\mathbb{E}(|Y|^q))^{1/q}. \quad (65)$$

**Beweis**

- Wenn  $\mathbb{E}(|X|^p) = 0$  oder  $\mathbb{E}(|Y|^q) = 0$ , dann ergibt sich aus Teilaussage 7 von Theorem 4.4, dass  $|X|^p = 0$  oder  $|X|^q = 0$ .
- In diesem Fall ergibt sich also, dass  $XY = 0$ .
- Es gilt somit  $XY \in L^1$  und  $\mathbb{E}|XY| = 0$ , d.h. in diesem Fall ist die Gültigkeit der Ungleichung (65) offensichtlich.
- Es gelte nun  $\mathbb{E}(|X|^p) > 0$  und  $\mathbb{E}(|Y|^q) > 0$ .
- Für  $\varphi(x) = x^{p-1}$  und  $a, b > 0$  ergibt sich aus der Ungleichung von Young (vgl. (63) in Lemma 4.3), dass

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

- Hieraus folgt für  $a = |X|/(\mathbb{E}(|X|^p))^{1/p}$  und  $b = |Y|/(\mathbb{E}(|Y|^q))^{1/q}$ , dass

$$\frac{|XY|}{(\mathbb{E}(|X|^p))^{1/p}(\mathbb{E}(|Y|^q))^{1/q}} \leq \frac{|X|^p}{p\mathbb{E}(|X|^p)} + \frac{|Y|^q}{q\mathbb{E}(|Y|^q)}.$$

- Wenn nun auf beiden Seiten dieser Ungleichung der Erwartungswert gebildet und die Bedingung (64) berücksichtigt wird, dann ergibt sich die Ungleichung (65) und damit auch, dass  $XY \in L^1$ .  $\square$

**Beachte**

- Als Spezialfall der Ungleichung (65) von Hölder ergibt sich für  $p = q = 2$  die Ungleichung (48) von Cauchy-Schwarz.
- Eine weitere Folgerung aus der Ungleichung (65) von Hölder ist das folgende Resultat.

**Korollar 4.6** (Ljapunow-Ungleichung) *Wenn  $1 \leq r \leq s$ , dann gilt*

$$L^s \subset L^r, \quad (66)$$

*und für jedes  $X \in L^s$  gilt*

$$(\mathbb{E}(|X|^r))^{1/r} \leq (\mathbb{E}(|X|^s))^{1/s}. \quad (67)$$

**Beweis**

- Sei  $1 \leq r \leq s$  und  $X \in L^s$ .
- Dann gilt auch  $X \in L^r$ , denn

$$\begin{aligned} \mathbb{E}(|X|^r) &= \mathbb{E}(|X|^r \mathbf{1}_{\{|X| \leq 1\}}) + \mathbb{E}(|X|^r \mathbf{1}_{\{|X| > 1\}}) \\ &\leq 1 + \mathbb{E}(|X|^s \mathbf{1}_{\{|X| > 1\}}) \\ &\leq 1 + \mathbb{E}(|X|^s) < \infty. \end{aligned}$$

- Damit ist (66) bewiesen.
- Um die Gültigkeit von (67) zu zeigen, genügt es, in die Ungleichung (65) von Hölder die Zufallsvariable  $|X|^r$  anstelle von  $X$  einzusetzen sowie  $p = s/r$  bzw.  $Y = 1$  zu setzen.
- Dann ergibt sich aus (65), dass

$$\mathbb{E}(|X|^r) \leq (\mathbb{E}(|X|^{rp}))^{1/p} = (\mathbb{E}(|X|^s))^{r/s}. \quad \square$$

Es gilt die folgende Abschätzung für das  $p$ -te absolute Moment der Summe von zwei Zufallsvariablen.

**Theorem 4.16** (Minkowski–Ungleichung) *Wenn  $p \geq 1$  und  $X, Y \in L^p$ , dann gilt  $X + Y \in L^p$  und*

$$(\mathbb{E}(|X + Y|^p))^{1/p} \leq (\mathbb{E}(|X|^p))^{1/p} + (\mathbb{E}(|Y|^p))^{1/p}. \quad (68)$$

**Beweis**

- Für  $p = 1$  ist (68) ein Spezialfall von Teilaussage 3 in Theorem 4.4.
- Sei nun  $p > 1$  und  $X, Y \in L^p$ .
- Für  $q = p/(p - 1)$  ist dann die Bedingung (64) erfüllt,
- und es gilt

$$\begin{aligned} \mathbb{E}(|X + Y|^p) &= \mathbb{E}(|X + Y|^{p-1}|X + Y|) \\ &\leq \mathbb{E}(|X + Y|^{p-1}|X|) + \mathbb{E}(|X + Y|^{p-1}|Y|) \\ &\leq (\mathbb{E}(|X|^p))^{1/p} (\mathbb{E}(|X + Y|^{q(p-1)}))^{1/q} \\ &\quad + (\mathbb{E}(|Y|^p))^{1/p} (\mathbb{E}(|X + Y|^{q(p-1)}))^{1/q} \\ &= (\mathbb{E}(|X + Y|^p))^{1/q} \left( (\mathbb{E}(|X|^p))^{1/p} + (\mathbb{E}(|Y|^p))^{1/p} \right), \end{aligned}$$

wobei in der vorletzten Gleichheit zweimal die Ungleichung (65) von Hölder bzw. in der letzten Gleichheit die Identität  $q(p - 1) = p$  angewendet wurde.

- Also ist

$$\mathbb{E}(|X + Y|^p) \leq (\mathbb{E}(|X + Y|^p))^{1/q} \left( (\mathbb{E}(|X|^p))^{1/p} + (\mathbb{E}(|Y|^p))^{1/p} \right). \quad (69)$$

- Hieraus folgt (68), wenn beide Seiten der Ungleichung (69) durch  $(\mathbb{E}(|X + Y|^p))^{1/q}$  dividiert werden und wenn dabei berücksichtigt wird, dass  $1 - q^{-1} = p^{-1}$ .  $\square$

#### 4.4.2 Jensen–Ungleichung

- Durch die folgende Ungleichung ist eine untere Schranke gegeben für den Erwartungswert von konvexen Funktionen von Zufallsvariablen.
- Dabei heißt die Funktion  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  *konvex*, falls es für jedes  $x_0 \in \mathbb{R}$  eine Zahl  $c = c(x_0)$  gibt, so dass

$$\varphi(x) \geq \varphi(x_0) + c(x - x_0) \quad (70)$$

für jedes  $x \in \mathbb{R}$  gilt.

**Theorem 4.17** (Jensen–Ungleichung) *Sei  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  eine konvexe Funktion. Wenn  $X \in L^1$  und  $\varphi(X) \in L^1$ , dann gilt*

$$\varphi(\mathbb{E} X) \leq \mathbb{E} \varphi(X). \quad (71)$$

**Beweis**

- Die Funktion  $\varphi$  sei derart, dass  $X \in L^1$  und  $\varphi(X) \in L^1$ .
- Wenn nun  $x = X$  und  $x_0 = \mathbb{E} X$  in (70) eingesetzt wird, dann ergibt sich die Ungleichung

$$\varphi(X) \geq \varphi(\mathbb{E} X) + c(X - \mathbb{E} X).$$

- Werden auf beiden Seiten dieser Ungleichung die Erwartungswerte gebildet, so ergibt dies wegen der Linearität des Erwartungswertes, dass

$$\mathbb{E} \varphi(X) \geq \mathbb{E} (\varphi(\mathbb{E} X) + c(X - \mathbb{E} X)) = \varphi(\mathbb{E} X).$$

□

### Beachte

- Die Ungleichung (67) von Ljapunow lässt sich auch als Folgerung aus der Ungleichung (71) von Jensen gewinnen.
- Hierfür genügt es, in (71) die Zufallsvariable  $|X|^r$  anstelle von  $X$  zu betrachten und  $\varphi(x) = |x|^p$  zu setzen, wobei  $p = s/r$  und  $1 \leq r \leq s$ .
- Dann ergibt sich aus (71), dass

$$(\mathbb{E}(|X|^r))^{s/r} \leq \mathbb{E}(|X|^s).$$

- Das ist äquivalent mit der Ungleichung (67) von Ljapunow.

### 4.4.3 Tschebyschew–Ungleichung; Markow–Ungleichung

- In vielen Fällen lässt sich die Wahrscheinlichkeit  $P(A)$  von interessierenden Ereignissen  $A \in \mathcal{F}$  nicht in geschlossenen Formeln ausdrücken.
- Manchmal ist es jedoch möglich, Ungleichungen herzuleiten, um (obere) Schranken, d.h. Abschätzungen für  $P(A)$  zu erhalten.
- In diesem Abschnitt wird die sogenannte *Tschebyschew–Ungleichung* und, in Verallgemeinerung hiervon, die *Markow–Ungleichung* diskutiert.
- Sie liefern obere Schranken für die Wahrscheinlichkeit, dass die Abweichungen  $|X(\omega) - \mathbb{E} X|$  der Werte  $X(\omega)$  einer Zufallsvariablen  $X$  von ihrem Erwartungswert  $\mathbb{E} X$  einen vorgegebenen Schwellenwert  $\varepsilon > 0$  überschreiten.

**Theorem 4.18** (Tschebyschew–Ungleichung) Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable mit

$$\mathbb{E}(X^2) < \infty.$$

Dann gilt für jedes  $\varepsilon > 0$

$$P(|X - \mathbb{E} X| > \varepsilon) \leq \frac{\text{Var } X}{\varepsilon^2}. \quad (72)$$

### Beweis

- Aus den Linearitäts- bzw. Monotonie-Eigenschaften des Erwartungswertes (vgl. Theorem 4.4) ergibt sich, dass für jedes  $\varepsilon > 0$

$$\begin{aligned} \text{Var } X &= \mathbb{E}((X - \mathbb{E} X)^2) \\ &= \mathbb{E}((X - \mathbb{E} X)^2(\mathbb{1}_{\{|X - \mathbb{E} X| > \varepsilon\}} + \mathbb{1}_{\{|X - \mathbb{E} X| \leq \varepsilon\}})) \\ &= \mathbb{E}((X - \mathbb{E} X)^2 \mathbb{1}_{\{|X - \mathbb{E} X| > \varepsilon\}}) + \mathbb{E}((X - \mathbb{E} X)^2 \mathbb{1}_{\{|X - \mathbb{E} X| \leq \varepsilon\}}) \\ &\geq \mathbb{E}((X - \mathbb{E} X)^2 \mathbb{1}_{\{|X - \mathbb{E} X| > \varepsilon\}}) \\ &\geq \mathbb{E}(\varepsilon^2 \mathbb{1}_{\{|X - \mathbb{E} X| > \varepsilon\}}) = \varepsilon^2 \mathbb{E} \mathbb{1}_{\{|X - \mathbb{E} X| > \varepsilon\}} \\ &= \varepsilon^2 P(|X - \mathbb{E} X| > \varepsilon). \end{aligned}$$

□

**Beachte**

- Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable mit  $\mathbb{E}(|X|^p) < \infty$  für ein  $p \geq 1$ . Dann lässt sich genauso wie im Beweis von Theorem 4.18 zeigen, dass für jedes  $\varepsilon > 0$  die sogenannte *Markow-Ungleichung* gilt:

$$P(|X| > \varepsilon) \leq \frac{\mathbb{E}(|X|^p)}{\varepsilon^p}. \quad (73)$$

- Die Tschebyschew-Ungleichung (72) bzw. die Markow-Ungleichung (73) sind *nicht* an spezielle Annahmen über die Form der Verteilung der Zufallsvariablen  $X$  gebunden.
- Der „Preis“ hierfür ist, dass (72) und (73) in vielen Fällen zu relativ groben Abschätzungen führen.
- Wenn zusätzliche Annahmen über die Verteilung von  $X$  gemacht werden, dann lassen sich genauere Abschätzungen herleiten bzw. die Wahrscheinlichkeit  $P(|X - \mathbb{E} X| > \varepsilon)$  lässt sich explizit bestimmen.

**Beispiele**1. *fehlerbehaftete Messungen*

- Von einem Messgerät sei bekannt, dass die Messergebnisse fehlerbehaftet sind.
- Die  $n$ -te Messung einer (unbekannten) Größe  $\mu \in \mathbb{R}$  liefere den Wert  $\mu + X_n(\omega)$  für  $\omega \in \Omega$ .
- Die Messfehler  $X_1, X_2, \dots$  seien unabhängige und identisch verteilte Zufallsvariablen.
- Über die Verteilung von  $X_n$  sei lediglich bekannt, dass

$$\mathbb{E} X_n = 0 \quad \text{und} \quad \text{Var } X_n = 1. \quad (74)$$

- Es soll nun die Frage diskutiert werden, wieviele Messungen erforderlich sind, um mit Hilfe der Tschebyschew-Ungleichung (72) schlussfolgern zu können, dass das arithmetische Mittel

$$Y_n = \frac{1}{n} \sum_{i=1}^n (\mu + X_i)$$

der zufälligen Messwerte  $\mu + X_i$  höchstens mit Wahrscheinlichkeit 0.1 um mehr als 1 vom „wahren“, jedoch unbekannten Wert  $\mu$  abweicht.

- Aus den elementaren Eigenschaften von Erwartungswert und Varianz der Summe von  $n$  unabhängigen Zufallsvariablen ergibt sich (vgl. Korollar 4.2 bzw. die Theoreme 4.6 und 4.10), dass

$$\begin{aligned} \mathbb{E} Y_n &= \mathbb{E} \frac{1}{n} \sum_{i=1}^n (\mu + X_i) \\ &= \frac{1}{n} \sum_{i=1}^n (\mu + \mathbb{E} X_i) = \mu \end{aligned}$$

und

$$\begin{aligned} \text{Var } Y_n &= \text{Var} \frac{1}{n} \sum_{i=1}^n (\mu + X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var} (\mu + X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var } X_i = \frac{1}{n}. \end{aligned}$$

- Hieraus und aus der Tschebyschew-Ungleichung (72) ergibt sich, dass

$$P(|Y_n - \mu| > 1) \leq \frac{1}{n}.$$

- Es gilt also  $P(|Y_n - \mu| > 1) \leq 0.1$ , falls  $n^{-1} \leq 0.1$ .
- Aus diesen Überlegungen folgt, dass die obengenannten Genauigkeitsvorgaben erfüllt sind, falls  $n \geq 10$  Messungen durchgeführt werden.

## 2. normalverteilte Messfehler

- Es wird nun zusätzlich angenommen, dass die Messfehler normalverteilt sind, d.h.  $X_n \sim N(0, 1)$ .
- Man kann zeigen, dass dann  $Y_n \sim N(\mu, 1/n)$  bzw.  $\sqrt{n}(Y_n - \mu) \sim N(0, 1)$ , vgl. das Beispiel in Abschnitt 3.4.2 bzw. den Kommentar nach Korollar 3.2.
- Es gilt also

$$\begin{aligned}
 P(|Y_n - \mu| > 1) &= P(\sqrt{n}|Y_n - \mu| > \sqrt{n}) \\
 &= P(\sqrt{n}(Y_n - \mu) < -\sqrt{n}) + P(\sqrt{n}(Y_n - \mu) > \sqrt{n}) \\
 &= \Phi(-\sqrt{n}) + (1 - \Phi(\sqrt{n})) \\
 &= 2(1 - \Phi(\sqrt{n})),
 \end{aligned}$$

wobei

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \quad (75)$$

die Verteilungsfunktion der Standardnormalverteilung ist.

- Somit ist  $P(|Y_n - \mu| > 1) \leq 0.1$  genau dann erfüllt, wenn  $\Phi(\sqrt{n}) \geq 0.95$ .
- Dies gilt dann, wenn  $\sqrt{n} \geq 1.645$  bzw.  $n \geq 3$ .

## Beachte

- Es gibt keine geschlossene Formel für die Stammfunktion des Integrals in (75).
- Die Werte  $\Phi(x)$  der Verteilungsfunktion der Standardnormalverteilung müssen deshalb *numerisch* berechnet werden.
- Für  $x \geq 0$  sind die Werte  $\Phi(x)$  in Tabelle 1 gegeben.
- Für  $x < 0$  erhält man  $\Phi(x)$  dann aus der folgenden Symmetrieeigenschaft, die sich unmittelbar aus der Definitionsgleichung (75) ergibt.
- Für jedes  $x \in \mathbb{R}$  gilt

$$\Phi(x) = 1 - \Phi(-x). \quad (76)$$

## 5 Konvergenzarten und Grenzwertsätze

- Oft ist es nützlich, das asymptotische (Grenz-) Verhalten der Wahrscheinlichkeit  $P(A)$  eines Ereignisses  $A \in \mathcal{F}$  zu kennen, wenn bestimmte Modellparameter unendlich groß bzw. klein werden.
- In diesem Zusammenhang diskutieren wir zunächst verschiedene Konvergenzarten von Zufallsvariablen.

### 5.1 Konvergenzarten

#### 5.1.1 Definitionen und elementare Eigenschaften

**Definition** Seien  $X, X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen. Man sagt: Die Folge

1.  $\{X_n\}$  konvergiert *fast sicher* gegen  $X$ , falls

$$P(\{\omega : \omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1. \quad (1)$$

Schreibweise:  $X_n \xrightarrow{\text{f.s.}} X$

2.  $\{X_n\}$  konvergiert *in Wahrscheinlichkeit* gegen  $X$ , falls für jedes  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0. \quad (2)$$

Schreibweise:  $X_n \xrightarrow{P} X$

3.  $\{X_n\}$  konvergiert im  $L^1$  gegen  $X$ , falls  $X, X_1, X_2, \dots \in L^1$  und

$$\lim_{n \rightarrow \infty} \mathbb{E} |X_n - X| = 0. \quad (3)$$

Schreibweise:  $X_n \xrightarrow{L^1} X$

4.  $\{X_n\}$  konvergiert *im quadratischen Mittel* gegen  $X$ , falls  $X, X_1, X_2, \dots \in L^2$  und

$$\lim_{n \rightarrow \infty} \mathbb{E} ((X_n - X)^2) = 0. \quad (4)$$

Schreibweise:  $X_n \xrightarrow{L^2} X$

5.  $\{X_n\}$  konvergiert *in Verteilung* gegen  $X$ , falls

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad (5)$$

für jeden Stetigkeitspunkt  $x \in \mathbb{R}$  der Verteilungsfunktion  $F_X$ .

Schreibweise:  $X_n \xrightarrow{d} X$

#### Beachte

- Wenn  $X_n \xrightarrow{\text{f.s.}} X$ , dann sagt man auch, dass die Folge  $\{X_n\}$  *mit Wahrscheinlichkeit 1* gegen  $X$  konvergiert.
- Wenn  $X_n \xrightarrow{P} X$ , dann sagt man auch, dass die Folge  $\{X_n\}$  *stochastisch* gegen  $X$  konvergiert.
- Wenn  $X_n \xrightarrow{L^2} X$ , dann sagt man auch, dass die Folge  $\{X_n\}$  im  $L^2$  gegen  $X$  konvergiert.

Das folgende *Konvergenzkriterium* ist ein nützliches Hilfsmittel beim Beweis tieferliegender Grenzwertsätze.

**Theorem 5.1** Sei  $Y_n = \sup_{k \geq n} |X_k - X|$ . Es gilt  $X_n \xrightarrow{\text{f.s.}} X$  genau dann, wenn

$$Y_n \xrightarrow{P} 0. \quad (6)$$

**Beweis**

- Für jedes  $\varepsilon > 0$  sei

$$A_\varepsilon = \limsup_n \{|X_n - X| > \varepsilon\} \quad (7)$$

- Die Gültigkeit von  $X_n \xrightarrow{\text{f.s.}} X$  impliziert dann, dass  $P(A_\varepsilon) = 0$  für jedes  $\varepsilon > 0$ .
- Außerdem ergibt sich aus der Stetigkeit von Wahrscheinlichkeitsmaßen (vgl. Theorem 2.3), dass

$$P(A_\varepsilon) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}\right) = \lim_{n \rightarrow \infty} P\left(\sup_{k \geq n} |X_k - X| > \varepsilon\right).$$

- Damit ist die Notwendigkeit von Bedingung (6) gezeigt.
- Es gelte nun (6), d.h.,  $P(A_\varepsilon) = 0$  für jedes  $\varepsilon > 0$ .
- Hieraus folgt, dass  $X_n \xrightarrow{\text{f.s.}} X$ , weil dann

$$P(\{\omega : X_n(\omega) \not\xrightarrow{\text{f.s.}} X(\omega)\}) = P\left(\bigcup_{m=1}^{\infty} A_{1/m}\right) \leq \sum_{m=1}^{\infty} P(A_{1/m}) = 0. \quad \square$$

Aus Theorem 5.1 ergibt sich sofort, dass die fast sichere Konvergenz stets die Konvergenz in Wahrscheinlichkeit impliziert.

**Korollar 5.1** Aus  $X_n \xrightarrow{\text{f.s.}} X$  folgt  $X_n \xrightarrow{P} X$ .

**Beweis** In Theorem 5.1 wurde gezeigt, dass  $X_n \xrightarrow{\text{f.s.}} X$  die Gültigkeit von  $\sup_{k \geq n} |X_k - X| \xrightarrow{P} 0$  und damit auch  $|X_n - X| \xrightarrow{P} 0$  impliziert. Dies ist aber gleichbedeutend mit  $X_n \xrightarrow{P} X$ .  $\square$

Außerdem gibt es die folgende hinreichende Bedingung für die fast sichere Konvergenz.

**Korollar 5.2** Wenn

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty \quad (8)$$

für jedes  $\varepsilon > 0$ , dann gilt  $X_n \xrightarrow{\text{f.s.}} X$ .

**Beweis**

- Aus (8) und aus dem Lemma von Borel–Cantelli (vgl. Korollar 2.3) ergibt sich, dass  $P(A_\varepsilon) = 0$  für jedes  $\varepsilon > 0$ ,
- wobei  $A_\varepsilon$  die in (7) eingeführte Menge ist.



- Die Behauptung ergibt sich nun genauso wie im zweiten Teil des Beweises von Theorem 5.1.  $\square$

### Beachte

- Es gilt also  $X_n \xrightarrow{P} X$  genau dann, wenn die Wahrscheinlichkeiten  $P(|X_n - X| > \varepsilon)$  für jedes  $\varepsilon > 0$  gegen Null konvergieren.
- Im Unterschied hierzu gilt  $X_n \xrightarrow{\text{f.s.}} X$ , wenn (jedoch nicht nur dann) diese Null-Konvergenz hinreichend schnell erfolgt, so dass die Summe in (8) endlich ist.

Eine andere Charakterisierung der Konvergenz in Wahrscheinlichkeit lässt sich durch die Betrachtung von Teilfolgen gewinnen.

**Theorem 5.2** *Es gilt  $X_n \xrightarrow{P} X$  genau dann, wenn es für jede Teilfolge  $\{X_{n_i}\}$  von  $\{X_n\}$  eine Teilfolge  $\{X_{n_{i_j}}\} \subset \{X_{n_i}\}$  gibt, so dass  $X_{n_{i_j}} \xrightarrow{\text{f.s.}} X$ .*

### Beweis

- Es gelte  $X_n \xrightarrow{P} X$ .
- Dann gilt auch  $X_{n_i} \xrightarrow{P} X$  für jede Teilfolge  $\{X_{n_i}\}$  von  $\{X_n\}$ .
- Um die Schreibweise abzukürzen, identifizieren wir nun  $\{X_{n_i}\}$  mit  $\{X_n\}$ .
- Es genügt dann zu zeigen, dass es eine Teilfolge  $\{X_{n_i}\}$  von  $\{X_n\}$  gibt, so dass  $X_{n_i} \xrightarrow{\text{f.s.}} X$ .
- Für jedes  $n \in \mathbb{N}$  sei  $n_i$  so gewählt, dass  $P(|X_{n_i} - X| > 2^{-n}) \leq 2^{-n}$ .
- Dann gilt für jedes  $\varepsilon > 0$

$$\begin{aligned} & \sum_{n_i} P(|X_{n_i} - X| > \varepsilon) \\ & \leq \sum_{n_i: 2^{-n} > \varepsilon} P(|X_{n_i} - X| > \varepsilon) + \sum_{n_i: 2^{-n} \leq \varepsilon} P(|X_{n_i} - X| > 2^{-n}) \\ & \leq \sum_{n_i: 2^{-n} > \varepsilon} P(|X_{n_i} - X| > \varepsilon) + \sum_{n=1}^{\infty} 2^{-n} < \infty. \end{aligned}$$

- Aus Korollar 5.2 ergibt sich nun, dass  $X_{n_i} \xrightarrow{\text{f.s.}} X$ .
- Wir nehmen nun umgekehrt an, dass  $X_n \xrightarrow{P} X$  nicht gilt. Dann gibt es Zahlen  $\varepsilon > 0$ ,  $\delta > 0$  und eine Teilfolge  $\{X_{n_i}\}$  von  $\{X_n\}$ , so dass  $P(|X_{n_i} - X| > \varepsilon) > \delta$  für jedes  $i$ .
- Wegen Korollar 5.1 kann es somit keine Teilfolge  $\{X_{n_{i_j}}\} \subset \{X_{n_i}\}$  geben, so dass  $X_{n_{i_j}} \xrightarrow{\text{f.s.}} X$ .  $\square$

Wir zeigen nun, dass ebenfalls die  $L^2$ -Konvergenz und die  $L^1$ -Konvergenz stets die Konvergenz in Wahrscheinlichkeit implizieren.

**Theorem 5.3** *Wenn  $X, X_1, X_2, \dots \in L^2$  und  $X_n \xrightarrow{L^2} X$ , dann gilt auch  $X_n \xrightarrow{L^1} X$ .*

**Beweis**

- Aus der Ungleichung (4.48) von Cauchy–Schwarz ergibt sich, dass für jedes  $n \in \mathbb{N}$

$$0 \leq \mathbb{E}|X_n - X| \leq \sqrt{\mathbb{E}((X_n - X)^2)}.$$

- Somit gilt  $\mathbb{E}|X_n - X| \rightarrow 0$ , falls  $\mathbb{E}((X_n - X)^2) \rightarrow 0$ . □

**Theorem 5.4** Wenn  $X, X_1, X_2, \dots \in L^1$  und  $X_n \xrightarrow{L^1} X$ , dann gilt  $X_n \xrightarrow{P} X$ .

**Beweis**

- Aus der Markow–Ungleichung (4.73) für  $p = 1$  ergibt sich, dass für jedes  $\varepsilon > 0$

$$P(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}|X_n - X|}{\varepsilon}.$$

- Somit gilt  $P(|X_n - X| > \varepsilon) \rightarrow 0$ , falls  $\mathbb{E}|X_n - X| \rightarrow 0$ . □

Als nächstes zeigen wir, dass die Konvergenz in Wahrscheinlichkeit stets die Konvergenz in Verteilung impliziert.

**Theorem 5.5** Wenn  $X_n \xrightarrow{P} X$ , dann gilt  $X_n \xrightarrow{d} X$ .

**Beweis**

- Es gelte  $X_n \xrightarrow{P} X$ , d.h.,  $P(|X_n - X| > \varepsilon) \rightarrow 0$  für jedes  $\varepsilon > 0$ .
- Für beliebige  $\varepsilon > 0$  und  $n \in \mathbb{N}$  gilt

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(X_n \leq x, |X_n - X| \leq \varepsilon) + P(X_n \leq x, |X_n - X| > \varepsilon) \\ &\leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon) \\ &= F_X(x + \varepsilon) + P(|X_n - X| > \varepsilon). \end{aligned}$$

- Hieraus und wegen  $P(|X_n - X| > \varepsilon) \rightarrow 0$  ergibt sich nun, dass

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \varepsilon).$$

- Weil  $\varepsilon > 0$  beliebig klein gewählt werden kann und weil die Verteilungsfunktion  $F_X$  rechtsseitig stetig ist, ergibt sich somit, dass für jedes  $x \in \mathbb{R}$

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x). \quad (9)$$

- Sei nun  $x \in \mathbb{R}$  ein Stetigkeitspunkt der Verteilungsfunktion  $F_X$ , d.h., es gelte

$$\lim_{\varepsilon \downarrow 0} F_X(x - \varepsilon) = F_X(x). \quad (10)$$

- Ähnlich wie im ersten Teil des Beweises ergibt sich für beliebige  $\varepsilon > 0$  und  $n \in \mathbb{N}$ , dass

$$\begin{aligned} F_X(x - \varepsilon) &= P(X \leq x - \varepsilon) \\ &= P(X \leq x - \varepsilon, |X_n - X| \leq \varepsilon) + P(X \leq x - \varepsilon, |X_n - X| > \varepsilon) \\ &\leq F_{X_n}(x) + P(|X_n - X| > \varepsilon). \end{aligned}$$

- Mit Hilfe von (10) ergibt sich somit, dass

$$F_X(x) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x).$$

- Hieraus und aus (9) folgt, dass  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ . □

### Beachte

- Die Umkehrung der Aussage von Theorem 5.5 gilt im allgemeinen nicht.
- Wenn jedoch die „Grenzzufallsvariable“  $X$  mit Wahrscheinlichkeit 1 *einunddenselben* (deterministischen) Wert annimmt, dann gilt auch die folgende umgekehrte Aussage.

**Theorem 5.6** Wenn  $X_n \xrightarrow{d} c$  für eine Konstante  $c \in \mathbb{R}$ , dann gilt  $X_n \xrightarrow{P} c$ .

### Beweis

- Es gelte  $X_n \xrightarrow{d} c$ .
- Für beliebige  $\varepsilon > 0$  und  $n \in \mathbb{N}$  gilt dann

$$\begin{aligned} P(|X_n - c| > \varepsilon) &= P(X_n < c - \varepsilon) + P(X_n > c + \varepsilon) \\ &\leq F_{X_n}(c - \varepsilon) + (1 - F_{X_n}(c + \varepsilon)) \\ &\longrightarrow F_c(c - \varepsilon) + (1 - F_c(c + \varepsilon)) \\ &= 0, \end{aligned}$$

weil sowohl  $c - \varepsilon$  als  $c + \varepsilon$  Stetigkeitspunkte der Verteilungsfunktion  $F_c$  sind und weil  $F_c(c - \varepsilon) = 0$  bzw.  $F_c(c + \varepsilon) = 1$ . □

### 5.1.2 Charakterisierung der Verteilungskonvergenz

- Sei  $\mathcal{C}$  die Familie aller beschränkten stetigen Funktionen  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ .
- Für jede Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  und für jedes  $\varphi \in \mathcal{C}$  ist dann die zusammengesetzte Abbildung  $\varphi(X)$  eine integrierbare Zufallsvariable, d.h. insbesondere, dass der Erwartungswert  $\mathbb{E} \varphi(X)$  wohldefiniert ist.

Ein sehr nützliches *Konvergenzkriterium* ist durch das folgende Theorem gegeben.

**Theorem 5.7** Seien  $X, X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen. Dann gilt  $X_n \xrightarrow{d} X$  genau dann, wenn für jedes  $\varphi \in \mathcal{C}$

$$\lim_{n \rightarrow \infty} \mathbb{E} \varphi(X_n) = \mathbb{E} \varphi(X). \quad (11)$$

**Beweis**

## 1. Notwendigkeit von (11)

- Es gelte  $X_n \xrightarrow{d} X$ . Wir zeigen, dass dann (11) für jedes  $\varphi \in \mathcal{C}$  gilt.
- Weil  $\varphi \in \mathcal{C}$  beschränkt ist, gilt  $b = \sup_{x \in \mathbb{R}} |\varphi(x)| < \infty$ .
- Für jedes  $\varepsilon > 0$  sei nun  $\nu > 0$  so gewählt, dass  $\nu$  und  $-\nu$  Stetigkeitspunkte von  $F_X$  sind und dass  $P(|X| > \nu) < \varepsilon b^{-1}$ .
- Dies ist möglich, weil  $F_X$  höchstens abzählbar viele Unstetigkeitspunkte hat und weil

$$\lim_{x \rightarrow \infty} P(|X| > x) = 0.$$

- Wegen  $X_n \xrightarrow{d} X$  gilt außerdem  $P(|X_n| > \nu) < 2\varepsilon b^{-1}$  für jedes hinreichend große  $n$ .
- Weil  $\varphi \in \mathcal{C}$  beschränkt und stetig ist, lässt sich  $\varphi$  für jedes Paar  $\varepsilon, \nu > 0$  durch eine Treppenfunktion  $g : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$g(x) = \sum_{i=1}^k a_i \mathbb{I}_{(x_{i-1}, x_i]}(x), \quad a_1, \dots, a_k \in \mathbb{R}, \quad -\nu = x_0 < \dots < x_k = \nu$$

approximieren, so dass  $x_0, \dots, x_k$  Stetigkeitspunkte von  $F_X$  sind und

$$\sup_{x \in [-\nu, \nu]} |\varphi(x) - g(x)| < \varepsilon$$

gilt.

- Für jedes hinreichend große  $n$  gilt dann

$$\begin{aligned} & |\mathbb{E} \varphi(X_n) - \mathbb{E} \varphi(X)| \\ & \leq |\mathbb{E}(\varphi(X_n) \mathbb{I}_{\{|X_n| \leq \nu\}}) - \mathbb{E}(\varphi(X) \mathbb{I}_{\{|X| \leq \nu\}})| \\ & \quad + |\mathbb{E}(\varphi(X_n) \mathbb{I}_{\{|X_n| > \nu\}}) - \mathbb{E}(\varphi(X) \mathbb{I}_{\{|X| > \nu\}})| \\ & \leq |\mathbb{E}(\varphi(X_n) \mathbb{I}_{\{|X_n| \leq \nu\}}) - \mathbb{E}(\varphi(X) \mathbb{I}_{\{|X| \leq \nu\}})| + 3\varepsilon \\ & \leq 3\varepsilon + |\mathbb{E}(\varphi(X_n) \mathbb{I}_{\{|X_n| \leq \nu\}}) - \mathbb{E} g(X_n)| \\ & \quad + |\mathbb{E}(\varphi(X) \mathbb{I}_{\{|X| \leq \nu\}}) - \mathbb{E} g(X)| \\ & \quad + |\mathbb{E} g(X_n) - \mathbb{E} g(X)| \\ & \leq 5\varepsilon + |\mathbb{E} g(X_n) - \mathbb{E} g(X)|. \end{aligned}$$

- Weil  $X_n \xrightarrow{d} X$  und weil  $x_0, \dots, x_k$  Stetigkeitspunkte von  $F_X$  sind, gilt außerdem

$$\begin{aligned} \mathbb{E} g(X_n) &= \sum_{i=1}^k a_i (F_{X_n}(x_i) - F_{X_n}(x_{i-1})) \\ &\rightarrow \sum_{i=1}^k a_i (F_X(x_i) - F_X(x_{i-1})) \\ &= \mathbb{E} g(X). \end{aligned}$$

- Also gilt  $|\mathbb{E} \varphi(X_n) - \mathbb{E} \varphi(X)| \leq 6\varepsilon$  für jedes hinreichend große  $n$ .
- Weil  $\varepsilon > 0$  beliebig klein gewählt werden kann, folgt hieraus die Gültigkeit von (11) für jedes  $\varphi \in \mathcal{C}$ .

## 2. Hinlänglichkeit von (11)

- Es gelte nun (11) für jedes  $\varphi \in \mathcal{C}$ .
- Sei  $x \in \mathbb{R}$  ein Stetigkeitspunkt von  $F_X$  und sei  $\varepsilon > 0$  eine beliebige, jedoch fest vorgegebene Zahl.

- Dann gibt es eine Funktion  $\varphi \in \mathcal{C}$ , so dass für jedes  $y \in \mathbb{R}$

$$\mathbb{I}_{(-\infty, x]}(y) \leq \varphi(y) \leq \mathbb{I}_{(-\infty, x+\varepsilon]}(y)$$

- und folglich

$$F_{X_n}(x) = \mathbb{E} \mathbb{I}_{(-\infty, x]}(X_n) \leq \mathbb{E} \varphi(X_n)$$

bzw.

$$\mathbb{E} \varphi(X) \leq \mathbb{E} \mathbb{I}_{(-\infty, x+\varepsilon]}(X) = F_X(x + \varepsilon).$$

- Hieraus und aus (11) ergibt sich, dass

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq \lim_{n \rightarrow \infty} \mathbb{E} \varphi(X_n) = \mathbb{E} \varphi(X) \leq F_X(x + \varepsilon).$$

- Weil  $\varepsilon > 0$  beliebig klein gewählt werden kann und weil  $F_X$  rechtsstetig ist, ergibt dies

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x). \quad (12)$$

- Sei nun  $\varphi \in \mathcal{C}$  eine Funktion, so dass für jedes  $y \in \mathbb{R}$

$$\mathbb{I}_{(-\infty, x-\varepsilon]}(y) \leq \varphi(y) \leq \mathbb{I}_{(-\infty, x]}(y)$$

- Dann gilt

$$F_X(x - \varepsilon) \leq \mathbb{E} \varphi(X) = \lim_{n \rightarrow \infty} \mathbb{E} \varphi(X_n) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x).$$

- Weil  $x \in \mathbb{R}$  ein Stetigkeitspunkt von  $F_X$  ist, ergibt dies

$$F_X(x) = \lim_{\varepsilon \downarrow 0} F_X(x - \varepsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x).$$

- Hieraus und aus (12) folgt, dass  $F_{X_n}(x) \rightarrow F_X(x)$ . □

### Beachte

- Aus dem zweiten Teil des Beweises von Theorem 5.7 ergibt sich, dass die Gültigkeit von (11) nicht notwendig für jede Funktion aus  $\mathcal{C}$  gefordert werden muss.
- Es genügt beispielsweise, anstelle von  $\mathcal{C}$  die kleinere Klasse  $\mathcal{C}^{(k)}$  der beschränkten und  $k$ -mal gleichmäßig stetig differenzierbaren Funktionen  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  zu betrachten, wobei  $k \in \mathbb{N}$  eine beliebige, jedoch fest vorgegebene natürliche Zahl ist.
- Dies führt zu der folgenden Aussage.

**Korollar 5.3** Seien  $X, X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen. Wenn (11) für jedes  $\varphi \in \mathcal{C}^{(k)}$  gilt, dann gilt  $X_n \xrightarrow{d} X$ .

### 5.1.3 Konvergenz zusammengesetzter Abbildungen; Satz von Slutsky

Wir zeigen zunächst, dass die fast sichere Konvergenz, die Konvergenz in Wahrscheinlichkeit, die  $L^1$ -Konvergenz und die Konvergenz im quadratischen Mittel bei der Addition von Zufallsvariablen erhalten bleiben.

**Theorem 5.8** Seien  $X, X_n, Y, Y_n : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen über einunddemselben Wahrscheinlichkeitsraum. Dann gilt

1.  $X_n + Y_n \xrightarrow{\text{f.s.}} X + Y$ , falls  $X_n \xrightarrow{\text{f.s.}} X$  und  $Y_n \xrightarrow{\text{f.s.}} Y$ ,

2.  $X_n + Y_n \xrightarrow{P} X + Y$ , falls  $X_n \xrightarrow{P} X$  und  $Y_n \xrightarrow{P} Y$ ,
3.  $X_n + Y_n \xrightarrow{L^1} X + Y$ , falls  $X, X_n, Y, Y_n \in L^1$ ,  $X_n \xrightarrow{L^1} X$  und  $Y_n \xrightarrow{L^1} Y$ ,
4.  $X_n + Y_n \xrightarrow{L^2} X + Y$ , falls  $X, X_n, Y, Y_n \in L^2$ ,  $X_n \xrightarrow{L^2} X$  und  $Y_n \xrightarrow{L^2} Y$ .

**Beweis**

Zu 1: Wenn  $X_n(\omega) \rightarrow X(\omega)$  und  $Y_n(\omega) \rightarrow Y(\omega)$  für ein  $\omega \in \Omega$ , dann gilt auch  $X_n(\omega) + Y_n(\omega) \rightarrow X(\omega) + Y(\omega)$ . Hieraus folgt die erste Teilaussage.

Zu 2: Für jedes  $\varepsilon > 0$  gilt

$$\{|X_n - X| \leq \varepsilon/2\} \cap \{|Y_n - Y| \leq \varepsilon/2\} \subset \{|(X_n + Y_n) - (X + Y)| \leq \varepsilon\}$$

bzw. nach Übergang zu den Komplementen

$$\{|(X_n + Y_n) - (X + Y)| > \varepsilon\} \subset \{|X_n - X| > \varepsilon/2\} \cup \{|Y_n - Y| > \varepsilon/2\}.$$

Hieraus folgt, dass

$$P(|(X_n + Y_n) - (X + Y)| > \varepsilon) \leq P(|X_n - X| > \varepsilon/2) + P(|Y_n - Y| > \varepsilon/2)$$

und somit die Gültigkeit der zweiten Teilaussage.

Zu 3: Die dritte Teilaussage ergibt sich unmittelbar aus der Monotonie und der Linearität des Erwartungswertes (vgl. Theorem 4.4), denn es gilt

$$\begin{aligned} \mathbb{E}|(X_n + Y_n) - (X + Y)| &= \mathbb{E}|(X_n - X) + (Y_n - Y)| \\ &\leq \mathbb{E}(|X_n - X| + |Y_n - Y|) \\ &= \mathbb{E}|X_n - X| + \mathbb{E}|Y_n - Y|. \end{aligned}$$

Zu 4: Für  $p = 2$  ergibt sich aus der Minkowski-Ungleichung (4.68), dass

$$\sqrt{\mathbb{E}(((X_n + Y_n) - (X + Y))^2)} \leq \sqrt{\mathbb{E}((X_n - X)^2)} + \sqrt{\mathbb{E}((Y_n - Y)^2)}.$$

Hieraus folgt die vierte Teilaussage. □

**Beachte**

- Eine zu Theorem 5.8 analoge Aussage gilt im allgemeinen *nicht* für die Konvergenz in Verteilung, d.h., aus  $X_n \xrightarrow{d} X$  und  $Y_n \xrightarrow{d} Y$  folgt im allgemeinen nicht, dass  $X_n + Y_n \xrightarrow{d} X + Y$ .
- Unter der zusätzlichen Annahme, dass  $X$  oder  $Y$  (mit Wahrscheinlichkeit 1) konstante Größen sind, kann man jedoch zeigen, dass auch die Konvergenz in Verteilung bei der Addition von Zufallsvariablen erhalten bleibt. Gemeint ist die folgende Aussage, die in der Literatur *Satz von Slutsky* genannt wird.

**Theorem 5.9** Seien  $X, X_n, Y_n : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen über einunddemselben Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ , und sei  $c \in \mathbb{R}$ . Dann gilt  $X_n + Y_n \xrightarrow{d} X + c$ , falls  $X_n \xrightarrow{d} X$  und  $Y_n \xrightarrow{d} c$ .

**Beweis**

- Wegen Korollar 5.3 genügt es zu zeigen, dass

$$\lim_{n \rightarrow \infty} \mathbb{E} \varphi(X_n + Y_n) = \mathbb{E} \varphi(X + c) \quad (13)$$

für jede (beschränkte, gleichmäßig stetige) Funktion  $\varphi \in \mathcal{C}^{(1)}$ .

- Für jede solche Funktion  $\varphi$  und für jedes  $\varepsilon > 0$  gibt es ein  $\delta > 0$ , so dass  $|\varphi(x) - \varphi(y)| \leq \varepsilon$  für beliebige  $x, y \in \mathbb{R}$  mit  $|x - y| \leq \delta$ .
- Dies ergibt die folgende Abschätzung

$$\begin{aligned}
& |\mathbb{E} \varphi(X_n + Y_n) - \mathbb{E} \varphi(X + c)| \\
& \leq \mathbb{E} (|\varphi(X_n + Y_n) - \varphi(X_n + c)| \mathbb{1}_{\{|Y_n - c| > \delta\}}) \\
& \quad + \mathbb{E} (|\varphi(X_n + Y_n) - \varphi(X_n + c)| \mathbb{1}_{\{|Y_n - c| \leq \delta\}}) \\
& \quad + |\mathbb{E} \varphi(X_n + c) - \mathbb{E} \varphi(X + c)| \\
& \leq 2bP(|Y_n - c| > \delta) + \varepsilon + |\mathbb{E} \varphi(X_n + c) - \mathbb{E} \varphi(X + c)|,
\end{aligned}$$

wobei  $b = \sup_{x \in \mathbb{R}} |\varphi(x)| < \infty$ .

- Der erste Summand in dieser Abschätzung konvergiert gegen Null, weil mit  $Y_n \xrightarrow{d} c$  auch  $Y_n \xrightarrow{P} c$  gilt; vgl. Theorem 5.6.
- Der dritte Summand in dieser Abschätzung konvergiert gegen Null, weil  $h : \mathbb{R} \rightarrow \mathbb{R}$  mit  $h(x) = \varphi(x + c)$  eine beschränkte und stetige Funktion und weil deshalb mit  $X_n \xrightarrow{d} X$  auch

$$\mathbb{E} \varphi(X_n + c) = \mathbb{E} h(X_n) \rightarrow \mathbb{E} h(X) = \mathbb{E} \varphi(X + c).$$

- Damit ist (13) bewiesen, weil  $\varepsilon > 0$  beliebig klein gewählt werden kann. □

### Beachte

- Ähnlich wie bei der Addition von Zufallsvariablen (vgl. Theorem 5.8) bleibt die fast sichere Konvergenz und die Konvergenz in Wahrscheinlichkeit bei der Multiplikation von Zufallsvariablen erhalten.
- Die Konvergenz im quadratischen Mittel geht jedoch im allgemeinen bei der Produktbildung verloren; vgl. das folgende Theorem 5.10.

**Theorem 5.10** Seien  $X, X_n, Y, Y_n : \Omega \rightarrow \mathbb{R}$  Zufallsvariablen über einunddemselben Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ . Dann gilt

1.  $X_n Y_n \xrightarrow{\text{f.s.}} XY$ , falls  $X_n \xrightarrow{\text{f.s.}} X$  und  $Y_n \xrightarrow{\text{f.s.}} Y$ ,
2.  $X_n Y_n \xrightarrow{P} XY$ , falls  $X_n \xrightarrow{P} X$  und  $Y_n \xrightarrow{P} Y$ ,
3.  $X_n Y_n \xrightarrow{L^1} XY$ , falls  $X, X_n, Y, Y_n \in L^2$ ,  $X_n \xrightarrow{L^2} X$  und  $Y_n \xrightarrow{L^2} Y$

### Beweis

Zu 1: Wenn  $X_n(\omega) \rightarrow X(\omega)$  und  $Y_n(\omega) \rightarrow Y(\omega)$  für ein  $\omega \in \Omega$ , dann gilt auch  $X_n(\omega)Y_n(\omega) \rightarrow X(\omega)Y(\omega)$ . Hieraus folgt die erste Teilaussage.

Zu 2:

- Es gelte  $X_n \xrightarrow{P} X$  und  $Y_n \xrightarrow{P} Y$ .
- Aus Theorem 5.2 folgt dann, dass es für jede Teilfolge  $\{n_i\}$  von  $\mathbb{N}$  eine Teilfolge  $\{n_{i_j}\} \subset \{n_i\}$  gibt, so dass  $X_{n_{i_j}} \xrightarrow{\text{f.s.}} X$  und  $Y_{n_{i_j}} \xrightarrow{\text{f.s.}} Y$ .
- Aus Teilaussage 1 ergibt sich somit, dass  $X_{n_{i_j}} Y_{n_{i_j}} \xrightarrow{\text{f.s.}} XY$ .
- Die erneute Anwendung von Theorem 5.2 zeigt nun, dass  $X_n Y_n \xrightarrow{P} XY$ .

Zu 3: • Es gelte  $X, X_n, Y, Y_n \in L^2$ ,  $X_n \xrightarrow{L^2} X$  und  $Y_n \xrightarrow{L^2} Y$ .

- Aus der Monotonie bzw. Linearität des Erwartungswertes und aus der Ungleichung (4.48) von Cauchy–Schwarz ergibt sich die Abschätzung

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

- Der zweite Summand konvergiert gegen Null für  $n \rightarrow \infty$ , weil  $X_n \xrightarrow{L^2} X$  und  $Y \in L^2$ .
- Der erste Summand konvergiert gegen Null für  $n \rightarrow \infty$ , weil  $Y_n \xrightarrow{L^2} Y$  bzw. weil  $\mathbb{E}(X_n^2) \rightarrow \mathbb{E}(X^2)$  aus  $X_n \xrightarrow{L^2} X$  folgt und somit  $\sup_{n \in \mathbb{N}} \mathbb{E}(X_n^2) < \infty$  gilt.  $\square$

Die folgende Aussage wird *Satz von Slutsky* über die Erhaltung der Verteilungskonvergenz bei der Multiplikation von Zufallsvariablen genannt.

**Theorem 5.11** *Seien  $X, X_n, Y_n : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen über einunddemselben Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ , und sei  $c \in \mathbb{R}$ . Dann gilt  $X_n Y_n \xrightarrow{d} cX$ , falls  $X_n \xrightarrow{d} X$  und  $Y_n \xrightarrow{d} c$ .*

### Beweis

- Ähnlich wie im Beweis von Theorem 5.9 genügt es wegen Korollar 5.3 zu zeigen, dass

$$\lim_{n \rightarrow \infty} \mathbb{E} \varphi(X_n Y_n) = \mathbb{E} \varphi(cX) \quad (14)$$

für jede (beschränkte, gleichmäßig stetige) Funktion  $\varphi \in \mathcal{C}^{(1)}$ .

- Für jede solche Funktion  $\varphi$  und für jedes  $\varepsilon > 0$  gibt es ein  $\delta > 0$ , so dass  $|\varphi(x) - \varphi(y)| \leq \varepsilon$  für beliebige  $x, y \in \mathbb{R}$  mit  $|x - y| \leq \delta$ .
- Außerdem sei  $\nu > 0$  so gewählt, dass  $\nu$  und  $-\nu$  Stetigkeitspunkte von  $F_X$  sind und dass  $P(|X| > \nu) < \varepsilon$ .
- Wegen  $X_n \xrightarrow{d} X$  gilt dann  $P(|X_n| > \nu) < 2\varepsilon$  für jedes hinreichend große  $n$ .
- Dies ergibt die folgende Abschätzung

$$\begin{aligned}&|\mathbb{E} \varphi(X_n Y_n) - \mathbb{E} \varphi(cX)| \\ &\leq \mathbb{E} (|\varphi(X_n Y_n) - \varphi(cX_n)| \mathbb{1}_{\{|Y_n - c| > \delta/\nu\}}) \\ &\quad + \mathbb{E} (|\varphi(X_n Y_n) - \varphi(cX_n)| \mathbb{1}_{\{|Y_n - c| \leq \delta/\nu\}} \mathbb{1}_{\{|X_n| > \nu\}}) \\ &\quad + \mathbb{E} (|\varphi(X_n Y_n) - \varphi(cX_n)| \mathbb{1}_{\{|Y_n - c| \leq \delta/\nu\}} \mathbb{1}_{\{|X_n| \leq \nu\}}) \\ &\quad + |\mathbb{E} \varphi(cX_n) - \mathbb{E} \varphi(cX)| \\ &\leq 2bP(|Y_n - c| > \delta/\nu) + 2bP(|X_n| > \nu) + \varepsilon \\ &\quad + |\mathbb{E} \varphi(cX_n) - \mathbb{E} \varphi(cX)| \\ &\leq 2bP(|Y_n - c| > \delta/\nu) + (4b + 1)\varepsilon + |\mathbb{E} \varphi(cX_n) - \mathbb{E} \varphi(cX)|\end{aligned}$$

für jedes hinreichend große  $n$ , wobei  $b = \sup_{x \in \mathbb{R}} |\varphi(x)| < \infty$ .

- Der erste Summand in dieser Abschätzung konvergiert gegen Null, weil mit  $Y_n \xrightarrow{d} c$  auch  $Y_n \xrightarrow{P} c$  gilt; vgl. Theorem 5.6.



- Der dritte Summand in dieser Abschätzung konvergiert gegen Null, weil  $h : \mathbb{R} \rightarrow \mathbb{R}$  mit  $h(x) = \varphi(cx)$  eine beschränkte und stetige Funktion und weil deshalb mit  $X_n \xrightarrow{d} X$  auch

$$\mathbb{E} \varphi(cX_n) = \mathbb{E} h(X_n) \rightarrow \mathbb{E} h(X) = \mathbb{E} \varphi(cX).$$

- Damit ist (14) bewiesen, weil  $\varepsilon > 0$  beliebig klein gewählt werden kann.  $\square$

Wir zeigen nun noch, dass die fast sichere Konvergenz, die Konvergenz in Wahrscheinlichkeit und die Konvergenz in Verteilung bei der stetigen Abbildung von Zufallsvariablen erhalten bleiben. Aussagen dieses Typs werden in der Literatur *Continuous Mapping Theorem* genannt.

**Theorem 5.12** *Seien  $X, X_n : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen über einunddemselben Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ , und sei  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  eine stetige Funktion. Dann gilt*

1.  $\varphi(X_n) \xrightarrow{\text{f.s.}} \varphi(X)$ , falls  $X_n \xrightarrow{\text{f.s.}} X$ ,
2.  $\varphi(X_n) \xrightarrow{P} \varphi(X)$ , falls  $X_n \xrightarrow{P} X$ ,
3.  $\varphi(X_n) \xrightarrow{d} \varphi(X)$ , falls  $X_n \xrightarrow{d} X$ .

### Beweis

Zu 1: Wenn  $X_n(\omega) \rightarrow X(\omega)$  für ein  $\omega \in \Omega$ , dann gilt wegen der Stetigkeit von  $\varphi$  auch  $\varphi(X_n)(\omega) \rightarrow \varphi(X)(\omega)$ . Hieraus folgt die erste Teilaussage.

Zu 2:

- Es gelte  $X_n \xrightarrow{P} X$ .
- Aus Theorem 5.2 folgt dann, dass es für jede Teilfolge  $\{n_i\}$  von  $\mathbb{N}$  eine Teilfolge  $\{n_{i_j}\} \subset \{n_i\}$  gibt, so dass  $X_{n_{i_j}} \xrightarrow{\text{f.s.}} X$ .
- Hieraus und aus Teilaussage 1 folgt, dass  $\varphi(X_{n_{i_j}}) \xrightarrow{\text{f.s.}} \varphi(X)$ .
- Durch die erneute Anwendung von Theorem 5.2 ergibt sich nun, dass  $\varphi(X_n) \xrightarrow{P} \varphi(X)$ .

Zu 3:

- Sei  $\varphi' : \mathbb{R} \rightarrow \mathbb{R}$  eine beschränkte, stetige Funktion.
- Dann hat auch die Superposition  $\varphi' \circ \varphi : \mathbb{R} \rightarrow \mathbb{R}$  mit  $(\varphi' \circ \varphi)(x) = \varphi'(\varphi(x))$  diese beiden Eigenschaften.
- Wenn  $X_n \xrightarrow{d} X$ , dann ergibt sich deshalb aus Theorem 5.7, dass

$$\mathbb{E} \varphi'(\varphi(X_n)) = \mathbb{E} (\varphi' \circ \varphi)(X_n) \rightarrow \mathbb{E} (\varphi' \circ \varphi)(X) = \mathbb{E} \varphi'(\varphi(X)).$$

- Hieraus ergibt sich die Gültigkeit von  $\varphi(X_n) \xrightarrow{d} \varphi(X)$  durch die erneute Anwendung von Theorem 5.7.  $\square$

## 5.2 Gesetz der großen Zahlen

Wir diskutieren nun allgemeinere Varianten des Gesetzes der großen Zahlen, das bereits in Abschnitt 4.1.1 im Zusammenhang mit dem Beispiel des wiederholten Würfels erwähnt wurde. Dabei

- betrachten wir eine beliebige Folge von Zufallsvariablen  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  mit  $X_i \in L^1$  und mit dem gleichen Erwartungswert  $\mu = \mathbb{E} X_i$  für alle  $i = 1, 2, \dots$  und

- untersuchen die Frage, unter welchen Bedingungen und in welchem Sinne das arithmetische Mittel

$$Y_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (15)$$

gegen  $\mu$  strebt, falls  $n$  unendlich groß wird.

### 5.2.1 Schwaches Gesetz der großen Zahlen

- Zunächst geben wir eine Bedingung dafür an, dass das in (15) betrachtete arithmetische Mittel  $Y_n$  in *Wahrscheinlichkeit* gegen den Erwartungswert  $\mu$  konvergiert, falls  $n \rightarrow \infty$ .
- In der Literatur nennt man Aussagen dieses Typs *schwaches Gesetz der großen Zahlen*.

**Theorem 5.13** Sei  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  eine Folge von Zufallsvariablen mit dem gleichen Erwartungswert  $\mu = \mathbb{E} X_i$  und mit  $\mathbb{E}(X_i^2) < \infty$  für alle  $i = 1, 2, \dots$

1. Wenn

$$\lim_{n \rightarrow \infty} \text{Var } Y_n = 0, \quad (16)$$

dann gilt  $Y_n \xrightarrow{\text{P}} \mu$ .

2. Die Bedingung (16) ist insbesondere dann erfüllt, wenn die Zufallsvariablen  $X_1, X_2, \dots$  unabhängig sind mit der gleichen Varianz  $\sigma^2 = \text{Var } X_i$  für alle  $i = 1, 2, \dots$

#### Beweis

- Zu 1: • Durch wiederholte Anwendung der Minkowski-Ungleichung (4.68) für  $p = 2$  ergibt sich, dass

$$\sqrt{\mathbb{E}(Y_n^2)} \leq \frac{1}{n} \sum_{i=1}^n \sqrt{\mathbb{E}(X_i^2)},$$

- d.h.,  $\mathbb{E}(Y_n^2) < \infty$  gilt für jedes  $n = 1, 2, \dots$
- Weil außerdem  $\mathbb{E} Y_n = \mu$  für jedes  $n = 1, 2, \dots$ , ergibt sich aus der Tschebyschew-Ungleichung (4.72), dass für jedes  $\varepsilon > 0$

$$\begin{aligned} P(|Y_n - \mu| > \varepsilon) &\leq \frac{\mathbb{E}((Y_n - \mu)^2)}{\varepsilon^2} \\ &= \frac{\text{Var } Y_n}{\varepsilon^2}. \end{aligned}$$

- Hieraus und aus (16) ergibt sich, dass  $Y_n \xrightarrow{\text{P}} \mu$ .

- Zu 2: • Die Zufallsvariablen  $X_1, X_2, \dots$  seien nun unabhängig mit der gleichen Varianz  $\sigma^2 = \text{Var } X_i$  für alle  $i = 1, 2, \dots$
- Dann gilt für  $n \rightarrow \infty$

$$\text{Var } Y_n = \frac{1}{n^2} \sum_{i=1}^n \text{Var } X_i = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n} \longrightarrow 0.$$

□

### 5.2.2 Starkes Gesetz der großen Zahlen

- Wir diskutieren nun Bedingungen dafür, dass das in (15) betrachtete arithmetische Mittel  $Y_n$ , nach einer geeignet gewählten Zentrierung, *fast sicher* gegen Null konvergiert, falls  $n \rightarrow \infty$ .
- In der Literatur nennt man Aussagen dieses Typs *starkes Gesetz der großen Zahlen*.
- Eines der ersten Resultate in dieser Richtung ist das folgende starke Gesetz der großen Zahlen, das auf Cantelli zurückgeht.

**Theorem 5.14** Sei  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  eine Folge von unabhängigen Zufallsvariablen mit  $\mathbb{E}(X_i^4) < \infty$  für alle  $i = 1, 2, \dots$ . Wenn es eine Konstante  $c < \infty$  gibt, so dass

$$\mathbb{E}((X_i - \mathbb{E} X_i)^4) \leq c \quad (17)$$

für jedes  $i \in \mathbb{N}$ , dann gilt für  $n \rightarrow \infty$

$$Y_n - \mathbb{E} Y_n \xrightarrow{\text{f.s.}} 0. \quad (18)$$

#### Beweis

- Ohne Einschränkung der Allgemeinheit können wir annehmen, dass  $\mathbb{E} X_i = 0$  für alle  $i \in \mathbb{N}$ .
- Wegen Korollar 5.2 genügt es dann zu zeigen, dass für jedes  $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P(|Y_n| > \varepsilon) < \infty. \quad (19)$$

- Aus der Markow–Ungleichung (4.73) für  $p = 4$  ergibt sich, dass

$$\sum_{n=1}^{\infty} \mathbb{E}(Y_n^4) < \infty \quad (20)$$

hinreichend für die Gültigkeit von (19) ist.

- Wir zeigen nun, dass unter den getroffenen Voraussetzungen die Bedingung (20) erfüllt ist.
- Es gilt

$$\begin{aligned} Y_n^4 &= \frac{1}{n^4} (X_1 + \dots + X_n)^4 \\ &= \frac{1}{n^4} \left( \sum_{i=1}^n X_i^4 + \sum_{1 \leq i < j \leq n} \frac{4!}{2!2!} X_i^2 X_j^2 + \sum_{i \neq j, i \neq k, j < k} \frac{4!}{2!1!1!} X_i^2 X_j X_k \right. \\ &\quad \left. + \sum_{i < j < k < l} 4! X_i X_j X_k X_l + \sum_{i \neq j} \frac{4!}{3!1!} X_i^3 X_j \right). \end{aligned}$$

- Weil  $X_1, X_2, \dots$  unabhängige Zufallsvariablen sind und weil  $\mathbb{E} X_i = 0$  für alle  $i \in \mathbb{N}$  vorausgesetzt wird, ergibt sich hieraus, dass

$$\begin{aligned} \mathbb{E}(Y_n^4) &= \frac{1}{n^4} \left( \sum_{i=1}^n \mathbb{E}(X_i^4) + 6 \sum_{1 \leq i < j \leq n} \mathbb{E}(X_i^2 X_j^2) \right) \\ &\leq \frac{1}{n^4} \left( nc + 6 \sum_{1 \leq i < j \leq n} \sqrt{\mathbb{E}(X_i^4)} \sqrt{\mathbb{E}(X_j^4)} \right) \\ &\leq \frac{1}{n^4} \left( nc + \frac{6n(n-1)}{2} c \right) \\ &= \frac{1}{n^4} (3n^2 - 2n) c < 3c \frac{1}{n^2}. \end{aligned}$$

- Folglich gilt

$$\sum_{n=1}^{\infty} \mathbb{E}(Y_n^4) \leq 3c \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

- D.h., die Bedingung (20) ist erfüllt.  $\square$

### Beachte

- Wenn  $X_1, X_2, \dots$  nicht nur unabhängige, sondern auch identisch verteilte Zufallsvariablen sind, dann lassen sich die in Theorem 5.14 formulierten Integrierbarkeitsbedingungen deutlich abschwächen.
- Der „Preis“ hierfür ist allerdings, dass nun ein wesentlich umfangreicherer Beweis erforderlich ist, im Vergleich zum Beweis von Theorem 5.14.
- Dies führt zu der folgenden Version des starken Gesetzes der großen Zahlen.

**Theorem 5.15** Sei  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit  $\mathbb{E}|X_i| < \infty$  für alle  $i = 1, 2, \dots$ ;  $\mu = \mathbb{E}X_i$ . Dann gilt  $Y_n \xrightarrow{\text{f.s.}} \mu$  für  $n \rightarrow \infty$ .

- Im Beweis von Theorem 5.15 benötigen wir mehrere Hilfssätze, die auch von eigenständigem Interesse sind.
- Zunächst leiten wir eine einfache untere bzw. obere Schranke für den Erwartungswert von nichtnegativen Zufallsvariablen her.

**Lemma 5.1** Sei  $X : \Omega \rightarrow [0, \infty)$  eine nichtnegative Zufallsvariable mit  $X \in L^1$ , d.h.  $0 \leq \mathbb{E}X < \infty$ . Dann gilt

$$\sum_{n=1}^{\infty} P(X \geq n) \leq \mathbb{E}X \leq 1 + \sum_{n=1}^{\infty} P(X \geq n). \quad (21)$$

**Beweis** Es gilt

$$\begin{aligned} \sum_{n=1}^{\infty} P(X \geq n) &= \sum_{n=1}^{\infty} \sum_{k \geq n} P(k \leq X < k+1) = \sum_{k=1}^{\infty} k P(k \leq X < k+1) \\ &= \sum_{k=0}^{\infty} \mathbb{E}(k \mathbb{I}_{[k, k+1)}(X)) \leq \sum_{k=0}^{\infty} \mathbb{E}(X \mathbb{I}_{[k, k+1)}(X)) = \mathbb{E}X \\ &\leq \sum_{k=0}^{\infty} \mathbb{E}((k+1) \mathbb{I}_{[k, k+1)}(X)) = \sum_{k=0}^{\infty} (k+1) P(k \leq X < k+1) \\ &= \sum_{n=1}^{\infty} P(X \geq n) + \sum_{k=0}^{\infty} P(k \leq X < k+1) = \sum_{n=1}^{\infty} P(X \geq n) + 1. \end{aligned} \quad \square$$

Ein weiteres nützliches Hilfsmittel ist die folgende *Ungleichung von Kolmogorow*.

**Lemma 5.2** Seien  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  unabhängige Zufallsvariablen mit  $\mathbb{E}(X_i^2) < \infty$  und  $\mathbb{E}X_i = 0$  für jedes  $i = 1, \dots, n$ . Für jedes  $\varepsilon > 0$  und  $n \in \mathbb{N}$  gilt dann

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon\right) \leq \frac{\mathbb{E}(S_n^2)}{\varepsilon^2}, \quad (22)$$

wobei  $S_n = X_1 + \dots + X_n$ .

**Beweis**

- Sei

$$A = \left\{ \max_{1 \leq k \leq n} |S_k| \geq \varepsilon \right\},$$

und für  $k = 1, \dots, n$  betrachten wir die paarweise disjunkten Mengen

$$A_k = \left\{ |S_i| < \varepsilon, i = 1, \dots, k-1, |S_k| \geq \varepsilon \right\}.$$

- Dann gilt  $A = \bigcup_{k=1}^n A_k$  und folglich

$$\mathbb{E}(S_n^2) \geq \mathbb{E}(S_n^2 \mathbb{1}_A) = \sum_{k=1}^n \mathbb{E}(S_n^2 \mathbb{1}_{A_k}). \quad (23)$$

- Weil  $X_1, \dots, X_n$  unabhängige Zufallsvariablen sind, sind wegen Theorem 3.17 auch  $S_k \mathbb{1}_{A_k}$  und  $X_{k+1} + \dots + X_n$  unabhängige Zufallsvariablen.
- Aus der Multiplikationsformel für den Erwartungswert des Produktes von unabhängigen Zufallsvariablen (vgl. Theorem 4.9) ergibt sich deshalb, dass

$$\begin{aligned} \mathbb{E}(S_k(X_{k+1} + \dots + X_n) \mathbb{1}_{A_k}) &= \mathbb{E}(S_k \mathbb{1}_{A_k}) \mathbb{E}(X_{k+1} + \dots + X_n) \\ &= \mathbb{E}(S_k \mathbb{1}_{A_k}) (\mathbb{E} X_{k+1} + \dots + \mathbb{E} X_n) = 0. \end{aligned}$$

- Hieraus ergibt sich, dass

$$\begin{aligned} \mathbb{E}(S_n^2 \mathbb{1}_{A_k}) &= \mathbb{E}((S_k + (X_{k+1} + \dots + X_n))^2 \mathbb{1}_{A_k}) \\ &= \mathbb{E}(S_k^2 \mathbb{1}_{A_k}) + 2\mathbb{E}(S_k(X_{k+1} + \dots + X_n) \mathbb{1}_{A_k}) + \mathbb{E}((X_{k+1} + \dots + X_n)^2 \mathbb{1}_{A_k}) \\ &\geq \mathbb{E}(S_k^2 \mathbb{1}_{A_k}). \end{aligned}$$

- Mit Hilfe von (23) ergibt dies, dass

$$\begin{aligned} \mathbb{E}(S_n^2) &\geq \sum_{k=1}^n \mathbb{E}(S_k^2 \mathbb{1}_{A_k}) \\ &\geq \varepsilon^2 \sum_{k=1}^n \mathbb{E} \mathbb{1}_{A_k} = \varepsilon^2 P\left(\bigcup_{k=1}^n A_k\right) = \varepsilon^2 P\left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon\right). \end{aligned} \quad \square$$

Außerdem benötigen wir den folgenden Hilfssatz.

**Lemma 5.3** Seien  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  unabhängige Zufallsvariablen mit  $X_n \in L^2$  und  $\mathbb{E} X_n = 0$  für alle  $n \in \mathbb{N}$ . Wenn

$$\sum_{n=1}^{\infty} \mathbb{E} X_n^2 < \infty, \quad (24)$$

dann gibt es eine Zufallsvariable  $S : \Omega \rightarrow \mathbb{R}$ , so dass  $S_n \xrightarrow{\text{f.s.}} S$ , wobei  $S_n = X_1 + \dots + X_n$ .

**Beweis**

- Es gilt  $S_n \xrightarrow{\text{f.s.}} S$  genau dann, wenn  $\{S_n\}$  mit Wahrscheinlichkeit 1 eine Cauchy-Folge ist.

- Dies gilt genau dann, wenn für jedes  $\varepsilon > 0$

$$P\left(\sup_{k \geq 1} |S_{n+k} - S_n| \geq \varepsilon\right) \rightarrow 0 \quad (25)$$

für  $n \rightarrow \infty$ ; vgl. Theorem 5.1.

- Außerdem ergibt sich aus Theorem 2.3 und aus der Ungleichung (4.22) von Kolmogorow, dass

$$\begin{aligned} P\left(\sup_{k \geq 1} |S_{n+k} - S_n| \geq \varepsilon\right) &= \lim_{m \rightarrow \infty} P\left(\max_{1 \leq k \leq m} |S_{n+k} - S_n| \geq \varepsilon\right) \\ &\leq \lim_{m \rightarrow \infty} \varepsilon^{-2} \mathbb{E}\left(\left(\sum_{k=n+1}^{n+m} X_k\right)^2\right) \\ &= \lim_{m \rightarrow \infty} \varepsilon^{-2} \sum_{k=n+1}^{n+m} \mathbb{E}(X_k^2) = \varepsilon^{-2} \sum_{k=n+1}^{\infty} \mathbb{E}(X_k^2), \end{aligned}$$

wobei sich die vorletzte Gleichheit aus der Unabhängigkeit von  $X_1, X_2, \dots$  und aus der Annahme, dass  $\mathbb{E} X_n = 0$  für alle  $n \in \mathbb{N}$ , ergibt.

- Hieraus folgt, dass die Bedingung (24) die Konvergenz (25) impliziert.  $\square$

Die nächsten beiden Hilfssätze betreffen zwei klassische Konvergenzeigenschaften von Folgen reeller Zahlen. Sie werden in der Literatur *Lemma von Toeplitz* bzw. *Lemma von Kronecker* genannt.

**Lemma 5.4** (Toeplitz) *Seien  $a_n, b, b_n \in \mathbb{R}$  reelle Zahlen mit*

$$0 < a_1 \leq a_2 \leq \dots \quad \text{und} \quad \lim_{n \rightarrow \infty} a_n = \infty. \quad (26)$$

*Wenn  $b_n \rightarrow b$ , dann gilt*

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} b_k = b. \quad (27)$$

**Beweis**

- Es gilt die Identität

$$\sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} b_k - b = \sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} (b_k - b) - \frac{a_1}{a_n} b. \quad (28)$$

- Außerdem ergibt sich aus (26), dass die Summe

$$\sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} = \frac{a_n - a_1}{a_n}$$

monoton gegen 1 konvergiert für  $n \rightarrow \infty$ .

- Es gelte  $b_n \rightarrow b$ , und für  $\varepsilon > 0$  sei  $n_0 \in \mathbb{N}$  so gewählt, dass  $|b_n - b| < \varepsilon$  für alle  $n \geq n_0$ .

- Wegen (28) gilt dann für  $n \geq n_0$

$$\begin{aligned} & \left| \sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} b_k - b \right| \\ & \leq \sum_{k=1}^{n_0-1} \frac{a_{k+1} - a_k}{a_n} |b_k - b| + \varepsilon \frac{a_n - a_{n_0}}{a_n} + \frac{a_1}{a_n} |b|. \end{aligned}$$

- Wegen  $a_n \uparrow \infty$  ergibt sich hieraus, dass für jedes hinreichend große  $n$

$$\left| \sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} b_k - b \right| \leq 2\varepsilon.$$

- Damit ist die Behauptung bewiesen, weil  $\varepsilon > 0$  beliebig klein gewählt werden kann.  $\square$

**Lemma 5.5** (Kronecker) *Seien  $a_n, c_n \in \mathbb{R}$  reelle Zahlen, so dass die Bedingung (26) erfüllt ist. Wenn die Reihe  $\sum_{k=1}^{\infty} c_k/a_k$  konvergiert, dann gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} \sum_{k=1}^n c_k = 0. \quad (29)$$

**Beweis**

- Wir setzen  $b_0 = 0$  und  $b_n = \sum_{k=1}^n c_k/a_k$ .
- Dann gilt  $c_k = a_k(b_k - b_{k-1})$  und

$$\frac{1}{a_n} \sum_{k=1}^n c_k = b_n - \sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} b_k. \quad (30)$$

- Wenn nun die Reihe  $\sum_{k=1}^{\infty} c_k/a_k$  konvergiert, d.h., falls  $b_n \rightarrow b$  für ein  $b \in \mathbb{R}$ , dann ergibt sich die Behauptung aus (30) und aus Lemma 5.4.  $\square$

**Beweis von Theorem 5.15**

- Die Behauptung gilt genau dann, wenn

$$\frac{1}{n} ((X_1 - \mathbb{E} X_1) + \dots + (X_n - \mathbb{E} X_n)) \rightarrow 0$$

mit Wahrscheinlichkeit 1.

- Ohne Einschränkung der Allgemeinheit setzen wir voraus, dass  $\mathbb{E} X_n = 0$  und zeigen, dass

$$\frac{1}{n} (X_1 + \dots + X_n) \rightarrow 0 \quad (31)$$

mit Wahrscheinlichkeit 1.

- Aus Lemma 5.1 ergibt sich, dass  $\mathbb{E} |X_1| < \infty$  genau dann, wenn  $\sum_{n=1}^{\infty} P(|X_1| \geq n) < \infty$ .
- Weil  $X_1, X_2, \dots$  identisch verteilt sind, gilt dies genau dann, wenn  $\sum_{n=1}^{\infty} P(|X_n| \geq n) < \infty$ .
- Wegen des Lemmas von Borel–Cantelli (vgl. Korollar 2.3 und Theorem 2.7) ist das gleichbedeutend mit  $P(\limsup_n \{|X_n| \geq n\}) = 0$ .
- Hieraus folgt, dass es für fast jedes  $\omega \in \Omega$  eine natürliche Zahl  $n_0 = n_0(\omega)$  gibt, so dass  $|X_n(\omega)| < n$  für jedes  $n \geq n_0$  gilt.

- Also gilt (31) genau dann, wenn

$$\frac{1}{n} (X'_1 + \dots + X'_n) \rightarrow 0 \quad (32)$$

mit Wahrscheinlichkeit 1, wobei für  $k = 1, \dots, n$

$$X'_k = \begin{cases} X_k, & \text{falls } |X_k| < k, \\ 0, & \text{falls } |X_k| \geq k. \end{cases}$$

- *Beachte:* Der Übergang von  $X_k$  zu  $X'_k$  wird *Abschneidetechnik* genannt.
- Weil

$$\mathbb{E} X'_k = \mathbb{E} (X_k \mathbb{1}_{\{|X_k| < k\}}) = \mathbb{E} (X_1 \mathbb{1}_{\{|X_1| < k\}}) \rightarrow \mathbb{E} X_1 = 0,$$

ergibt sich aus Lemma 5.4 mit  $a_k = k$ , dass

$$\frac{1}{n} \sum_{k=1}^n \mathbb{E} X'_k \rightarrow 0.$$

- Deshalb gilt (32) und damit auch (31) genau dann, wenn

$$\frac{1}{n} (Z_1 + \dots + Z_n) \rightarrow 0 \quad (33)$$

mit Wahrscheinlichkeit 1, wobei  $Z_k = X'_k - \mathbb{E} X'_k$ .

- Aus Lemma 5.5 mit  $a_k = k$  und  $c_k = Z_k(\omega)$  ergibt sich die Gültigkeit von (33), wenn gezeigt wird, dass es eine Zufallsvariable  $Z$  gibt, so dass mit Wahrscheinlichkeit 1

$$\lim_{m \rightarrow \infty} \sum_{n=1}^m \frac{Z_n}{n} = Z.$$

- Wegen Lemma 5.3 genügt es nun zu zeigen, dass  $\sum_{n=1}^{\infty} n^{-2} \mathbb{E} (Z_n^2) < \infty$ .
- Dies ergibt sich aus den folgenden Abschätzungen:

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{E} \left( \frac{Z_n^2}{n^2} \right) &\leq \sum_{n=1}^{\infty} \mathbb{E} \left( \frac{(X'_n)^2}{n^2} \right) = \sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{E} (X_n^2 \mathbb{1}_{\{|X_n| < n\}}) = \sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{E} (X_1^2 \mathbb{1}_{\{|X_1| < n\}}) \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^n \mathbb{E} (X_1^2 \mathbb{1}_{\{k-1 \leq |X_1| < k\}}) = \sum_{k=1}^{\infty} \mathbb{E} (X_1^2 \mathbb{1}_{\{k-1 \leq |X_1| < k\}}) \sum_{n=k}^{\infty} \frac{1}{n^2} \\ &\leq 2 \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E} (X_1^2 \mathbb{1}_{\{k-1 \leq |X_1| < k\}}) \\ &\leq 2 \sum_{k=1}^{\infty} \mathbb{E} (|X_1| \mathbb{1}_{\{k-1 \leq |X_1| < k\}}) \\ &= 2 \mathbb{E} |X_1| < \infty, \end{aligned}$$

wobei in der vorletzten Abschätzung die Ungleichung

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{2}{k} \quad \forall k \geq 1$$

verwendet wurde, die sich ergibt aus  $\sum_{n=1}^{\infty} n^{-2} = 1 + \sum_{n=2}^{\infty} n^{-2}$  und

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \int_{k-1}^{\infty} \frac{1}{x^2} dx = \frac{1}{k-1} \leq \frac{2}{k} \quad \forall k \geq 2.$$

- Damit ist Theorem 5.15 bewiesen. □



### 5.2.3 Anwendungsbeispiele

#### 1. Buffonsches Nadelexperiment

- Betrachten das System

$$K = \{(x, y) : (x, y) \in \{\dots, -1, 0, 1, \dots\} \times \mathbb{R}\} \subset \mathbb{R}^2$$

von parallelen und äquidistanten (vertikalen) Geraden in der euklidischen Ebene  $\mathbb{R}^2$ ; vgl. auch Abschnitt 2.5.

- Werfen eine Nadel mit der Länge 1 „willkürlich“ in die Ebene  $\mathbb{R}^2$ , wobei mit „willkürlich“ das folgende stochastische Modell gemeint ist.
- Betrachten zwei Zufallsvariablen  $S$  und  $T$ , die die zufällige Lage der Nadel beschreiben, wobei
  - $S$  der (orthogonale) Abstand des Nadelmittelpunktes zur nächsten linksliegenden Nachbargeraden von  $K$  ist,
  - $T$  der Winkel ist, den die Nadel zum Lot auf die Geraden von  $K$  bildet, und
  - die Zufallsvariablen  $S$  und  $T$  unabhängig und gleichverteilt seien auf den Intervallen  $[0, 1]$  bzw.  $[-\pi/2, \pi/2]$ .
- Bestimmen die Wahrscheinlichkeit des Ereignisses

$$A = \left\{0 < S < \frac{1}{2} \cos T\right\} \cup \left\{1 - \frac{1}{2} \cos T < S < 1\right\},$$

dass die willkürlich geworfene Nadel eine der Geraden von  $K$  schneidet.

- Es gilt

$$\begin{aligned} P(A) &= P\left(0 < S < \frac{1}{2} \cos T\right) + P\left(1 - \frac{1}{2} \cos T < S < 1\right) \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} P\left(0 < S < \frac{1}{2} \cos t\right) dt + \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} P\left(1 - \frac{1}{2} \cos t < S < 1\right) dt \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t dt + \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t dt \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos t dt = \frac{2}{\pi}. \end{aligned}$$

- Aus der Gleichung  $P(A) = 2/\pi$  ergibt sich nun eine Methode zur experimentellen Bestimmung der Zahl  $\pi$ , die auf dem Gesetz der großen Zahlen beruht.
- Seien  $(S_1, T_1), \dots, (S_n, T_n)$  unabhängige und identisch verteilte Zufallsvektoren (mit der gleichen Verteilung wie  $(S, T)$ ), die wir als das Ergebnis von  $n$  (unabhängig durchgeführten) Nadelexperimenten auffassen.
- Dann sind  $X_1, X_2, \dots, X_n$  mit

$$X_i = \begin{cases} 1, & \text{falls } S_i < \frac{1}{2} \cos T_i \text{ oder } 1 - \frac{1}{2} \cos T_i < S_i \\ 0 & \text{sonst} \end{cases}$$

unabhängige und identisch verteilte Zufallsvariablen mit dem Erwartungswert  $\mathbb{E} X_i = 2/\pi$ .

- Aus Theorem 5.15 ergibt sich also, dass das arithmetische Mittel  $Y_n = n^{-1} \sum_{i=1}^n X_i$  fast sicher gegen die Zahl  $2/\pi$  strebt.
- D.h., für große  $n$  ist  $2/Y_n$  mit hoher Wahrscheinlichkeit eine gute Näherung der Zahl  $\pi$ .

2. Computer-Algorithmus zur Bestimmung von  $\pi$ 

- Ein einfacher Algorithmus zur Monte-Carlo-Simulation der Zahl  $\pi$  hängt mit dem folgenden geometrischen Sachverhalt zusammen.
- Betrachten das Quadrat

$$B = [-1, 1] \times [-1, 1] \subset \mathbb{R}^2$$

und den Kreis

$$C = \{(x, y) : (x, y) \in B, x^2 + y^2 < 1\}.$$

- Werfen einen Punkt willkürlich in die Menge  $B$ .
- D.h., wir betrachten zwei unabhängige Zufallsvariablen  $S$  und  $T$ , die jeweils gleichverteilt auf dem Intervall  $[-1, 1]$  sind.
- Bestimmen die Wahrscheinlichkeit des Ereignisses

$$A = \{(S, T) \in C\} = \{S^2 + T^2 < 1\},$$

dass der „zufällige Punkt“  $(S, T)$  in  $C \subset B$  liegt.

- Es gilt

$$P(A) = P(S^2 + T^2 < 1) = \frac{|C|}{|B|} = \frac{\pi}{4},$$

wobei  $|B|$ ,  $|C|$  den Flächeninhalt von  $B$  bzw.  $C$  bezeichnet.

- Ähnlich wie beim Buffonschen Nadelexperiment ergibt sich nun aus der Gleichung  $P(A) = \pi/4$  eine weitere Methode zur experimentellen Bestimmung der Zahl  $\pi$ , die auf dem Gesetz der großen Zahlen beruht und die sich leicht implementieren lässt.
- Seien  $(S_1, T_1), \dots, (S_n, T_n)$  unabhängige und identisch verteilte Zufallsvektoren (mit der gleichen Verteilung wie  $(S, T)$ ), die wir als das Ergebnis von  $n$  (unabhängig durchgeführten) Experimenten auffassen.
- Dann sind  $X_1, X_2, \dots, X_n$  mit

$$X_i = \begin{cases} 1, & \text{falls } S_i^2 + T_i^2 < 1 \\ 0 & \text{sonst} \end{cases}$$

unabhängige und identisch verteilte Zufallsvariablen mit dem Erwartungswert  $\mathbb{E} X_i = \pi/4$ .

- Aus Theorem 5.15 ergibt sich also, dass das arithmetische Mittel  $Y_n = n^{-1} \sum_{i=1}^n X_i$  fast sicher gegen die Zahl  $\pi/4$  strebt.
- D.h., für große  $n$  ist  $4Y_n$  mit hoher Wahrscheinlichkeit eine gute Näherung der Zahl  $\pi$ .
- *Beachte:* Bei der Implementierung dieser Monte-Carlo-Simulation kann man wie folgt vorgehen.
  - Erzeuge  $2n$  auf  $[0, 1]$  gleichverteilte Pseudozufallszahlen  $z_1, \dots, z_{2n}$  mit einem Zufallszahlengenerator.
  - Setze  $s_i = 2z_i - 1$  und  $t_i = 2z_{n+i} - 1$  für  $i = 1, \dots, n$ .
  - Setze

$$x_i = \begin{cases} 1, & \text{falls } s_i^2 + t_i^2 < 1 \\ 0 & \text{sonst} \end{cases}$$

- Berechne  $4(x_1 + \dots + x_n)/n$ .

## 3. Numerische Berechnung von Integralen durch Monte-Carlo-Simulation

- Sei  $\varphi : [0, 1] \rightarrow [0, 1]$  eine stetige Funktion.

- Mittels Monte-Carlo-Simulation soll der Wert des Integrals  $\int_0^1 \varphi(x) dx$  bestimmt werden.
- Seien  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  unabhängige und identisch verteilte Zufallsvariable, die auf dem Intervall  $[0, 1]$  gleichverteilt sind.
- Außerdem sei  $Z_k = \varphi(X_k)$  für jedes  $k = 1, 2, \dots$
- Wegen Theorem 3.17 sind dann auch  $Z_1, Z_2, \dots$  unabhängige und identisch verteilte Zufallsvariable,
- und es gilt

$$\mathbb{E} Z_1 = \int_0^1 \varphi(x) f_{X_1}(x) dx = \int_0^1 \varphi(x) dx.$$

- Aus dem starken Gesetz der großen Zahlen (vgl. Theorem 5.15) ergibt sich nun, dass für  $n \rightarrow \infty$

$$\frac{1}{n} \sum_{k=1}^n Z_k \xrightarrow{\text{f.s.}} \int_0^1 \varphi(x) dx.$$

- D.h., für große  $n$  ist das arithmetische Mittel  $\frac{1}{n} \sum_{k=1}^n Z_k$  mit hoher Wahrscheinlichkeit eine gute Näherung des Integrals  $\int_0^1 \varphi(x) dx$ .
- *Beachte:* Bei der Implementierung dieser Monte-Carlo-Simulation kann man wie folgt vorgehen.
  - Erzeuge  $n$  auf  $[0, 1]$  gleichverteilte Pseudozufallszahlen  $x_1, \dots, x_n$  mit einem Zufallszahlengenerator.
  - Setze  $z_k = \varphi(x_k)$  für  $k = 1, \dots, n$  und berechne  $\frac{1}{n} \sum_{k=1}^n z_k$ .

#### 4. Probabilistischer Beweis des Approximationssatzes von Weierstrass

- Sei  $\varphi : [0, 1] \rightarrow [0, 1]$  eine stetige (und somit beschränkte) Funktion.
- Mit Hilfe des starken Gesetzes der großen Zahlen (vgl. Theorem 5.15) zeigen wir, dass sich die Funktion  $\varphi$  gleichmäßig durch Polynome approximieren lässt.
- Seien  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  unabhängige und identisch verteilte Zufallsvariablen mit  $X_k \sim \text{Bin}(1, p)$ , wobei  $0 \leq p \leq 1$ .
- Aus Theorem 5.15 folgt dann, dass  $Y_n \xrightarrow{\text{f.s.}} p$ , wobei  $Y_n = \frac{1}{n}(X_1 + \dots + X_n)$ .
- Wir zeigen nun, dass darüber hinaus für  $n \rightarrow \infty$

$$\mathbb{E} \varphi(Y_n) \rightarrow \varphi(p) \tag{34}$$

gleichmäßig in  $p \in [0, 1]$  gilt.

- Für  $\varepsilon > 0$  sei  $\delta > 0$  so gewählt, dass  $|\varphi(x) - \varphi(p)| < \varepsilon$ , falls  $|x - p| \leq \delta$ .
- Mit der Schreibweise  $b = \sup_{x \in [0, 1]} \varphi(x)$  gilt dann die Abschätzung

$$\begin{aligned} |\mathbb{E}(\varphi(Y_n) - \varphi(p))| &\leq \mathbb{E} |\varphi(Y_n) - \varphi(p)| \\ &= \mathbb{E} (|\varphi(Y_n) - \varphi(p)| \mathbb{I}_{\{|Y_n - p| \leq \delta\}}) + \mathbb{E} (|\varphi(Y_n) - \varphi(p)| \mathbb{I}_{\{|Y_n - p| > \delta\}}) \\ &\leq \varepsilon + 2bP(|Y_n - p| > \delta) \\ &\leq \varepsilon + 2b\varepsilon, \end{aligned}$$

wobei sich die letzte Ungleichung aus der Tschebyschew-Ungleichung (4.72) ergibt.

- Denn wegen (4.72) gilt (gleichmäßig in  $p \in [0, 1]$ )

$$P(|Y_n - p| > \delta) \leq \frac{p(1-p)}{n\delta^2} \leq \frac{1}{n\delta^2} \leq \varepsilon$$

für jedes hinreichend große  $n \in \mathbb{N}$ .

- Damit ist (34) bewiesen, weil  $\varepsilon > 0$  beliebig klein gewählt werden kann.
- Weil  $nY_n$  binomialverteilt ist, gilt

$$\mathbb{E} \varphi(Y_n) = \sum_{k=0}^n \varphi(k/n) \binom{n}{k} p^k (1-p)^{n-k}, \quad (35)$$

und diese Polynome in  $p \in [0, 1]$  konvergieren wegen (34) gleichmäßig in  $p \in [0, 1]$  gegen  $\varphi(p)$ .

- *Beachte:* Die Polynome in (35) werden *Bernstein-Polynome* genannt.

### 5. Eine Anwendung in der Zahlentheorie

- Wir betrachten den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P) = ([0, 1], \mathcal{B}([0, 1]), P)$ , wobei  $P$  die Gleichverteilung auf  $\mathcal{B}([0, 1])$  sei.
- Die Dezimalbruchentwicklung

$$\omega = 0.x_1x_2x_3 \dots$$

ist (bis auf eine abzählbare Ausnahmемenge) für fast jedes  $\omega \in [0, 1]$  eindeutig festgelegt.

- Dabei heißt  $\omega \in [0, 1]$  *normal*, wenn in der Dezimalbruchentwicklung

$$\omega = 0.x_1x_2x_3 \dots$$

jede endliche Ziffernfolge  $a = (a_1, \dots, a_k)$  mit der relativen Häufigkeit  $10^{-k}$  vorkommt.

- Wir zeigen, dass fast jede Zahl  $\omega \in [0, 1]$  normal ist,
- d.h., dass für jedes  $k \in \mathbb{N}$  und für jedes  $a \in (0, 1, \dots, 9)^k$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{a\}}(x_i, x_{i+1}, \dots, x_{i+k-1}) = 10^{-k} \quad (36)$$

für fast jedes  $\omega = 0.x_1x_2x_3 \dots$  gilt.

- Sei  $X_i(\omega) = x_i$  die  $i$ -te Ziffer in der Dezimalbruchentwicklung von  $\omega$ .
- Weil für jedes  $m \geq 1$  und jedes  $b = (b_1, \dots, b_m) \in (0, \dots, 9)^m$  die Menge

$$\{X_1 = b_1, \dots, X_m = b_m\} \subset [0, 1]$$

ein Intervall der Länge  $10^{-m}$  ist, gilt

$$P(X_1 = b_1, \dots, X_m = b_m) = 10^{-m} = P(X_1 = b_1) \dots P(X_m = b_m).$$

- Deshalb sind  $X_1, X_2, \dots$  unabhängige und identisch verteilte Zufallsvariable, deren Verteilung die (discrete) Gleichverteilung auf  $\{0, 1, \dots, 9\}$  ist.
- Aus Theorem 5.15 ergibt sich dann sofort die Gültigkeit von (36) für  $k = 1$ .
- Sei nun  $k > 1$ .
- Für  $a = (a_1, \dots, a_k)$ , für  $j \in \{1, \dots, k\}$  und für  $i \geq 0$  setzen wir

$$Z_i^{(j)}(\omega) = \mathbb{1}_{\{X_{ik+j}=a_1, \dots, X_{ik+j+k-1}=a_k\}}(\omega).$$

- Für jedes  $j \in \{1, \dots, k\}$  sind  $Z_1^{(j)}, Z_2^{(j)}, \dots$  unabhängige (wegen Theorem 3.17) und identisch verteilte Zufallsvariablen mit

$$P(Z_i^{(j)} = 1) = P(X_{ik+j} = a_1, \dots, X_{ik+j+k-1} = a_k) = 10^{-k}$$

und somit  $\mathbb{E} Z_i^{(j)} = 10^{-k}$ .

- Aus Theorem 5.15 ergibt sich also, dass

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Z_i^{(j)}(\omega) = 10^{-k} \quad (37)$$

für jedes  $\omega \in [0, 1] \setminus B_j(a)$ , wobei  $B_j(a) \in \mathcal{B}([0, 1])$  eine Ausnahmemenge ist mit

$$P(B_j(a)) = 0. \quad (38)$$

- Für jedes  $k \geq 1$  gibt es nur endlich viele  $a$  und endlich viele  $j$ .
- Weil deshalb

$$B = \bigcup_{k,a,j} B_j(a)$$

die Vereinigung von abzählbar vielen Mengen ist, ergibt sich aus (38), dass auch  $P(B) = 0$ .

- Wegen (37) gilt nun

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{a\}}(x_{ik+j}, \dots, x_{ik+j+k-1}) = 10^{-k}$$

für beliebige  $\omega = 0.x_1x_2 \dots \in B^c$ ,  $k \geq 1$ ,  $a = (a_1, \dots, a_k)$  und  $j \in \{1, \dots, k\}$ .

- Hieraus folgt, dass

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{a\}}(x_i, \dots, x_{i+k-1}) = 10^{-k}$$

für beliebige  $\omega = 0.x_1x_2 \dots \in B^c$ ,  $k \geq 1$  und  $a \in (0, 1, \dots, 9)^k$ .

## 6. Erneuerungsprozesse

- Seien  $X_1, X_2, \dots : \Omega \rightarrow (0, \infty)$  unabhängige und identisch verteilte Zufallsvariable, die nur positive Werte annehmen können;  $0 < \mu = \mathbb{E} X_1 < \infty$ .
- Die Zufallsvariable  $X_n$  kann man als Modell für die zufällige Zeitdauer deuten, die zwischen dem  $(n-1)$ -ten und  $n$ -ten Eintreten eines biologischen, ökonomischen oder technischen Systems in einen bestimmten (kritischen) Systemzustand vergeht.
- Im Englischen spricht man dann von *interoccurrence time*.
- Beispielsweise kann  $X_n$  die zufällige Zeitdauer zwischen dem  $(n-1)$ -ten und  $n$ -ten Ausfallzeitpunkt eines technischen Systems sein.
- Wenn das System unmittelbar nach jedem Ausfall vollständig repariert wird, dann kann man  $S_n = X_1 + \dots + X_n$  als den  $n$ -ten *Erneuerungszeitpunkt* des Systems auffassen.
- Für jedes  $t \geq 0$  ist  $N(t) = \sum_{n=1}^{\infty} \mathbb{I}_{\{S_n \leq t\}}$  die zufällige Anzahl von Erneuerungen im Intervall  $(0, t]$ .
- Die Familie  $\{N(t), t \geq 0\}$  von Zufallsvariablen heißt *Erneuerungszählprozess*.
- Aus Theorem 5.15 folgt, dass mit Wahrscheinlichkeit 1

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \mu^{-1}. \quad (39)$$

- Dies ergibt sich aus den folgenden Überlegungen.
- Für jedes  $t > 0$  gilt

$$P(N(t) < \infty) = 1, \quad (40)$$

denn aus Theorem 5.15 folgt, dass  $\lim_{n \rightarrow \infty} S_n = \infty$  mit Wahrscheinlichkeit 1.

- *Beachte.* Man kann (40) auch auf direktem Wege beweisen, denn es gilt

$$\begin{aligned}
 P(N(t) = \infty) &= P(S_n \leq t, \forall n \geq 1) \\
 &\leq P(X_{k(n-1)+1} + \dots + X_{kn} \leq t, \forall n \geq 1) \\
 &= \lim_{m \rightarrow \infty} \prod_{n=1}^m P(X_{k(n-1)+1} + \dots + X_{kn} \leq t) \\
 &= \lim_{m \rightarrow \infty} (P(X_1 + \dots + X_k \leq t))^m = 0,
 \end{aligned}$$

weil  $P(X_1 + \dots + X_k \leq t) < 1$  für jedes hinreichend große  $k \in \mathbb{N}$ .

- Außerdem ist  $N(t, \omega)$  für jedes  $\omega \in \Omega$  monoton nichtfallend in  $t \geq 0$ , d.h., der Grenzwert  $\lim_{t \rightarrow \infty} N(t, \omega)$  existiert für jedes  $\omega \in \Omega$ .
- Darüber hinaus gilt mit Wahrscheinlichkeit 1

$$\lim_{t \rightarrow \infty} N(t) = \infty, \quad (41)$$

weil

$$\begin{aligned}
 P\left(\lim_{t \rightarrow \infty} N(t) < \infty\right) &= \lim_{k \rightarrow \infty} P\left(\lim_{t \rightarrow \infty} N(t) < k\right) \\
 &= \lim_{k \rightarrow \infty} \lim_{t \rightarrow \infty} P(N(t) < k) \\
 &= \lim_{k \rightarrow \infty} \lim_{t \rightarrow \infty} P(S_k > t) \\
 &\leq \lim_{k \rightarrow \infty} \lim_{t \rightarrow \infty} (P(X_1 > k^{-1}t) + \dots + P(X_k > k^{-1}t)) = 0.
 \end{aligned}$$

- Aus Theorem 5.15 folgt also, dass mit Wahrscheinlichkeit 1

$$\lim_{t \rightarrow \infty} \frac{S_{N(t)}}{N(t)} = \mu. \quad (42)$$

- Außerdem gilt für beliebige  $t \geq 0$  und  $n \in \mathbb{N}$

$$\{N(t) = n\} = \{S_n \leq t < S_{n+1}\}. \quad (43)$$

- Folglich gilt  $S_{N(t)} \leq t < S_{N(t)+1}$  bzw.

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)+1} \frac{N(t)+1}{N(t)} \quad \forall t \geq X_1.$$

- Hieraus und aus (42) ergibt sich nun die Gültigkeit von (39).

### 5.3 Zentraler Grenzwertsatz

- Wir diskutieren nun eine weitere Kategorie von Grenzwertsätzen.
- Dabei wird die Summe  $X_1 + \dots + X_n$  jetzt auf andere Weise als beim Gesetz der großen Zahlen normiert.
- Während die Normierung  $1/n$  beim Gesetz der großen Zahlen zu dem *deterministischen* Grenzwert  $\mu$  führt, wird nun die (kleinere) Normierung  $1/\sqrt{n}$  betrachtet, die zu einem nichtdeterministischen, d.h. zufälligen Grenzwert führt.
- Anstelle der fast sicheren Konvergenz bzw. der Konvergenz in Wahrscheinlichkeit, die beim Gesetz der großen Zahlen betrachtet wird, erfolgt jetzt die *Konvergenz in Verteilung* gegen eine standardnormalverteilte Zufallsvariable.
- Aussagen dieses Typs heißen *zentraler Grenzwertsatz*.

### 5.3.1 Zentraler Grenzwertsatz für Summen von unabhängigen Zufallsvariablen

In Verallgemeinerung des zentralen Grenzwertsatzes von DeMoivre–Laplace, der bereits in Abschnitt 3.2.3 erwähnt wurde, leiten wir den folgenden zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen her.

**Theorem 5.16** *Sei  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit  $\mathbb{E}(X_i^2) < \infty$  und  $\text{Var } X_i > 0$  für alle  $i = 1, 2, \dots$ ;  $\mu = \mathbb{E} X_i$ ,  $\sigma^2 = \text{Var } X_i$ . Dann gilt für jedes  $x \in \mathbb{R}$*

$$\lim_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), \quad (44)$$

wobei  $\Phi : \mathbb{R} \rightarrow [0, 1]$  die Verteilungsfunktion der Standardnormalverteilung ist.

Im *Beweis* von Theorem 5.16 benutzen wir eine Approximationsmethode, die auf G. Kersting zurückgeht. Dabei benötigen wir mehrere Hilfssätze, die auch von eigenständigem Interesse sind.

Zunächst betrachten wir eine Reihe von analytischen Eigenschaften der Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung, die gegeben ist durch

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{v^2}{2}\right) dv \quad \forall x \in \mathbb{R}. \quad (45)$$

#### Lemma 5.6

1. Die Funktion  $\Phi : \mathbb{R} \rightarrow [0, 1]$  ist unendlich oft differenzierbar.
2. Sämtliche Ableitungen  $\Phi^{(n)}$  von  $\Phi$  sind beschränkte Funktionen ( $n \in \mathbb{N}$ ), und es gilt

$$\sup_{x \in \mathbb{R}} \Phi^{(1)}(x) = \Phi^{(1)}(0) = \frac{1}{\sqrt{2\pi}} \leq 1 \quad (46)$$

und

$$\sup_{x \in \mathbb{R}} |x \Phi^{(1)}(x)| < \infty, \quad \sup_{x \in \mathbb{R}} |x^2 \Phi^{(2)}(x)| < \infty. \quad (47)$$

3. Außerdem gilt die Identität

$$\Phi^{(2)}(x) = -x \Phi^{(1)}(x) \quad \forall x \in \mathbb{R}. \quad (48)$$

Der *Beweis* von Lemma 5.6 ergibt sich unmittelbar aus der Definitionsgleichung (45) der Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung.

Der nächste Hilfssatz enthält eine nützliche (gleichmäßige) Stetigkeitseigenschaft von  $\Phi$ .

**Lemma 5.7** *Sei  $X \in L^2$  eine beliebige Zufallsvariable mit*

$$\mathbb{E} X = 0, \quad \text{Var } X = 1 \quad (49)$$

und

$$P(|X| \leq c) = 1 \quad (50)$$

für eine Konstante  $c < \infty$ . Mit der Schreibweise  $\alpha_n = \sqrt{(n+1)/n}$  und  $\beta_n = 1/\sqrt{n}$  gilt dann für  $n \rightarrow \infty$

$$\sup_{x \in \mathbb{R}} |\mathbb{E}(\Phi(\alpha_n x - \beta_n X) - \Phi(x))| = O\left(\frac{1}{n^{3/2}}\right). \quad (51)$$

**Beweis**

- Für  $x, y \in \mathbb{R}$  mit  $|y| \leq c$  entwickeln wir die Verteilungsfunktion  $\Phi$  im Punkt  $\alpha_n x$  in eine Taylor-Reihe und erhalten

$$\Phi(\alpha_n x - \beta_n y) = \Phi(\alpha_n x) - \beta_n y \Phi^{(1)}(\alpha_n x) + \frac{1}{2}(\beta_n y)^2 \Phi^{(2)}(\alpha_n x) - \frac{1}{6}(\beta_n y)^3 \Phi^{(3)}(\alpha_n x - \theta \beta_n y) \quad (52)$$

mit  $|\theta| \leq 1$ .

- Weil  $\Phi^{(3)}$  beschränkt ist und  $|y| \leq c$ , hat der letzte Summand auf der rechten Seite von (52) die Größenordnung  $O(n^{-3/2})$ .
- Durch Taylorentwicklung der Funktion  $f(x) = \sqrt{x}$  im Punkt  $x_0 = 1$  ergibt sich, dass

$$\alpha_n = 1 + \frac{1}{2n} + O(n^{-2}).$$

- Deshalb ergibt die Taylorentwicklung von  $\Phi$  im Punkt  $x$  den Ausdruck

$$\begin{aligned} \Phi(\alpha_n x) &= \Phi(x) + \left( \frac{1}{2n} + O(n^{-2}) \right) x \Phi^{(1)}(x) \\ &\quad + \left( \frac{1}{2n} + O(n^{-2}) \right)^2 \frac{x^2}{2} \Phi^{(2)}\left(x + \theta x \left( \frac{1}{2n} + O(n^{-2}) \right)\right) \end{aligned}$$

mit  $|\theta| \leq 1$ .

- Wegen (47) ergibt sich hieraus, dass

$$\Phi(\alpha_n x) = \Phi(x) + \frac{x}{2n} \Phi^{(1)}(x) + O(n^{-2}). \quad (53)$$

- Auf ähnliche Weise ergibt sich, dass

$$\Phi^{(2)}(\alpha_n x) = \Phi^{(2)}(x) + \frac{x}{2n} \Phi^{(3)}(x) + O(n^{-2}), \quad (54)$$

wobei

$$\sup_{x \in \mathbb{R}} x \Phi^{(3)}(x) < \infty.$$

- Wenn nun (53) und (54) in die Taylor-Reihe (52) eingesetzt werden, ergibt sich der Ausdruck

$$\Phi(\alpha_n x - \beta_n y) - \Phi(x) = \frac{x}{2n} \Phi^{(1)}(x) - \beta_n y \Phi^{(1)}(\alpha_n x) + \frac{y^2}{2n} \Phi^{(2)}(x) + O(n^{-3/2}).$$

- Weil  $\mathbb{E} X = 0$  und  $\text{Var } X = 1$  vorausgesetzt wird (vgl. (49)), folgt hieraus, dass

$$\mathbb{E} \Phi(\alpha_n x - \beta_n X) - \Phi(x) = \frac{x}{2n} \Phi^{(1)}(x) + \frac{1}{2n} \Phi^{(2)}(x) + O(n^{-3/2}).$$

- Wegen der Identität

$$\frac{x}{2n} \Phi^{(1)}(x) + \frac{1}{2n} \Phi^{(2)}(x) = 0$$

für jedes  $x \in \mathbb{R}$  (vgl. (48)), ergibt dies die Behauptung.  $\square$

**Definition** Für beliebige Verteilungsfunktionen  $F, G : \mathbb{R} \rightarrow [0, 1]$  sei

$$d(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|. \quad (55)$$

Die in (55) gegebene Abstandsfunktion heißt *Supremum-Metrik*.



**Beachte** Anstelle  $d(F, G)$  schreiben wir manchmal auch  $d(X, Y)$  oder  $d(X, G)$ , falls  $X$  und  $Y$  Zufallsvariablen mit den Verteilungsfunktionen  $F$  bzw.  $G$  sind.

**Lemma 5.8** Seien  $X, Y : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen mit  $Y \in L^2$ , und sei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung  $N(0, 1)$ . Wenn  $\mathbb{E}(Y^2) < c$ , dann gilt

$$d(X + Y, \Phi) \leq d(X, \Phi) + 2c^{1/3}. \quad (56)$$

**Beweis**

- Für die erste Ableitung  $\Phi^{(1)}$  von  $\Phi$  gilt  $\Phi^{(1)}(x) \leq 1$  für jedes  $x \in \mathbb{R}$ ; vgl. (46).
- Hieraus folgt, dass für beliebige  $x \in \mathbb{R}$  und  $\delta > 0$

$$|\Phi(x \pm \delta) - \Phi(x)| \leq \delta. \quad (57)$$

- Außerdem gilt

$$\{X \leq x - \delta\} \subset \{Y > \delta\} \cup \{X + Y \leq x\}$$

und

$$\{X + Y \leq x\} \subset \{X \leq x + \delta\} \cup \{Y < -\delta\}.$$

- Also gilt

$$P(X \leq x - \delta) - P(Y > \delta) \leq P(X + Y \leq x) \leq P(X \leq x + \delta) + P(Y < -\delta).$$

- Wegen (57) ergibt sich nun hieraus, dass

$$\begin{aligned} P(X \leq x - \delta) - \Phi(x - \delta) - \delta - P(Y > \delta) &\leq P(X + Y \leq x) - \Phi(x) \\ &\leq P(X \leq x + \delta) - \Phi(x + \delta) + \delta + P(Y < -\delta). \end{aligned}$$

- Wenn die erste Ungleichung mit  $-1$  multipliziert wird, dann impliziert dies, dass

$$\Phi(x) - P(X + Y \leq x) \leq |\Phi(x - \delta) - P(X \leq x - \delta)| + \delta + P(Y > \delta)$$

und

$$P(X + Y \leq x) - \Phi(x) \leq |P(X \leq x + \delta) - \Phi(x + \delta)| + \delta + P(Y < -\delta).$$

- Hieraus folgt, dass

$$d(X + Y, \Phi) \leq d(X, \Phi) + \delta + P(|Y| > \delta), \quad (58)$$

weil

$$\max\{P(Y > \delta), P(Y < -\delta)\} \leq P(Y > \delta) + P(Y < -\delta) = P(|Y| > \delta).$$

- Genauso wie im Beweis der Tschebyschew–Ungleichung (4.72) kann man zeigen, dass für jedes  $\delta > 0$

$$\delta^2 P(|Y| > \delta) \leq \mathbb{E}(Y^2).$$

- Für  $\delta = c^{1/3}$  ergibt sich hieraus und aus  $\mathbb{E}(Y^2) < c$ , dass  $P(|Y| > c^{1/3}) \leq c^{1/3}$ .
- Dies und (58) liefert nun die Behauptung. □

In dem folgenden Hilfssatz leiten wir Bedingungen her, unter denen die Summe von unabhängigen (nichtnotwendig identisch verteilten) diskreten Zufallsvariablen näherungsweise normalverteilt ist.

**Lemma 5.9** Seien  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  unabhängige diskrete Zufallsvariable, die jeweils nur endlich viele verschiedene Werte mit positiver Wahrscheinlichkeit annehmen. Wenn

$$\mathbb{E} X_n = 0, \quad \text{Var } X_n = 1 \quad \forall n \in \mathbb{N} \quad (59)$$

und falls es eine Konstante  $c < \infty$  gibt, so dass

$$|X_n| \leq c \quad \forall n \in \mathbb{N}, \quad (60)$$

dann gilt

$$d(Y_n^*, \Phi) \rightarrow 0 \quad (61)$$

für  $n \rightarrow \infty$ , wobei  $Y_n^* = (X_1 + \dots + X_n)/\sqrt{n}$ .

**Beweis**

- Sei  $Y$  eine  $N(0, 1)$ -verteilte Zufallsvariable, die unabhängig von der Folge  $X_1, X_2, \dots$  ist.
- Für eine beliebige, jedoch fest vorgegebene natürliche Zahl  $n_0$  und für jedes  $n \geq n_0$  betrachten wir die Zufallsvariable

$$Z_n = \sqrt{\frac{n_0}{n}} Y + \frac{1}{\sqrt{n}} (S_n - S_{n_0}),$$

wobei  $S_n = X_1 + \dots + X_n$ .

- Dann gilt

$$Z_{n+1} = \sqrt{\frac{n}{n+1}} Z_n + \frac{1}{\sqrt{n+1}} X_{n+1}. \quad (62)$$

- Die Zufallsvariable  $X_{n+1}$  möge die Werte  $x_1, \dots, x_m$  mit den Wahrscheinlichkeiten  $p_1, \dots, p_m > 0$  annehmen;  $p_1 + \dots + p_m = 1$ .
- Weil die beiden Summanden auf der rechten Seite von (62) unabhängig sind, lässt sich die Verteilungsfunktion  $F_{n+1}(x) = P(Z_{n+1} \leq x)$  von  $Z_{n+1}$  wie folgt durch  $F_n(x) = P(Z_n \leq x)$  ausdrücken:

$$\begin{aligned} F_{n+1}(x) &= P\left(\sqrt{\frac{n}{n+1}} Z_n \leq x - \frac{1}{\sqrt{n+1}} X_{n+1}\right) \\ &= \sum_{i=1}^m P(X_{n+1} = x_i) P\left(\sqrt{\frac{n}{n+1}} Z_n \leq x - \frac{1}{\sqrt{n+1}} X_{n+1} \mid X_{n+1} = x_i\right) \\ &= \sum_{i=1}^m p_i P\left(\sqrt{\frac{n}{n+1}} Z_n \leq x - \frac{1}{\sqrt{n+1}} x_i\right) \\ &= \sum_{i=1}^m p_i F_n\left(\sqrt{\frac{n+1}{n}} x - \frac{1}{\sqrt{n}} x_i\right) \\ &= \mathbb{E} \left( F_n\left(\sqrt{\frac{n+1}{n}} x - \frac{1}{\sqrt{n}} X_{n+1}\right) \right). \end{aligned}$$

- Mit der Schreibweise  $\alpha_n = \sqrt{(n+1)/n}$  und  $\beta_n = 1/\sqrt{n}$  ergibt sich hieraus, dass

$$F_{n+1}(x) - \Phi(x) = \mathbb{E} (F_n(\alpha_n x - \beta_n X_{n+1}) - \Phi(\alpha_n x - \beta_n X_{n+1})) + \mathbb{E} (\Phi(\alpha_n x - \beta_n X_{n+1}) - \Phi(x)).$$

- Also gilt

$$d(F_{n+1}, \Phi) \leq d(F_n, \Phi) + \sup_{x \in \mathbb{R}} |\mathbb{E} (\Phi(\alpha_n x - \beta_n X_{n+1}) - \Phi(x))|,$$

weil

$$\begin{aligned} &\sup_{x \in \mathbb{R}} |\mathbb{E} (F_n(\alpha_n x - \beta_n X_{n+1}) - \Phi(\alpha_n x - \beta_n X_{n+1}))| \\ &\leq \mathbb{E} \sup_{x \in \mathbb{R}} |F_n(\alpha_n x - \beta_n X_{n+1}) - \Phi(\alpha_n x - \beta_n X_{n+1})| \\ &= d(F_n, \Phi). \end{aligned}$$

- Wegen Lemma 5.7 gibt es deshalb eine Konstante  $c' < \infty$ , so dass für jedes  $n_0 \in \mathbb{N}$  und für jedes  $n \geq n_0$

$$\begin{aligned} d(F_n, \Phi) &\leq d(F_{n-1}, \Phi) + c'(n-1)^{-3/2} \\ &\vdots \\ &\leq d(F_{n_0}, \Phi) + c' \sum_{i=n_0-1}^{n-1} i^{-3/2}. \end{aligned}$$

- Weil  $d(F_{n_0}, \Phi) = 0$ , ergibt sich hieraus, dass

$$d(Z_n, \Phi) = d(F_n, \Phi) \leq c' \sum_{i=n_0}^{\infty} i^{-3/2}.$$

- Weil die Reihe in der letzten Abschätzung gegen einen endlichen Grenzwert konvergiert, gibt es für jedes  $\varepsilon > 0$  ein  $n_0 \in \mathbb{N}$ , so dass für jedes  $n \geq n_0$

$$d(Z_n, \Phi) \leq \frac{\varepsilon}{2}. \quad (63)$$

- Außerdem ergibt sich aus der Rekursionsformel (62), dass für jedes  $n_0 \in \mathbb{N}$

$$Y_n^* - Z_n = \frac{1}{\sqrt{n}} S_{n_0} - \sqrt{\frac{n_0}{n}} Y$$

und dass somit

$$\lim_{n \rightarrow \infty} \mathbb{E}((Y_n^* - Z_n)^2) = \lim_{n \rightarrow \infty} \mathbb{E}\left(\left(\sqrt{\frac{n_0}{n}} Y - \frac{1}{\sqrt{n}} S_{n_0}\right)^2\right) = 0.$$

- Hieraus und aus (63) ergibt sich wegen Lemma 5.8, dass für jedes hinreichend große  $n$

$$d(Y_n^*, \Phi) \leq \varepsilon.$$

- Weil  $\varepsilon > 0$  beliebig klein gewählt werden kann, ist damit die Behauptung bewiesen.  $\square$

### Beweis von Theorem 5.16

- Für jedes  $m \in \mathbb{N}$  setzen wir

$$\varphi_m(x) = \begin{cases} \frac{k}{m}, & \text{falls } \frac{k}{m} \leq x < \frac{k+1}{m} \text{ mit } -m^2 \leq k \leq m^2, \\ 0 & \text{sonst.} \end{cases}$$

- Außerdem sei

$$Z_{m,i} = \varphi_m(X_i), \quad \mu_m = \mathbb{E} Z_{m,i}, \quad \sigma_m^2 = \mathbb{E}((Z_{m,i} - \mu_m)^2).$$

- Dann gilt für  $m \rightarrow \infty$

$$\mu_m \rightarrow \mu = \mathbb{E} X_1, \quad \sigma_m^2 \rightarrow \sigma^2, \quad \mathbb{E}((Z_{m,i} - X_i)^2) \rightarrow 0. \quad (64)$$

- Für beliebige  $i, m \in \mathbb{N}$  betrachten wir nun die Zufallsvariable

$$X_{m,i} = \frac{Z_{m,i} - \mu_m}{\sigma_m}.$$

- Weil dann die Folge  $X_{m,1}, X_{m,2}, \dots$  den Bedingungen von Lemma 5.9 für jedes beliebige, jedoch fest vorgegebene  $m \in \mathbb{N}$  genügt, gilt also

$$\lim_{n \rightarrow \infty} d(Y_{m,n}^*, \Phi) = 0, \quad (65)$$

wobei

$$Y_{m,n}^* = \frac{1}{\sqrt{n}} (X_{m,1} + \dots + X_{m,n}).$$

- Wir betrachten nun die Differenz

$$Y_n^* - Y_{m,n}^* = \left( \frac{\sigma_m}{\sigma} - 1 \right) Y_{m,n}^* + \frac{1}{\sigma \sqrt{n}} \sum_{i=1}^n ((X_i - \mu) - (Z_{m,i} - \mu_m)) \quad (66)$$

und schätzen ihr zweites Moment nach oben ab, wobei

$$Y_n^* = \frac{1}{\sigma \sqrt{n}} (X_1 + \dots + X_n - n\mu).$$

- Wenn wir die beiden Summanden auf der rechten Seite der Identität (66) mit  $U_{m,n}$  bzw.  $V_{m,n}$  bezeichnen, dann erhalten wir

$$\mathbb{E}(U_{m,n}^2) = \left( \frac{\sigma_m}{\sigma} - 1 \right)^2$$

und

$$\mathbb{E}(V_{m,n}^2) = \frac{1}{\sigma^2} \text{Var}(X_1 - Z_{m,1}) \leq \frac{1}{\sigma^2} \mathbb{E}((X_1 - Z_{m,1})^2),$$

weil die Summanden von  $V_{m,n}$  unabhängig und identisch verteilt sind und weil ihr Erwartungswert gleich Null ist.

- Aus (64) und aus der Minkowski-Ungleichung (4.68) ergibt sich nun, dass

$$\lim_{m \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E}((Y_n^* - Y_{m,n}^*)^2) = 0.$$

- Hieraus und aus (65) ergibt sich mit Hilfe von Lemma 5.8, dass

$$\lim_{n \rightarrow \infty} d(Y_n^*, \Phi) = 0. \quad \square$$

**Korollar 5.4** *Unter den Voraussetzungen von Theorem 5.16 gilt*

$$\lim_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma \sqrt{n}} < x\right) = \Phi(x) \quad (67)$$

für jedes  $x \in \mathbb{R}$ , und

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{X_1 + \dots + X_n - n\mu}{\sigma \sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a) \quad (68)$$

für beliebige  $a, b \in \mathbb{R}$  mit  $a \leq b$ .

**Beweis**

- Die Behauptung (67) ergibt sich aus Theorem 5.16, weil

$$\limsup_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma \sqrt{n}} < x\right) \leq \limsup_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma \sqrt{n}} \leq x\right) = \Phi(x),$$

weil für jedes  $h > 0$

$$\Phi(x - h) = \liminf_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma \sqrt{n}} \leq x - h\right) \leq \liminf_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma \sqrt{n}} < x\right)$$

und weil die Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung stetig ist.

- Die Behauptung (68) ergibt sich nun aus Theorem 5.16 und aus (67), denn es gilt

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P\left(a \leq \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \\
&= \lim_{n \rightarrow \infty} \left(P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) - P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} < a\right)\right) \\
&= \lim_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) - \lim_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} < a\right) \\
&= \Phi(b) - \Phi(a).
\end{aligned}$$

□

### 5.3.2 Anwendungsbeispiele

Wir diskutieren nun eine Reihe von Anwendungsbeispielen des zentralen Grenzwertsatzes, der in Theorem 5.16 bzw. Korollar 5.4 gegeben ist.

#### 1. maximaler Gewinn bei $n$ -maligem Münzwurf

- Seien  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  unabhängige und identisch verteilte Zufallsvariablen mit

$$P(X_1 = 1) = P(X_1 = -1) = \frac{1}{2}.$$

- Wir deuten  $X_i$  als den zufälligen Gewinn beim  $i$ -ten Münzwurf und betrachten die maximale (kumulative) Gewinnsumme

$$M_n = \max_{j=1, \dots, n} S_j,$$

die beim  $n$ -maligen Werfen einer Münze erzielt wird, wobei  $S_j = \sum_{i=1}^j X_i$ .

- Es gilt

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n}{\sqrt{n}} \leq x\right) = \sqrt{\frac{2}{\pi}} \int_0^x e^{-y^2/2} dy \quad \forall x \geq 0. \quad (69)$$

- D.h., die normierten Maxima  $M_n/\sqrt{n}$  streben in Verteilung gegen die sogenannte *asymmetrische (Standard-) Normalverteilung*.
- Dies kann man sich wie folgt überlegen: Für jedes  $m \in \mathbb{N}$  gilt

$$\begin{aligned}
P(M_n \geq m) &= P(M_n \geq m, S_n < m) + P(M_n \geq m, S_n = m) + P(M_n \geq m, S_n > m) \\
&= P(M_n \geq m, S_n < m) + P(S_n = m) + P(S_n > m).
\end{aligned}$$

- Außerdem gilt  $P(M_n \geq m, S_n > m) = P(M_n \geq m, S_n < m)$ , denn

$$\begin{aligned}
P(M_n \geq m, S_n > m) &= \sum_{j=1}^{n-1} P(S_1 < m, \dots, S_{j-1} < m, S_j = m, S_n > m) \\
&= \sum_{j=1}^{n-1} P(S_1 < m, \dots, S_{j-1} < m, S_j = m, S_n - S_j > 0) \\
&= \sum_{j=1}^{n-1} P(S_1 < m, \dots, S_{j-1} < m, S_j = m) P(S_n - S_j > 0) \\
&= \sum_{j=1}^{n-1} P(S_1 < m, \dots, S_{j-1} < m, S_j = m) P(S_n - S_j < 0) \\
&\vdots \\
&= P(M_n \geq m, S_n < m).
\end{aligned}$$

- Also gilt für jedes  $m \in \mathbb{N}$

$$P(M_n \geq m) = 2P(S_n > m) + P(S_n = m) = 2P(S_n \geq m) - P(S_n = m).$$

- Hieraus folgt, dass für jedes  $x \geq 0$  und  $m_n = \min\{i \in \mathbb{N} : i > x\sqrt{n}\}$

$$P\left(\frac{M_n}{\sqrt{n}} > x\right) = P(M_n \geq m_n) = 2P\left(\frac{S_n}{\sqrt{n}} > x\right) - P(S_n = m_n). \quad (70)$$

- Weil  $\mu = \mathbb{E} X_1 = 0$  und  $\sigma^2 = \mathbb{E} X_1^2 = 1$ , ergibt sich aus Theorem 5.16, dass

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n}{\sqrt{n}} > x\right) = 1 - \Phi(x)$$

und dass es für jedes  $\varepsilon > 0$  ein  $n_0 \in \mathbb{N}$  gibt, so dass

$$P(S_n = m_n) \leq P\left(x < \frac{S_n}{\sqrt{n}} \leq x + \varepsilon\right) \leq 2\varepsilon$$

für jedes  $n \geq n_0$ , wobei sich die letzte Ungleichung aus (46) und durch die erneute Anwendung von Theorem 5.16 ergibt.

- Hieraus und aus (70) ergibt sich nun die Behauptung (69).

## 2. fehlerbehaftete Messungen

- So wie in dem Beispiel, das bereits in Abschnitt 4.4.3 betrachtet wurde, nehmen wir an, dass die  $n$ -te Messung einer (unbekannten) Größe  $\mu \in \mathbb{R}$  den Wert  $\mu + X_n(\omega)$  liefert für  $\omega \in \Omega$ .
- Die Messfehler  $X_1, X_2, \dots$  seien unabhängige und identisch verteilte Zufallsvariablen.
- Über die Verteilung von  $X_n$  sei lediglich bekannt, dass  $\mathbb{E} X_n = 0$  und  $\text{Var } X_n = \sigma^2$ .
- Ein Ansatz zur „Schätzung“ der unbekannten Größe  $\mu$  ist durch  $Y_n = n^{-1} \sum_{i=1}^n (\mu + X_i)$ , d.h. durch das arithmetische Mittel der zufälligen Messwerte  $\mu + X_i$  gegeben.
- Mit Hilfe von Korollar 5.4 lässt sich die Wahrscheinlichkeit  $P(|Y_n - \mu| > \varepsilon)$ , dass der Schätzfehler  $|Y_n - \mu|$  größer als  $\varepsilon$  ist, näherungsweise bestimmen.
- Und zwar gilt für große  $n$

$$\begin{aligned} P(|Y_n - \mu| > \varepsilon) &= P\left(\left|\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}\right| > \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\ &= 1 - P\left(\left|\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}\right| \leq \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\ &= 1 - P\left(-\frac{\varepsilon\sqrt{n}}{\sigma} \leq \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}} \leq \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\ &\stackrel{(68)}{\approx} 1 - \left(\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right) \\ &= 2\left(1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right). \end{aligned}$$

## 3. Summen mit einer zufälligen Anzahl von Summanden

- Mit Hilfe des Satzes von Slutsky (vgl. die Theoreme 5.9 und 5.11) zeigen wir nun, wie die Aussage von Theorem 5.16 auf den Fall von Summen mit einer zufälligen Anzahl von Summanden übertragen werden kann.
- Seien  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  unabhängige und identisch verteilte Zufallsvariablen mit  $X_n \in L^2$  und  $\text{Var } X_n > 0$ .

- Ohne Einschränkung der Allgemeinheit setzen wir voraus, dass  $\mathbb{E} X_n = 0$  und  $\text{Var } X_n = 1$ .
- Außerdem sei  $N_1, N_2, \dots : \Omega \rightarrow \mathbb{N}$  eine Folge von Zufallsvariablen, so dass  $N_1 \leq N_2 \leq \dots$  und  $N_n \rightarrow \infty$  mit Wahrscheinlichkeit 1.
- Wenn es Konstanten  $a_1, a_2, \dots > 0$  und  $c > 0$  mit  $a_1 \leq a_2 \leq \dots$  und  $a_n \rightarrow \infty$  gibt, so dass

$$\frac{N_n}{a_n} \xrightarrow{\text{P}} c \quad (71)$$

für  $n \rightarrow \infty$ , dann gilt

$$\frac{S_{N_n}}{\sqrt{N_n}} \xrightarrow{\text{d}} Z \quad \text{und} \quad \frac{S_{N_n}}{\sqrt{a_n}} \xrightarrow{\text{d}} \sqrt{c}Z, \quad (72)$$

wobei  $S_n = X_1 + \dots + X_n$  und  $Z$  eine  $N(0, 1)$ -verteilte Zufallsvariable ist.

- Die Gültigkeit von (72) kann man sich wie folgt überlegen.
- Sei  $k_n = \lfloor ca_n \rfloor$  der ganzzahlige Anteil von  $ca_n$ .
- Wegen (71) gilt dann auch

$$\frac{N_n}{k_n} \xrightarrow{\text{P}} 1 \quad \text{bzw.} \quad \frac{k_n}{N_n} \xrightarrow{\text{P}} 1 \quad (73)$$

für  $n \rightarrow \infty$ .

- Außerdem gilt

$$\frac{S_{N_n}}{\sqrt{N_n}} = \sqrt{\frac{k_n}{N_n}} \left( \frac{S_{k_n}}{\sqrt{k_n}} + \frac{S_{N_n} - S_{k_n}}{\sqrt{k_n}} \right)$$

und, wegen Theorem 5.16,

$$\frac{S_{k_n}}{\sqrt{k_n}} \xrightarrow{\text{d}} Z.$$

- Wegen (73) und wegen des Satzes von Slutsky (vgl. die Theoreme 5.9 und 5.11) genügt es nun noch zu zeigen, dass

$$\frac{S_{N_n} - S_{k_n}}{\sqrt{k_n}} \xrightarrow{\text{P}} 0 \quad (74)$$

für  $n \rightarrow \infty$ .

- Für beliebige  $\varepsilon, \delta > 0$  gilt

$$\begin{aligned} P(|S_{N_n} - S_{k_n}| > \varepsilon \sqrt{k_n}) &= P(|S_{N_n} - S_{k_n}| > \varepsilon \sqrt{k_n}, |N_n - k_n| \leq \delta k_n) \\ &\quad + P(|S_{N_n} - S_{k_n}| > \varepsilon \sqrt{k_n}, |N_n - k_n| > \delta k_n) \\ &\leq P(|S_{N_n} - S_{k_n}| > \varepsilon \sqrt{k_n}, |N_n - k_n| \leq \delta k_n) + P(|N_n - k_n| > \delta k_n). \end{aligned}$$

- Wegen (73) konvergiert der zweite Summand gegen 0 für  $n \rightarrow \infty$ .
- Der erste Summand lässt sich wie folgt weiter nach oben abschätzen:

$$\begin{aligned} P(|S_{N_n} - S_{k_n}| > \varepsilon \sqrt{k_n}, |N_n - k_n| \leq \delta k_n) &\leq P\left(\max_{k_n \leq j \leq (1+\delta)k_n} |S_j - S_{k_n}| > \varepsilon \sqrt{k_n}\right) \\ &\quad + P\left(\max_{(1-\delta)k_n \leq j \leq k_n} |S_j - S_{k_n}| > \varepsilon \sqrt{k_n}\right) \\ &\leq \frac{(1+\delta)k_n - k_n + 1}{k_n \varepsilon^2} + \frac{k_n - (1-\delta)k_n + 1}{k_n \varepsilon^2} \\ &\leq \frac{2(\delta+1)}{\varepsilon^2}, \end{aligned}$$

wobei sich die vorletzte Abschätzung aus der Ungleichung von Kolmogorow (vgl. Lemma 5.2) ergibt.

- Weil  $\delta > 0$  beliebig klein gewählt werden kann, ergibt sich hieraus die Gültigkeit von (74).

- Damit ist die erste Teilaussage in (72) bewiesen. Die zweite Teilaussage ergibt sich aus der ersten Teilaussage in (72) und aus dem Satz von Slutsky (vgl. die Theorem 5.11).

#### 4. Erneuerungsprozesse

- Seien  $X_1, X_2, \dots : \Omega \rightarrow (0, \infty)$  unabhängige und identisch verteilte Zufallsvariable, die nur positive Werte annehmen können;  $0 < \mu = \mathbb{E} X_1 < \infty$ ,  $0 < \text{Var } X_1 < \infty$ .
- Dann kann man  $S_n = X_1 + \dots + X_n$  als den  $n$ -ten *Erneuerungszeitpunkt* eines (biologischen, ökonomischen, technischen) Systems auffassen; vgl. Beispiel 6 in Abschnitt 5.2.3.
- Wir betrachten die zufällige Anzahl  $N(t) = \sum_{n=1}^{\infty} \mathbb{1}_{\{S_n \leq t\}}$  von Erneuerungen im Intervall  $(0, t]$ ;  $t \geq 0$ .
- Für den Erneuerungszählprozess  $\{N(t), t \geq 0\}$  kann man nun den folgenden zentralen Grenzwertsatz beweisen.
- Für jedes  $x \in \mathbb{R}$  gilt

$$\lim_{t \rightarrow \infty} P\left(\frac{N(t) - t\mu^{-1}}{\sqrt{ct}} \leq x\right) = \Phi(x), \quad (75)$$

wobei  $c = \mu^{-3} \text{Var } X_1$  und  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung ist.

- Die Gültigkeit von (75) ergibt sich aus den folgenden Überlegungen.
- Zunächst ergibt sich unmittelbar aus dem Beweis von Theorem 5.16, dass die Konvergenz

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sqrt{n \text{Var } X_1}} \leq x\right) = \Phi(x) \quad (76)$$

gleichmäßig in  $x \in \mathbb{R}$  erfolgt.

- Mit der Schreibweise  $m(t) = \lfloor x\sqrt{ct} + t\mu^{-1} \rfloor$  gilt außerdem

$$\begin{aligned} P\left(\frac{N(t) - t\mu^{-1}}{\sqrt{ct}} \leq x\right) &= P(N(t) \leq x\sqrt{ct} + t\mu^{-1}) \\ &= P(S_{m(t)+1} > t) \\ &= P\left(\frac{S_{m(t)+1} - \mu(m(t)+1)}{\sqrt{(m(t)+1)\text{Var } X_1}} > \frac{t - \mu(m(t)+1)}{\sqrt{(m(t)+1)\text{Var } X_1}}\right). \end{aligned}$$

- Also gilt

$$\begin{aligned} &\left| P\left(\frac{N(t) - t\mu^{-1}}{\sqrt{ct}} \leq x\right) - \Phi(x) \right| \\ &\leq \left| P\left(\frac{S_{m(t)+1} - \mu(m(t)+1)}{\sqrt{(m(t)+1)\text{Var } X_1}} > \frac{t - \mu(m(t)+1)}{\sqrt{(m(t)+1)\text{Var } X_1}}\right) - \left(1 - \Phi\left(\frac{t - \mu(m(t)+1)}{\sqrt{(m(t)+1)\text{Var } X_1}}\right)\right) \right| \\ &\quad + \left| \left(1 - \Phi\left(\frac{t - \mu(m(t)+1)}{\sqrt{(m(t)+1)\text{Var } X_1}}\right)\right) - \Phi(x) \right|. \end{aligned}$$

- Außerdem gilt für jedes  $x \in \mathbb{R}$

$$\lim_{t \rightarrow \infty} m(t) = \lim_{t \rightarrow \infty} \lfloor x\sqrt{ct} + t\mu^{-1} \rfloor = \infty.$$

- Wegen (76) und wegen der Symmetrieeigenschaft

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

der Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung genügt es somit noch zu zeigen, dass

$$\lim_{t \rightarrow \infty} \frac{t - \mu m(t)}{\sqrt{m(t)\text{Var } X_1}} = -x. \quad (77)$$



- Weil  $m(t) = x\sqrt{ct} + t\mu^{-1} + \varepsilon(t)$  mit  $0 \leq |\varepsilon(t)| < 1$  für jedes  $t \geq 0$  gilt

$$\begin{aligned} \frac{t - \mu m(t)}{\sqrt{m(t)\text{Var } X_1}} &= \frac{t - \mu x\sqrt{ct} - t - \mu\varepsilon(t)}{\sqrt{m(t)\text{Var } X_1}} \\ &= -x \frac{\sqrt{\mu^{-1}\text{Var } X_1 t}}{\sqrt{\mu^{-1}\text{Var } X_1 t + x\text{Var } X_1\sqrt{ct} + \text{Var } X_1\varepsilon(t)}} - \frac{\mu\varepsilon(t)}{\sqrt{m(t)\text{Var } X_1}}, \end{aligned}$$

wobei der erste Summand gegen  $-x$  und der zweite Summand gegen 0 strebt für  $t \rightarrow \infty$ .

- Damit ist die Behauptung (75) bewiesen.

### 5.3.3 Charakteristische Funktionen

Charakteristische Funktionen sind ein wichtiges analytisches Hilfsmittel in der Stochastik, insbesondere bei der Herleitung des zentralen Grenzwertsatzes für Summen von unabhängigen, jedoch nichtnotwendig identisch verteilten Zufallsvariablen; vgl. Abschnitt 5.3.4.

- In diesem Zusammenhang betrachten wir Funktionen, die Werte in der Menge  $\mathbb{C}$  der komplexen Zahlen annehmen.
- Für jedes  $z = x + iy \in \mathbb{C}$  bezeichne dabei  $\mathcal{R}z = x$  bzw.  $\mathcal{I}z = y$  den Realteil bzw. den Imaginärteil von  $z$ .
- Außerdem sei  $\bar{z} = x - iy \in \mathbb{C}$  die zu  $z = x + iy \in \mathbb{C}$  konjugiert komplexe Zahl.
- Es gilt  $\bar{\bar{z}} = z$  genau dann, wenn  $\mathcal{I}z = 0$ , d.h., wenn  $z \in \mathbb{R}$ .
- Der Betrag  $|z|$  von  $z \in \mathbb{C}$  ist gegeben durch

$$|z| = \sqrt{x^2 + y^2},$$

d.h.  $|z|^2 = z\bar{z}$ .

- Eine wichtige Rolle spielt die komplexe Exponentialfunktion  $e^{it}$ , die gegeben ist durch  $e^{it} = \cos t + i \sin t$  für jedes  $t \in \mathbb{R}$ .
- Dabei gilt  $e^{-it} = \overline{e^{it}}$  und  $|e^{it}| = 1$  für jedes  $t \in \mathbb{R}$  bzw.  $e^{z+z'} = e^z e^{z'}$  für beliebige  $z, z' \in \mathbb{C}$ .
- Außerdem gilt die Taylor-Reihenentwicklung

$$e^{it} = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \quad \forall t \in \mathbb{R}. \quad (78)$$

**Definition** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable. Die charakteristische Funktion  $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$  von  $X$  ist dann gegeben durch

$$\varphi_X(t) = \mathbb{E} e^{itX} = \int_{\mathbb{R}} e^{itx} dF_X(x), \quad (79)$$

wobei das Integral als Lebesgue–Stieltjes–Integral aufgefasst wird.

#### Beachte

- Der Erwartungswert  $\mathbb{E} e^{itX}$  der Zufallsvariablen  $e^{itX} : \Omega \rightarrow \mathbb{C}$  in (79) wird jeweils separat für den Realteil bzw. den Imaginärteil gebildet, d.h.

$$\mathbb{E} e^{itX} = \int_{\mathbb{R}} \cos(tx) dF_X(x) + i \int_{\mathbb{R}} \sin(tx) dF_X(x).$$

- Weil die Funktionen  $\sin : \mathbb{R} \rightarrow [-1, 1]$  und  $\cos : \mathbb{R} \rightarrow [-1, 1]$  beschränkt und stetig sind, ist der Erwartungswert  $\mathbb{E} e^{itX}$  für jede Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  wohldefiniert, ohne dass es einer zusätzlichen Integrierbarkeitsbedingung bedarf.

Wir zeigen zunächst einige elementare Eigenschaften von charakteristischen Funktionen.

**Theorem 5.17** *Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable, und seien  $a, b \in \mathbb{R}$  beliebige reelle Zahlen. Dann gilt:*

1. Für jedes  $t \in \mathbb{R}$  ist

$$|\varphi_X(t)| \leq \varphi_X(0) = 1 \quad (80)$$

und

$$\varphi_X(-t) = \overline{\varphi_X(t)}. \quad (81)$$

2. Die charakteristische Funktion  $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$  ist gleichmäßig stetig.
3. Für die charakteristische Funktion  $\varphi_Y$  von  $Y = aX + b$  gilt

$$\varphi_Y(t) = e^{itb} \varphi_X(at) \quad \forall t \in \mathbb{R}. \quad (82)$$

### Beweis

- Die Gültigkeit von (80) ergibt sich aus der Jensen-Ungleichung (4.71), denn für jedes  $t \in \mathbb{R}$  gilt

$$\begin{aligned} |\varphi_X(t)|^2 &= \left| \int_{\mathbb{R}} \cos(tx) dF_X(x) + i \int_{\mathbb{R}} \sin(tx) dF_X(x) \right|^2 \\ &= \left( \int_{\mathbb{R}} \cos(tx) dF_X(x) \right)^2 + \left( \int_{\mathbb{R}} \sin(tx) dF_X(x) \right)^2 \\ &\leq \int_{\mathbb{R}} \cos^2(tx) dF_X(x) + \int_{\mathbb{R}} \sin^2(tx) dF_X(x) \\ &= \int_{\mathbb{R}} (\cos^2(tx) + \sin^2(tx)) dF_X(x) = 1 \quad \left( = \varphi_X(0) \right). \end{aligned}$$

- Die Gleichung (81) ergibt sich auf ähnliche Weise, denn es gilt für jedes  $t \in \mathbb{R}$

$$\begin{aligned} \varphi_X(-t) &= \mathbb{E} e^{-itX} \\ &= \int_{\mathbb{R}} \cos(-tx) dF_X(x) + i \int_{\mathbb{R}} \sin(-tx) dF_X(x) \\ &= \int_{\mathbb{R}} \cos(tx) dF_X(x) - i \int_{\mathbb{R}} \sin(tx) dF_X(x) \\ &= \overline{\mathbb{E} e^{itX}} = \overline{\varphi_X(t)}. \end{aligned}$$

- Außerdem gilt für beliebige  $t, h \in \mathbb{R}$

$$\begin{aligned} |\varphi_X(t+h) - \varphi_X(t)| &= |\mathbb{E} e^{i(t+h)X} - \mathbb{E} e^{itX}| \\ &= |\mathbb{E} (e^{itX} (e^{ihX} - 1))| \\ &\leq \mathbb{E} (|e^{itX}| |e^{ihX} - 1|) \\ &= \mathbb{E} |e^{ihX} - 1|. \end{aligned}$$

- Hieraus ergibt sich die gleichmäßige Stetigkeit von  $\varphi_X$ , weil  $\lim_{h \rightarrow 0} \mathbb{E} |e^{ihX} - 1| = 0$  wegen des Satzes von Lebesgue über die beschränkte Konvergenz.

- Für die charakteristische Funktion  $\varphi_Y$  von  $Y = aX + b$  gilt

$$\varphi_Y(t) = \mathbb{E} e^{it(aX+b)} = \mathbb{E} (e^{itb} e^{itaX}) = e^{itb} \mathbb{E} e^{itaX} = e^{itb} \varphi_X(at). \quad \square$$

### Beispiel

- Sei  $X$  eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable;  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$ .
- Dann gilt

$$\varphi_X(t) = e^{it\mu} e^{-\sigma^2 t^2/2} \quad \forall t \in \mathbb{R}. \quad (83)$$

- Wegen (82) genügt es zu zeigen, dass

$$\varphi_X(t) = e^{-t^2/2} \quad \forall t \in \mathbb{R}, \quad (84)$$

falls  $\mu = 0$  und  $\sigma^2 = 1$ .

- Sei also  $X$  nun eine  $N(0, 1)$ -verteilte Zufallsvariable. Dann kann man leicht zeigen, dass

$$\mathbb{E}(X^{2n-1}) = 0 \quad (85)$$

und

$$\mathbb{E}(X^{2n}) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^{2n} e^{-x^2/2} dx = \frac{(2n)!}{2^n n!} \quad (86)$$

für jedes  $n \in \mathbb{N}$ .

- Aus der Taylor-Reihenentwicklung (78) und aus (85)–(86) ergibt sich nun, dass

$$\begin{aligned} \varphi_X(t) &= \mathbb{E} e^{itX} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \sum_{n=0}^{\infty} \frac{(itx)^n}{n!} e^{-x^2/2} dx \\ &= \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^n e^{-x^2/2} dx = \sum_{n=0}^{\infty} \frac{(it)^{2n} (2n)!}{(2n)! 2^n n!} = \sum_{n=0}^{\infty} \left(-\frac{t^2}{2}\right)^n \frac{1}{n!} = e^{-t^2/2}. \end{aligned}$$

- Damit ist (84) und folglich auch (83) bewiesen.

Wir leiten nun eine einfache Formel für die charakteristische Funktion von Summen unabhängiger Zufallsvariablen her.

**Theorem 5.18** *Seien  $X, Y : \Omega \rightarrow \mathbb{R}$  unabhängige Zufallsvariablen. Dann gilt für jedes  $t \in \mathbb{R}$*

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t). \quad (87)$$

**Beweis** Weil  $X$  und  $Y$  unabhängig sind, sind wegen Theorem 3.17 auch die Zufallsvariablen  $\cos(tX)$  und  $\cos(tY)$ ,  $\sin(tX)$  und  $\sin(tY)$  bzw.  $\sin(tX)$  und  $\cos(tY)$  jeweils unabhängig, und für jedes  $t \in \mathbb{R}$  gilt somit

$$\begin{aligned} \varphi_{X+Y}(t) &= \mathbb{E} e^{it(X+Y)} = \mathbb{E} (e^{itX} e^{itY}) \\ &= \mathbb{E} ((\cos(tX) + i \sin(tX))(\cos(tY) + i \sin(tY))) \\ &= \mathbb{E} (\cos(tX) \cos(tY) - \sin(tX) \sin(tY) \\ &\quad + i \sin(tX) \cos(tY) + i \cos(tX) \sin(tY)) \\ &= \mathbb{E} \cos(tX) \mathbb{E} \cos(tY) - \mathbb{E} \sin(tX) \mathbb{E} \sin(tY) \\ &\quad + i \mathbb{E} \sin(tX) \mathbb{E} \cos(tY) + i \mathbb{E} \cos(tX) \mathbb{E} \sin(tY) \\ &= \mathbb{E} (\cos(tX) + i \sin(tX)) \mathbb{E} (\cos(tY) + i \sin(tY)) = \varphi_X(t) \varphi_Y(t). \quad \square \end{aligned}$$

In Theorem 5.19 beweisen wir eine nützliche *Umkehrformel* für charakteristische Funktionen. Hierfür benötigen wir die folgende *trigonometrische Identität*.

**Lemma 5.10** Für jedes  $c \in \mathbb{R}$  gilt

$$\int_0^\infty \frac{\sin(cx)}{x} dx = (\operatorname{sgn} c) \frac{\pi}{2}, \quad (88)$$

wobei

$$\operatorname{sgn} c = \begin{cases} 1, & \text{falls } c > 0, \\ 0, & \text{falls } c = 0, \\ -1, & \text{falls } c < 0. \end{cases}$$

Einen *Beweis* von Lemma 5.10 kann man beispielsweise in dem Buch von K.L. Chung (*A Course in Probability Theory*, Academic Press, New York 1974) finden.

**Theorem 5.19** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable mit der Verteilungsfunktion  $F_X$ . Für beliebige Stetigkeitspunkte  $a, b \in \mathbb{R}$  von  $F_X$  mit  $a < b$  gilt

$$P(a < X < b) = \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-i ta} - e^{-i tb}}{i t} \varphi_X(t) dt, \quad (89)$$

wobei  $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$  die charakteristische Funktion von  $X$  ist.

**Beweis**

- Für beliebige  $a, b, c \in \mathbb{R}$  mit  $a < b$  und  $c > 0$  gilt

$$\begin{aligned} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-i ta} - e^{-i tb}}{i t} \varphi_X(t) dt &= \int_{\mathbb{R}} \left( \frac{1}{2\pi} \int_{-c}^c \frac{e^{i t(x-a)} - e^{i t(x-b)}}{i t} dt \right) dF_X(x) \\ &= \int_{\mathbb{R}} \left( \frac{1}{\pi} \int_0^c \frac{\sin(t(x-a))}{t} dt - \frac{1}{\pi} \int_0^c \frac{\sin(t(x-b))}{t} dt \right) dF_X(x) \end{aligned}$$

- Wegen Lemma 5.10 ist die Funktion  $\varphi : [0, \infty) \rightarrow \mathbb{R}$  mit

$$\varphi(x) = \frac{1}{\pi} \int_0^x \frac{\sin(t(x-a))}{t} dt$$

beschränkt für jedes  $a \in \mathbb{R}$ .

- Aus dem Satz von Lebesgue über die beschränkte Konvergenz ergibt sich also, dass

$$\begin{aligned} &\lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-i ta} - e^{-i tb}}{i t} \varphi_X(t) dt \\ &= \lim_{c \rightarrow \infty} \int_{\mathbb{R}} \left( \frac{1}{\pi} \int_0^c \frac{\sin(t(x-a))}{t} dt - \frac{1}{\pi} \int_0^c \frac{\sin(t(x-b))}{t} dt \right) dF_X(x) \\ &= \int_{\mathbb{R}} \left( \frac{1}{\pi} \lim_{c \rightarrow \infty} \int_0^c \frac{\sin(t(x-a))}{t} dt - \frac{1}{\pi} \lim_{c \rightarrow \infty} \int_0^c \frac{\sin(t(x-b))}{t} dt \right) dF_X(x) \\ &= \int_{\mathbb{R}} \left( \frac{1}{\pi} \int_0^\infty \frac{\sin(t(x-a))}{t} dt - \frac{1}{\pi} \int_0^\infty \frac{\sin(t(x-b))}{t} dt \right) dF_X(x). \end{aligned}$$

- Aus Lemma 5.10 ergibt sich nun, dass

$$\begin{aligned}
& \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-it} - e^{-itb}}{it} \varphi_X(t) dt \\
&= \int_{(-\infty, a)} \left( \frac{1}{\pi} \frac{-\pi}{2} - \frac{1}{\pi} \frac{-\pi}{2} \right) dF_X(x) + \int_{\{a\}} \left( 0 - \frac{1}{\pi} \frac{-\pi}{2} \right) dF_X(x) \\
&\quad + \int_{(a, b)} \left( \frac{1}{\pi} \frac{\pi}{2} - \frac{1}{\pi} \frac{-\pi}{2} \right) dF_X(x) \\
&\quad + \int_{\{b\}} \left( \frac{1}{\pi} \frac{\pi}{2} - 0 \right) dF_X(x) + \int_{(b, \infty)} \left( \frac{1}{\pi} \frac{\pi}{2} - \frac{1}{\pi} \frac{\pi}{2} \right) dF_X(x) \\
&= \int_{(a, b)} \left( \frac{1}{\pi} \frac{\pi}{2} - \frac{1}{\pi} \frac{-\pi}{2} \right) dF_X(x) \\
&= P(a < X < b),
\end{aligned}$$

wobei sich die vorletzte Gleichheit aus der Annahme ergibt, dass  $a$  und  $b$  Stetigkeitspunkte von  $F_X$  sind.  $\square$

### Beachte

- Aus Theorem 5.19 ergibt sich der folgende *Eindeutigkeitssatz* für charakteristische Funktionen.
- D.h., die Verteilung  $P_X$  einer Zufallsvariablen  $X$  ist eindeutig durch die charakteristische Funktion  $\varphi_X$  von  $X$  bestimmt.

**Korollar 5.5** Seien  $X, Y : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen. Dann gilt  $X \stackrel{d}{=} Y$ , falls

$$\varphi_X(t) = \varphi_Y(t) \quad \forall t \in \mathbb{R}. \quad (90)$$

### Beweis

- Aus (90) und aus Theorem 5.19 ergibt sich, dass

$$F_X(b) - F_X(a) = F_Y(b) - F_Y(a), \quad (91)$$

falls  $a$  und  $b$  Stetigkeitspunkte von  $F_X$  und  $F_Y$  sind.

- Weil jede Verteilungsfunktion höchstens abzählbar unendlich viele Sprungstellen hat, gibt es eine Folge  $\{a_n\}$  von Stetigkeitspunkten von  $F_X$  und  $F_Y$  mit  $a_n \rightarrow -\infty$ .
- Hieraus und aus (91) folgt dass für jeden Stetigkeitspunkt  $b$  von  $F_X$  und  $F_Y$

$$F_X(b) = F_Y(b).$$

- Damit gilt auch  $F_X(x) = F_Y(x)$  für jedes  $x \in \mathbb{R}$ , weil  $F_X$  und  $F_Y$  rechtsstetige Funktionen sind.
- Wegen des eindeutigen Zusammenhanges zwischen Verteilung und Verteilungsfunktion (vgl. Theorem 3.3) ergibt sich hieraus die Behauptung.  $\square$

Aus Theorem 5.19 ergibt sich schließlich der folgende *Stetigkeitssatz* für charakteristische Funktionen.

**Theorem 5.20** Seien  $X, X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  beliebige Zufallsvariablen. Es gilt  $X_n \xrightarrow{d} X$  genau dann, wenn

$$\lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \varphi_X(t) \quad \forall t \in \mathbb{R}. \quad (92)$$

**Beweis**

- Die Notwendigkeit der Behauptung ergibt sich unmittelbar aus dem Kriterium der Verteilungskonvergenz, das in Theorem 5.7 hergeleitet wurde.
- Denn für jedes  $t \in \mathbb{R}$  sind  $\cos(tx)$  und  $\sin(tx)$  stetige und beschränkte Funktionen in  $x$ .
- Wegen Theorem 5.7 folgt also aus  $X_n \xrightarrow{d} X$ , dass für jedes  $t \in \mathbb{R}$

$$\begin{aligned}\varphi_{X_n}(t) &= \mathbb{E} \cos(tX_n) + i \mathbb{E} \sin(tX_n) \\ &\rightarrow \mathbb{E} \cos(tX) + i \mathbb{E} \sin(tX) \\ &= \varphi_X(t).\end{aligned}$$

- Um die Hinlänglichkeit der Bedingung zu beweisen, zeigen wir zunächst, dass

$$F_{X_n}(b) - F_{X_n}(a) \rightarrow F_X(b) - F_X(a) \quad (93)$$

für alle  $a, b \in \mathbb{R}$  mit  $a < b$  gilt, die Stetigkeitspunkte von  $F_{X_n}$  für jedes  $n \in \mathbb{N}$  und Stetigkeitspunkte von  $F_X$  sind.

- Für solche  $a, b \in \mathbb{R}$  ergibt sich aus der Umkehrformel (89) in Theorem 5.19, dass

$$\begin{aligned}F_X(b) - F_X(a) &= \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{-ib}}{it} \varphi_X(t) dt \\ &= \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{-ib}}{it} \left( \lim_{n \rightarrow \infty} \varphi_{X_n}(t) \right) dt \\ &= \lim_{n \rightarrow \infty} \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{-ib}}{it} \varphi_{X_n}(t) dt \\ &= \lim_{n \rightarrow \infty} (F_{X_n}(b) - F_{X_n}(a)),\end{aligned}$$

wobei sich die vorletzte Gleichheit aus dem Satz von Lebesgue über die beschränkte Konvergenz und die letzte Gleichheit durch die erneute Anwendung der Umkehrformel (89) in Theorem 5.19 ergibt.

- Weil dann (93) für alle  $a, b \in \mathbb{R}$  mit  $a < b$  bis auf höchstens abzählbar unendlich viele Ausnahmepunkte gilt, ergibt sich genauso wie im ersten Teil des Beweises von Theorem 5.7, dass (93) auch für *alle*  $a, b \in \mathbb{R}$  mit  $a < b$  gilt, die Stetigkeitspunkte von  $F_X$  sind.  $\square$

Wir betrachten nun noch den Zusammenhang zwischen den Momenten und der charakteristischen Funktion von Zufallsvariablen. Dabei zeigen wir, wie die charakteristische Funktion in eine *Taylor-Reihe* entwickelt werden kann.

**Theorem 5.21** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable mit der Verteilungsfunktion  $F_X$ . Wenn  $\mathbb{E}(|X|^n) < \infty$  für ein  $n \in \mathbb{N}$ , dann gilt:

1. Die charakteristische Funktion  $\varphi_X$  von  $X$  ist  $n$ -mal stetig differenzierbar, und
2. für jedes  $k \in \{1, \dots, n\}$

$$\varphi_X^{(k)}(t) = \int_{\mathbb{R}} (ix)^k e^{itx} dF_X(x) \quad \forall t \in \mathbb{R} \quad (94)$$

bzw.

$$\mathbb{E}(X^k) = \frac{\varphi_X^{(k)}(0)}{i^k}. \quad (95)$$

3. Außerdem gilt die Taylor-Reihenentwicklung

$$\varphi_X(t) = \sum_{k=0}^n \frac{(it)^k}{k!} \mathbb{E}(X^k) + \frac{(it)^n}{n!} \mathbb{E}\varepsilon_n(t) \quad \forall t \in \mathbb{R}, \quad (96)$$

wobei  $\varepsilon_n(t) : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable ist mit  $|\varepsilon_n(t)| \leq 3|X|^n$  für jedes  $t \in \mathbb{R}$  und  $\lim_{t \rightarrow 0} \mathbb{E}\varepsilon_n(t) = 0$ .

**Beweis**

- Aus  $\mathbb{E}(|X|^n) < \infty$  und aus der Ljapunow-Ungleichung (4.67) folgt, dass  $\mathbb{E}(|X|^k) < \infty$  für jedes  $k \in \{1, \dots, n\}$ .
- Außerdem gilt für beliebige  $t, h \in \mathbb{R}$  mit  $h \neq 0$

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \mathbb{E} \left( e^{itX} \frac{e^{ihX} - 1}{h} \right). \quad (97)$$

- Weil für beliebige  $h, x \in \mathbb{R}$

$$\left| \frac{e^{ihx} - 1}{h} \right| \leq |x|$$

und weil  $\mathbb{E}|X| < \infty$  gilt, ergibt sich aus (97) und aus dem Satz von Lebesgue über die beschränkte Konvergenz, dass

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\varphi_X(t+h) - \varphi_X(t)}{h} &= \mathbb{E} \left( e^{itX} \lim_{h \rightarrow 0} \frac{e^{ihX} - 1}{h} \right) \\ &= \mathbb{E}((iX)e^{itX}) \\ &= \int_{\mathbb{R}} (ix)e^{itx} dF_X(x). \end{aligned}$$

- Damit ist (94) bzw. (95) für  $k = 1$  bewiesen.
- Der Beweis von (94) bzw. (95) für jedes  $k \in \{1, \dots, n\}$  erfolgt nun mittels vollständiger Induktion.
- Um (96) zu beweisen, betrachten wir die folgende Taylor-Reihenentwicklung von  $\cos$  bzw.  $\sin$ .
- Für jedes  $x \in \mathbb{R}$  gilt

$$e^{ix} = \cos x + i \sin x = \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} + \frac{(ix)^n}{n!} (\cos(\theta_1 x) + i \sin(\theta_2 x))$$

mit  $|\theta_1|, |\theta_2| \leq 1$ .

- Hieraus folgt, dass für jedes  $t \in \mathbb{R}$

$$\mathbb{E}e^{itX} = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \mathbb{E}(X^k) + \frac{(it)^n}{n!} (\mathbb{E}(X^n) + \mathbb{E}\varepsilon_n(t)),$$

wobei

$$\varepsilon_n(t) = X^n (\cos(\Theta_1 tX) + i \sin(\Theta_2 tX) - 1) \quad (98)$$

und  $\Theta_1, \Theta_2 : \Omega \rightarrow \mathbb{R}$  Zufallsvariablen sind mit  $P(|\Theta_1|, |\Theta_2| \leq 1) = 1$ .

- Aus (98) ergibt sich, dass für jedes  $t \in \mathbb{R}$

$$|\varepsilon_n(t)| \leq 3|X|^n.$$

- Hieraus und aus dem Satz von Lebesgue über die majorisierte Konvergenz folgt, dass  $\lim_{t \rightarrow 0} \mathbb{E}\varepsilon_n(t) = 0$ .
- Damit ist (96) bewiesen.  $\square$

**Beachte**

- Mit Hilfe der Eigenschaften charakteristischer Funktionen, die in den Theoremen 5.18, 5.20 und 5.21 bewiesen wurden, kann nun der zentrale Grenzwertsatz in Theorem 5.16 für Summen von unabhängigen und identisch verteilten Zufallsvariablen ohne weiteres hergeleitet werden.
- *Zur Erinnerung:* In Theorem 5.16 hatten wir gezeigt, dass für jede Folge  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  von unabhängigen und identisch verteilten Zufallsvariablen mit  $\mathbb{E}(X_1^2) < \infty$  und  $\text{Var } X_1 > 0$  gilt:

$$\lim_{n \rightarrow \infty} P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \quad \forall x \in \mathbb{R}, \quad (99)$$

wobei  $\mu = \mathbb{E} X_k$  und  $\sigma^2 = \text{Var } X_k$  für jedes  $k \in \mathbb{N}$ .

- Weil  $\mathbb{E}(X_k - \mu) = 0$  und  $\text{Var}((X_k - \mu)/\sigma) = 1$  können wir o.B.d.A. beim Beweis von Theorem 5.16 voraussetzen, dass  $\mathbb{E} X_k = 0$  und  $\text{Var } X_k = 1$ , indem wir die Gültigkeit von Theorem 5.16 zunächst für die (unabhängigen und identisch verteilten) Zufallsvariablen  $X'_k = (X_k - \mu)/\sigma$  zeigen.
- Sei also

$$\mathbb{E} X_k = 0 \quad \text{und} \quad \text{Var } X_k = 1.$$

- Aus Theorem 5.21 ergibt sich dann, dass für beliebige  $t \in \mathbb{R}$  und  $n \in \mathbb{N}$

$$\varphi_{X_k/\sqrt{n}}(t) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

- Aus Theorem 5.18 ergibt sich nun, dass für  $S_n = X_1 + \dots + X_n$

$$\varphi_{S_n/\sqrt{n}}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n,$$

wobei  $o(n^{-1})/n^{-1} \rightarrow 0$  für  $n \rightarrow \infty$ . Hieraus folgt, dass

$$\lim_{n \rightarrow \infty} \varphi_{S_n/\sqrt{n}}(t) = e^{-t^2/2} \quad \forall t \in \mathbb{R}.$$

- Aus Theorem 5.20 ergibt sich somit, dass

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n}{\sqrt{n}} \leq x\right) = \Phi(x) \quad \forall x \in \mathbb{R}. \quad \square$$

**5.3.4 Bedingungen von Lindeberg und Ljapunow**

- Wir diskutieren nun die Frage, unter welchen Bedingungen der zentrale Grenzwertsatz für Summen von unabhängigen, jedoch nicht notwendig identisch verteilten Zufallsvariablen gilt.
- Dabei betrachten wir für jedes  $n \in \mathbb{N}$  eine Folge von unabhängigen Zufallsvariablen  $X_{n1}, \dots, X_{nn} : \Omega \rightarrow \mathbb{R}$ , wobei wir (o.B.d.A.) voraussetzen, dass für jedes  $k \in \{1, \dots, n\}$

$$\mathbb{E} X_{nk} = 0, \quad 0 < \sigma_{nk}^2 = \text{Var } X_{nk} < \infty \quad \text{und} \quad \sum_{k=1}^n \sigma_{nk}^2 = 1. \quad (100)$$

- Die Verteilungsfunktion von  $X_{nk}$  bezeichnen wir mit  $F_{nk}$ .
- *Beachte.* Es wird nicht ausgeschlossen, dass  $F_{nk}$  für jedes  $k \in \{1, \dots, n\}$  auch von der Anzahl  $n$  der insgesamt betrachteten Zufallsvariablen  $X_{n1}, \dots, X_{nn}$  abhängen kann.



Der folgende *zentrale Grenzwertsatz von Lindeberg* kann sowohl als Verallgemeinerung von Theorem 5.16 als auch von Lemma 5.9 aufgefasst werden, wenn dabei

$$X_{nk} = \frac{X_k - \mathbb{E} X_k}{\sqrt{n \operatorname{Var} X_k}} \quad (101)$$

gesetzt wird.

**Theorem 5.22** Für jedes  $n \in \mathbb{N}$  sei  $X_{n1}, \dots, X_{nn} : \Omega \rightarrow \mathbb{R}$  eine Folge von unabhängigen Zufallsvariablen, die den Bedingungen (100) genügen. Wenn für jedes  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \int_{\mathbb{R} \setminus (-\varepsilon, \varepsilon)} x^2 dF_{nk}(x) = 0, \quad (102)$$

dann gilt für jedes  $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P(X_{n1} + \dots + X_{nn} \leq x) = \Phi(x), \quad (103)$$

wobei  $\Phi : \mathbb{R} \rightarrow [0, 1]$  die Verteilungsfunktion der Standardnormalverteilung ist.

### Beachte

- Der zentrale Grenzwertsatz in Theorem 5.16 für Summen von unabhängigen und identisch verteilten Summanden  $X_1, X_2, \dots$  ist ein Spezialfall von Theorem 5.22.
- Mit dem Ansatz (101) gilt nämlich für jedes  $\varepsilon > 0$

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=1}^n \int_{\mathbb{R} \setminus (-\varepsilon, \varepsilon)} x^2 dF_{nk}(x) &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{n\sigma^2} \int_{\mathbb{R} \setminus (-\sqrt{n}\sigma\varepsilon, \sqrt{n}\sigma\varepsilon)} x^2 dF(x) \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sigma^2} \int_{\mathbb{R} \setminus (-\sqrt{n}\sigma\varepsilon, \sqrt{n}\sigma\varepsilon)} x^2 dF(x) = 0, \end{aligned}$$

wobei

$$\sigma^2 = \operatorname{Var} X_k \quad \text{und} \quad F(x) = P(X_k - \mathbb{E} X_k < x).$$

- Auf ähnliche Weise lässt sich zeigen, dass auch Lemma 5.9 als Spezialfall von Theorem 5.22 aufgefasst werden kann.

### Beachte

- Die Bedingung (102) bedeutet, dass

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E} (X_{nk}^2 \mathbb{1}_{\{|X_{nk}| > \varepsilon\}}) = 0. \quad (104)$$

- Weil

$$\max_{1 \leq k \leq n} \mathbb{E} (X_{nk}^2) \leq \varepsilon^2 + \sum_{k=1}^n \mathbb{E} (X_{nk}^2 \mathbb{1}_{\{|X_{nk}| > \varepsilon\}})$$

folgt aus (102), dass

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \mathbb{E} (X_{nk}^2) = 0. \quad (105)$$

- Weil  $\mathbb{E} X_{nk} = 0$  vorausgesetzt wird, ergibt sich aus (105) und aus der Tschebyschew-Ungleichung (4.72), dass für jedes  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} P(|X_{nk}| > \varepsilon) = 0. \quad (106)$$

- Wenn die Bedingung (106) erfüllt ist, dann spricht man von der (gleichmäßigen) *asymptotischen Kleinheit der Summanden*  $X_{nk}$ .

Im *Beweis* von Theorem 5.22 benötigen wir die folgenden Hilfssätze.

**Lemma 5.11** Für jedes  $n \in \mathbb{N}$  und für beliebige  $y_1, \dots, y_n \in \mathbb{C}$ ,  $z_1, \dots, z_n \in \mathbb{C}$  mit  $|y_k|, |z_k| \leq 1$  für jedes  $k \in \{1, \dots, n\}$  gilt

$$\left| \prod_{k=1}^n z_k - \prod_{k=1}^n y_k \right| \leq \sum_{k=1}^n |z_k - y_k|. \quad (107)$$

**Beweis**

- Wir beweisen (107) mit vollständiger Induktion.
- Für  $n = 1$  ist die Gültigkeit von (107) offensichtlich.
- Es gelte nun (107) für ein  $n \in \mathbb{N}$ . Dann gilt auch

$$\begin{aligned} \left| \prod_{k=1}^{n+1} z_k - \prod_{k=1}^{n+1} y_k \right| &\leq \left| \prod_{k=1}^{n+1} z_k - z_{n+1} \prod_{k=1}^n y_k \right| + \left| z_{n+1} \prod_{k=1}^n y_k - \prod_{k=1}^{n+1} y_k \right| \\ &= \underbrace{|z_{n+1}|}_{\leq 1} \left| \prod_{k=1}^n z_k - \prod_{k=1}^n y_k \right| + |z_{n+1} - y_{n+1}| \underbrace{\left| \prod_{k=1}^n y_k \right|}_{\leq 1} \\ &\leq \sum_{k=1}^n |z_k - y_k| + |z_{n+1} - y_{n+1}| \\ &= \sum_{k=1}^{n+1} |z_k - y_k|, \end{aligned}$$

wobei sich die letzte Ungleichung aus der Induktionsannahme ergibt.  $\square$

Ähnlich wie in Theorem 5.21 entwickeln wir nun die charakteristische Funktion  $\varphi_X$  einer Zufallsvariablen  $X$  in eine Taylor-Reihe, wobei wir jetzt aber eine *andere Abschätzung des Restgliedes* als in (96) betrachten.

**Lemma 5.12** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable mit  $\mathbb{E}(|X|^n) < \infty$  für ein  $n \in \mathbb{N}$ . Dann gilt:

$$\varphi_X(t) = \sum_{k=0}^n \frac{(it)^k}{k!} \mathbb{E}(X^k) + R_n(t) \quad \forall t \in \mathbb{R}, \quad (108)$$

wobei

$$|R_n(t)| \leq \mathbb{E} \left( \min \left\{ \frac{|tX|^{n+1}}{(n+1)!}, 2 \frac{|tX|^n}{n!} \right\} \right).$$

**Beweis**

- Mit vollständiger Induktion zeigen wir zunächst, dass für beliebige  $n \in \mathbb{N}$  und  $t \in \mathbb{R}$

$$\left| e^{it} - \left( 1 + \frac{it}{1!} + \dots + \frac{(it)^n}{n!} \right) \right| \leq \min \left\{ \frac{|t|^{n+1}}{(n+1)!}, 2 \frac{|t|^n}{n!} \right\}. \quad (109)$$

- Für  $n = 0$  gilt offenbar

$$|e^{it} - 1| \leq |e^{it}| + 1 = 2$$

und

$$|e^{it} - 1| = \left| \int_0^t e^{is} ds \right| \leq \int_0^{|t|} |e^{is}| ds = |t|.$$

- Es gelte nun (109) für ein  $n \in \mathbb{N}$ . Dann gilt auch

$$\begin{aligned} \left| e^{it} - \sum_{k=0}^{n+1} \frac{(it)^k}{k!} \right| &= \left| \int_0^t \left( e^{is} - \sum_{k=0}^n \frac{(is)^k}{k!} \right) ds \right| \\ &\stackrel{\text{Induktionsannahme}}{\leq} \int_0^{|t|} \min \left\{ \frac{|s|^{n+1}}{(n+1)!}, 2 \frac{|s|^n}{n!} \right\} ds \\ &\leq \min \left\{ \int_0^{|t|} \frac{|s|^{n+1}}{(n+1)!} ds, 2 \int_0^{|t|} \frac{|s|^n}{n!} ds \right\} \\ &= \min \left\{ \frac{|t|^{n+2}}{(n+2)!}, 2 \frac{|t|^{n+1}}{(n+1)!} \right\}. \end{aligned}$$

- Damit ist die Gültigkeit von (109) für jedes  $n \in \mathbb{N}$  bewiesen.
- Wenn nun in (109) die Zahl  $t$  durch  $tX$  ersetzt wird und wenn auf beiden Seiten von (109) der Erwartungswert gebildet wird, dann ergibt sich für jedes  $t \in \mathbb{R}$

$$\begin{aligned} \left| \varphi_X(t) - \sum_{k=0}^n \frac{(it)^k}{k!} \mathbb{E}(X^k) \right| &= \left| \mathbb{E} e^{itX} - \mathbb{E} \left( 1 + \frac{itX}{1!} + \dots + \frac{(itX)^n}{n!} \right) \right| \\ &\leq \mathbb{E} \left| e^{itX} - \left( 1 + \frac{itX}{1!} + \dots + \frac{(itX)^n}{n!} \right) \right| \\ &\leq \mathbb{E} \min \left\{ \frac{|tX|^{n+1}}{(n+1)!}, 2 \frac{|tX|^n}{n!} \right\}. \end{aligned}$$

- Damit ist (108) bewiesen. □

### Beweis von Theorem 5.22

- Für beliebige  $n \in \mathbb{N}$  und  $k \in \{1, \dots, n\}$  betrachten wir die charakteristische Funktion  $\varphi_{nk}(t) = \mathbb{E} e^{itX_{nk}}$  von  $X_{nk}$ .
- Wegen Theorem 5.18 gilt für die charakteristische Funktion  $\varphi_{X_{n1}+\dots+X_{nn}}(t)$  der Summe  $X_{n1}+\dots+X_{nn}$

$$\varphi_{X_{n1}+\dots+X_{nn}}(t) = \prod_{k=1}^n \varphi_{nk}(t) \quad \forall t \in \mathbb{R}.$$

- Wegen des Stetigkeitssatzes für charakteristische Funktionen (vgl. Theorem 5.20) genügt es also zu zeigen, dass für jedes  $t \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} \left| \prod_{k=1}^n \varphi_{nk}(t) - e^{-t^2/2} \right| = 0. \quad (110)$$

- Aus (100) und aus Lemma 5.11 ergibt sich, dass

$$\begin{aligned} \left| \prod_{k=1}^n \varphi_{nk}(t) - e^{-t^2/2} \right| &= \left| \prod_{k=1}^n \varphi_{nk}(t) - \prod_{k=1}^n e^{-\sigma_{nk}^2 t^2/2} \right| \\ &\leq \sum_{k=1}^n \left| \varphi_{nk}(t) - \left( 1 - \frac{\sigma_{nk}^2 t^2}{2} \right) \right| + \sum_{k=1}^n \left| 1 - \frac{\sigma_{nk}^2 t^2}{2} - e^{-\sigma_{nk}^2 t^2/2} \right|. \end{aligned}$$

- Um (110) zu beweisen, genügt es nun zu zeigen, dass beide Summen in dieser Abschätzung für  $n \rightarrow \infty$  gegen Null konvergieren.
- Aus Lemma 5.12 ergibt sich, dass für jedes  $\varepsilon > 0$

$$\begin{aligned}
\left| \varphi_{nk}(t) - \left( 1 - \frac{\sigma_{nk}^2 t^2}{2} \right) \right| &\leq \mathbb{E} \min\{|tX_{nk}|^3, 2|tX_{nk}|^2\} \\
&\leq \int_{(-\varepsilon, \varepsilon)} |t|^3 |x|^3 dF_{nk}(x) + 2 \int_{\mathbb{R} \setminus (-\varepsilon, \varepsilon)} t^2 x^2 dF_{nk}(x) \\
&\leq \varepsilon |t|^3 \sigma_{nk}^2 + 2t^2 \int_{\mathbb{R} \setminus (-\varepsilon, \varepsilon)} x^2 dF_{nk}(x).
\end{aligned}$$

- Weil  $\varepsilon > 0$  beliebig klein gewählt werden kann, ergibt sich nun aus (100) und (102), dass

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \left| \varphi_{nk}(t) - \left( 1 - \frac{\sigma_{nk}^2 t^2}{2} \right) \right| = 0.$$

- Um zu zeigen, dass auch die zweite Summe für  $n \rightarrow \infty$  gegen Null strebt, benutzen wir die Tatsache, dass für  $|x| \leq 1/2$

$$|e^{-x} - 1 + x| \leq x^2$$

und dass wegen (105)

$$t^2 \max_{1 \leq k \leq n} \sigma_{nk}^2 \leq \frac{1}{2}$$

für jedes hinreichend große  $n \geq n_0$  gilt.

- Hieraus folgt, dass für  $n \geq n_0$

$$\begin{aligned}
\sum_{k=1}^n \left| 1 - \frac{\sigma_{nk}^2 t^2}{2} - e^{-\sigma_{nk}^2 t^2 / 2} \right| &\leq \frac{t^4}{4} \sum_{k=1}^n \sigma_{nk}^4 \\
&\leq \frac{t^4}{4} \max_{1 \leq k \leq n} \sigma_{nk}^2 \sum_{k=1}^n \sigma_{nk}^2 = \frac{t^4}{4} \max_{1 \leq k \leq n} \sigma_{nk}^2.
\end{aligned}$$

- Aus (105) ergibt sich nun, dass auch die zweite Summe für  $n \rightarrow \infty$  gegen Null strebt. □

Der folgende *zentrale Grenzwertsatz von Ljapunow* enthält eine Bedingung, die zwar schärfer, jedoch einfacher handhabbar ist als die Lindeberg-Bedingung (102) in Theorem 5.22.

**Theorem 5.23** Für jedes  $n \in \mathbb{N}$  sei  $X_{n1}, \dots, X_{nn} : \Omega \rightarrow \mathbb{R}$  eine Folge von unabhängigen Zufallsvariablen, die den Bedingungen (100) genügen. Wenn

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E} |X_{nk}|^{2+\delta} = 0 \tag{111}$$

für ein  $\delta > 0$ , dann gilt für jedes  $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P(X_{n1} + \dots + X_{nn} \leq x) = \Phi(x). \tag{112}$$

**Beweis**

- Für beliebige  $\delta, \varepsilon > 0$  gilt

$$\begin{aligned}
\int_{\mathbb{R} \setminus (-\varepsilon, \varepsilon)} x^2 dF_{nk}(x) &\leq \int_{\mathbb{R} \setminus (-\varepsilon, \varepsilon)} \frac{|x|^{2+\delta}}{\varepsilon^\delta} dF_{nk}(x) \\
&\leq \varepsilon^{-\delta} \mathbb{E} |X_{nk}|^{2+\delta}.
\end{aligned}$$

- Aus (111) folgt also die Gültigkeit der Lindeberg-Bedingung (102) in Theorem 5.22.  $\square$

### Beachte

- In den Theoremen 5.22 bzw. 5.23 haben wir gezeigt, dass die Lindeberg-Bedingung (102) bzw. die Ljapunow-Bedingung (111) hinreichend für die asymptotische Normalverteiltetheit der Summe  $X_{n1} + \dots + X_{nn}$  sind.
- Die Lindeberg-Bedingung (102) und damit auch die Ljapunow-Bedingung (111) sind im allgemeinen jedoch *nicht notwendig* für die asymptotische Normalverteiltetheit der Summe  $X_{n1} + \dots + X_{nn}$ .
- Wir illustrieren diesen Sachverhalt durch das folgende *Beispiel*.
- Seien  $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$  unabhängige normalverteilte Zufallsvariablen mit  $\mathbb{E} X_1 = \mathbb{E} X_2 = \dots = 0$  und

$$\text{Var } X_1 = 1, \quad \text{Var } X_k = 2^{k-2} \quad \forall k \geq 2.$$

- Dann ist die Summe  $X_{n1} + \dots + X_{nn}$  der Zufallsvariablen  $X_{n1}, \dots, X_{nn}$  mit

$$X_{nk} = \frac{X_k}{\sqrt{\sum_{j=1}^n \text{Var } X_j}} \quad (113)$$

$N(0, 1)$ -verteilt wegen der Faltungsstabilität der Normalverteilung (vgl. Korollar 3.2)).

- D.h., es gilt insbesondere die asymptotische Normalverteiltetheit (112).
- Die Lindeberg-Bedingung (102) ist jedoch *nicht* erfüllt, denn mit der Schreibweise

$$c_{nk} = \frac{\sum_{j=1}^n \text{Var } X_j}{\text{Var } X_k}$$

gilt für die in (113) eingeführten Zufallsvariablen

$$\begin{aligned} \sum_{k=1}^n \int_{\mathbb{R} \setminus (-\varepsilon, \varepsilon)} x^2 dF_{nk}(x) &= \sum_{k=1}^n \frac{1}{c_{nk}} \int_{\mathbb{R} \setminus (-\sqrt{c_{nk}}\varepsilon, \sqrt{c_{nk}}\varepsilon)} x^2 d\Phi(x) \\ &\geq \frac{1}{c_{nn}} \int_{\mathbb{R} \setminus (-\sqrt{c_{nn}}\varepsilon, \sqrt{c_{nn}}\varepsilon)} x^2 d\Phi(x) \\ &\geq \frac{2^{n-2}}{1 + 1 + 2 + \dots + 2^{n-2}} \int_{\mathbb{R} \setminus (-\sqrt{2}\varepsilon, \sqrt{2}\varepsilon)} x^2 d\Phi(x), \end{aligned}$$

wobei diese untere Schranke offenbar nicht gegen 0 strebt.

- Der Grund hierfür ist, dass in diesem Beispiel die Summanden  $X_{nk}$  mit großem Index  $k$  *nicht* asymptotisch klein im Sinne von (106) sind, sondern die übrigen Summanden *dominieren*.

Der folgende zentrale Grenzwertsatz, den wir hier ohne Beweis erwähnen, enthält ein hinreichende *und* notwendige Bedingung für die Gültigkeit der asymptotischen Normalverteiltetheit (112).

**Theorem 5.24** Für jedes  $n \in \mathbb{N}$  sei  $X_{n1}, \dots, X_{nn} : \Omega \rightarrow \mathbb{R}$  eine Folge von unabhängigen Zufallsvariablen, die den Bedingungen (100) genügen. Es gilt (112) genau dann, wenn für jedes  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \int_{\mathbb{R} \setminus (-\varepsilon, \varepsilon)} |x| |F_{nk}(x) - \Phi_{nk}(x)| dx = 0, \quad (114)$$

wobei  $\Phi_{nk}(x) = \Phi(x/\sigma_{nk})$ .

Einen *Beweis* von Theorem 5.24 kann man beispielsweise in Abschnitt III.4 des Buches A.N. Sirjaev ( *Wahrscheinlichkeit*, Deutscher Verlag der Wissenschaften, Berlin 1988) finden.

**Beachte** Für das in (113) betrachtete Beispiel normalverteilter Summanden ist die Bedingung (114) offenbar erfüllt, denn es gilt  $F_{nk}(x) - \Phi_{nk}(x) = 0$  für jedes  $x \in \mathbb{R}$ .

## 6 Stichproben und Stichprobenfunktionen

Zu den Zielen der mathematischen Statistik gehört die Beschreibung und Untersuchung der Eigenschaften bzw. Gesetzmäßigkeiten von (großen) Datensätzen. Dabei kann es sich einerseits um sogenannte

- *reale Daten*

handeln, die sich z.B. bei der Beobachtung (Messung) von Vorgängen bzw. Strukturen in Natur, Technik oder Wirtschaft ergeben, oder es können

- *synthetische Daten*

sein, die bei der Simulation solcher Vorgänge bzw. Strukturen durch Computeralgorithmen erzeugt werden.

Die grundlegende Idee der Statistik, um diese Zielstellung zu erreichen, ist die *stochastische Modellierung* der vorliegenden Daten.

Dabei nutzt die Statistik Begriffe und Ergebnisse der *Wahrscheinlichkeitsrechnung*, wie zum Beispiel:

- Ereignis und Wahrscheinlichkeit, stochastische Unabhängigkeit,
- Zufallsvariable, ihre Verteilung, Erwartungswert und Varianz,
- Gesetz der großen Zahlen bzw. zentraler Grenzwertsatz.

### 6.1 Zufallsstichprobe

Der Vektor der vorliegenden Daten  $(x_1, \dots, x_n)$  kann im allgemeinen eine komplizierte Struktur aufweisen.

- Dabei muss der „Wert“  $x_i$  nicht unbedingt eine Zahl sein, sondern  $x_i$  kann für jedes  $i = 1, \dots, n$  selbst ein Vektor sein, der beispielsweise die Lage, Größe, Form und Orientierung eines geometrischen Objektes beschreiben kann.
- In dieser einführenden Vorlesung setzen wir jedoch meistens voraus, dass  $x_i \in \mathbb{R}$  für jedes  $i = 1, \dots, n$ .
- Eine Ausnahme bilden die in der zweiten Hälfte des Semesters diskutierten *Zwei-Stichproben-Probleme*, bei denen der Fall  $x_i \in \mathbb{R}^2$  für jedes  $i = 1, \dots, n$  betrachtet wird.

Wir nehmen an, dass die Daten  $x_1, \dots, x_n$  die *Realisierung* eines stochastischen Modells sind.

- Und zwar sei  $x_1, \dots, x_n$  die Realisierung einer Folge von unabhängigen und identisch verteilten Zufallsvariablen  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ , die über einem (im allgemeinen nicht näher spezifizierten) Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  definiert sind.
- D.h. insbesondere, dass für ein  $\omega \in \Omega$

$$X_i(\omega) = x_i \quad \forall i \in \{1, \dots, n\}.$$

- Außerdem setzen wir stets voraus, dass  $\mathbb{E}(X_i^2) < \infty$  für  $i = 1, \dots, n$ .

#### Definition

1. Der Vektor  $(x_1, \dots, x_n)$  heißt (konkrete) *Stichprobe*.

2. Die Menge  $B \subset \mathbb{R}^n$  aller (potentiell möglichen) Stichproben  $(x_1, \dots, x_n)$  heißt *Stichprobenraum*, wobei wir zur Vereinfachung der Notation annehmen, dass  $B = \mathbb{R}^n$ .
3. Der Zufallsvektor  $(X_1, \dots, X_n)$  heißt *Zufallsstichprobe*.
4. Für jedes  $i = 1, \dots, n$  heißt  $x_i$  *Stichprobenwert* von  $(x_1, \dots, x_n)$ . Analog hierzu nennt man  $X_i$  *Stichprobenvariable* von  $(X_1, \dots, X_n)$ .
5. Die Dimension  $n$  von  $(x_1, \dots, x_n)$  bzw.  $(X_1, \dots, X_n)$  heißt *Stichprobenumfang*.

### Beachte

- Die allgemeine Zielstellung der statistischen Datenanalyse, die in der Einleitung von Kapitel 6 diskutiert wurde, kann nun wie folgt präzisiert werden: Aus den vorliegenden Daten  $x_1, \dots, x_n$  sollen Schlussfolgerungen über Eigenschaften der (unbekannten) Verteilung der Zufallsstichprobe  $(X_1, \dots, X_n)$  gezogen werden.
- Weil wir voraussetzen, dass die Stichprobenvariablen  $X_1, \dots, X_n$  unabhängig und identisch verteilt sind, wird die Verteilung von  $(X_1, \dots, X_n)$  eindeutig durch die (Rand-) Verteilungsfunktion  $F = F_{X_i}$  einer (einzeln) Stichprobenvariablen  $X_i$  bestimmt, vgl. Abschnitt 3.3.5.
- Aus den Daten  $x_1, \dots, x_n$  sollen deshalb Schlussfolgerungen über Eigenschaften der unbekannten Verteilungsfunktion  $F$  gezogen werden.

## 6.2 Stichprobenfunktionen

Um Eigenschaften der Verteilungsfunktion  $F$  zu bestimmen, werden Funktionen  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$  betrachtet, die der Stichprobe  $(x_1, \dots, x_n)$  die „Bewertung“  $\varphi(x_1, \dots, x_n) \in \mathbb{R}^m$  zuordnen, vgl. auch Abschnitt 3.4. Dies führt zu der folgenden Begriffsbildung.

**Definition** Sei  $m \geq 1$  eine beliebige natürliche Zahl. Eine Borel-messbare Abbildung  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$  heißt *Stichprobenfunktion*.

### Beachte

- Im allgemeinen ist es üblich, die zusammengesetzte Abbildung  $\varphi(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^m$  mit

$$\varphi(X_1, \dots, X_n)(\omega) = \varphi(X_1(\omega), \dots, X_n(\omega))$$

*Statistik* zu nennen, d.h.,  $\varphi$  ist dann eine Funktion der Zufallsstichprobe  $(X_1, \dots, X_n)$ .

- Manchmal spricht man auch von einer *zufälligen Stichprobenfunktion*.
- Im Zusammenhang mit der Schätzung von Parametern oder anderen Modellcharakteristiken nennt man  $\varphi(X_1, \dots, X_n)$  *Schätzer*. Beim Testen von Hypothesen spricht man dagegen von (Test-) *Statistiken* bzw. von *Testgrößen*.

### 6.2.1 Stichprobenmittel

Wir diskutieren zunächst die Frage, wie der Erwartungswert  $\mu = \mathbb{E} X_i$  der Stichprobenvariablen  $X_1, \dots, X_n$  aus den beobachteten Daten  $x_1, \dots, x_n$  bestimmt werden kann.

- Hierfür betrachten wir die Stichprobenfunktion  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  mit  $\varphi(x_1, \dots, x_n) = n^{-1} \sum_{i=1}^n x_i$ .
- D.h., wir betrachten das arithmetische Mittel

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \tag{1}$$

der Stichprobenwerte  $x_1, \dots, x_n$ .



- Die Zahl  $\bar{x}_n$  wird *Stichprobenmittel* der (konkreten) Stichprobe  $(x_1, \dots, x_n)$  genannt.
- Außerdem betrachten wir das arithmetische Mittel der Stichprobenvariablen  $X_1, \dots, X_n$ , dessen Eigenschaften bereits in Abschnitt 5.2 im Zusammenhang mit dem Gesetz der großen Zahlen untersucht worden sind.

**Definition** Die Zufallsvariable

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (2)$$

heißt *Stichprobenmittel* der Zufallsstichprobe  $(X_1, \dots, X_n)$ .

Der Erwartungswert und die Varianz von  $\bar{X}_n$  lassen sich wie folgt darstellen.

**Theorem 6.1** Es gilt

$$\mathbb{E} \bar{X}_n = \mu \quad (3)$$

und

$$\text{Var} \bar{X}_n = \frac{\sigma^2}{n}, \quad (4)$$

wobei  $\mu = \mathbb{E} X_i$  und  $\sigma^2 = \text{Var} X_i$  den Erwartungswert bzw. die Varianz der Stichprobenvariablen  $X_1, \dots, X_n$  bezeichnen.

**Beweis**

- Die Formeln (3) und (4) ergeben sich wie folgt aus den Theoremen 4.4 bzw. 4.10.
- Wegen der Linearität des Erwartungswertes gilt nämlich (vgl. Theorem 4.4)

$$\begin{aligned} \mathbb{E} \bar{X}_n &= \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n X_i \right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} X_i \\ &= \frac{1}{n} \sum_{i=1}^n \mu = \mu. \end{aligned}$$

- Aus Theorem 4.10 über die Varianz der Summe von unabhängigen Zufallsvariablen ergibt sich, dass

$$\begin{aligned} \text{Var} \bar{X}_n &= \frac{1}{n^2} \text{Var} \left( \sum_{i=1}^n X_i \right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var} X_i \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}. \end{aligned}$$

□

**Beachte**

- Weil das Stichprobenmittel  $\bar{X}_n$  den Erwartungswert  $\mu$  hat (vgl. (3)), kann man  $\bar{X}_n$  als einen geeigneten „Schätzer“ der (im allgemeinen unbekannten) Modellcharakteristik  $\mu$  ansehen.
- Wegen (3) sagt man, dass bei der Schätzung von  $\mu$  durch  $\bar{X}_n$  kein „systematischer Fehler“ begangen wird.

- Der Schätzer  $\bar{X}_n$  kann dennoch sehr ungenau sein, wobei man den in (4) gegebenen Wert  $\sigma^2/n$  als Kennzahl für die Schätzgenauigkeit von  $\bar{X}_n$  auffassen kann.
- Dabei bedeutet (4), dass die Schätzgenauigkeit mit wachsendem Stichprobenumfang  $n$  verbessert wird.
- Es ist jedoch zu beachten, dass  $\sigma^2$  im allgemeinen ebenfalls unbekannt ist.
- Um eine Vorstellung über die Größenordnung der Schätzgenauigkeit von  $\bar{X}_n$  bei vorgegebenem Stichprobenumfang  $n$  zu erlangen, muss deshalb auch  $\sigma^2$  aus den beobachteten Daten  $x_1, \dots, x_n$  geschätzt werden.
- Diese Fragestellung werden wir in Abschnitt 6.2.2 diskutieren.

Neben den Formeln (3) und (4) für Erwartungswert und Varianz des Stichprobenmittels  $\bar{X}_n$  sind noch weitere Aussagen über die Verteilung von  $\bar{X}_n$  von Interesse bzw. über deren asymptotisches Verhalten für große  $n$ .

**Theorem 6.2** *Es gilt*

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1 \quad (5)$$

und

$$\lim_{n \rightarrow \infty} P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq x\right) = \Phi(x) \quad (6)$$

für jedes  $x \in \mathbb{R}$ , wobei  $\Phi(x)$  die Verteilungsfunktion der Standardnormalverteilung ist.

Der *Beweis* von Theorem 6.2 ergibt sich unmittelbar aus dem starken Gesetz der großen Zahlen bzw. aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen (vgl. die Theoreme 5.15 bzw. 5.16).

### Beachte

- Theorem 6.2 bietet eine Möglichkeit, um die Wahrscheinlichkeit des Ereignisses  $\{|\bar{X}_n - \mu| > \varepsilon\}$  zu bestimmen, dass das Stichprobenmittel  $\bar{X}_n$  um mehr als einen vorgegebenen Schwellenwert  $\varepsilon > 0$  von dem zu schätzenden Wert  $\mu$  abweicht.
- Aus (6) ergibt sich hierfür die folgende Näherungsformel

$$P(|\bar{X}_n - \mu| > \varepsilon) \approx 2\left(1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right) \quad (7)$$

für große  $n$ , denn es gilt

$$\begin{aligned} P(|\bar{X}_n - \mu| > \varepsilon) &= P\left(\sqrt{n} \left|\frac{\bar{X}_n - \mu}{\sigma}\right| > \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\ &= P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} < -\frac{\varepsilon\sqrt{n}}{\sigma}\right) + P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} > \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\ &\stackrel{(6)}{\approx} \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right) + 1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) \\ &= 2\left(1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right). \end{aligned}$$

- Falls  $\sigma^2$  unbekannt ist, dann kann man anstelle der Näherungsformel (7) ein (zufälliges) Intervall angeben, in dem die Abweichung  $\bar{X}_n - \mu$  des Stichprobenmittels  $\bar{X}_n$  von dem zu schätzenden Wert  $\mu$  mit einer (näherungsweise) vorgegebenen Wahrscheinlichkeit liegt, vgl. die Anmerkungen am Ende von Abschnitt 6.2.2.

### 6.2.2 Stichprobenvarianz

Wir untersuchen nun die Frage, wie die Varianz  $\sigma^2 = \text{Var } X_i$  der Stichprobenvariablen  $X_1, \dots, X_n$  aus den beobachteten Daten  $x_1, \dots, x_n$  bestimmt werden kann. Dabei gehen wir ähnlich wie in Abschnitt 6.2.1 vor.

- Wir betrachten die Stichprobenfunktion  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  mit

$$\varphi(x_1, \dots, x_n) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2, \quad (8)$$

wobei  $\bar{x}_n$  in (1) gegeben ist.

- Die in (8) eingeführte Größe wird *Stichprobenvarianz* der (konkreten) Stichprobe  $(x_1, \dots, x_n)$  genannt und mit  $s_n^2$  bezeichnet.

**Definition** Die Zufallsvariable

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (9)$$

heißt *Stichprobenvarianz* der Zufallsstichprobe  $(X_1, \dots, X_n)$ .

Der Erwartungswert und die Varianz von  $S_n^2$  lassen sich wie folgt darstellen.

**Theorem 6.3** *Es gilt*

$$\mathbb{E}(S_n^2) = \sigma^2. \quad (10)$$

Falls  $\mathbb{E}(X_i^4) < \infty$  für  $i = 1, \dots, n$ , dann gilt außerdem

$$\text{Var}(S_n^2) = \frac{1}{n} \left( \mu'_4 - \frac{n-3}{n-1} \sigma^4 \right), \quad (11)$$

wobei  $\mu'_4 = \mathbb{E}((X_i - \mu)^4)$  und  $\sigma^4 = (\text{Var } X_i)^2$  das 4-te zentrale Moment bzw. die quadrierte Varianz der Stichprobenvariablen  $X_1, \dots, X_n$  bezeichnen.

**Beweis**

- Aus der Definitionsgleichung (9) von  $S_n^2$  ergibt sich, dass

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n \left( (X_i - \mu) - \frac{1}{n} \sum_{j=1}^n (X_j - \mu) \right)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (X'_i - \bar{X}'_n)^2, \end{aligned}$$

d.h.,

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X'_i - \bar{X}'_n)^2, \quad (12)$$

wobei  $X'_i = X_i - \mu$  mit  $\mathbb{E} X'_i = 0$ .

- Weil die Formel (12) die gleiche Form hat wie die Definitionsgleichung (9) von  $S_n^2$ , können (und werden) wir o.B.d.A. voraussetzen, dass

$$\mathbb{E} X_i = 0. \quad (13)$$

- Hieraus folgt insbesondere, dass

$$\mathbb{E} \bar{X}_n = 0. \quad (14)$$

- Außerdem ergibt sich aus der Definitionsgleichung (9) von  $S_n^2$ , dass

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) \\ &= \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right), \end{aligned}$$

d.h.,

$$S_n^2 = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right). \quad (15)$$

- Wegen der Linearität des Erwartungswertes ergibt sich somit bei Berücksichtigung von (13) und (14), dass

$$\begin{aligned} \mathbb{E} S_n^2 &= \frac{1}{n-1} \left( \sum_{i=1}^n \mathbb{E} (X_i^2) - n\mathbb{E} (\bar{X}_n^2) \right) \\ &= \frac{1}{n-1} \left( \sum_{i=1}^n \text{Var } X_i - n\text{Var } \bar{X}_n \right) \\ &= \frac{1}{n-1} \left( n\sigma^2 - n \frac{\sigma^2}{n} \right) = \sigma^2, \end{aligned}$$

wobei sich die vorletzte Gleichheit aus (4) ergibt.

- Damit ist (10) bewiesen.
- Um die Gültigkeit von (11) zu zeigen, berechnen wir zunächst den Erwartungswert  $\mathbb{E} (S_n^4)$  der quadrierten Stichprobenvarianz  $S_n^4 = (S_n^2)^2$ .
- Durch Quadrierung beider Seiten von (15) ergibt sich, dass

$$(n-1)^2 S_n^4 = \left( \sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right)^2 = \left( \sum_{i=1}^n X_i^2 \right)^2 - 2n\bar{X}_n^2 \sum_{i=1}^n X_i^2 + n^2 \bar{X}_n^4.$$

- Für den Erwartungswert des ersten Summanden dieser Summe gilt

$$\begin{aligned} \mathbb{E} \left( \left( \sum_{i=1}^n X_i^2 \right)^2 \right) &= \mathbb{E} \left( \sum_{i=1}^n X_i^2 \sum_{j=1}^n X_j^2 \right) \\ &= \mathbb{E} \left( \sum_{i=1}^n X_i^4 + \sum_{i \neq j} X_i^2 X_j^2 \right) \\ &= \sum_{i=1}^n \mathbb{E} (X_i^4) + \sum_{i \neq j} \mathbb{E} (X_i^2 X_j^2) \\ &= \sum_{i=1}^n \mathbb{E} (X_i^4) + \sum_{i \neq j} \mathbb{E} (X_i^2) \mathbb{E} (X_j^2) \\ &= n\mu_4 + n(n-1)\sigma^4, \end{aligned}$$

wobei  $\mu_4 = \mathbb{E}(X_i^4)$ , d.h.,

$$\mathbb{E}\left(\left(\sum_{i=1}^n X_i^2\right)^2\right) = n\mu_4 + n(n-1)\sigma^4. \quad (16)$$

- Für den Erwartungswert des zweiten Summanden gilt

$$\begin{aligned} \mathbb{E}\left(\overline{X}_n^2 \sum_{i=1}^n X_i^2\right) &= \mathbb{E}\left(\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2 \sum_{j=1}^n X_j^2\right) \\ &= \frac{1}{n^2} \mathbb{E}\left(\left(\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j\right) \sum_{j=1}^n X_j^2\right) \\ &= \frac{1}{n^2} \mathbb{E}\left(\sum_{i=1}^n X_i^2 \sum_{j=1}^n X_j^2\right) + \frac{1}{n^2} \mathbb{E}\left(\sum_{i \neq j} X_i X_j \sum_{k=1}^n X_k^2\right) \\ &= \frac{1}{n^2} \mathbb{E}\left(\left(\sum_{i=1}^n X_i^2\right)^2\right) + \frac{1}{n^2} \sum_{i \neq j} \sum_{k=1}^n \underbrace{\mathbb{E}(X_i X_j X_k^2)}_{=0} \\ &\stackrel{(16)}{=} \frac{1}{n^2} (n\mu_4 + n(n-1)\sigma^4) \\ &= \frac{\mu_4 + (n-1)\sigma^4}{n}, \end{aligned}$$

wobei die vorletzte Gleichheit aus (16), aus der Unabhängigkeit der Stichprobenvariablen  $X_1, \dots, X_n$  und aus der Annahme folgt, dass  $\mathbb{E} X_i = 0$  für jedes  $i = 1, \dots, n$ .

- Schließlich ergibt sich auf ähnliche Weise für den Erwartungswert des dritten Summanden

$$\begin{aligned} \mathbb{E}\left(\overline{X}_n^4\right) &= \mathbb{E}\left(\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2 \left(\frac{1}{n} \sum_{j=1}^n X_j\right)^2\right) \\ &= \frac{1}{n^4} \mathbb{E}\left(\left(\sum_{k=1}^n X_k^2 + \sum_{i \neq j} X_i X_j\right) \left(\sum_{r=1}^n X_r^2 + \sum_{s \neq t} X_s X_t\right)\right) \\ &= \frac{1}{n^4} \mathbb{E}\left(\sum_{k=1}^n X_k^4 + \sum_{k \neq r} X_k^2 X_r^2 + 2 \sum_{i \neq j} X_i^2 X_j^2\right) \\ &= \frac{1}{n^4} (n\mu_4 + 3n(n-1)\sigma^4). \end{aligned}$$

- Insgesamt ergibt sich also, dass

$$\begin{aligned} \mathbb{E}(S_n^4) &= \frac{n\mu_4}{(n-1)^2} - \frac{2\mu_4}{(n-1)^2} + \frac{\mu_4}{n(n-1)^2} + \sigma^4 \left( \frac{n}{n-1} - \frac{2}{n-1} + \frac{3}{n(n-1)} \right) \\ &= \frac{(n^2 - 2n + 1)\mu_4}{n(n-1)^2} + \sigma^4 \frac{n^2 - 2n + 3}{n(n-1)} \\ &= \frac{1}{n} \mu_4 + \frac{n^2 - 2n + 3}{n(n-1)} \sigma^4. \end{aligned}$$

- In Theorem 4.6 hatten wir gezeigt, dass

$$\text{Var } Z = \mathbb{E}(Z^2) - (\mathbb{E} Z)^2$$

für jede Zufallsvariable  $Z$  mit  $\mathbb{E}(Z^2) < \infty$ .

- Hieraus und aus (10) folgt nun, dass

$$\begin{aligned}
 \text{Var}(S_n^2) &= \mathbb{E}(S_n^4) - (\mathbb{E}(S_n^2))^2 \\
 &= \frac{1}{n} \mu_4 + \frac{n^2 - 2n + 3}{n(n-1)} \sigma^4 - \sigma^4 \\
 &= \frac{1}{n} \left( \mu_4 - \frac{n-3}{n-1} \sigma^4 \right). \quad \square
 \end{aligned}$$

### Beachte

- Weil die Stichprobenvarianz  $S_n^2$  den Erwartungswert  $\sigma^2$  hat (vgl. (10)), kann  $S_n^2$  als ein geeigneter Schätzer der (im allgemeinen unbekannten) Modellcharakteristik  $\sigma^2$  angesehen werden.
- Wegen (10) sagt man, dass bei der Schätzung von  $\sigma^2$  durch  $S_n^2$  kein systematischer Fehler begangen wird.
- Darüber hinaus bedeutet (11), dass die Schätzgenauigkeit mit wachsendem Stichprobenumfang  $n$  verbessert wird, falls das 4-te zentrale Moment der Stichprobenvariablen endlich ist.

Neben den Formeln (10) und (11) für Erwartungswert und Varianz der Stichprobenvarianz  $S_n^2$  sind erneut weitere Aussagen über die Verteilung von  $S_n^2$  bzw. über deren asymptotisches Verhalten für große  $n$  von Interesse.

**Theorem 6.4** *Es gilt*

$$P\left(\lim_{n \rightarrow \infty} S_n^2 = \sigma^2\right) = 1. \quad (17)$$

Falls  $\mathbb{E}(X_i^4) < \infty$  und  $P((X_i - \mu)^2 = c) < 1$  für beliebige  $i = 1, 2, \dots$  und  $c \in \mathbb{R}$ , dann gilt außerdem  $\mu'_4 > \sigma^4$  und

$$\lim_{n \rightarrow \infty} P\left(\sqrt{n} \frac{S_n^2 - \sigma^2}{\sqrt{\mu'_4 - \sigma^4}} \leq x\right) = \Phi(x) \quad (18)$$

für jedes  $x \in \mathbb{R}$ , wobei  $\mu'_4 = \mathbb{E}((X_i - \mu)^4)$  und  $\sigma^4 = (\text{Var } X_i)^2$  das 4-te zentrale Moment bzw. die quadrierte Varianz der Stichprobenvariablen  $X_1, \dots, X_n$  bezeichnen und  $\Phi(x)$  die Verteilungsfunktion der Standardnormalverteilung ist.

### Beweis

- Aus der Voraussetzung, dass die Stichprobenvariablen  $X_1, \dots, X_n$  unabhängig und identisch verteilt sind, ergibt sich, dass auch die Zufallsvariablen  $X_1^2, \dots, X_n^2$  unabhängig und identisch verteilt sind, vgl. Theorem 3.17.
- Deshalb ergibt sich aus dem starken Gesetz der großen Zahlen (vgl. Theorem 5.15), dass mit Wahrscheinlichkeit 1

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^2 = \mathbb{E}(X_1^2).$$

- Außerdem ergibt sich aus (5), dass mit Wahrscheinlichkeit 1

$$\lim_{n \rightarrow \infty} \overline{X}_n^2 = \lim_{n \rightarrow \infty} \left( \frac{1}{n} \sum_{i=1}^n X_i^2 \right) = \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i \right)^2 = (\mathbb{E} X_1)^2.$$

- Hieraus und aus (15) ergibt sich nun, dass mit Wahrscheinlichkeit 1

$$\begin{aligned}
\lim_{n \rightarrow \infty} S_n^2 &= \lim_{n \rightarrow \infty} \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n \bar{X}_n^2 \right) \\
&= \lim_{n \rightarrow \infty} \frac{n}{n-1} \left( \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \right) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^2 - \lim_{n \rightarrow \infty} \bar{X}_n^2 \\
&= \mathbb{E}(X_1^2) - (\mathbb{E} X_1)^2 = \sigma^2.
\end{aligned}$$

- Damit ist (17) bewiesen.
- Um die Gültigkeit von (18) zu zeigen, benutzen wir die Darstellungsformeln (12) und (15) der Stichprobenvarianz  $S_n^2$ .
- Dabei ergibt sich, dass

$$S_n^2 = \frac{1}{n-1} \left( \sum_{i=1}^n (X'_i)^2 - n(\bar{X}'_n)^2 \right) = \frac{n}{n-1} \left( \frac{1}{n} \sum_{i=1}^n (X'_i)^2 - (\bar{X}'_n)^2 \right),$$

wobei  $X'_i = X_i - \mu$ .

- Aus Formel (5) in Theorem 6.2 ergibt sich, dass  $\sqrt{n}(\bar{X}'_n)^2$  in Verteilung gegen 0 konvergiert.
- Deshalb ergibt sich aus dem Satz von Slutsky für die Addition bzw. Multiplikation (vgl. die Theoreme 5.9 und 5.11), dass für jedes  $x \in \mathbb{R}$

$$\begin{aligned}
\lim_{n \rightarrow \infty} P\left(\sqrt{n} \frac{S_n^2 - \sigma^2}{\sqrt{\mu'_4 - \sigma^4}} \leq x\right) &= \lim_{n \rightarrow \infty} P\left(\sqrt{n} \frac{\frac{1}{n} \sum_{i=1}^n (X'_i)^2 - \sigma^2}{\sqrt{\mu'_4 - \sigma^4}} \leq x\right) \\
&= \lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n (X'_i)^2 - n\sigma^2}{\sqrt{n(\mu'_4 - \sigma^4)}} \leq x\right) \\
&= \lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n (X'_i)^2 - n\mathbb{E}((X'_1)^2)}{\sqrt{n\text{Var}((X'_1)^2)}} \leq x\right) \\
&= \Phi(x),
\end{aligned}$$

wobei sich die letzte Gleichheit aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen ergibt (vgl. Theorem 5.16).  $\square$

**Korollar 6.1** *Es gilt*

$$\lim_{n \rightarrow \infty} P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq x\right) = \Phi(x) \quad (19)$$

für jedes  $x \in \mathbb{R}$ , wobei  $\Phi(x)$  die Verteilungsfunktion der Standardnormalverteilung ist.

**Beweis**

- Aus (17) in Theorem 6.4 ergibt sich, dass  $S_n \xrightarrow{\text{f.s.}} \sigma$ .
- Aus (6) in Theorem 6.2 ergibt sich nun die Behauptung mit Hilfe des Satzes von Slutsky für die Multiplikation (vgl. Theorem 5.11).  $\square$

**Beachte**

- Korollar 6.1 bietet die Möglichkeit, ein (zufälliges) Intervall anzugeben, in dem die Abweichung  $\bar{X}_n - \mu$  des Stichprobenmittels  $\bar{X}_n$  von dem zu schätzenden Wert  $\mu$  mit einer (näherungsweise) vorgegebenen Wahrscheinlichkeit liegt, vgl. Formel (20).
- Sei  $\alpha \in (0, 1)$ , und sei  $z_\alpha \in \mathbb{R}$  die (eindeutig bestimmte) Lösung der Gleichung  $\Phi(z_\alpha) = \alpha$ .
- Dann heißt  $z_\alpha$  das  $\alpha$ -Quantil der Standardnormalverteilung  $N(0, 1)$ .
- Für  $\alpha \geq 0.5$  kann man das  $\alpha$ -Quantil der  $N(0, 1)$ -Verteilung aus Tabelle 1 entnehmen, vgl. Abschnitt 7.
- Aus (19) ergibt sich nun, dass

$$P\left(\frac{-z_{1-\alpha/2}S_n}{\sqrt{n}} < \bar{X}_n - \mu < \frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\right) \approx 1 - \alpha, \quad (20)$$

denn

$$\begin{aligned} P\left(\frac{-z_{1-\alpha/2}S_n}{\sqrt{n}} < \bar{X}_n - \mu < \frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\right) &= P\left(-z_{1-\alpha/2} < \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} < z_{1-\alpha/2}\right) \\ &= P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} < z_{1-\alpha/2}\right) - P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq -z_{1-\alpha/2}\right) \\ &\stackrel{(19)}{\approx} \Phi(z_{1-\alpha/2}) - \Phi(-z_{1-\alpha/2}) \\ &= \Phi(z_{1-\alpha/2}) - \Phi(z_{\alpha/2}) = 1 - \alpha, \end{aligned}$$

wobei in der vorletzten Gleichheit die Symmetrieeigenschaft

$$z_{\alpha/2} = -z_{1-\alpha/2} \quad (21)$$

der Quantile der  $N(0, 1)$ -Verteilung genutzt wurde.

### 6.3 Beispiel: Normalverteilte Stichprobenvariablen

- In Abschnitt 6.3.3 werden wir voraussetzen, dass die Stichprobenvariablen  $X_1, \dots, X_n$  normalverteilt sind, d.h.,  $X_i \sim N(\mu, \sigma^2)$  für jedes  $i \in \{1, \dots, n\}$ , wobei  $\mu \in \mathbb{R}$  und  $\sigma^2 > 0$ .
- Anstelle der Näherungsformeln (7) und (20) können wir dann „exakte“ Formeln für die Wahrscheinlichkeit herleiten, dass die Abweichung  $\bar{X}_n - \mu$  des Stichprobenmittels  $\bar{X}_n$  von dem zu schätzenden Wert  $\mu$  in einem bestimmten Intervall liegt.
- Hierfür betrachten wir zunächst eine Klasse von Verteilungen, die mit der Gammafunktion eng zusammenhängt.

#### 6.3.1 Gammaverteilung und $\chi^2$ -Verteilung

- Sei  $\Gamma : (0, \infty) \rightarrow (0, \infty)$  die *Gammafunktion* mit

$$\Gamma(p) = \int_0^\infty e^{-y} y^{p-1} dy, \quad p > 0. \quad (22)$$

- Durch partielle Integration ergibt sich, dass für jedes  $p > 0$

$$\Gamma(p+1) = \int_0^\infty e^{-y} y^p dy = \left[-e^{-y} y^p\right]_0^\infty + p \int_0^\infty e^{-y} y^{p-1} dy,$$



- Also gilt für jedes  $p > 0$

$$\Gamma(p+1) = p\Gamma(p). \quad (23)$$

- Weil

$$\Gamma(1) = \int_0^{\infty} e^{-y} dy = \left[ -e^{-y} \right]_0^{\infty} = 1,$$

ergibt sich aus (23), dass  $\Gamma(n+1) = n!$  für jedes  $n \in \mathbb{N}$ .

### Beachte

- Mit der Substitution  $y' = b^{-1}y$ , wobei  $b > 0$ , geht (22) über in

$$\Gamma(p) = b^p \int_0^{\infty} e^{-by} y^{p-1} dy$$

bzw.

$$1 = \int_0^{\infty} \frac{b^p}{\Gamma(p)} e^{-by} y^{p-1} dy, \quad (24)$$

wobei der Integrand als Wahrscheinlichkeitsdichte aufgefasst werden kann.

- Dies führt zu der folgenden Begriffsbildung.

**Definition** Man sagt, dass die Zufallsvariable  $Y : \Omega \rightarrow \mathbb{R}$  *gammaverteilt* ist mit den Parametern  $b > 0$  und  $p > 0$ , wenn  $Y$  absolutstetig ist und wenn die Dichte  $f_Y : \mathbb{R} \rightarrow [0, \infty)$  von  $Y$  gegeben ist durch

$$f_Y(y) = \begin{cases} \frac{b^p}{\Gamma(p)} e^{-by} y^{p-1}, & \text{falls } y > 0, \\ 0, & \text{falls } y \leq 0. \end{cases} \quad (25)$$

Schreibweise:  $Y \sim \Gamma(b, p)$ .

**Theorem 6.5** Die Zufallsvariable  $Y : \Omega \rightarrow \mathbb{R}$  sei gammaverteilt mit den Parametern  $b > 0$  und  $p > 0$ . Für die momenterzeugende Funktion  $\psi_Y : (-\infty, b) \rightarrow \mathbb{R}$  mit  $\psi_Y(t) = \mathbb{E} e^{tY}$  bzw. für die charakteristische Funktion  $\varphi_Y : \mathbb{R} \rightarrow \mathbb{C}$  mit  $\varphi_Y(t) = \mathbb{E} e^{itY}$  gilt dann

$$\psi_Y(t) = \frac{1}{\left(1 - \frac{t}{b}\right)^p}, \quad \forall t \in (-\infty, b) \quad \text{bzw.} \quad \varphi_Y(t) = \frac{1}{\left(1 - \frac{it}{b}\right)^p}, \quad \forall t \in \mathbb{R}. \quad (26)$$

Außerdem gilt für jedes  $k \in \mathbb{N}$

$$\mathbb{E}(Y^k) = \frac{p(p+1) \dots (p+k-1)}{b^k}. \quad (27)$$

### Beweis

- Aus der Definition der momenterzeugenden Funktion und aus (25) ergibt sich, dass

$$\psi_Y(t) = \mathbb{E} e^{tY} = \int_0^{\infty} e^{ty} f_Y(y) dy = \int_0^{\infty} \frac{b^p}{\Gamma(p)} e^{ty} e^{-by} y^{p-1} dy = \int_0^{\infty} \frac{b^p}{\Gamma(p)} e^{-(b-t)y} y^{p-1} dy.$$

- Ähnlich wie bei der Herleitung von (24) betrachten wir nun die Substitution  $y = (b-t)y'$ , wobei  $b > 0$  und  $t < b$ , und erhalten die Gleichung

$$1 = \int_0^\infty \frac{(b-t)^p}{\Gamma(p)} e^{-(b-t)y} y^{p-1} dy.$$

- Hieraus folgt, dass für jedes  $t < b$

$$\psi_Y(t) = \frac{b^p}{(b-t)^p} = \frac{1}{\left(1 - \frac{t}{b}\right)^p},$$

d.h., die in dem Gebiet  $\{z = t_1 + it_2 : t_1 < b\}$  holomorphen Funktionen  $\varphi_1(z) = \mathbb{E} e^{zY}$  und  $\varphi_2(z) = (1 - z/b)^{-p}$  besitzen für jedes  $z \in (-\infty, b)$  jeweils die gleichen Funktionswerte.

- Aus dem Identitätssatz für holomorphe Funktionen (vgl. Abschnitt 8.1 in R. Remmert (1992) *Funktionentheorie 1*, Springer, Berlin) ergibt sich nun, dass  $\varphi_1(it) = \varphi_2(it)$  für jedes  $t \in \mathbb{R}$  gilt.
- Damit ist (26) bewiesen.
- Aus (26) ergibt sich nun, dass  $\varphi_Y(t)$  unendlich oft differenzierbar ist und dass die  $k$ -te Ableitung  $\varphi_Y^{(k)}(t)$  von  $\varphi_Y(t)$  gegeben ist durch

$$\varphi_Y^{(k)}(t) = \frac{p(p+1) \dots (p+k-1)}{b^k} i^k \frac{1}{\left(1 - \frac{it}{b}\right)^{p+k}}, \quad \forall t \in \mathbb{R}.$$

- Weil außerdem  $\mathbb{E}(|Y|^k) < \infty$  für jedes  $k \in \mathbb{N}$ , gilt (vgl. Theorem 5.21)

$$\mathbb{E}(Y^k) = \frac{\varphi_Y^{(k)}(0)}{i^k}, \quad \forall k \in \mathbb{N}.$$

- Hieraus folgt die Gültigkeit von (27). □

Aus Theorem 6.5 ergibt sich insbesondere, dass die Familie der Gammaverteilungen die folgende Eigenschaft der *Faltungsstabilität* besitzt.

**Korollar 6.2** Seien  $Y_1, Y_2 : \Omega \rightarrow \mathbb{R}$  unabhängige Zufallsvariablen mit  $Y_1 \sim \Gamma(b, p_1)$  und  $Y_2 \sim \Gamma(b, p_2)$ , wobei  $b, p_1, p_2 > 0$ . Dann gilt  $Y_1 + Y_2 \sim \Gamma(b, p_1 + p_2)$ .

#### Beweis

- Für die charakteristische Funktion  $\varphi_{Y_1+Y_2}(t) = \mathbb{E} e^{it(Y_1+Y_2)}$  der Summe  $Y_1 + Y_2$  gilt wegen der Unabhängigkeit von  $Y_1$  und  $Y_2$  (vgl. Theorem 5.18), dass

$$\varphi_{Y_1+Y_2}(t) = \varphi_{Y_1}(t) \varphi_{Y_2}(t), \quad \forall t \in \mathbb{R}.$$

- Aus (26) ergibt sich also, dass

$$\begin{aligned} \varphi_{Y_1+Y_2}(t) &= \frac{1}{\left(1 - \frac{it}{b}\right)^{p_1}} \frac{1}{\left(1 - \frac{it}{b}\right)^{p_2}} \\ &= \frac{1}{\left(1 - \frac{it}{b}\right)^{p_1+p_2}}. \end{aligned}$$

- Die charakteristische Funktion  $\varphi_{Y_1+Y_2}(t)$  der Summe  $Y_1 + Y_2$  besitzt somit die in (26) hergeleitete Form der charakteristischen Funktion der Gammaverteilung  $\Gamma(b, p_1 + p_2)$ .
- Aus dem Eindeutigkeitssatz für charakteristische Funktionen (vgl. Korollar 5.5) folgt nun, dass  $Y_1 + Y_2 \sim \Gamma(b, p_1 + p_2)$ .  $\square$

### Beachte

- Aus der Definitionsgleichung (25) der Dichte der Gammaverteilung ergibt sich, dass die Exponentialverteilung  $\text{Exp}(b)$  eine spezielle Gammaverteilung ist, wobei

$$\text{Exp}(b) = \Gamma(b, 1).$$

- Wegen Korollar 6.2 ist die Summe  $Y_1 + \dots + Y_n$  von  $n$  unabhängigen und identisch verteilten Zufallsvariablen  $Y_1, \dots, Y_n$  mit  $Y_i \sim \text{Exp}(b)$  für jedes  $i \in \{1, \dots, n\}$  ebenfalls gammaverteilt, und zwar gilt

$$Y_1 + \dots + Y_n \sim \Gamma(b, n).$$

Ein solche Gammaverteilung heißt *Erlangverteilung* der Ordnung  $n$ .

- Außer den Klassen der Exponential- bzw. Erlangverteilungen werden noch weitere Teilklassen von Gammaverteilungen betrachtet.
- Eine solche Teilklasse von Gammaverteilungen spielt bei der Bestimmung der Verteilung der Stichprobenvarianz  $S_n^2$  von normalverteilten Stichprobenvariablen eine wichtige Rolle.
- Es ist dies die Familie der  $\chi^2$ -Verteilungen, die zu den sogenannten *statistischen Prüfverteilungen* gehören und die wie folgt definiert sind.

### Definition

- Sei  $r \in \mathbb{N}$  eine beliebige natürliche Zahl, und seien  $X_1, \dots, X_r : \Omega \rightarrow \mathbb{R}$  unabhängige und  $N(0, 1)$ -verteilte Zufallsvariablen.
- Dann sagt man, dass die Zufallsvariable  $U_r = \sum_{i=1}^r X_i^2$  eine  $\chi^2$ -Verteilung mit  $r$  Freiheitsgraden hat. (Schreibweise:  $U_r \sim \chi_r^2$ )

**Theorem 6.6** Sei  $r \geq 1$  eine beliebige natürliche Zahl, und sei  $U_r$  eine  $\chi^2$ -verteilte Zufallsvariable mit  $r$  Freiheitsgraden. Dann ist die Dichte von  $U_r$  gegeben durch

$$f_{U_r}(x) = \begin{cases} \frac{x^{(r-2)/2} e^{-x/2}}{2^{r/2} \Gamma(r/2)}, & \text{falls } x > 0, \\ 0, & \text{sonst,} \end{cases} \quad (28)$$

wobei  $\Gamma(1) = 1$ ,  $\Gamma(1/2) = \sqrt{\pi}$  und  $\Gamma(p+1) = p\Gamma(p)$ , d.h., es gilt  $U_r \sim \Gamma(1/2, r/2)$ .

### Beweis

- Wir betrachten zunächst den Fall  $r = 1$ , d.h., es gelte  $U_1 = X^2$  mit  $X \sim N(0, 1)$ .
- Für die Dichte  $f_{U_1}(x)$  von  $U_1$  ergibt sich dann aus Theorem WR-3.15, dass

$$f_{U_1}(x) = \begin{cases} \frac{1}{2\sqrt{x}} (f_X(\sqrt{x}) + f_X(-\sqrt{x})), & \text{falls } x > 0, \\ 0, & \text{falls } x \leq 0. \end{cases}$$

- Weil

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

ergibt sich somit, dass

$$f_{U_1}(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{x}{2}\right), & \text{falls } x > 0, \\ 0, & \text{falls } x \leq 0. \end{cases} \quad (29)$$

- Damit ist (28) für den Fall  $r = 1$  bewiesen.
- Ein Vergleich von Formel (29) mit der Definitionsgleichung (25) der Gammaverteilung zeigt außerdem, dass

$$U_1 = X^2 \sim \Gamma(1/2, 1/2). \quad (30)$$

- Seien nun  $X_1, \dots, X_r : \Omega \rightarrow \mathbb{R}$  unabhängige und  $N(0, 1)$ -verteilte Zufallsvariablen.
- Aus (30) und aus Korollar 6.2 folgt dann, dass

$$X_1^2 + \dots + X_r^2 \sim \Gamma(1/2, r/2).$$

- Hieraus und aus der Definitionsgleichung (25) der Gammaverteilung ergibt sich nun die Gültigkeit von (28) für jedes  $r \in \mathbb{N}$ .  $\square$

### Beachte

- Für  $U_r \sim \chi_r^2$  und  $\alpha \in (0, 1)$  sei  $\chi_{r,\alpha}^2$  die (eindeutig bestimmte) Lösung der Gleichung  $F_{U_r}(\chi_{r,\alpha}^2) = \alpha$ .
- Dann heißt  $\chi_{r,\alpha}^2$  das  $\alpha$ -Quantil der  $\chi^2$ -Verteilung mit  $r$  Freiheitsgraden.
- Quantile der  $\chi^2$ -Verteilung mit  $r$  Freiheitsgraden sind in Tabelle 2 gegeben, vgl. Abschnitt 7.

### 6.3.2 Unabhängigkeit und Transformation von Zufallsvektoren

- In Abschnitt 6.3.3 benötigen wir einige Eigenschaften von Zufallsvektoren, die wir hier lediglich erwähnen (ohne sie im einzelnen zu beweisen).
- In den Abschnitten 3 und 4 sind solche Eigenschaften für den Fall reellwertiger Zufallsvariablen hergeleitet worden.

**Definition** In Verallgemeinerung des Begriffes der Unabhängigkeit von reellwertigen Zufallsvariablen, der in Abschnitt 3.3.5 eingeführt wurde, sagen wir, dass die Zufallsvektoren  $X_1 : \Omega \rightarrow \mathbb{R}^{m_1}, \dots, X_n : \Omega \rightarrow \mathbb{R}^{m_n}$  *unabhängig* sind, falls

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n) \quad \forall x_1 \in \mathbb{R}^{m_1}, \dots, x_n \in \mathbb{R}^{m_n}. \quad (31)$$

Analog zu Theorem 3.11 ergibt sich dann die folgende Charakterisierung der Unabhängigkeit absolutstetiger Zufallsvektoren.

**Theorem 6.7** Seien  $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^{m_1 + \dots + m_n}$  und damit auch  $X_1 : \Omega \rightarrow \mathbb{R}^{m_1}, \dots, X_n : \Omega \rightarrow \mathbb{R}^{m_n}$  absolutstetige Zufallsvektoren. Die Komponenten  $X_1, \dots, X_n$  des Vektors  $(X_1, \dots, X_n)$  sind genau dann unabhängig, wenn für fast alle  $(x_1, \dots, x_n) \in \mathbb{R}^{m_1 + \dots + m_n}$

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n). \quad (32)$$

Analog zu Theorem 3.17 ergibt sich der folgende Satz über die Unabhängigkeit zusammengesetzter Abbildungen.

**Theorem 6.8** *Die Zufallsvektoren  $X_1 : \Omega \rightarrow \mathbb{R}^{m_1}, \dots, X_n : \Omega \rightarrow \mathbb{R}^{m_n}$  seien unabhängig. Für beliebige Borel-messbare Funktionen  $\varphi_1 : \mathbb{R}^{m_1} \rightarrow \mathbb{R}^{m'_1}, \dots, \varphi_n : \mathbb{R}^{m_n} \rightarrow \mathbb{R}^{m'_n}$  sind die Zufallsvektoren  $\varphi_1(X_1), \dots, \varphi_n(X_n)$  dann auch unabhängig.*

Schließlich gilt der folgende Transformationssatz für die Dichte von absolutstetigen Zufallsvektoren.

**Theorem 6.9**

- Sei  $X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$  ein absolutstetiger Zufallsvektor mit der Dichte  $f_X : \mathbb{R}^n \rightarrow [0, \infty)$ , sei  $B \in \mathcal{B}(\mathbb{R}^n)$  ein offener  $n$ -dimensionaler Quader, und sei  $\varphi = (\varphi_1, \dots, \varphi_n) : \mathbb{R}^n \rightarrow \mathbb{R}^n$  eine auf  $B$  differenzierbare Abbildung mit den (stetigen) partiellen Ableitungen  $\partial\varphi_i/\partial x_j(x_1, \dots, x_n)$ .

- Außerdem gelte

$$\{x \in \mathbb{R}^n : f_X(x) \neq 0\} \subset B$$

und

$$\det\left(\frac{\partial\varphi_i}{\partial x_j}(x_1, \dots, x_n)\right) \neq 0 \quad \forall (x_1, \dots, x_n) \in B,$$

so dass die Einschränkung  $\varphi : B \rightarrow C$  von  $\varphi$  auf die Menge  $B$  eine eindeutige Abbildung ist, wobei  $C = \{\varphi(x) : x \in B\}$ .

- Sei  $\varphi^{-1} = (\varphi_1^{-1}, \dots, \varphi_n^{-1}) : C \rightarrow B$  die Umkehrung der Abbildung  $\varphi : B \rightarrow C$ .
- Dann ist auch der Zufallsvektor  $Y = \varphi(X)$  absolutstetig, und für die Dichte  $f_Y(y)$  von  $Y$  gilt

$$f_Y(y) = \begin{cases} f_X(\varphi_1^{-1}(y), \dots, \varphi_n^{-1}(y)) \left| \det\left(\frac{\partial\varphi_i^{-1}}{\partial y_j}(y_1, \dots, y_n)\right) \right|, & \text{falls } y = (y_1, \dots, y_n) \in C, \\ 0, & \text{falls } y \notin C. \end{cases} \quad (33)$$

### 6.3.3 Verteilung von Stichprobenmittel und Stichprobenvarianz

- Wir bestimmen nun die (gemeinsame) Verteilung des Stichprobenmittels  $\bar{X}_n$  und der Stichprobenvarianz  $S_n^2$  bei normalverteilten Stichprobenvariablen  $X_1, \dots, X_n$ .
- Zunächst zeigen wir, dass  $\bar{X}_n$  und  $S_n^2$  unabhängig sind.

**Theorem 6.10** *Sei  $(X_1, \dots, X_n)$  eine normalverteilte Zufallsstichprobe mit  $X_i \sim N(\mu, \sigma^2)$ . Dann sind  $\bar{X}_n$  und  $S_n^2$  unabhängige Zufallsvariablen.*

**Beweis**

- Zur Erinnerung: Mit der Schreibweise  $X'_i = X_i - \mu$  gilt

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X'_i + \mu \quad \text{und} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X'_i - \bar{X}'_n)^2. \quad (34)$$

- Wegen Theorem 6.8 können (und werden) wir deshalb o.B.d.A. voraussetzen, dass  $\mu = 0$  und  $\sigma^2 = 1$ , d.h.  $X_i \sim N(0, 1)$ .

- Um die Unabhängigkeit von  $\bar{X}_n$  und  $S_n^2$  zu zeigen, nutzen wir die Tatsache, dass sich die Stichprobenvarianz  $S_n^2$  wie folgt darstellen lässt:

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \left( (X_1 - \bar{X}_n)^2 + \sum_{i=2}^n (X_i - \bar{X}_n)^2 \right) \\ &= \frac{1}{n-1} \left( \left( \sum_{i=2}^n (X_i - \bar{X}_n) \right)^2 + \sum_{i=2}^n (X_i - \bar{X}_n)^2 \right), \end{aligned}$$

wobei sich die letzte Gleichheit aus der Identität  $\sum_{i=1}^n (X_i - \bar{X}_n) = 0$  ergibt.

- Die Stichprobenvarianz  $S_n^2$  ist also eine Funktion des Zufallsvektors  $(X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ , d.h., es gilt

$$S_n^2 = \tilde{\varphi}(X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n), \quad (35)$$

wobei

$$\tilde{\varphi}(x_2, \dots, x_n) = \frac{1}{n-1} \left( \left( \sum_{i=2}^n x_i \right)^2 + \sum_{i=2}^n x_i^2 \right). \quad (36)$$

- Wir zeigen nun zunächst, dass der Zufallsvektor  $(X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$  unabhängig von  $\bar{X}_n$  ist.
- Wegen der Unabhängigkeit und  $N(0,1)$ -Verteiltheit der Stichprobenvariablen  $X_1, \dots, X_n$  ist die (gemeinsame) Dichte  $f_X(x_1, \dots, x_n)$  des Zufallsvektors  $X = (X_1, \dots, X_n)$  gegeben durch

$$f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

- Wir betrachten die lineare Abbildung  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^n$  mit

$$\varphi_1(x) = \bar{x}_n, \quad \varphi_2(x) = x_2 - \bar{x}_n, \quad \dots, \quad \varphi_n(x) = x_n - \bar{x}_n.$$

- Für die Umkehrabbildung  $\varphi^{-1}: \mathbb{R}^n \rightarrow \mathbb{R}^n$  gilt dann für jedes  $y = (y_1, \dots, y_n) \in \mathbb{R}^n$

$$\varphi_1^{-1}(y) = y_1 - \sum_{i=2}^n y_i, \quad \varphi_2^{-1}(y) = y_1 + y_2, \quad \dots, \quad \varphi_n^{-1}(y) = y_1 + y_n.$$

- Für die Jacobi-Determinante der Abbildung  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^n$  bzw. der Umkehrabbildung  $\varphi^{-1}: \mathbb{R}^n \rightarrow \mathbb{R}^n$  gilt

$$\det\left(\frac{\partial \varphi_i}{\partial x_j}(x_1, \dots, x_n)\right) = \frac{1}{n}$$

für jedes  $(x_1, \dots, x_n) \in \mathbb{R}^n$  bzw.

$$\det\left(\frac{\partial \varphi_i^{-1}}{\partial y_j}(y_1, \dots, y_n)\right) = n$$

für jedes  $(y_1, \dots, y_n) \in \mathbb{R}^n$ .

- Aus Theorem 6.9 ergibt sich somit für die Dichte  $f_{(\bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)}(y_1, \dots, y_n)$  des Zufallsvektors

$$(\bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n) = \varphi(X_1, \dots, X_n)$$

die folgende Darstellungsformel.

- Für jedes  $(y_1, \dots, y_n) \in \mathbb{R}^n$  gilt

$$\begin{aligned} f_{(\bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)}(y_1, \dots, y_n) &= \frac{n}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \left(y_1 - \sum_{i=2}^n y_i\right)^2\right) \exp\left(-\frac{1}{2} \sum_{i=2}^n (y_i + y_1)^2\right) \\ &= \left(\left(\frac{n}{2\pi}\right)^{1/2} \exp\left(-\frac{1}{2} n y_1^2\right)\right) \left(\left(\frac{n}{(2\pi)^{n-1}}\right)^{1/2} \exp\left(-\frac{1}{2} \left(\sum_{i=2}^n y_i^2 + \left(\sum_{i=2}^n y_i\right)^2\right)\right)\right). \end{aligned}$$

- Wegen dieser Produktdarstellung der Dichte  $f_{(\bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)}(y_1, \dots, y_n)$  ergibt sich nun aus Theorem 6.7, dass  $(X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$  unabhängig von  $\bar{X}_n$  ist.
- Wegen (35) ergibt sich somit aus Theorem 6.8, dass auch die Zufallsvariablen  $S_n^2$  und  $\bar{X}_n$  unabhängig sind.  $\square$

**Theorem 6.11** Sei  $(X_1, \dots, X_n)$  eine normalverteilte Zufallsstichprobe mit  $X_i \sim N(\mu, \sigma^2)$  für  $i \in \{1, \dots, n\}$ . Dann gilt

$$\bar{X}_n \sim N(\mu, \sigma^2/n) \quad (37)$$

und

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2. \quad (38)$$

**Beweis**

- Aus Theorem 3.13 (bzw. aus dessen vektorieller Version in Theorem 6.9) ergibt sich, dass

$$\frac{X_i}{n} \sim N(\mu/n, \sigma^2/n^2)$$

für jedes  $i \in \{1, \dots, n\}$ , vgl. auch das Beispiel in Abschnitt 3.4.2.

- Weil mit  $X_1, \dots, X_n$  auch die Zufallsvariablen  $X_1/n, \dots, X_n/n$  unabhängig sind, ergibt sich nun aus der Faltungsstabilität der Normalverteilung (vgl. Korollar 3.2), dass

$$\bar{X}_n = \frac{X_1}{n} + \dots + \frac{X_n}{n} \sim N(\mu, \sigma^2/n).$$

- Damit ist (37) bewiesen.
- Um (38) zu beweisen, betrachten wir die Identität

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n ((X_i - \bar{X}_n) + (\bar{X}_n - \mu))^2 \\ &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + 2(\bar{X}_n - \mu) \cdot \underbrace{\sum_{i=1}^n (X_i - \bar{X}_n)}_{=0} + n(\bar{X}_n - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2. \end{aligned}$$

- Weil die Zufallsvariablen  $\sigma^{-1}(X_1 - \mu), \dots, \sigma^{-1}(X_n - \mu)$  unabhängig und  $N(0, 1)$ -verteilt sind, ergibt sich somit aus der Definition der  $\chi^2$ -Verteilung, dass

$$\underbrace{\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2}_{\sim \chi_n^2} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma^2} + \underbrace{\left( \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \right)^2}_{\sim \chi_1^2}. \quad (39)$$

- Wegen Theorem 6.10 sind die beiden Summanden  $RS_1$  und  $RS_2$  auf der rechten Seite dieser Gleichung unabhängig.
- Wegen Theorem 5.18 gilt deshalb für die charakteristische Funktion  $\varphi_{LS}(t)$  der linken Seite LS von (39), dass

$$\varphi_{LS}(t) = \varphi_{RS_1}(t) \varphi_{RS_2}(t)$$

für jedes  $t \in \mathbb{R}$ , wobei  $\varphi_{RS_i}(t)$  die charakteristische Funktion von  $RS_i$  bezeichnet.

- Aus Theorem 6.5 folgt somit, dass für jedes  $t \in \mathbb{R}$

$$\begin{aligned}\varphi_{\text{RS}_1}(t) &= \frac{\varphi_{\text{LS}}(t)}{\varphi_{\text{RS}_2}(t)} \\ &= \frac{1}{(1 - 2it)^{(n-1)/2}}.\end{aligned}$$

- Die erneute Anwendung von Theorem 6.5 und des Eindeutigkeitsatzes für charakteristische Funktionen (vgl. Korollar 5.5) ergibt nun, dass

$$\frac{(n-1)S_n^2}{\sigma^2} = \text{RS}_1 \sim \chi_{n-1}^2.$$

□

### 6.3.4 t-Verteilung

Wir führen nun eine weitere Klasse von statistischen Prüfverteilungen ein, die wie folgt definiert sind.

**Definition** Sei  $r \in \mathbb{N}$  eine beliebige natürliche Zahl, und seien  $X$  und  $U_r$  unabhängige Zufallsvariablen mit  $X \sim N(0, 1)$  und  $U_r \sim \chi_r^2$ . Dann sagt man, dass die Zufallsvariable

$$V_r = X / \sqrt{\frac{U_r}{r}} \quad (40)$$

t-verteilt ist mit  $r$  Freiheitsgraden. (Schreibweise:  $V_r \sim t_r$ )

**Theorem 6.12** Sei  $V_r \sim t_r$ . Für die Dichte  $f_{V_r}(x)$  von  $V_r$  gilt dann

$$f_{V_r}(x) = \frac{\Gamma((r+1)/2)}{\Gamma(r/2)} \frac{1}{\sqrt{r\pi} (1 + x^2/r)^{(r+1)/2}} \quad (41)$$

für jedes  $x \in \mathbb{R}$ .

### Beweis

- Seien  $X$  und  $U_r$  unabhängige Zufallsvariablen mit  $X \sim N(0, 1)$  und  $U_r \sim \chi_r^2$ .
- Für die gemeinsame Dichte  $f_{(X, U_r)}(x, u)$  von  $(X, U_r)$  gilt dann:

$$f_{(X, U_r)}(x, u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \frac{u^{(r-2)/2} e^{-u/2}}{2^{r/2} \Gamma(r/2)} \quad (42)$$

für beliebige  $x \in \mathbb{R}$  und  $u > 0$ .

- Wir betrachten die Abbildung  $(x, u) \rightarrow (v, w)$  mit

$$v = \frac{x}{\sqrt{u/r}}, \quad w = u.$$

- Für die Jacobi-Determinante der Umkehrabbildung gilt  $(w/r)^{1/2}$ .



- Aus Theorem 6.9 ergibt sich somit für die (Rand-) Dichte  $f_{V_r}(v)$  von  $V_r$ , dass

$$\begin{aligned}
 f_{V_r}(v) &= \int_0^\infty f_{(X, U_r)}(v(w/r)^{1/2}, w) (w/r)^{1/2} dw \\
 &= \frac{1}{\sqrt{2\pi}} \frac{1}{2^{r/2} \Gamma(r/2)} \int_0^\infty \exp\left(-\frac{v^2 w}{2r}\right) w^{(r-2)/2} \exp\left(-\frac{w}{2}\right) \left(\frac{w}{r}\right)^{1/2} dw \\
 &= \frac{1}{\sqrt{2\pi}} \frac{1}{2^{r/2} r^{1/2} \Gamma(r/2)} \int_0^\infty \exp\left(-\frac{(1+v^2/r)w}{2}\right) w^{((r+1)/2)-1} dw.
 \end{aligned}$$

- Für das Integral ergibt sich aus der Darstellungsformel (24) der Gammafunktion, dass

$$\int_0^\infty \exp\left(-\frac{(1+v^2/r)w}{2}\right) w^{((r+1)/2)-1} dw = \frac{\Gamma((r+1)/2)}{\left(\frac{(1+v^2/r)}{2}\right)^{(r+1)/2}}.$$

- Hieraus folgt, dass für jedes  $v \in \mathbb{R}$

$$f_{V_r}(v) = \frac{1}{\sqrt{2\pi}} \frac{1}{2^{r/2} r^{1/2} \Gamma(r/2)} \frac{\Gamma((r+1)/2)}{\left(\frac{(1+v^2/r)}{2}\right)^{(r+1)/2}}.$$

□

### Beachte

- Sei  $V_r \sim t_r$ . Für  $\alpha \in (0, 1)$  wird dann die (eindeutige) Lösung  $t_{r,\alpha}$  der Gleichung  $F_{V_r}(t_{r,\alpha}) = \alpha$  das  $\alpha$ -Quantil der t-Verteilung mit  $r$  Freiheitsgraden genannt, vgl. Tabelle 3 in Abschnitt 7.
- Analog zu der Symmetrieeigenschaft  $z_\alpha = -z_{1-\alpha}$  der Quantile  $z_\alpha$  der Standardnormalverteilung gilt auch

$$t_{r,\alpha} = -t_{r,1-\alpha} \quad (43)$$

für beliebige  $r \in \mathbb{N}$  und  $\alpha \in (0, 1)$ , weil die in (41) gegebene Dichte der t-Verteilung eine bezüglich des Nullpunktes symmetrische Funktion ist.

**Theorem 6.13** Sei  $(X_1, \dots, X_n)$  eine normalverteilte Zufallsstichprobe mit  $X_i \sim N(\mu, \sigma^2)$  für  $i \in \{1, \dots, n\}$ . Dann gilt

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}. \quad (44)$$

### Beweis

- In Theorem 6.10 hatten wir gezeigt, dass die Zufallsvariablen  $\bar{X}_n$  und  $S_n^2$  unabhängig sind.
- Aus Theorem 3.17 (bzw. aus dessen vektorieller Version in Theorem 6.8) ergibt sich somit, dass auch die Zufallsvariablen

$$\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) \quad \text{und} \quad \sqrt{\frac{1}{n-1} \frac{(n-1)S_n^2}{\sigma^2}}$$

unabhängig sind.

- Außerdem ergibt sich aus Theorem 6.11, dass

$$\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) \sim N(0, 1)$$

und

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2.$$

- Deshalb ergibt sich unmittelbar aus der Definition der t-Verteilung, dass

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}}{\sqrt{\frac{1}{n-1} \frac{(n-1)S_n^2}{\sigma^2}}} \sim t_{n-1}.$$

□

### Beachte

- Theorem 6.13 bietet die Möglichkeit, ein (zufälliges) Intervall anzugeben, in dem die Abweichung  $\bar{X}_n - \mu$  des Stichprobenmittels  $\bar{X}_n$  von dem zu schätzenden Wert  $\mu$  mit einer vorgegebenen Wahrscheinlichkeit liegt.
- Von besonderer Bedeutung ist dabei die Tatsache, dass die Kenntnis der (im allgemeinen unbekannten) Parameter  $\mu$  und  $\sigma^2$  *nicht* zur Konstruktion dieses Intervalls erforderlich ist.
- Aus Theorem 6.13 ergibt sich nämlich auf die gleiche Weise wie bei der Herleitung von (20), dass

$$P\left(\frac{-t_{n-1, 1-\alpha/2} S_n}{\sqrt{n}} < \bar{X}_n - \mu < \frac{t_{n-1, 1-\alpha/2} S_n}{\sqrt{n}}\right) = 1 - \alpha \quad (45)$$

für jedes  $\alpha \in (0, 1)$ .

## 7 Tabellen für Verteilungsfunktionen und Quantile

**Tabelle 1** Verteilungsfunktion  $\Phi(x)$  der Standardnormalverteilung

$x$	0	1	2	3	4	5	6	7	8	9
0,0	0,500000	0,503989	0,507978	0,511967	0,515953	0,519939	0,523922	0,527903	0,531881	0,535856
0,1	0,539828	0,543795	0,547758	0,551717	0,555670	0,559618	0,563559	0,567495	0,571424	0,575345
0,2	0,579260	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,606420	0,610261	0,614092
0,3	0,617911	0,621719	0,625516	0,629300	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,691462	0,694974	0,698468	0,701944	0,705402	0,708840	0,712260	0,715661	0,719043	0,722405
0,6	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,758036	0,761148	0,764238	0,767305	0,770350	0,773373	0,776373	0,779350	0,782305	0,785236
0,8	0,788145	0,791030	0,793892	0,796731	0,799546	0,802338	0,805106	0,807850	0,810570	0,813267
0,9	0,815940	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913
1,0	0,841345	0,843752	0,846136	0,848495	0,850830	0,853141	0,855428	0,857690	0,859929	0,862143
1,1	0,864334	0,866500	0,868643	0,870762	0,872857	0,874928	0,876976	0,878999	0,881000	0,882977
1,2	0,884930	0,886860	0,888767	0,890651	0,892512	0,894350	0,896165	0,897958	0,899727	0,901475
1,3	0,903199	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914656	0,916207	0,917736
1,4	0,919243	0,920730	0,922196	0,923641	0,925066	0,926471	0,927855	0,929219	0,930563	0,931888
1,5	0,933193	0,934478	0,935744	0,936992	0,938220	0,939429	0,940620	0,941792	0,942947	0,944083
1,6	0,945201	0,946301	0,947384	0,948449	0,949497	0,950529	0,951543	0,952540	0,953521	0,954486
1,7	0,955435	0,956367	0,957284	0,958185	0,959071	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,964070	0,964852	0,965621	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,971284	0,971933	0,972571	0,973197	0,973810	0,974412	0,975002	0,975581	0,976148	0,976705
2,0	0,977250	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,986097	0,986447	0,986791	0,987126	0,987455	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,989276	0,989556	0,989830	0,990097	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576
2,4	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995339	0,995473	0,995603	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605
3,0	0,998650	0,998694	0,998736	0,998777	0,998817	0,998856	0,998893	0,998930	0,998965	0,998999
3,5	0,999767	0,999776	0,999784	0,999792	0,999800	0,999807	0,999815	0,999821	0,999828	0,999835
4,0	0,999968	0,999970	0,999971	0,999972	0,999973	0,999974	0,999975	0,999976	0,999977	0,999978

**Tabelle 2**  $\gamma$ -Quantil  $\chi^2_{r,\gamma}$  der  $\chi^2$ -Verteilung mit  $r$  Freiheitsgraden

$r \backslash \gamma$	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
1	.00004	.00016	.00098	.0039	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.1026	.2107	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.064	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.6	5.23	6.26	7.26	8.55	22.31	25	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

**Tabelle 3**  $\gamma$ -Quantil  $t_{r,\alpha}$  der t-Verteilung mit  $r$  Freiheitsgraden

$r \setminus \alpha$	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995
1	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656
2	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
30	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
60	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660
70	0,387	0,527	0,678	0,847	1,044	1,294	1,667	1,994	2,381	2,648
80	0,387	0,526	0,678	0,846	1,043	1,292	1,664	1,990	2,374	2,639
90	0,387	0,526	0,677	0,846	1,042	1,291	1,662	1,987	2,368	2,632
100	0,386	0,526	0,677	0,845	1,042	1,290	1,660	1,984	2,364	2,626
150	0,386	0,526	0,676	0,844	1,040	1,287	1,655	1,976	2,351	2,609
200	0,386	0,525	0,676	0,843	1,039	1,286	1,653	1,972	2,345	2,601
500	0,386	0,525	0,675	0,842	1,038	1,283	1,648	1,965	2,334	2,586
1000	0,385	0,525	0,675	0,842	1,037	1,282	1,646	1,962	2,330	2,581