

# Einführung in die Statistik

Skript zur Vorlesung  
im  
Herbstsemester 2018

Helmut Harbrecht

Stand: 19. Dezember 2018

# Vorwort

Diese Mitschrift kann und soll nicht ganz den Wortlaut der Vorlesung wiedergeben. Speziell soll sie kein Lehrbuch ersetzen. Vielmehr soll diese Mitschrift das Vorarbeiten des Vorlesungsstoffes ermöglichen. Dabei sei angemerkt, dass es sich bei den mit Sternchen versehenen Kapitel um ergänzenden Stoff handelt, der aus Zeitgründen nicht in der Vorlesung behandelt wird, jedoch ebenfalls sehr interessant und praxisrelevant ist.

**Literatur zur Vorlesung:** Maßgeblich ist das Stoff der Vorlesung. Der interessierte Leser findet in den nachfolgenden Lehrbüchern zusätzliche Informationen und vertiefende Einblicke in das Thema:

- Hans-Otto Georgii: *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik* (De Gruyter-Verlag)
- Ulrich Krengel: *Einführung in die Wahrscheinlichkeitstheorie und Statistik* (Teubner-Verlag)
- Norbert Henze: *Stochastik für Einsteiger* (Springer-Verlag)

# Inhaltsverzeichnis

<b>1</b>	<b>Wahrscheinlichkeitsräume</b>	<b>6</b>
1.1	Zufällige Ereignisse . . . . .	6
1.2	Rechnen mit zufälligen Ereignissen . . . . .	7
1.3	Rechnen mit Wahrscheinlichkeiten . . . . .	10
1.4	Grundformeln der Kombinatorik . . . . .	13
<b>2</b>	<b>Bedingte Wahrscheinlichkeiten und Unabhängigkeit</b>	<b>17</b>
2.1	Definition der bedingten Wahrscheinlichkeit . . . . .	17
2.2	Multiplikationsregeln . . . . .	19
2.3	Stochastische Unabhängigkeit . . . . .	23
2.4	Produktexperimente . . . . .	25
<b>3</b>	<b>Diskrete Verteilungen</b>	<b>27</b>
3.1	Zufallsgrößen . . . . .	27
3.2	Verteilungsfunktion . . . . .	29
3.3	Erwartungswert . . . . .	30
3.4	Varianz . . . . .	32
3.5	Schwaches Gesetz der großen Zahlen . . . . .	35
3.6	Binomialverteilung . . . . .	37
3.7	Poisson-Verteilung* . . . . .	39
3.8	Hypergeometrische Verteilung . . . . .	42
<b>4</b>	<b>Stetige Verteilungen</b>	<b>44</b>
4.1	Dichtefunktion . . . . .	44
4.2	Erwartungswert und Varianz . . . . .	45
4.3	Verteilungsfunktion . . . . .	48
4.4	Exponentialverteilung* . . . . .	49
4.5	Normalverteilung . . . . .	52
<b>5</b>	<b>Simulationsverfahren</b>	<b>58</b>
5.1	Pseudozufallszahlen . . . . .	58
5.2	Simulation beliebiger Verteilungen . . . . .	61
5.2.1	Diskrete Verteilungen . . . . .	61
5.2.2	Verwerfungsmethode . . . . .	62
5.2.3	Inversionsmethode . . . . .	63
5.3	Monte-Carlo-Verfahren . . . . .	65
<b>6</b>	<b>Beschreibende Statistik</b>	<b>67</b>

---

<b>7</b>	<b>Mathematische Statistik</b>	<b>73</b>
7.1	Grundbegriffe . . . . .	73
7.2	Punktschätzungen . . . . .	74
7.3	Maximum-Likelihood-Methode . . . . .	77
7.4	Verteilungen wichtiger Stichproben . . . . .	79
7.4.1	Binomialverteilte Grundgesamtheit . . . . .	79
7.4.2	Normalverteilte Grundgesamtheit . . . . .	79
7.5	Konfidenzintervalle . . . . .	82
7.5.1	Konfidenzintervalle bei binomialverteilter Grundgesamtheit . . . . .	83
7.5.2	Konfidenzintervalle bei normalverteilter Grundgesamtheit . . . . .	84
7.6	Tests . . . . .	86
7.6.1	Einführung . . . . .	86
7.6.2	Allgemeines Testschema . . . . .	87
7.6.3	Tests bei normalverteilter Grundgesamtheit . . . . .	87
7.6.4	$\chi^2$ -Test . . . . .	90
<b>8</b>	<b>Markov-Ketten</b>	<b>93</b>

# Einführung

## Was ist Stochastik?

Stochastik ist die “Mathematik des Zufalls” und wird in die Bereiche Wahrscheinlichkeitsrechnung und Statistik unterteilt. Bei der Wahrscheinlichkeitsrechnung werden zufällige Prozesse mit bekannten Wahrscheinlichkeiten untersucht. Die Statistik hingegen zieht Rückschlüsse aus beobachteten Daten. Die Wahrscheinlichkeitsrechnung ist die Grundlage für die Statistik.

## Beispiele

- Werde ich in der nächsten Ziehung “6 aus 49” einen Hauptgewinn haben?
- Wird die Mittagstemperatur morgen über 5° Celsius liegen?
- Wird mein Computer mehr als zwei Jahre fehlerfrei arbeiten?
- Wie groß ist die Wahrscheinlichkeit, dass wenigstens zwei von  $n$  Personen am gleichen Tag Geburtstag haben?

# 1. Wahrscheinlichkeitsräume

## 1.1 Zufällige Ereignisse

Um die Konzepte in der Wahrscheinlichkeitstheorie zu motivieren, werden wir zunächst eine umgangssprachliche Erklärung des Begriffs *Wahrscheinlichkeitsraum* geben. Am Ende des Kapitels wird dann eine mathematische Definition folgen.

**Wahrscheinlichkeitsraum:** Ein Wahrscheinlichkeitsraum ist die Zusammenfassung aller Teile eines mathematischen Modells zur Beschreibung einer Zufallssituation. Verschiedene Zufallssituationen führen auf verschiedene Wahrscheinlichkeitsräume. Ein Wahrscheinlichkeitsraum wird charakterisiert durch ein Tupel  $(\Omega, \mathcal{A}, \mathbb{P})$ , wobei wir die Bedeutung von  $\Omega$ ,  $\mathcal{A}$  und  $\mathbb{P}$  im folgenden klären werden.

**Zufallssituation:** Eine Zufallssituation ist gekennzeichnet durch zwei Eigenschaften:

- sie ist beliebig oft und gleichartig wiederholbar (zumindest gedanklich),
- ihr Ergebnis ist absolut nicht vorhersagbar.

Beispielsweise stellt das Werfen eines Würfels, das Lottospiel oder auch der Lebenszyklus eines Produktes eine Zufallssituation dar.

**Versuch:** Ein Versuch ist eine Realisierung einer Zufallssituation, wobei wir mit

- $\omega$  das Ergebnis des Versuchs und
- $\Omega$  die Menge aller möglichen Ergebnisse

bezeichnen. Dabei wollen wir annehmen, dass sich jedes Ergebnis des Versuchs eindeutig einem Element  $\omega$  der Ergebnismenge  $\Omega$  zuordnen lässt.

### Beispiele 1.1 (Zufallssituationen und Ergebnismengen)

1. Beim Werfen eines Würfels ist  $\Omega = \{1, 2, 3, 4, 5, 6\}$  eine endliche Ergebnismenge mit 6 möglichen verschiedenen Ergebnissen.
2. Wir betrachten die Lebensdauer einer Glühlampe. Die Ergebnismenge  $\Omega = \{\omega \in \mathbb{R} : \omega \geq 0\}$  ist überabzählbar unendlich, wobei das Ergebnis  $\omega \in \Omega$  der Lebensdauer der Glühlampe entspricht.
3. Bei der Überprüfung von  $n$  verschiedenen Geräten sei das Ergebnis  $\omega_i$  definiert gemäß

$$\omega_i = \begin{cases} 0, & i\text{-tes Gerät defekt,} \\ 1, & i\text{-tes Gerät in Ordnung.} \end{cases}$$

Dann ist  $\Omega = \{(\omega_1, \dots, \omega_n) \in \{0, 1\}^n\}$  eine endliche Ergebnismenge mit  $2^n$  Elementen.

△

**Definition 1.2 (Ereignis)** Ein **zufälliges Ereignis** ist eine Teilmenge  $A \subset \Omega$ . Wenn  $\omega \in A$  gilt, so sagt man, dass das Ereignis  $A$  **eingetreten** ist. Nicht jede Teilmenge  $A \subset \Omega$  muss sich als zufälliges Ereignis betrachten lassen, aber alle zufälligen Ereignisse sind Teilmengen von  $\Omega$ .

### Beispiele 1.3 (Zufällige Ereignisse (Fortsetzung der Beispiele 1.1))

1. Beim Werfen eines Würfels entspricht das Ereignis

$A \hat{=} \text{“es wird eine gerade Zahl gewürfelt”}$

der Menge  $A = \{2, 4, 6\}$ .

2. Für eine Glühlampe wird das Ereignis

$A \hat{=} \text{“Brenndauer liegt zwischen 500 und 5000 Stunden”}$

beschrieben durch

$$A = \{\omega \in \mathbb{R} : 500 \leq \omega \leq 5000\} \subset \Omega = \mathbb{R}_{\geq 0}.$$

3. Bei der Überprüfung von  $n$  Geräten gilt für das Ereignis

$A \hat{=} \text{“es funktionieren mindestens 2 Geräte”}$

die Beziehung

$$A = \left\{ (\omega_1, \dots, \omega_n) \in \{0, 1\}^n : \sum_{i=1}^n \omega_i \geq 2 \right\}.$$

△

## 1.2 Rechnen mit zufälligen Ereignissen

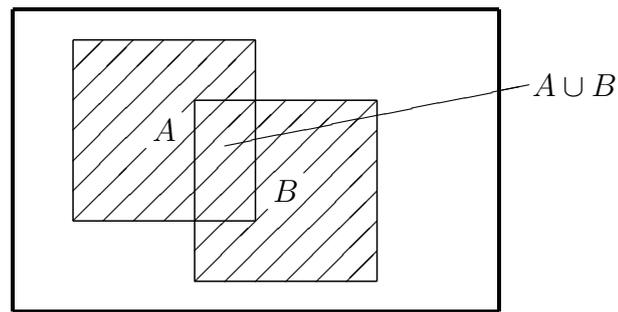
Wir haben zufällige Ereignisse als Mengen modelliert. Oftmals interessiert jedoch neben dem Eintreten eines Ereignisses  $A$  auch das Nicht-Eintreten von  $A$ . Bei zwei Ereignissen  $A$  und  $B$  können wir uns fragen, wann beide eintreten oder mindestens eines von beiden. Derartige Operationen mit Ereignissen können auf Operationen mit Mengen zurückgeführt werden, wobei diese im Kontext der Wahrscheinlichkeitsrechnung eine eigene Bedeutung besitzen. Nachfolgend findet sich eine Zusammenstellung der mengentheoretischen Sprechweisen und Operationen in Sprechweisen und Operationen mit Ereignissen.

### Bezeichnungen:

- “ $A$  oder  $B$ ”: Dieses Ereignis tritt ein, wenn entweder  $A$  oder  $B$  oder beide Ereignisse  $A$  und  $B$  eintreten, kurz

$$\omega \in A \cup B.$$

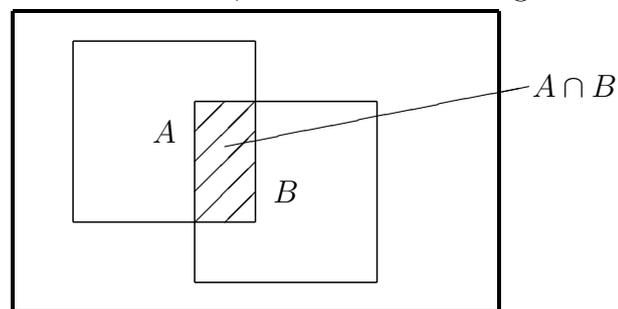
Graphisch lässt sich dies wie folgt darstellen:



- “*A und B*”: Dieses Ereignis tritt ein, wenn  $A$  und  $B$  gleichzeitig eintreten, kurz

$$\omega \in A \cap B.$$

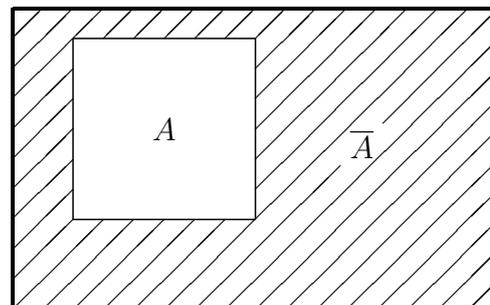
Für die Mengen  $A$  und  $B$  von oben, erhalten wir die folgende Illustration:



- “*nicht A bzw.  $\bar{A}$* ”: Dieses Ereignis tritt ein, wenn  $A$  nicht eintritt, kurz

$$\omega \notin A \Leftrightarrow \omega \in \Omega \setminus A =: \bar{A}.$$

Hierbei nennen wir  $\bar{A}$  das *Komplementärereignis* von  $A$ , das wie folgt veranschaulicht werden kann:



- “*A zieht B nach sich*”: Ist  $A$  eine Teilmenge von  $B$ , kurz  $A \subset B$ , dann tritt mit dem Ereignis  $A$  stets auch das Ereignis  $B$  ein:

$$\omega \in A \Rightarrow \omega \in B.$$

- “*sicheres Ereignis*”:  $A$  heißt das sichere Ereignis, falls gilt

$$A = \Omega.$$

- “*unmögliches Ereignis*”:  $A$  heißt das unmögliche Ereignis, falls gilt

$$A = \bar{\Omega} = \emptyset.$$

- “*Elementarereignis*”: Ein Elementarereignis ist eine einelementige Menge mit dem Ergebnis  $\omega \in \Omega$ , kurz

$$A = \{\omega\}.$$

- “*unvereinbares Ereignis*”:  $A$  und  $B$  heißen unvereinbar, wenn gilt

$$A \cap B = \emptyset.$$

Die “und”- beziehungsweise “oder”-Operationen können auch auf endlich viele oder abzählbar unendlich viele Ereignisse  $A_i$  angewandt werden im Sinne von

$$\bigcup_{i=1}^n A_i, \quad \bigcap_{i=1}^n A_i \quad \text{beziehungsweise} \quad \bigcup_{i=1}^{\infty} A_i, \quad \bigcap_{i=1}^{\infty} A_i.$$

Hierbei ist ein Ergebnis  $\omega \in \Omega$  in  $\bigcup_{i=1}^{\infty} A_i$  enthalten, wenn es ein  $j \in \mathbb{N}$  gibt, so dass  $\omega \in A_j$  gilt. Umgekehrt ist das Ergebnis  $\omega \in \Omega$  in  $\bigcap_{i=1}^{\infty} A_i$  enthalten, wenn  $\omega \in A_i$  für alle  $i \in \mathbb{N}$  gilt.

**Satz 1.4 (Rechenregeln für zufällige Ereignisse)** Für das Rechnen mit zufälligen Ereignissen  $A$ ,  $B$  und  $C$  gelten die folgenden Regeln.

1. Kommutativgesetz:

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

2. Assoziativgesetz:

$$(A \cup B) \cup C = A \cup (B \cup C)$$

$$(A \cap B) \cap C = A \cap (B \cap C)$$

3. Distributivgesetz:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

4. De Morgansche Regeln:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

5. Für das unmögliche Ereignis und das sichere Ereignis gelten die Rechenregeln:

$$A \cup \emptyset = A, \quad A \cup \Omega = \Omega$$

$$A \cap \emptyset = \emptyset, \quad A \cap \Omega = A$$

Die De Morganschen Regeln können für endlich viele oder abzählbar unendlich viele Ereignisse  $A_i$  verallgemeinert werden. Es gilt

$$\overline{A_1 \cup \dots \cup A_n} = \bar{A}_1 \cap \dots \cap \bar{A}_n \quad \text{und} \quad \overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \bar{A}_i.$$

Die verbale Version hierzu lautet: Es tritt genau dann *nicht* mindestens eines der Ereignisse  $A_1, A_2, \dots$  ein, wenn keines dieser Ereignisse eintritt. Dies bedeutet, weder  $A_1$  noch  $A_2$  noch  $\dots$  treten ein. Analog ist natürlich auch

$$\overline{A_1 \cap \dots \cap A_n} = \overline{A_1} \cup \dots \cup \overline{A_n} \quad \text{und} \quad \overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \overline{A_i}.$$

Wir wollen nun die nächste Zutat für den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$  einführen, nämlich die  $\sigma$ -Algebra  $\mathcal{A}$ . Dabei handelt es sich um ein System von Mengen, das alle zufälligen Ereignisse enthält.

**Definition 1.5 ( $\sigma$ -Algebra)** Die Menge  $\mathcal{A}$  von Teilmengen der Ergebnismenge  $\Omega$ , welche die zufälligen Ereignisse beschreiben, heißt  **$\sigma$ -Algebra** beziehungsweise **Ereignisalgebra** bezogen auf eine feste Zufallssituation, wenn gilt:

1.  $\Omega \in \mathcal{A}$ ,
2.  $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$ ,
3.  $A_i \in \mathcal{A} \forall i \in \mathbb{N} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ .

Wie man sich leicht überlegt, gelten für eine  $\sigma$ -Algebra  $\mathcal{A}$  zusätzlich die Aussagen:

- Aus  $\Omega \in \mathcal{A}$  folgt wegen der zweiten Eigenschaft, dass auch  $\emptyset = \overline{\Omega} \in \mathcal{A}$ .
- Sind die Mengen  $A_i \in \mathcal{A}$  für alle  $i \in \mathbb{N}$ , so ergibt sich aus der zweiten Eigenschaft auch  $\overline{A_i} \in \mathcal{A}$  für alle  $i \in \mathbb{N}$ . Daher folgt aus der dritten Eigenschaft

$$\overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \overline{A_i} \in \mathcal{A}.$$

Dies bedeutet:  $A_i \in \mathcal{A} \forall i \in \mathbb{N} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$ .

- Gilt  $A, B \in \mathcal{A}$ , so ist wegen der zweiten Eigenschaft und dem soeben gezeigten auch  $A \setminus B = A \cap \overline{B} \in \mathcal{A}$ .

Zusammen mit den Eigenschaften 1–3 aus Definition 1.5 ist daher offensichtlich sichergestellt, dass jedes mögliche Ergebnis von endlich vielen oder abzählbar unendlich vielen Mengenoperationen wieder in der  $\sigma$ -Algebra liegt.

## 1.3 Rechnen mit Wahrscheinlichkeiten

$A \in \mathcal{A}$  sei ein festes Ereignis innerhalb einer Zufallssituation. Wir betrachten  $n$  unabhängige, das heißt sich gegenseitig nicht beeinflussende, Versuche. Mit  $h_n(A)$  bezeichnen wir die *absolute Häufigkeit* des Ereignisses  $A$ , dies ist die Anzahl des Eintretens von  $A$  bei  $n$  Versuchen.

**Definition 1.6 (relative Häufigkeit)** Wir bezeichnen mit

$$H_n(A) = \frac{h_n(A)}{n}$$

die **relative Häufigkeit** für das Eintreten von  $A$  bei  $n$  Versuchen.

Die Erfahrung lehrt, dass für  $n \rightarrow \infty$  die relative Häufigkeit gegen eine feste Zahl  $\mathbb{P}(A)$  strebt. Dies nennt man den Stabilisierungseffekt der Folge  $H_n(A)$  der relativen Häufigkeiten.

**Eigenschaften:** Offensichtlich erfüllt die relative Häufigkeit die folgenden Eigenschaften:

1. *Positivität:* Die relative Häufigkeit ist nichtnegativ. Es gilt demnach stets  $0 \leq H_n(A)$  für alle  $A$ .
2. *Normierung:*  $H_n(\Omega) = 1$ .
3. *Additivität:*  $A, B$  seien unvereinbar, das heißt  $A \cap B = \emptyset$ , dann gilt

$$H_n(A \cup B) = \frac{h_n(A) + h_n(B)}{n} = \frac{h_n(A)}{n} + \frac{h_n(B)}{n} = H_n(A) + H_n(B).$$

Dies bedeutet, dass sich für unvereinbare Ereignisse die relativen Häufigkeiten addieren.

Diese Beobachtungen bilden die Grundlage des Kolmogorovschen Axiomsystems zum Rechnen mit Wahrscheinlichkeiten.

**Definition 1.7 (Kolmogorovsches Axiomsystem, 1933)** Gegeben sei eine Ergebnismenge  $\Omega$  und eine  $\sigma$ -Algebra  $\mathcal{A}$ , welche eine Zufallssituation beschreiben. Dann ist jedem Ereignis  $A \in \mathcal{A}$  eine reelle Zahl  $\mathbb{P}(A)$  zugeordnet, die **Wahrscheinlichkeit** des Ereignisses  $A$ . Dabei gelten die folgenden Axiome:

- (A1) *Positivität:*  $0 \leq \mathbb{P}(A)$  für alle  $A \in \mathcal{A}$ .
- (A2) *Normierung:*  $\mathbb{P}(\Omega) = 1$ .
- (A3)  *$\sigma$ -Additivität:* Für eine Folge von Ereignissen  $A_i \in \mathcal{A}$ ,  $i \in \mathbb{N}$ , welche paarweise unvereinbar sind, das heißt  $A_i \cap A_j = \emptyset$  für  $i \neq j$ , ist

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

**Folgerungen aus den Axiomen (Rechenregeln):**

1.  $\mathbb{P}(\emptyset) = 0$

*Beweis.* Die Ereignisse  $A_1 := \Omega$  und  $A_i := \emptyset$ ,  $i > 1$ , sind paarweise unvereinbar wegen  $A_i \cap A_j = \emptyset$ ,  $i \neq j$ . Aus den Axiomen A1 und A3 folgt, dass

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = 1 + \sum_{i=2}^{\infty} \mathbb{P}(\emptyset).$$

Hieraus ergibt sich die Behauptung  $\mathbb{P}(\emptyset) = 0$ . □

2. Für jede endliche Folge von paarweise unvereinbaren Ereignissen  $A_i \in \mathcal{A}$  gilt

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

*Beweis.* Setzen wir  $A_i := \emptyset$  für alle  $i > n$ , so folgt die Behauptung sofort aus Axiom A3.  $\square$

$$3. \mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$$

*Beweis.* Wegen  $A \cup \bar{A} = \Omega$  und  $A \cap \bar{A} = \emptyset$  folgt nach den Axiomen A2 und A3

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(\bar{A}).$$

 $\square$ 

$$4. \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

*Beweis.* Wir definieren  $C := B \cap \bar{A} \subset B$ . Es folgt  $A \cup B = A \cup C$  und  $A \cap C = \emptyset$ , und daher mit Axiom A3

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(C). \quad (1.1)$$

Weiter gilt für  $D := A \cap B$ , dass  $B = C \cup D$  und  $C \cap D = \emptyset$ , und folglich

$$\mathbb{P}(B) = \mathbb{P}(C) + \mathbb{P}(D) = \mathbb{P}(C) + \mathbb{P}(A \cap B) \quad (1.2)$$

gemäß Axiom A3. Aus (1.1) und (1.2) ergibt sich dann die Behauptung.  $\square$

5. Bilden die Ereignisse  $A_i \in \mathcal{A}$  eine *Zerlegung von  $\Omega$* , dies bedeutet die  $A_i$  sind jeweils paarweise unvereinbare Ereignisse und  $\Omega = \bigcup_{i=1}^{\infty} A_i$ , dann gilt

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

*Beweis.* Aus den Axiomen A2 und A3 folgt

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i).$$

 $\square$ 

Den *klassischen* Wahrscheinlichkeitsbegriff werden wir dann zur Berechnung von Wahrscheinlichkeiten heranziehen, wenn ein Versuch nur endlich viele gleichmögliche Elementarereignisse besitzt. Wir sprechen hier von *Laplace-Modellen*:

**Satz 1.8 (Laplace-Experiment)** Die Ergebnismenge  $\Omega$  erfülle die folgenden beiden Voraussetzungen:

- a)  $\Omega = \{\omega_1, \dots, \omega_n\}$  besteht aus  $n$  Elementarereignissen,
- b) alle Elementarereignisse sind gleich wahrscheinlich, dies bedeutet

$$\mathbb{P}(\{\omega_1\}) = \dots = \mathbb{P}(\{\omega_n\}) = \frac{1}{n}.$$

Dann gilt für ein Ereignis  $A = \{\omega_{i_1}, \dots, \omega_{i_m}\} \subset \Omega$

$$\mathbb{P}(A) = \frac{m}{n}.$$

*Beweis.* Die Behauptung folgt sofort aus den Kolmogorovschen Axiomen.  $\square$

**Bemerkung 1.9** Die Aussage von Satz 1.8 kann auch kurz dargestellt werden durch

$$\mathbb{P}(A) = \frac{\text{Anzahl der für } A \text{ günstigen Ereignisse}}{\text{Anzahl aller möglichen Ereignisse}}.$$

$\triangle$

**Beispiel 1.10 (Werfen eines idealem Würfel)** Für die Ergebnismenge beim einmaligen Werfen eines idealen Würfels gilt  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Das Ereignis

$$A \hat{=} \text{“Primzahl wird gewürfelt”}$$

ist durch die Menge  $A = \{2, 3, 5\}$  gegeben. Damit gilt also

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

$\triangle$

Wir wollen die bisher vorgestellten Begriffe zusammenführen und den Wahrscheinlichkeitsraum streng mathematisch definieren. Zu diesem Zweck erinnern wir uns, dass folgende drei Komponenten zum Beschreiben des mathematische Modells einer Zufallssituation benötigt wurden:

- *Ergebnismenge*  $\Omega$ : Dies ist eine nichtleere Menge, deren Elemente die möglichen Versuchsausgänge darstellen.
- *$\sigma$ -Algebra*  $\mathcal{A}$ : Dies ist die Menge aller Teilmengen von  $\Omega$ , welche zufällige Ereignisse bilden und für die die Eigenschaften 1–3 aus Definition 1.5 gelten.
- *Wahrscheinlichkeitsmaß*  $\mathbb{P}$ : Jedem Ereignis  $A \in \mathcal{A}$  ist in eindeutiger Weise eine Zahl  $\mathbb{P}(A)$  zugeordnet, die Wahrscheinlichkeit genannt wird.

Durch die Zusammenfassung dieser drei Zutaten ergibt sich schließlich der Wahrscheinlichkeitsraum:

**Definition 1.11 (Wahrscheinlichkeitsraum)** Gegeben sei die Ergebnismenge  $\Omega$ , eine  $\sigma$ -Algebra  $\mathcal{A}$  über  $\Omega$  und ein Wahrscheinlichkeitsmaß  $\mathbb{P}$ . Das Tupel  $(\Omega, \mathcal{A}, \mathbb{P})$  heißt **Wahrscheinlichkeitsraum** der gegebenen Zufallssituation.

## 1.4 Grundformeln der Kombinatorik

Wir betrachten eine Urne mit  $n$  unterscheidbaren (zum Beispiel durch Numerierung) Kugeln. Wir wollen die Anzahl der Möglichkeiten beim Ziehen von  $m$  Kugeln bestimmen. Dabei müssen wir berücksichtigen, ob eine entnommene Kugel vor der Entnahme der nächsten Kugel wieder zurückgelegt wird und ob die Reihenfolge der Kugeln eine Rolle spielt.

Wenn die Reihenfolge der Entnahme wichtig ist, so spricht man von *Variationen*. Spiel die Reihenfolge der Entnahme aber keine Rolle, so spricht man von *Kombinationen*.

**Variationen:** Für die Anzahl der Variationen mit Zurücklegen und ohne Zurücklegen der Kugeln erhalten wir:

- Anzahl der Möglichkeiten mit Zurücklegen: Beim ersten Ziehen können wir aus  $n$  Kugeln auswählen. Da wir die Kugel wieder in die Urne zurücklegen, können wir beim zweiten Ziehen ebenfalls aus  $n$  Kugeln auswählen und so weiter. Es gibt demnach

$$\underbrace{n \cdot n \cdot \dots \cdot n}_{m \text{ Faktoren}} = n^m$$

verschiedene Variationen.

- Anzahl der Möglichkeiten ohne Zurücklegen: Beim ersten Ziehen können wir aus  $n$  Kugeln auswählen. Da wir nun aber die Kugel nicht wieder in die Urne zurücklegen, können wir beim zweiten Ziehen nur aus  $n - 1$  Kugeln auswählen. Da sich beim  $k$ -ten Ziehen gerade  $n - k + 1$  Kugeln in der Urne befinden, folgern wir, dass es beim Ziehen ohne Zurücklegen insgesamt

$$\underbrace{n \cdot (n - 1) \cdot \dots \cdot (n - m + 1)}_{m \text{ Faktoren}} = \frac{n!}{(n - m)!}$$

verschiedene Variationen gibt.

**Kombinationen:** Wir bestimmen nun die Anzahl der Kombinationen mit Zurücklegen und ohne Zurücklegen.

- Anzahl der Möglichkeiten ohne Zurücklegen:  $\binom{n}{m} = \frac{n!}{(n - m)!m!}$

*Beweis.* Mit  $C_n^m$  bezeichnen wir die Anzahl von Möglichkeiten beim Ziehen von  $m$  aus  $n$  Kugeln ohne Beachtung der Reihenfolge und ohne Zurücklegen.

Es gilt  $C_1^0 = 1$  und  $C_n^n = 1$ , ferner setzen wir  $C_n^m := 0$  im Falle  $m < 0$ . Wir erhalten

$$\begin{aligned} C_{n+1}^m &= |\{ \underbrace{(a_1, \dots, a_m)}_{a_i \hat{=} \text{Nummer der } i\text{-ten Kugel}} : 1 \leq a_1 < a_2 < \dots < a_m \leq n + 1 \}| \\ &= |\{ \underbrace{(a_1, \dots, a_m) : 1 \leq a_1 < a_2 < \dots < a_m \leq n}_{=C_n^m} \}| \\ &\quad + |\{ \underbrace{(a_1, \dots, a_{m-1}, n + 1) : 1 \leq a_1 < \dots < a_{m-1} \leq n}_{=C_n^{m-1}} \}| \\ &= C_n^{m-1} + C_n^m. \end{aligned}$$

Wie man leicht mit vollständiger Induktion zeigt, entspricht dies der Rekursion

$$\binom{n}{m-1} + \binom{n}{m} = \binom{n+1}{m},$$

dies bedeutet  $C_n^m = \binom{n}{m}$ . □

- Anzahl der Möglichkeiten mit Zurücklegen:  $\binom{n+m-1}{m}$

*Beweis.* Die Menge aller  $\{(a_1, \dots, a_m) : 1 \leq a_1 \leq \dots \leq a_m \leq n\}$  wird durch  $b_i := a_i + i - 1$  bijektiv auf die Menge aller Tupel

$$\{(b_1, \dots, b_m) : 1 \leq b_1 < b_2 < \dots < b_m \leq n + m - 1\}$$

abgebildet. Letztere Menge hat aber die Mächtigkeit  $\binom{n+m-1}{m}$ .  $\square$

### Zusammenfassung:

Anzahl der Möglichkeiten	Variationen (mit Beachtung der Reihenfolge)	Kombinationen (ohne Beachtung der Reihenfolge)
mit Zurücklegen	$n^m$	$\binom{n+m-1}{m}$
ohne Zurücklegen	$\frac{n!}{(n-m)!}$	$\binom{n}{m}$

**Beispiel 1.12 (Geburtstagsparadoxon)** Wir wollen die Wahrscheinlichkeit dafür bestimmen, dass zwei Personen in einem Raum am gleichen Tag Geburtstag haben. Dabei setzen wir die folgenden Annahmen voraus:

- $n$  Personen sind im Raum,
- keiner hat an einem 29. Februar Geburtstag,
- die übrigen 365 Tage seien als Geburtstag gleichwahrscheinlich.

Wir interessieren uns für das Ereignis

$$A \hat{=} \text{“mindestens 2 Personen haben am gleichen Tag Geburtstag”}$$

bzw. für die Wahrscheinlichkeit  $\mathbb{P}(A)$ .  $A$  tritt also ein, wenn 2, 3, 4, ... Personen am selben Tag Geburtstag haben. Um die Anzahl dieser Möglichkeiten einzugrenzen, bietet es sich an, mit dem Komplementäreignis zu arbeiten, dies ist

$$\begin{aligned} \bar{A} &\hat{=} \text{“keiner hat am gleichen Tag Geburtstag”} \\ &\hat{=} \text{“alle Geburtstage sind verschieden”}. \end{aligned}$$

Die dritte Annahme sichert uns zu, dass ein Laplace-Experiment vorliegt, womit sich

$$\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A}) = 1 - \frac{\text{Anzahl der günstigen Ereignisse}}{\text{Anzahl aller Ereignisse}}$$

ergibt.

Als nächstes stellt sich die Frage nach dem zu wählenden Urnenmodell. Das Modell “Ziehen mit Zurücklegen unter Beachtung der Reihenfolge” ist das passende, um den Ergebnisraum  $\Omega$  zu charakterisieren, während das Modell “Ziehen ohne Zurücklegen unter Beachtung der Reihenfolge” auf die für  $\bar{A}$  günstigen Ereignisse zutrifft. Es folgt daher

$$\mathbb{P}(A) = 1 - \frac{\overbrace{365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - n + 1)}^{n \text{ Faktoren}}}{365^n}.$$

Für  $n = 23$  Personen folgt beispielsweise  $\mathbb{P}(A) = 0.507$ , während sich für  $n = 70$  Personen schon  $\mathbb{P}(A) = 0.999$  ergibt.  $\triangle$

**Beispiel 1.13 (Lotto (6 aus 49))** Beim Lotto werden 6 Kugeln aus einer Urne mit 49 Kugeln ohne Zurücklegen und ohne Beachtung der Reihenfolge gezogen. Für das Ereignis

$$A \hat{=} \text{“genau } k \text{ Richtige getippt”}$$

gilt

$$\mathbb{P}(A) = \frac{\text{Anzahl der günstigen Ereignisse}}{\text{Anzahl aller Ereignisse}} = \frac{\overbrace{\binom{6}{k}}^{\text{Richtige}} \overbrace{\binom{43}{6-k}}^{\text{Nieten}}}{\binom{49}{6}}.$$

△

## 2. Bedingte Wahrscheinlichkeiten und Unabhängigkeit

### 2.1 Definition der bedingten Wahrscheinlichkeit

Gegeben sei ein Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$  zu einer Zufallssituation und ein festes Ereignis  $A \in \mathcal{A}$  mit der Wahrscheinlichkeit  $\mathbb{P}(A)$ . Es kann sein, dass die Wahrscheinlichkeit sich verändert, wenn man beachtet, dass ein anderes Ereignis  $B \in \mathcal{A}$  bereits eingetreten ist.

**Definition 2.1** (bedingte Wahrscheinlichkeit) Es seien  $A$  und  $B$  mit  $\mathbb{P}(B) > 0$  zufällige Ereignisse, die zu einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$  gehören. Dann heißt die Größe

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (2.1)$$

bedingte Wahrscheinlichkeit des Ereignisses  $A$  unter der Bedingung, dass  $B$  bereits eingetreten ist.

**Motivation über relative Häufigkeiten:** Es werden  $n$  Versuche durchgeführt, wobei  $h_n(A)$ ,  $h_n(B)$  und  $h_n(A \cap B)$  jeweils die Anzahl der Versuche angeben, bei denen  $A$ ,  $B$  bzw. beide Ereignisse eintreten. Die relative Häufigkeit für  $A$  unter der Bedingung  $B$ , kurz  $A|B$ , ergibt sich dann zu

$$H_n(A|B) = \frac{h_n(A \cap B)}{h_n(B)} = \frac{\frac{h_n(A \cap B)}{n}}{\frac{h_n(B)}{n}} = \frac{H_n(A \cap B)}{H_n(B)}.$$

**Beispiel 2.2** (Würfelproblem) Mit einem idealen Würfel werden zwei Würfel ausgeführt. Dabei bezeichnen  $A$  und  $B$  die folgenden Ereignisse

- $A \hat{=}$  "Ereignis, dass beim ersten Wurf eine 6 gewürfelt wird",
- $B \hat{=}$  "Ereignis, dass die Augensumme beider Würfel 8 ist".

Die Augensumme zweier Würfel kann durch folgendes Tableau veranschaulicht werden:

	1	2	3	4	5	6	(2. Wurf)
1	2	3	4	5	6	7	
2	3	4	5	6	7	8	
3	4	5	6	7	8	9	
4	5	6	7	8	9	10	
5	6	7	8	9	10	11	
6	7	8	9	10	11	12	
(1. Wurf)							

Wir sehen, dass fünf der insgesamt 36 Elementarereignisse günstig für  $B$  sind, das heißt, es gilt  $\mathbb{P}(B) = \frac{5}{36}$ . Weiter ist  $\mathbb{P}(A) = \frac{1}{6}$ . Nur im Fall, dass erst eine 6 und dann eine 2 gewürfelt wird, treten  $A$  und  $B$  ein, dies bedeutet  $\mathbb{P}(A \cap B) = \frac{1}{36}$  und  $\mathbb{P}(A|B) = \frac{1}{5}$ . Die Gleichung

$$\mathbb{P}(A|B) = \frac{\frac{1}{36}}{\frac{5}{36}} = \frac{1}{5}$$

bestätigt Definition 2.1. △

### Rechenregeln für bedingte Wahrscheinlichkeiten:

1.  $\mathbb{P}(C|C) = 1$

*Beweis.* Es gilt

$$\mathbb{P}(C|C) = \frac{\mathbb{P}(C \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(C)}{\mathbb{P}(C)} = 1.$$

□

2.  $\mathbb{P}(A|C) = 1 - \mathbb{P}(\bar{A}|C)$

*Beweis.* Wegen  $A \cup \bar{A} = \Omega$  folgt  $(A \cap C) \cup (\bar{A} \cap C) = C$  und weiter

$$\begin{aligned} 1 &= \frac{\mathbb{P}(C)}{\mathbb{P}(C)} = \frac{\mathbb{P}((A \cap C) \cup (\bar{A} \cap C))}{\mathbb{P}(C)} = \frac{\mathbb{P}(A \cap C) + \mathbb{P}(\bar{A} \cap C)}{\mathbb{P}(C)} \\ &= \mathbb{P}(A|C) + \mathbb{P}(\bar{A}|C) \end{aligned}$$

□

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

*Beweis.*

$$\begin{aligned} \mathbb{P}(A \cup B|C) &= \frac{\mathbb{P}(\overbrace{(A \cup B) \cap C}^{=(A \cap C) \cup (B \cap C)}})}{\mathbb{P}(C)} \\ &= \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} + \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} - \frac{\mathbb{P}(\overbrace{(A \cap C) \cap (B \cap C)}^{=(A \cap B) \cap C})}{\mathbb{P}(C)} \\ &= \mathbb{P}(A|C) + \mathbb{P}(B|C) - \mathbb{P}(A \cap B|C). \end{aligned}$$

□

**Bemerkung 2.3** Die Rechenregeln für die bedingte Wahrscheinlichkeit  $\mathbb{P}(A|C)$  sind wie für  $\mathbb{P}(A)$ , das heißt, im ersten Argument darf wie bisher umgeformt werden. Nichts an der Bedingung umformen, dies geht i.a. schief!  $\triangle$

**Beispiel 2.4 (Stapelsuchproblem)** Thorsten durchsucht 7 gleichgroße Stapel von CDs nach einer ganz bestimmten. Die Wahrscheinlichkeit, dass die CD überhaupt in einem Stapel vorhanden ist, sei 0.8. Es wurden bereits 6 Stapel erfolglos durchsucht. Wie groß ist die Wahrscheinlichkeit, die CD im 7. Stapel zu finden?

Wir legen zunächst die Ereignisse fest:

$$A_i \hat{=} \text{“CD ist im } i\text{-ten Stapel”},$$

wobei  $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \dots = \mathbb{P}(A_7) = p$  gilt.

Aus

$$\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_7) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_7) = 7p \stackrel{!}{=} 0.8$$

folgt  $p = \frac{0.8}{7} = 0.114$  und daher

$$\begin{aligned} \mathbb{P}(A_7 | \overline{A_1} \cup \overline{A_2} \cup \dots \cup \overline{A_6}) &= \mathbb{P}(A_7 | \overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_6}) = \frac{\overbrace{\mathbb{P}(A_7 \cap \overline{A_1} \cap \dots \cap \overline{A_6})}^{=A_7}}{\mathbb{P}(\overline{A_1} \cap \dots \cap \overline{A_6})} \\ &= \frac{\mathbb{P}(A_7)}{1 - \mathbb{P}(A_1 \cup \dots \cup A_6)} = \frac{p}{1 - 6p} = 0.3636. \end{aligned}$$

$\triangle$

## 2.2 Multiplikationsregeln

Durch Umstellen der die bedingte Wahrscheinlichkeit  $\mathbb{P}(A|B)$  definierenden Quotientenbeziehung (2.1) erhält man eine einfache Multiplikationsregel:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A). \quad (2.2)$$

**Beispiel 2.5** 70% der Studenten eines Jahrgangs schließen das Fach Mathematik wenigstens mit der Note 3 ab. Unter diesen Studenten erreichen 25% sogar eine der Noten 1 oder 2. Mit welcher Wahrscheinlichkeit schließt ein beliebig ausgewählter Student das Fach mit 1 oder 2 ab?

Um diese Frage zu beantworten, sei

$$\begin{aligned} A &\hat{=} \text{“Student schließt das Fach Mathematik mit 1 oder 2 ab”}, \\ B &\hat{=} \text{“Student schließt das Fach Mathematik mit 1,2 oder 3 ab”}. \end{aligned}$$

Gemäß (2.2) folgt dann

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) = \underbrace{\mathbb{P}(A|B)}_{=0.25} \cdot \underbrace{\mathbb{P}(B)}_{=0.7} = 0.175.$$

$\triangle$

**Satz 2.6 (erweiterte Multiplikationsregel)** Es seien  $A_1, A_2, \dots, A_n$  Ereignisse aus der  $\sigma$ -Algebra  $\mathcal{A}$  zu einer festen Zufallssituation, wobei  $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$  gelte. Dann ist

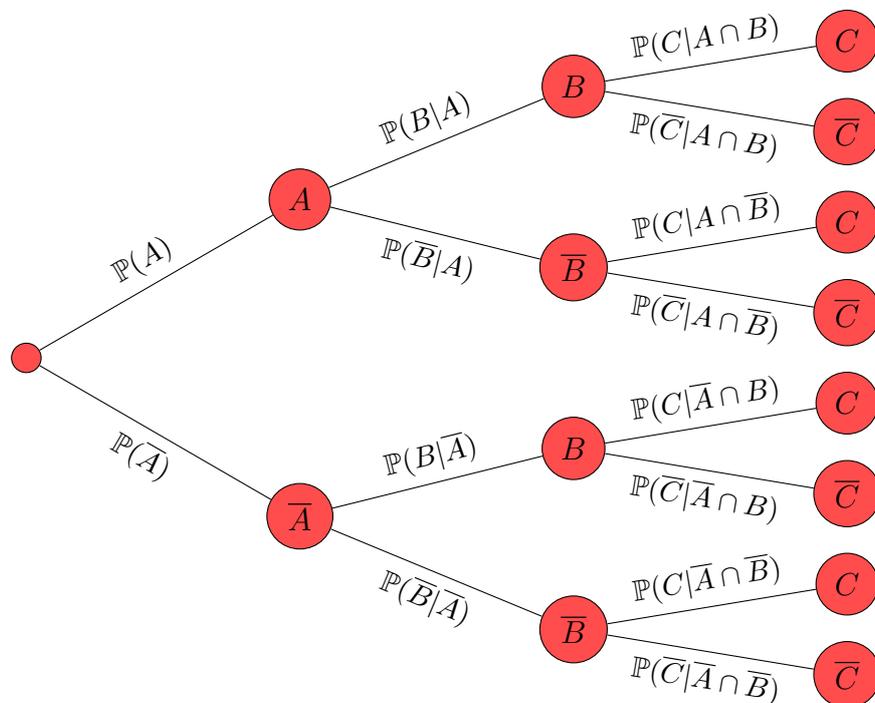
$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}).$$

*Beweis.* Aufgrund von  $A_1 \cap A_2 \cap \dots \cap A_m \subset A_1 \cap A_2 \cap \dots \cap A_{m-1}$  für alle  $m < n$ , folgt aus  $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$  auch  $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_m) > 0$ . Deshalb ergibt sich die Behauptung durch vollständige Induktion aus der Multiplikationsregel (2.2), weil

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

□

**Beispiel 2.7 (Baumdiagramm und Pfadregel)** Ein *Baumdiagramm* ist ein praktisches Hilfsmittel, um die verschiedenen Ergebnisse eines Zufallsexperiments übersichtlich darzustellen. Sinnvollerweise wird es verwendet, wenn ein Versuch aus mehreren Schritten besteht. Interessiert man sich etwa nacheinander für die Ereignisse  $A$ ,  $B$  und  $C$ , so können alle möglichen Ergebnisse durch den folgenden Baum erfasst werden:



Die Kanten des Baumes sind gewichtet mit den Übergangswahrscheinlichkeiten von einem Knoten zum nächsten. Die Übergangswahrscheinlichkeit von einem gegebenen Knoten zum Folgeknoten ist offensichtlich die bedingte Wahrscheinlichkeit, dass das Ereignis im Folgeknoten eintritt unter der Bedingung, dass man sich im gegebenen Knoten befindet. Beim

Entlanggehen eines Pfades muss man die Wahrscheinlichkeiten miteinander multiplizieren, was gerade der Multiplikationsregel aus Satz 2.6 entspricht. Diese heißt daher auch *Pfadregel*.  $\triangle$

**Satz 2.8 (Satz von der totalen Wahrscheinlichkeit)** Es seien  $B_1, B_2, \dots$  eine Zerlegung von  $\Omega$ , das heißt

$$1. \bigcup_{i=1}^{\infty} B_i = \Omega,$$

$$2. B_i \cap B_j = \emptyset \text{ für } i \neq j.$$

Dann gilt für ein beliebiges Ereignis  $A$  die Formel

$$\mathbb{P}(A) = \mathbb{P}(A|B_1) \cdot \mathbb{P}(B_1) + \mathbb{P}(A|B_2) \cdot \mathbb{P}(B_2) + \dots = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i).$$

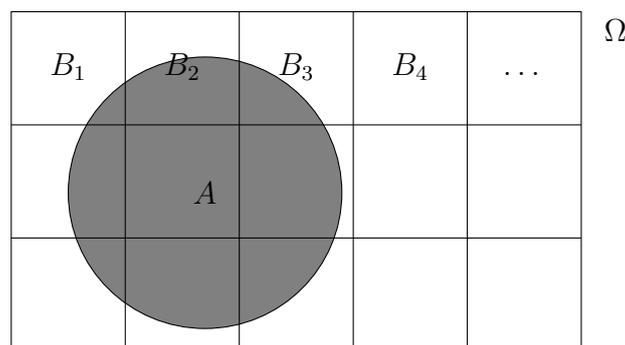
*Beweis.* Da die  $B_i$  eine Zerlegung von  $\Omega$  bilden, gilt

$$A = A \cap \Omega = A \cap \left( \bigcup_{i=1}^{\infty} B_i \right) = \bigcup_{i=1}^{\infty} (A \cap B_i).$$

Wegen der paarweisen Unvereinbarkeit von  $A \cap B_i$  für alle  $i = 1, 2, \dots$  folgt hieraus aber

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i) \stackrel{(2.2)}{=} \sum_{i=1}^{\infty} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i).$$

□



Veranschaulichung des Satzes von der totalen Wahrscheinlichkeit.

**Satz 2.9 (Formel von Bayes)** Unter den Voraussetzungen von Satz 2.8 gilt für  $\mathbb{P}(A) > 0$  die Bayessche Formel

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}{\sum_{i=1}^{\infty} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}, \quad j = 1, 2, \dots$$

*Beweis.* Aus (2.2) folgt

$$\mathbb{P}(A \cap B_j) = \mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j) = \mathbb{P}(B_j|A) \cdot \mathbb{P}(A),$$

dies bedeutet

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}{\mathbb{P}(A)}.$$

Mit dem Satz von der totalen Wahrscheinlichkeit folgt dann auch die letzte Gleichung.  $\square$

**Beispiel 2.10 (Pannenhilfe-Problem)** Die Statistik der ambulanten Pannenhelfer des Automobilclubs ADAC wies für ein Jahr aus, dass bei vorgefundenen Schäden im Bereich der Motorausfälle folgende Schadenstypverteilung zu verzeichnen war:

- 50% Störungen an der Zündanlage,
- 30% Störungen an der Kraftstoffzufuhr,
- 20% andere Störungen.

Der Pannenhelfer konnte vor Ort den Schaden beheben

- in 50% aller Fälle bei Störungen der Zündanlage,
- in 30% aller Fälle bei Störungen der Kraftstoffzufuhr,
- in 5% aller Fälle bei sonstigen Störungen.

Wir wollen zwei Fragen beantworten:

1. In wieviel Prozent der Fälle konnte bei Motorausfällen vor Ort geholfen werden?
2. Wie wahrscheinlich sind die drei Schadenstypen, wenn geholfen werden konnte?

Um die erste Frage zu beantworten, beachten wir, dass die Ereignisse

$$\begin{aligned} B_1 &\hat{=} \text{“Fehler an der Zündanlage”}, \\ B_2 &\hat{=} \text{“Fehler and der Kraftstoffzufuhr”}, \\ B_3 &\hat{=} \text{“sonstiger Fehler”} \end{aligned}$$

eine Zerlegung von  $\Omega$  bilden, wobei

$$\mathbb{P}(B_1) = 0.5, \quad \mathbb{P}(B_2) = 0.3, \quad \mathbb{P}(B_3) = 0.2$$

gilt. Damit erhalten wir für das Ereignis

$$A \hat{=} \text{“Schaden konnte vor Ort behoben werden”}$$

gemäß dem Satz von der totalen Wahrscheinlichkeit

$$\mathbb{P}(A) = \underbrace{\mathbb{P}(A|B_1) \cdot \mathbb{P}(B_1)}_{=0.5} + \underbrace{\mathbb{P}(A|B_2) \cdot \mathbb{P}(B_2)}_{=0.3} + \underbrace{\mathbb{P}(A|B_3) \cdot \mathbb{P}(B_3)}_{=0.05} = 0.35.$$

Es konnte also in 35% aller Fälle vor Ort geholfen werden.

Als nächstes wollen wir uns mit der zweiten Fragestellung befassen. Diese lässt sich mit Hilfe der Bayesschen Formel beantworten, denn es gilt mit Satz 2.9

$$\begin{aligned} \mathbb{P}(B_1|A) &= \frac{\mathbb{P}(A|B_1) \cdot \mathbb{P}(B_1)}{\mathbb{P}(A)} = 0.714, \\ \mathbb{P}(B_2|A) &= \frac{\mathbb{P}(A|B_2) \cdot \mathbb{P}(B_2)}{\mathbb{P}(A)} = 0.257, \\ \mathbb{P}(B_3|A) &= \frac{\mathbb{P}(A|B_3) \cdot \mathbb{P}(B_3)}{\mathbb{P}(A)} = 0.029. \end{aligned}$$

$\triangle$

## 2.3 Stochastische Unabhängigkeit

**Definition 2.11** (stochastische Unabhängigkeit) Zwei Ereignisse  $A, B \in \mathcal{A}$  heißen **stochastisch unabhängig**, wenn gilt

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

**Bemerkung 2.12** Stochastisch unabhängig bedeutet  $\mathbb{P}(A|B) = \mathbb{P}(A)$  und  $\mathbb{P}(B|A) = \mathbb{P}(B)$ , das heißt, der Zufallscharakter der Ereignisse  $A$  und  $B$  beeinflusst sich nicht gegenseitig.  $\triangle$

**Satz 2.13** Falls  $A$  und  $B$  stochastisch unabhängig sind, so sind auch die Ereignisse  $\bar{A}$  und  $\bar{B}$ ,  $A$  und  $\bar{B}$ , sowie  $\bar{A}$  und  $B$  stochastisch unabhängig.

*Beweis.* Wegen

$$\begin{aligned} \mathbb{P}(A \cap \bar{B}) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A) \cdot \mathbb{P}(B) \\ &= \mathbb{P}(A) \cdot (1 - \mathbb{P}(B)) = \mathbb{P}(A) \cdot \mathbb{P}(\bar{B}) \end{aligned}$$

sind mit  $A$  und  $B$  auch  $A$  und  $\bar{B}$  stochastisch unabhängig. Analog zeigt man die verbliebenen Fälle.  $\square$

Auf der stochastischen Unabhängigkeit beruhen die meisten Aussagen der Zuverlässigkeitstheorie, die nicht nur in der Technik, sondern auch in anderen Disziplinen Anwendung finden.

**Beispiel 2.14** (Serien- und Parallelschaltung von Bauteilen) Ein Gerät bestehe aus 2 Bauteilen  $T_1$  und  $T_2$ , bei denen unabhängig voneinander Defekte auftreten können. Die stochastisch unabhängigen Ereignisse  $A_1$  mit  $\mathbb{P}(A_1) = p_1$  und  $A_2$  mit  $\mathbb{P}(A_2) = p_2$  treten auf, wenn die Bauteile  $T_1$  und  $T_2$  funktionieren. Bei zwei Bauteilen können wir zwischen einer Serienschaltung und einer Reihenschaltung unterscheiden.

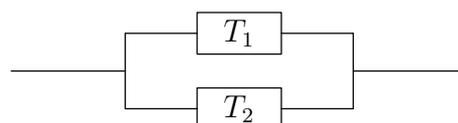
1. *Serienschaltung:*



Eine Serienschaltung funktioniert, wenn sowohl  $T_1$  als auch  $T_2$  funktionieren. Die Wahrscheinlichkeit hierfür ist

$$\mathbb{P}(A_1 \cap A_2) \stackrel{\text{stoch. unabh.}}{=} \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) = p_1 p_2.$$

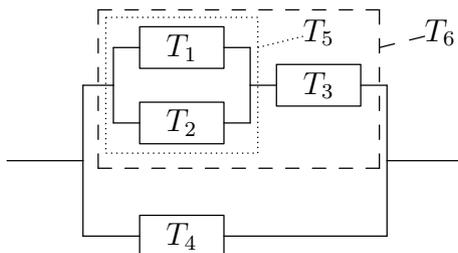
2. *Parallelschaltung:*



Eine Parallelschaltung funktioniert, wenn  $T_1$  oder  $T_2$  funktionieren. Die Wahrscheinlichkeit hierfür ist

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2) &= 1 - \mathbb{P}(\overline{A_1 \cup A_2}) = 1 - \mathbb{P}(\overline{A_1} \cap \overline{A_2}) \stackrel{\text{stoch.}}{\text{unabh.}} 1 - \mathbb{P}(\overline{A_1}) \cdot \mathbb{P}(\overline{A_2}) \\ &= 1 - (1 - p_1)(1 - p_2).\end{aligned}$$

Eine Kombinationen von Serien- und Parallelschaltungen kann man durch geeignetes Zusammenfassen behandeln:



In der oben gezeigten Schaltung kann zunächst die Parallelschaltung der Bauteile  $T_1$  und  $T_2$  durch ein neues Bauteil  $T_5$  ersetzt werden. Fasst man nun die Serienschaltung der Bauteile  $T_3$  und  $T_5$  in einem neuen Bauteil  $T_6$  zusammen, so kann die Baugruppe, welche aus den Bauteilen  $T_1$ ,  $T_2$  und  $T_3$  besteht, durch ein einziges Bauteil ausgedrückt werden. Schliesslich muss man nur noch die Parallelschaltung der Bauteile  $T_4$  und  $T_6$  berücksichtigen, um zu bestimmen mit welcher Wahrscheinlichkeit die Schaltung funktioniert.  $\triangle$

Wir betrachten nun Unabhängigkeitsfragen für  $n$  Ereignisse  $A_1, A_2, \dots, A_n \in \mathcal{A}$ .

**Definition 2.15** (vollständige stochastische Unabhängigkeit) Die Ereignisse  $A_1, \dots, A_n$  heißen **vollständig stochastisch unabhängig**, wenn für jede Auswahl von  $m \leq n$  Ereignissen gilt:

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdot \dots \cdot \mathbb{P}(A_{i_m}).$$

Man beachte, dass es in dieser Definition nicht genügt, nur die Gleichheit

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

zu fordern. Insbesondere können die Ereignisse  $A_1, \dots, A_n$  paarweise stochastisch unabhängig sein, ohne dass sie *vollständig* stochastisch unabhängig sind. Wir wollen diese Feststellung durch ein Beispiel illustrieren.

**Beispiel 2.16** Sei  $\Omega = \{1, 2, 3, 4\}$  mit

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \frac{1}{4}.$$

Die Ereignisse  $A = \{1, 2\}$ ,  $B = \{1, 3\}$  und  $C = \{2, 3\}$  besitzen die Wahrscheinlichkeit

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}.$$

Wegen

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(\{1\}) = \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(B), \\ \mathbb{P}(A \cap C) &= \mathbb{P}(\{2\}) = \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(C), \\ \mathbb{P}(B \cap C) &= \mathbb{P}(\{3\}) = \frac{1}{4} = \mathbb{P}(B) \cdot \mathbb{P}(C),\end{aligned}$$

sind  $A$ ,  $B$  und  $C$  paarweise stochastisch unabhängig. Jedoch ergibt sich

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0 \neq \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) = \frac{1}{8},$$

das heißt, es liegt keine vollständige stochastische Unabhängigkeit vor!  $\triangle$

Eine für die Praxis sehr wichtige Folgerung aus der vollständigen stochastischen Unabhängigkeit ist die folgende. Sind die Ereignisse  $A_1, A_2, \dots, A_n$  vollständig stochastisch unabhängig, so gilt

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) &= 1 - \mathbb{P}(\overline{A}_1 \cap \overline{A}_2 \cap \dots \cap \overline{A}_n) \\ &= 1 - \mathbb{P}(\overline{A}_1) \cdot \mathbb{P}(\overline{A}_2) \cdot \dots \cdot \mathbb{P}(\overline{A}_n) \\ &= 1 - (1 - \mathbb{P}(A_1)) \cdot (1 - \mathbb{P}(A_2)) \cdot \dots \cdot (1 - \mathbb{P}(A_n)).\end{aligned}$$

Mit anderen Worten, die Wahrscheinlichkeit dafür, dass wenigstens eines der Ereignisse  $A_1, A_2, \dots, A_n$  eintritt, ist gleich 1 minus dem Produkt aus den Einzelwahrscheinlichkeiten für das Eintreten von  $\overline{A}_i$ .

## 2.4 Produktexperimente

Wir nehmen an, wir kennen schon Modelle  $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1), \dots, (\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ , und wollen nun ein Modell für das Experiment konstruieren, welches aus der unabhängigen Hintereinanderausführung dieser Telexperimente besteht.

**Definition und Satz 2.17 (Produktexperiment)** Werden die Wahrscheinlichkeitsräume  $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ ,  $i = 1, 2, \dots, n$ , **unabhängig gekoppelt**, so entsteht das **Produkt** der Wahrscheinlichkeitsräume  $(\Omega, \mathcal{A}, \mathbb{P})$  gemäß

$$\Omega = \Omega_1 \times \dots \times \Omega_n, \quad \mathcal{A} = \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n, \quad \mathbb{P} = \mathbb{P}_1 \otimes \dots \otimes \mathbb{P}_n.$$

Die Ergebnismenge  $\Omega$  besteht demnach aus allen  $n$ -Tupeln  $\omega = (\omega_1, \dots, \omega_n)$  mit  $\omega_i \in \Omega_i$ . Die Produkt- $\sigma$ -Algebra  $\mathcal{A}$  ist die kleinste  $\sigma$ -Algebra, die alle Produktmengen  $A = A_1 \times \dots \times A_n$  mit  $A_i \in \mathcal{A}_i$  enthält. Insbesondere ist für die Produktmenge  $A = A_1 \times \dots \times A_n \in \mathcal{A}$  die Wahrscheinlichkeit  $\mathbb{P}(A)$  gegeben als

$$\mathbb{P}(A) = \mathbb{P}_1(A_1) \cdot \mathbb{P}_2(A_2) \cdot \dots \cdot \mathbb{P}_n(A_n).$$

Ist  $X_i(\boldsymbol{\omega})$  die  $i$ -te Koordinate von  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$ , so wird in  $\Omega$  das Ereignis, dass sich im  $i$ -ten Telexperiment  $A_i \subset \Omega_i$  ereignet, durch  $\{\boldsymbol{\omega} \in \Omega : X_i(\boldsymbol{\omega}) \in A_i\}$  beschrieben, kurz  $\{X_i \in A_i\}$ . Es ist

$$A = A_1 \times \dots \times A_n = \bigcap_{i=1}^n \{X_i \in A_i\}$$

das Ereignis, das sich für alle  $i = 1, \dots, n$  im  $i$ -ten Telexperiment  $A_i$  ereignet. Unter dem *Produktmaß*  $\mathbb{P}$  ist die Wahrscheinlichkeit dafür

$$\mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) = \mathbb{P}(A) = \mathbb{P}_1(A_1) \cdot \mathbb{P}_2(A_2) \cdot \dots \cdot \mathbb{P}_n(A_n).$$

Hält man  $1 \leq k \leq n$  fest und setzt  $A_i := \Omega_i$  für  $i \neq k$ , so folgt

$$\bigcap_{i=1}^n \{X_i \in A_i\} = \{X_k \in A_k\},$$

und weiter

$$\mathbb{P}(\{X_k \in A_k\}) = \mathbb{P}_k(A_k).$$

Dies entspricht der selbstverständlichen Forderung, dass die Wahrscheinlichkeit dafür, dass im  $k$ -ten Telexperiment  $A_k \subset \Omega_k$  eintritt, mit der Wahrscheinlichkeit übereinstimmen soll, die  $A_k$  im  $k$ -ten Teilmodell  $(\Omega_k, \mathcal{A}_k, \mathbb{P}_k)$  besitzt. Aus der Rechnung folgt insbesondere

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

Da hierin beliebig viele  $A_i = \Omega_i$  gesetzt werden können, gilt

$$\mathbb{P}\left(\bigcap_{i \in \mathcal{I}} \{X_i \in A_i\}\right) = \prod_{i \in \mathcal{I}} \mathbb{P}(\{X_i \in A_i\}), \quad \mathcal{I} \subseteq \{1, 2, \dots, n\}.$$

Das Modell hat also wirklich die geforderte Eigenschaft, dass darin Ereignisse, die etwas über die Ausgänge verschiedener Telexperimente aussagen, unabhängig sind.

**Beispiel 2.18** ( *$n$ -facher Münzwurf*) Wir betrachten eine Folge von  $n$  unabhängigen Einzelexperimenten, die für jedes  $i = 1, 2, \dots, n$  jeweils durch die Ergebnismenge  $\Omega_i = \{0, 1\}$  und das Wahrscheinlichkeitsmaß

$$\mathbb{P}_i(\omega_i) = \begin{cases} p, & \text{falls } \omega_i = 1, \\ 1 - p, & \text{falls } \omega_i = 0, \end{cases}$$

beschrieben sind. Hierbei ist  $0 \leq p \leq 1$ . Die Ergebnismenge ist

$$\Omega = \{0, 1\}^n = \{(\omega_1, \dots, \omega_n) : \omega_i \in \{0, 1\}, 1 \leq i \leq n\},$$

und das Wahrscheinlichkeitsmaß ist gegeben durch

$$\mathbb{P}(\boldsymbol{\omega}) = \mathbb{P}(\omega_1, \dots, \omega_n) = \prod_{i=1}^n \mathbb{P}_i(\omega_i) = p^k (1 - p)^{n-k},$$

wobei  $k$  die Anzahl der Indizes  $i$  mit  $\omega_i = 1$  ist. △

## 3. Diskrete Verteilungen

### 3.1 Zufallsgrößen

Vielfach sind die Ergebnisse von zufälligen Versuchen von Natur aus Zahlenwerte. Häufig möchte man aber auch in Fällen, wo dies nicht der Fall ist, Zahlenwerte zur Charakterisierung der Ergebnisse von Zufallssituationen verwenden. Dies geschieht mit Hilfe von *Zufallsgrößen*  $X$ , wobei jedem Ergebnis  $\omega \in \Omega$  eine reelle Zahl  $X(\omega)$  als Wert der Zufallsgröße zugeordnet wird.

#### Beispiele 3.1 (Zufallsgrößen)

1. Beim Werfen von zwei Würfeln gilt  $\Omega = \{(i, j) : i, j \in \{1, 2, \dots, 6\}\}$ , wobei  $i$  und  $j$  die gewürfelte Augenzahlen des ersten bzw. zweiten Würfels bezeichnen. Die Augensumme ergibt sich dann als  $X(\omega) = i + j$ .
2. Für  $n$  betrachtete Glühlampen bezeichne  $\omega_i$  die zufällige Lebensdauer der  $i$ -ten Lampe in Stunden. Dann ist  $\Omega = \{(\omega_1, \dots, \omega_n) : \omega_i \geq 0 \text{ für alle } i\}$  die Ergebnismenge. Sowohl  $X(\omega) = \omega_k$  (Lebensdauer der  $k$ -ten Lampe) als auch  $X(\omega) = \frac{1}{n}(\omega_1 + \omega_2 + \dots + \omega_n)$  (mittlere Lebensdauer von  $n$  Lampen) stellen Zufallsgrößen dar.

△

**Definition 3.2 (Zufallsgröße)** Sei  $(\Omega, \mathcal{A}, \mathbb{P})$  ein Wahrscheinlichkeitsraum zu einer festen Zufallssituation. Dann heißt die Abbildung  $X : \Omega \rightarrow \mathbb{R}$  **Zufallsgröße** oder **Zufallsvariable** über  $(\Omega, \mathcal{A}, \mathbb{P})$ , wenn

$$\{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{A}$$

für alle Intervalle  $I$  der reellen Achse.

Der durchaus gebräuchliche Begriff Zufallsvariable ist etwas irreführend, denn strenggenommen ist eine Zufallsgröße keine Variable, sondern eine *Abbildung*. Wir werden daher im folgenden den Begriff Zufallsgröße verwenden.

**Bemerkung 3.3** Für die Wahrscheinlichkeit  $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in I\})$  schreiben wir verkürzt  $\mathbb{P}(X \in I)$ . △

**Definition 3.4 (diskrete Zufallsgröße)** Eine Zufallsgröße heißt **diskret**, wenn sie nur endlich oder abzählbar unendlich viele Werte annehmen kann.

Die erste Zufallsgröße aus Beispiel 3.1 ist eine diskrete Zufallsgröße. Bei der zweiten handelt es sich um eine stetige Zufallsgröße, auf die wir im nächsten Kapitel eingehen werden.

**Definition 3.5 (Wahrscheinlichkeitsfunktion)** Ist  $X$  eine diskrete Zufallsgröße mit den Werten  $x_1, x_2, \dots$ , so bezeichnet die Zuordnung

$$p_i := \mathbb{P}(X = x_i), \quad i = 1, 2, \dots$$

die **Wahrscheinlichkeitsfunktion** der diskreten Zufallsgröße  $X$ .

**Beispiel 3.6 (Werfen mit einem Würfel)** Die Zufallsgröße  $X$  entspreche der gewürfelten Augenzahl beim Werfen mit einem Würfel. Die endliche Zahl der Ereignisse erlaubt eine Darstellung der Wahrscheinlichkeitsfunktion als Tabelle:

$i$	1	2	3	4	5	6
$x_i$	1	2	3	4	5	6
$p_i$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

△

Eine diskrete Zufallsgröße  $X$  wird offensichtlich vollständig durch ihre Wahrscheinlichkeitsfunktion charakterisiert. Dabei gelten für die Wahrscheinlichkeitsfunktion die folgenden Aussagen:

- Es gilt  $0 \leq p_i \leq 1$  für alle  $i$ .
- Die Summe aller Einzelwahrscheinlichkeiten ist identisch 1, kurz  $\sum_i p_i = 1$ .
- Die Wahrscheinlichkeit dafür, dass die Zufallsgröße  $X$  im Intervall  $[a, b]$  liegt, berechnet sich gemäß

$$\mathbb{P}(a \leq X \leq b) = \sum_{a \leq x_i \leq b} p_i.$$

**Beispiel 3.7 (geometrische Verteilung)** Ein Automat sei so eingerichtet, dass er sofort anhält, sobald ein defektes Teil produziert wird. Die Wahrscheinlichkeit dafür, dass ein defektes Teil erzeugt wird, sei  $p$ . Die Ausfälle (Defekte) sind von Teil zu Teil unabhängig. Die diskrete Zufallsgröße

$$X \hat{=} \text{“Anzahl der produzierten einwandfreien Teile”}$$

besitzt die *geometrische Verteilung*. Dabei gilt für das Ereignis

$$A_i \hat{=} \text{“}i\text{-tes produziertes Teil defekt”}$$

$\mathbb{P}(A_i) = p$  und  $\mathbb{P}(\overline{A_i}) = 1 - p$ . Dies liefert die Wahrscheinlichkeitsfunktion für  $X$ :

$$\begin{aligned} \mathbb{P}(X = 0) &= \mathbb{P}(A_1) &&= p, \\ \mathbb{P}(X = 1) &= \mathbb{P}(\overline{A_1} \cap A_2) &&= (1 - p) \cdot p, \\ \mathbb{P}(X = 2) &= \mathbb{P}(\overline{A_1} \cap \overline{A_2} \cap A_3) &&= (1 - p)^2 \cdot p, \\ &\vdots &&\vdots \\ \mathbb{P}(X = i) &= \mathbb{P}(\overline{A_1} \cap \dots \cap \overline{A_i} \cap A_{i+1}) &&= (1 - p)^i \cdot p, \quad i = 0, 1, 2, \dots \end{aligned}$$

Offensichtlich gilt  $0 \leq p \cdot (1-p)^i \leq 1$  für alle  $i = 0, 1, 2, \dots$  und

$$\sum_{i=0}^{\infty} (1-p)^i \cdot p = p \cdot \sum_{i=0}^{\infty} (1-p)^i = p \cdot \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1.$$

△

## 3.2 Verteilungsfunktion

**Definition 3.8 (Verteilungsfunktion)** Es sei  $X$  eine Zufallsgröße. Dann heißt die zu  $X$  gehörige Funktion

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

**Verteilungsfunktion** von  $X$ . Für diskrete Zufallsgrößen  $X$  mit den Werten  $x_1, x_2, \dots$  gilt

$$F(x) = \sum_{x_i \leq x} p_i.$$

**Eigenschaften der Verteilungsfunktion:**

1.  $\lim_{x \rightarrow -\infty} F(x) = 0$
2.  $\lim_{x \rightarrow \infty} F(x) = 1$
3.  $F(x)$  ist monoton wachsend (nicht notwendigerweise streng), das heißt

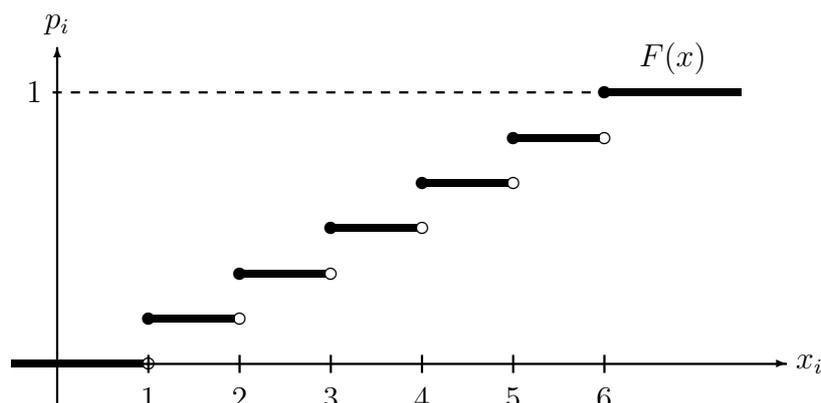
$$x < y \Rightarrow F(x) \leq F(y).$$

4.  $F$  ist rechtsseitig stetig, das heißt

$$\lim_{y \searrow x} F(y) = F(x).$$

**Bemerkung 3.9** Bei diskreten Zufallsgrößen ist die Verteilungsfunktion immer eine reine Treppenfunktion. Die Punkte  $x_i$  kennzeichnen die Sprungpunkte und die Werte  $p_i$  die dazugehörige Sprunghöhe. Rechtsseitige Stetigkeit bedeutet, dass der Funktionswert an der Sprungstelle dem rechten Teil des Funktionsgraphen zugeordnet wird. △

**Beispiel 3.10 (Fortsetzung von Beispiel 3.6)** Entspricht  $X$  der gewürfelten Augenzahl beim Werfen eines Würfels, so erhalten wir die folgende Verteilungsfunktion:



△

### 3.3 Erwartungswert

Wir nutzen zur Erklärung das Maschinenmodell der geometrischen Verteilung. Die Maschine wird dabei  $n$  mal gestartet mit dem folgenden Ergebnis:

$$\begin{aligned} h_n(0) &\hat{=} \text{“Anzahl der Versuche, die 0 einwandfreie Teile liefern”}, \\ h_n(1) &\hat{=} \text{“Anzahl der Versuche, die 1 einwandfreie Teile liefern”}, \\ &\vdots \\ h_n(i) &\hat{=} \text{“Anzahl der Versuche, die } i \text{ einwandfreie Teile liefern”}. \\ &\vdots \end{aligned}$$

Wir fragen, wieviel funktionstüchtige Teile die Maschine im Mittel auswirft. Es sind

$$\sum_{i=0}^{\infty} \frac{h_n(i)}{n} \cdot i = \sum_{i=0}^{\infty} H_n(i) \cdot i \xrightarrow{n \rightarrow \infty} \sum_{i=0}^{\infty} \mathbb{P}(X = i) \cdot i$$

funktionstüchtige Teile.

Für die geometrische Verteilung ergibt sich beispielsweise

$$\begin{aligned} \sum_{i=0}^{\infty} \mathbb{P}(X = i) \cdot i &= \sum_{i=0}^{\infty} p(1-p)^i \cdot i = p(1-p) \sum_{i=0}^{\infty} i(1-p)^{i-1} \\ &= p \cdot (1-p) \underbrace{\frac{1}{(1-(1-p))^2}}_{=\frac{1}{p^2}} = \frac{1-p}{p}, \end{aligned}$$

wobei wir die Summenformel

$$\sum_{i=0}^{\infty} ix^{i-1} = \frac{1}{(1-x)^2} \quad (3.1)$$

benutzt haben, die sich durch Differentiation nach  $x$  aus der bekannten Identität

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$$

ergibt.

**Definition 3.11** Sei  $p_i = \mathbb{P}(X = x_i)$  die Wahrscheinlichkeitsfunktion einer diskreten Zufallsgröße  $X$ . Dann wird

$$\mathbb{E}(X) = \sum_i x_i p_i = \sum_i x_i \mathbb{P}(X = x_i) \quad (3.2)$$

**Erwartungswert** oder **Mittelwert** der Zufallsgröße  $X$  genannt, falls gilt  $\sum_i |x_i| p_i < \infty$ . Konvergiert die Reihe (3.2) nicht absolut, so existiert kein Erwartungswert.

**Bemerkung 3.12** Es gilt offensichtlich die Identität

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) \cdot X(\omega). \quad (3.3)$$

△

**Beispiel 3.13 (Würfeln mit einem Würfel)** Für die Zufallsgröße

$$X \hat{=} \text{“gewürfelte Augenzahl”}$$

ergibt sich

$$\mathbb{E}(X) = \sum_{i=1}^6 i \cdot \frac{1}{6} = 3.5.$$

△

### Eigenschaften des Erwartungswertes:

1. Falls  $\mathbb{E}(X)$  existiert, so gilt für alle  $a, b \in \mathbb{R}$

$$\mathbb{E}(aX + b) = a \cdot \mathbb{E}(X) + b. \quad (3.4)$$

*Beweis.* Es gilt

$$\mathbb{E}(aX + b) = \sum_i (a \cdot x_i + b)p_i = a \cdot \underbrace{\sum_i x_i p_i}_{=\mathbb{E}(X)} + b \cdot \underbrace{\sum_i p_i}_{=1} = a \cdot \mathbb{E}(X) + b.$$

□

2. Falls  $\mathbb{E}(X)$  und  $\mathbb{E}(Y)$  existieren, so gilt

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y). \quad (3.5)$$

*Beweis.* Nach (3.3) folgt

$$\begin{aligned} \mathbb{E}(X + Y) &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \cdot \mathbb{P}(\{\omega\}) \\ &= \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega) \cdot \mathbb{P}(\{\omega\}) \\ &= \mathbb{E}(X) + \mathbb{E}(Y). \end{aligned}$$

□

3. Falls  $\mathbb{E}(X)$  und  $\mathbb{E}(Y)$  existieren, gilt

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

*Beweis.* Die Behauptung folgt wieder mit (3.3):

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} \underbrace{X(\omega)}_{\leq Y(\omega)} \cdot \mathbb{P}(\{\omega\}) \leq \sum_{\omega \in \Omega} Y(\omega) \cdot \mathbb{P}(\{\omega\}) = \mathbb{E}(Y).$$

□

**Bemerkung 3.14** Die Beziehungen (3.4) und (3.5) implizieren die Linearität des Erwartungswerts, dies bedeutet

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y). \quad (3.6)$$

△

Neben  $\mathbb{E}(X)$  kann auch der Erwartungswert von Funktionen einer Zufallsgröße  $X$  betrachtet werden, zum Beispiel  $g(X) = \sin(X)$  oder  $g(X) = X^2$ .

**Satz 3.15** Sei  $p_i = \mathbb{P}(X = x_i)$  die Wahrscheinlichkeitsfunktion der diskreten Zufallsgröße  $X$ . Falls  $\mathbb{E}(X)$  und  $\mathbb{E}(g(X))$  existieren, gilt

$$\mathbb{E}(g(X)) = \sum_i g(x_i) \cdot p_i = \sum_i g(x_i) \mathbb{P}(X = x_i).$$

*Beweis.* Für die Zufallsgröße  $Y := g(X)$  folgt

$$\begin{aligned} \mathbb{E}(Y) &= \sum_i y_i \mathbb{P}(Y = y_i) \\ &= \sum_i y_i \left[ \sum_{j: g(x_j)=y_i} \mathbb{P}(X = x_j) \right] \\ &= \sum_k g(x_k) \cdot \mathbb{P}(X = x_k). \end{aligned}$$

□

## 3.4 Varianz

**Definition 3.16** Existiert  $\mathbb{E}(X^2)$ , so heißt die Größe  $\sigma^2 = \mathbb{V}(X)$  mit

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

**Varianz** oder **Streuung** der Zufallsgröße  $X$ . Die Wurzel  $\sigma := \sqrt{\mathbb{V}(X)}$  heißt **Standardabweichung** der Zufallsgröße  $X$ .

**Bemerkung 3.17**

1. Die Varianz  $\mathbb{V}(X)$  ist die mittlere quadratische Abweichung einer Zufallsgröße  $X$  von ihrem Erwartungswert  $\mathbb{E}(X)$ .
2. Wegen

$$|X| \leq 1 + X^2$$

folgt aus der Existenz von  $\mathbb{E}(X^2)$  auch die Existenz von  $\mathbb{E}(X)$ .

△

**Lemma 3.18** Für eine Zufallsgröße  $X$  gilt mit  $a, b \in \mathbb{R}$

$$\mathbb{V}(aX + b) = a^2\mathbb{V}(X).$$

*Beweis.* Gemäß der Definition der Varianz folgt

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

das ist die Behauptung.  $\square$

**Satz 3.19** Es gilt

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

*Beweis.* Wir haben

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}([X - \mathbb{E}(X)]^2) = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}^2(X)) \\ &\stackrel{(3.6)}{=} \mathbb{E}(X^2) - 2 \cdot \underbrace{\mathbb{E}(X) \cdot \mathbb{E}(X)}_{=\mathbb{E}^2(X)} + \mathbb{E}^2(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X).\end{aligned}$$

$\square$

**Beispiel 3.20 (Varianz der geometrischen Verteilung)** Wir berechnen die Varianz  $\mathbb{V}(X)$  für die geometrische Verteilung auf Grundlage der Formeln

$$p_i = p \cdot (1 - p)^i, \quad \mathbb{E}(X) = \frac{1 - p}{p}, \quad \mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X).$$

Hierzu benötigen wir die folgende Summenformel

$$\sum_{i=0}^{\infty} i(i-1)x^{i-2} = \frac{2}{(1-x)^3},$$

die durch Differentiation nach  $x$  aus (3.1) folgt. Damit erhalten wir

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{i=0}^{\infty} i^2 p (1-p)^i \\ &= \sum_{i=0}^{\infty} i(i-1) \underbrace{p(1-p)^i}_{=p(1-p)^2(1-p)^{i-2}} + \underbrace{\sum_{i=0}^{\infty} ip(1-p)^i}_{=\mathbb{E}(X)} \\ &= p \cdot \underbrace{\frac{2}{(1-(1-p))^3}}_{=2/p^3} (1-p)^2 + \frac{1-p}{p} \\ &= \frac{2(1-p)^2}{p^2} + \frac{1-p}{p}\end{aligned}$$

und weiter

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{(1-p)^2}{p^2} + \frac{1-p}{p}.$$

△

Mit Hilfe von Erwartungswert und Varianz einer Zufallsgröße  $X$  kann man bereits eine obere Abschätzung für die Wahrscheinlichkeit angeben, dass die Zufallsgröße  $X$  Werte außerhalb eines symmetrisch um den Erwartungswert gelegenen Intervalls annimmt. Man beachte, dass die konkrete Verteilung der Zufallsgröße hierbei nicht bekannt sein muss!

**Satz 3.21 (Tschebyscheffsche Ungleichung)** Für alle  $\varepsilon > 0$  gilt

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}$$

oder anders geschrieben

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq k \cdot \sigma) \leq \frac{1}{k^2}, \quad k > 0.$$

*Beweis.* Für eine diskrete Zufallsgröße ergibt sich die Behauptung aus

$$\begin{aligned} \mathbb{V}(X) &= \sum_i (x_i - \mathbb{E}(X))^2 p_i \geq \sum_{i: |x_i - \mathbb{E}(X)| \geq \varepsilon} \underbrace{(x_i - \mathbb{E}(X))^2}_{\geq \varepsilon^2} p_i \\ &\geq \varepsilon^2 \sum_{i: |x_i - \mathbb{E}(X)| \geq \varepsilon} p_i = \varepsilon^2 \cdot \mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \end{aligned}$$

durch Auflösen nach  $\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon)$ . Der Beweis im Falle stetiger Zufallsgrößen wird später nachgeliefert. □

**Bemerkung 3.22** Wie der Beweis zeigt, gilt die Tschebyscheffsche Ungleichung auch im Falle strenger Ungleichheitszeichen: Für alle  $\varepsilon > 0$  gilt also auch

$$\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) < \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

△

Durch Übergang zum Komplement erhält man natürlich durch die Tschebyscheffsche Ungleichung auch eine untere Abschätzung für die Wahrscheinlichkeit, dass Zufallsgröße  $X$  Werte innerhalb eines symmetrisch um den Erwartungswert gelegenen Intervalls annimmt:

$$\mathbb{P}(|X - \mathbb{E}(X)| \leq \varepsilon) = 1 - \mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) > 1 - \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

## 3.5 Schwaches Gesetz der großen Zahlen

**Definition 3.23** Eine Folge von Zufallsgrößen  $X_1, X_2, \dots, X_n$  heißt **stochastisch unabhängig**, wenn sich der zufällige Charakter aller beteiligten Zufallsgrößen gegenseitig nicht beeinflusst, das heißt, wenn gilt

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_2 = x_2) \cdot \dots \cdot \mathbb{P}(X_n = x_n).$$

**Bemerkung 3.24** Im Gegensatz zu Ereignissen (vgl. Definition 2.15) ist hier die Definition ausreichend. Die Unabhängigkeit von  $X_1, X_2, \dots, X_n$  impliziert aufgrund von

$$\begin{aligned} & \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) \\ &= \sum_{x_n} \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= \sum_{x_n} \mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_2 = x_2) \cdot \dots \cdot \mathbb{P}(X_n = x_n) \\ &= \mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_2 = x_2) \cdot \dots \cdot \mathbb{P}(X_{n-1} = x_{n-1}) \underbrace{\sum_{x_n} \mathbb{P}(X_n = x_n)}_{=1} \end{aligned}$$

auch die Unabhängigkeit von  $X_1, X_2, \dots, X_{n-1}$ . Rekursiv erhält man hieraus

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m) = \mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_2 = x_2) \cdot \dots \cdot \mathbb{P}(X_m = x_m)$$

für alle  $m \leq n$ . Weil aber die Bezeichnung der  $X_i$  beliebig ist, folgt sogar für alle  $m \leq n$  die Beziehung

$$\begin{aligned} & \mathbb{P}(X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}, \dots, X_{i_m} = x_{i_m}) \\ &= \mathbb{P}(X_{i_1} = x_{i_1}) \cdot \mathbb{P}(X_{i_2} = x_{i_2}) \cdot \dots \cdot \mathbb{P}(X_{i_m} = x_{i_m}). \end{aligned}$$

△

**Lemma 3.25** Die Zufallsgrößen  $X_1, X_2, \dots, X_n$  seien stochastisch unabhängig und  $\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n)$  mögen existieren. Dann gilt

$$\mathbb{E}(X_1 \cdot X_2 \cdot \dots \cdot X_n) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2) \cdot \dots \cdot \mathbb{E}(X_n).$$

*Beweis.* Die Behauptung folgt aus

$$\begin{aligned} \mathbb{E}(X_1 \cdot X_2 \cdot \dots \cdot X_n) &= \sum_{x_1, \dots, x_n} x_1 \cdot x_2 \cdot \dots \cdot x_n \cdot \underbrace{\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)}_{=\mathbb{P}(X_1=x_1) \cdot \mathbb{P}(X_2=x_2) \cdot \dots \cdot \mathbb{P}(X_n=x_n)} \\ &= \left( \underbrace{\sum_{x_1} x_1 \cdot \mathbb{P}(X_1 = x_1)}_{=\mathbb{E}(X_1)} \right) \cdot \dots \cdot \left( \underbrace{\sum_{x_n} x_n \cdot \mathbb{P}(X_n = x_n)}_{=\mathbb{E}(X_n)} \right). \end{aligned}$$

□

**Lemma 3.26** Gegeben seien die Zufallsgrößen  $X_1, X_2, \dots, X_n$  mit endlichen Erwartungswerten und Varianzen. Sind  $X_1, X_2, \dots, X_n$  unabhängig, so gilt für alle  $a_1, \dots, a_n \in \mathbb{R}$

$$\mathbb{V}(a_1X_1 + a_2X_2 + \dots + a_nX_n) = a_1^2\mathbb{V}(X_1) + a_2^2\mathbb{V}(X_2) + \dots + a_n^2\mathbb{V}(X_n).$$

*Beweis.* Nach Satz 3.19 folgt

$$\begin{aligned} \mathbb{V}(a_1X_1 + a_2X_2 + \dots + a_nX_n) &= \mathbb{E}((a_1X_1 + a_2X_2 + \dots + a_nX_n)^2) \\ &\quad - \mathbb{E}^2(a_1X_1 + a_2X_2 + \dots + a_nX_n). \end{aligned}$$

Durch Ausmultiplizieren unter Berücksichtigung der Linearität des Erwartungswertes erhalten wir

$$\begin{aligned} &\mathbb{V}(a_1X_1 + a_2X_2 + \dots + a_nX_n) \\ &= E\left(a_1^2X_1^2 + a_2^2X_2^2 + \dots + a_n^2X_n^2 + \sum_{i \neq j} a_i a_j X_i X_j\right) \\ &\quad - (a_1\mathbb{E}(X_1) + a_2\mathbb{E}(X_2) + \dots + a_n\mathbb{E}(X_n))^2 \\ &= a_1^2\mathbb{E}(X_1^2) + a_2^2\mathbb{E}(X_2^2) + \dots + a_n^2\mathbb{E}(X_n^2) + \sum_{i \neq j} a_i a_j \mathbb{E}(X_i)\mathbb{E}(X_j) \\ &\quad - \left(a_1^2\mathbb{E}^2(X_1) + a_2^2\mathbb{E}^2(X_2) + \dots + a_n^2\mathbb{E}^2(X_n) + \sum_{i \neq j} a_i a_j \mathbb{E}(X_i)\mathbb{E}(X_j)\right). \end{aligned}$$

Die gemischten Terme heben sich gerade gegenseitig auf. Deshalb ergibt sich durch Umstellen das Behauptete:

$$\begin{aligned} &\mathbb{V}(a_1X_1 + a_2X_2 + \dots + a_nX_n) \\ &= a_1^2 \underbrace{(\mathbb{E}(X_1^2) - \mathbb{E}^2(X_1))}_{=\mathbb{V}(X_1)} + a_2^2 \underbrace{(\mathbb{E}(X_2^2) - \mathbb{E}^2(X_2))}_{=\mathbb{V}(X_2)} + \dots + a_n^2 \underbrace{(\mathbb{E}(X_n^2) - \mathbb{E}^2(X_n))}_{=\mathbb{V}(X_n)} \\ &= a_1^2\mathbb{V}(X_1) + a_2^2\mathbb{V}(X_2) + \dots + a_n^2\mathbb{V}(X_n). \end{aligned}$$

□

**Satz 3.27 (schwaches Gesetz der großen Zahlen)** Es sei  $X_1, X_2, \dots, X_n$  eine Folge von unabhängigen Zufallsgrößen mit

$$\mathbb{E}(X_i) = \mu, \quad \mathbb{V}(X_i) = \sigma_i^2 \leq M < \infty, \quad i = 1, 2, \dots, n.$$

Dann gilt für alle  $\varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{1}{n}(X_1 + X_2 + \dots + X_n) - \mu\right| \geq \varepsilon\right) \leq \frac{M}{\varepsilon^2 n} \xrightarrow{n \rightarrow \infty} 0.$$

*Beweis.* Setzen wir

$$\bar{X}_n := \frac{1}{n}(X_1 + X_2 + \dots + X_n),$$

dann gilt aufgrund der Linearität des Erwartungswerts

$$\mathbb{E}(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

und wegen Lemma 3.26

$$\mathbb{V}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \leq \frac{1}{n^2} \sum_{i=1}^n M = \frac{M}{n}.$$

Die Behauptung folgt nun sofort aus der Tschebyscheffschen Ungleichung.  $\square$

**Beispiel 3.28 (Konvergenz der relativen Häufigkeit)** Gegeben sei eine Zufallssituation mit dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$ . Wir betrachten ein festes, zufälliges Ereignis  $A \in \mathcal{A}$  und realisieren die Zufallssituation in  $n$  unabhängigen Versuchen. Definieren wir für  $i = 1, 2, \dots, n$  die Zufallsgrößen

$$X_i = \begin{cases} 1, & A \text{ tritt im } i\text{-ten Versuch ein,} \\ 0, & A \text{ tritt im } i\text{-ten Versuch nicht ein,} \end{cases}$$

so gilt

$$\mathbb{E}(X_i) = \mathbb{P}(A) := p, \quad \mathbb{V}(X_i) = p(1-p) \leq \frac{1}{4}.$$

Das schwache Gesetz der großen Zahlen liefert für die relative Häufigkeit  $H_n(A) = \frac{1}{n} \sum_{i=1}^n X_i$  die Abschätzung

$$\mathbb{P}(|H_n(A) - p| \geq \varepsilon) \leq \frac{1}{4\varepsilon^2 n} \xrightarrow{n \rightarrow \infty} 0.$$

$\triangle$

## 3.6 Binomialverteilung

Wir betrachten eine Zufallssituation mit dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$  und ein festes Ereignis  $A \in \mathcal{A}$ . Ein dieser Situation entsprechender Versuch wird unabhängig  $n$ -mal wiederholt. Dazu sei

$$X_i = \begin{cases} 1, & \text{wenn } A \text{ im } i\text{-ten Versuch eintritt,} \\ 0, & \text{wenn } A \text{ im } i\text{-ten Versuch nicht eintritt,} \end{cases}$$

und

$$\mathbb{P}(A) = \mathbb{P}(X_i = 1) := p, \quad \mathbb{P}(\bar{A}) = \mathbb{P}(X_i = 0) = 1 - p.$$

Für ein Tupel  $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$  mit  $\sum_{i=1}^n x_i = k$  gilt

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i) = p^k (1-p)^{n-k}.$$

Da es insgesamt  $\binom{n}{k}$  Tupel mit  $\sum_{i=1}^n x_i = k$  gibt, folgt für die Zufallsgröße

$$\bar{X}_n = X_1 + X_2 + \dots + X_n,$$

dass

$$\begin{aligned} \mathbb{P}(\bar{X}_n = k) &= \mathbb{P}\left(\left\{(X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n) \in \{0, 1\}^n : \sum_{i=1}^n x_i = k\right\}\right) \\ &= \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Insbesondere gilt

$$\mathbb{E}(\bar{X}_n) = \sum_{i=1}^n \underbrace{\mathbb{E}(X_i)}_{=p} = np, \quad \mathbb{V}(\bar{X}_n) = \sum_{i=1}^n \underbrace{\mathbb{V}(X_i)}_{=p(1-p)} = np(1-p).$$

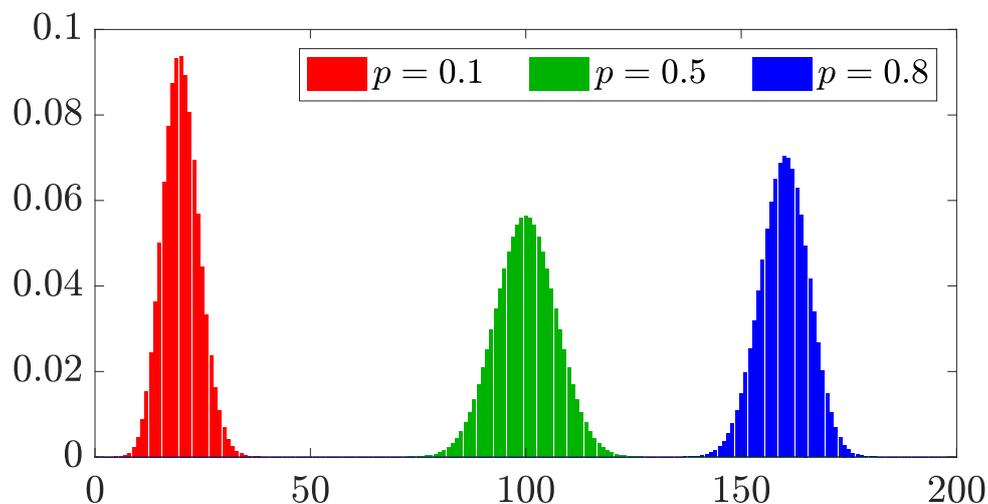
**Definition 3.29 (Binomialverteilung)** Eine Zufallsgröße  $X$  mit der Wahrscheinlichkeitsfunktion

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

heißt **binomialverteilt** mit den Parametern  $n$  (Zahl der Freiheitsgrade) und  $p$  (Fehler-rate), kurz

$$X \sim \text{Bin}(n, p).$$

Im Fall  $n = 200$  ist die Wahrscheinlichkeitsfunktion der Binomialverteilung für verschiedene Werte von  $p$  in nachfolgender Abbildung dargestellt. Wir bemerken, dass die Verteilung im Fall  $p = 0.5$  symmetrisch um den Mittelwert  $np = 100$  ist. Im Fall von  $p \neq 0.5$  ergibt sich hingegen kein Symmetrie bezüglich des Mittelwerts  $np$ .



**Beispiel 3.30 (Werfen eines idealen Würfels)** Ein idealer Würfel wird  $n = 20$  mal geworfen. Mit welcher Wahrscheinlichkeit werden mindestens zwei Sechsen gewürfelt? Dazu sei

$$A \hat{=} \text{“Es wird eine 6 gewürfelt”},$$

wobei  $p = \mathbb{P}(A) = 1/6$ , und

$X \hat{=}$  “Anzahl der geworfenen Sechsen bei  $n = 20$  Würfeln”.

Wegen  $X \sim \text{Bin}(n, p) = \text{Bin}(20, 1/6)$  folgt

$$\begin{aligned} \mathbb{P}(X \geq 2) &= 1 - \underbrace{\mathbb{P}(X < 2)}_{\hat{=} X \leq 1} \\ &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \\ &= 1 - \binom{20}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{20} - \binom{20}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{19} \\ &= 0.8696. \end{aligned}$$

Um die Frage zu beantworten, mit welcher Wahrscheinlichkeit die tatsächliche bei  $n = 20$  Würfeln erzielte Anzahl von Sechsen um mehr als 3 vom Mittelwert  $\mathbb{E}(X) = n \cdot p = 20 \cdot 1/6 = 3.\bar{3}$  abweicht, kann man die Tschebyscheffsche Ungleichung

$$\mathbb{P}(|X - \mathbb{E}(X)| > 3) < \frac{\mathbb{V}(X)}{9}$$

benutzen. Für unser Beispiel erhalten wir

$$\mathbb{V}(X) = n \cdot p \cdot (1 - p) = 20 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{25}{9}.$$

Die Tschebyscheffsche Ungleichung sichert mit

$$\mathbb{P}(|X - 3.\bar{3}| > 3) < \frac{25}{9 \cdot 9} = 0.31,$$

dass in höchstens 31% aller Fälle eine solch große Abweichung der gewürfelten Anzahl von Sechsen zum Mittelwert auftritt. Summiert man aber exakt alle Wahrscheinlichkeiten der Fälle mit einer so großen Abweichung, so ergibt sich:

$$\mathbb{P}(X = 0) + \mathbb{P}(X = 7) + \mathbb{P}(X = 8) + \dots + \mathbb{P}(X = 20) = 0.0632.$$

Es sind also in Wirklichkeit nur reichlich 6% aller Fälle, in denen eine Abweichung von mehr als 3 vom Mittelwert auftritt. Die Tschebyscheffsche Ungleichung liefert insofern eine recht grobe Abschätzung.  $\triangle$

## 3.7 Poisson-Verteilung\*

Als Referenzmodell für eine Poisson-verteilte Zufallsgröße wollen wir eine Telefonzentrale betrachten: Innerhalb eines Zeitintervalls der Länge  $t$  kommen  $X_t$  Anrufe an. Dabei ist die Zufallsgröße  $X_t$  *Poisson-verteilt*, wenn die folgenden drei *Poissonschen Annahmen* gelten:

1. *Stationarität*: Die Wahrscheinlichkeit für das Eintreten einer bestimmten Anzahl von Ereignissen im betrachteten Zeitintervall hängt nur von der Länge  $t$  des Zeitintervalls ab, nicht aber von seiner konkreten Lage auf der Zeitachse.
2. *Homogenität*: Die Ereignisfolge sei ohne Nachwirkungen, das heißt, die Anzahl von Ereignissen im Zeitintervall  $[t_0, t_1]$  hat keinen Einfluss auf die Anzahl von Ereignissen in einem späteren Zeitintervall  $[t_2, t_3]$  mit  $t_2 > t_1$ .

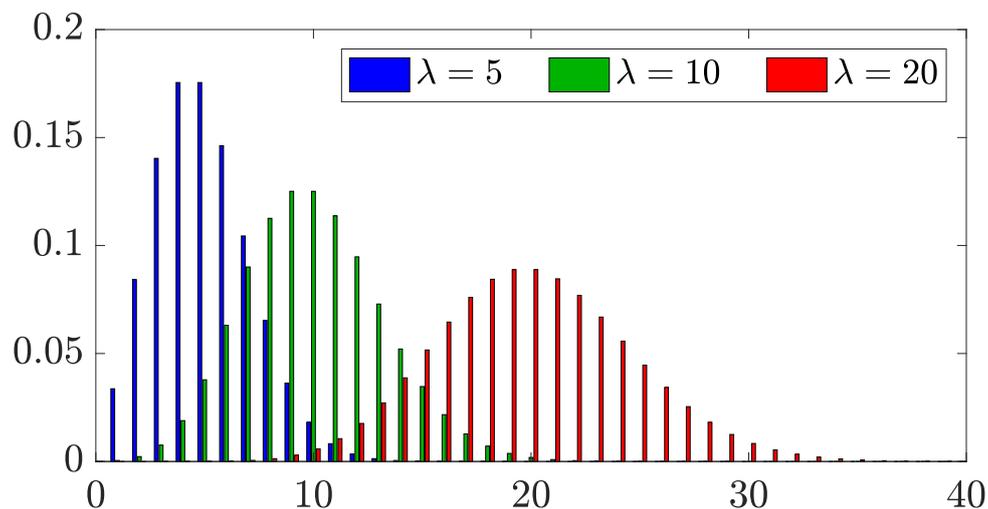
3. *Ordinarität*: Die Ereignisse treten für hinreichend kleine Zeitintervalle einzeln auf, d.h. für  $\Delta t$  genügend klein gilt entweder  $X_{\Delta t} = 0$  oder  $X_{\Delta t} = 1$ . Weiter gelte  $\mathbb{P}(X_{\Delta t} = 1) = \mu \cdot \Delta t$ . Der Parameter  $\mu$  mit  $0 < \mu < \infty$  heißt *Intensität*.

**Definition 3.31 (Poisson-Verteilung)** Eine Zufallsgröße  $X_t$ , welche der Wahrscheinlichkeitsfunktion

$$\mathbb{P}(X_t = k) = \frac{(\mu t)^k}{k!} e^{-\mu t}, \quad k = 0, 1, 2, \dots \quad (3.7)$$

genügt, heißt **Poisson-verteilt**. Oft wird  $\lambda = \mu \cdot t$  gesetzt ( $\lambda$  ist der Parameter der Poisson-Verteilung) und  $X_t \sim \pi_\lambda = \pi_{\mu t}$  geschrieben.

Die Wahrscheinlichkeitsfunktion der Poisson-Verteilung ist nachfolgend für verschiedene Werte von  $\lambda$  visualisiert. Die Verteilung steigt für kleinere Werte von  $\lambda$  schnell an und fällt dann auch schnell ab. Für grössere Werte von  $\lambda$  ist die Wahrscheinlichkeitsfunktion hingegen breiter und folglich auch flacher.



**Begründung von Formel (3.7):** Wir teilen das Zeitintervall der Länge  $t$  in  $n$  gleichlange, hinreichend kleine Teilintervalle  $\Delta t = t/n$  mit  $\mathbb{P}(X_{\Delta t} = 1) = \mu \cdot \Delta t$  auf. Dann liefert die Binomialverteilung

$$\mathbb{P}(X_t = k) = \binom{n}{k} (\mu \cdot \Delta t)^k (1 - \mu \Delta t)^{n-k},$$

da in  $k$  aus  $n$  Teilintervallen ein Anruf eingeht, während in  $n - k$  Teilintervallen kein Anruf registriert wird. Es folgt:

$$\begin{aligned} \mathbb{P}(X_t = k) &= \frac{n!}{k!(n-k)!} \left(\frac{\mu t}{n}\right)^k \left(1 - \frac{\mu t}{n}\right)^{n-k} \\ &= \frac{(\mu t)^k}{k!} \underbrace{\left(1 - \frac{\mu t}{n}\right)^n}_{\xrightarrow{n \rightarrow \infty} e^{-\mu t}} \overbrace{\frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{n^k}}^{k \text{ Faktoren}} \underbrace{\left(1 - \frac{\mu t}{n}\right)^k}_{\xrightarrow{n \rightarrow \infty} 1} \\ &\xrightarrow{n \rightarrow \infty} \frac{(\mu t)^k}{k!} e^{-\mu t}. \end{aligned}$$

Dies zeigt insbesondere, dass die Poisson-Verteilung als Grenzverteilung der Binomialverteilung im Fall  $p_n \cdot n \rightarrow \lambda$  und  $n \rightarrow \infty$  aufgefasst werden kann.

**Beispiel 3.32 (Telefonzentrale)** Wir suchen für eine Telefonzentrale die Wahrscheinlichkeit dafür, dass innerhalb einer Viertelstunde wenigstens 3 und höchstens 7 Anrufe ankommen. Dabei sei die gewählte Zeiteinheit “Minuten” und die entsprechende Intensität  $\mu = 1/3$ . Es gilt dann mit  $\lambda = \mu \cdot t = 1/3 \cdot 15 = 5$  für die Zufallsgröße  $X_t$ , welche die Anzahl der ankommenden Anrufe charakterisiert

$$X_t \sim \pi_{\mu t} = \pi_5,$$

also

$$\mathbb{P}(3 \leq X_t \leq 7) = \sum_{k=3}^7 \frac{5^k}{k!} e^{-5} = 0.742.$$

△

**Berechnung von  $\mathbb{E}(X_t)$  im Falle  $X_t \sim \pi_\lambda$ :** Es gilt

$$\mathbb{E}(X_t) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \stackrel{j:=k-1}{=} \lambda e^{-\lambda} \cdot \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}}_{=e^\lambda} = \lambda.$$

**Bemerkung 3.33** Für eine Poisson-verteilte Zufallsgröße  $X_t \sim \pi_{\mu t}$  gibt die Intensität  $\mu$  mit  $\mu = \mathbb{E}(X_t)/t$  die mittlere Anzahl von auftretenden Ereignissen pro Zeiteinheit an. △

**Berechnung von  $\mathbb{V}(X_t)$  im Falle von  $X_t \sim \pi_{\mu t}$ :** Mit

$$\begin{aligned} \mathbb{E}(X_t^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \left( \sum_{k=0}^{\infty} k \cdot (k-1) \frac{\lambda^k}{k!} e^{-\lambda} \right) + \underbrace{\left( \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \right)}_{=\mathbb{E}(X_t)} \\ &= \left( \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \right) + \lambda \\ &\stackrel{j:=k-2}{=} \lambda^2 e^{-\lambda} \left( \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}}_{=e^\lambda} \right) + \lambda \\ &= \lambda^2 + \lambda \end{aligned}$$

folgt

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

## 3.8 Hypergeometrische Verteilung

Als Referenzmodell zur hypergeometrischen Verteilung wird das schon aus der Kombinatorik bekannte Urnenmodell herangezogen. Eine Urne enthalte  $N$  Kugeln, von denen  $M$  schwarz sind. Der Rest der Kugeln sei weiß. Wir ziehen ohne Zurücklegen  $n$  Kugeln aus der Urne und zählen die dabei gezogenen schwarzen Kugeln, das heißt

$X \hat{=} \text{“Anzahl der entnommenen schwarzen Kugeln”}$ .

Es gilt

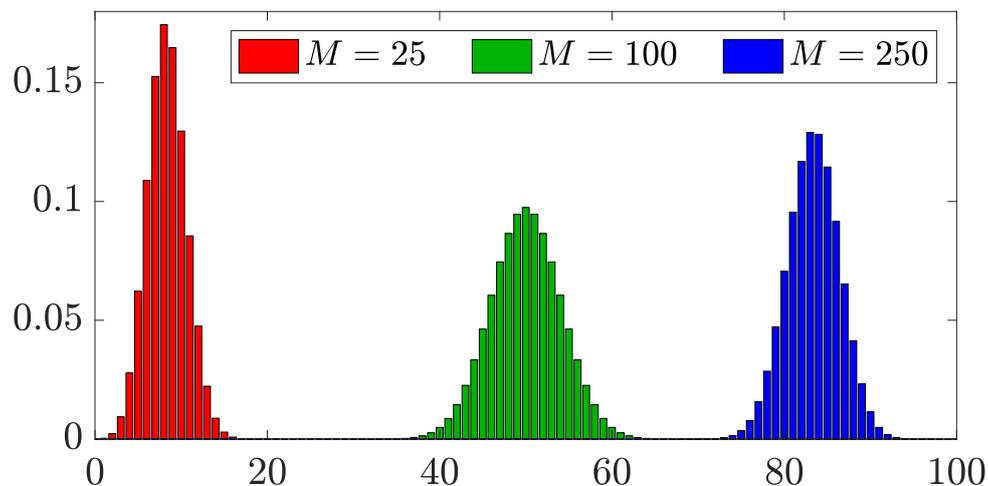
$$\mathbb{P}(X = m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}, \quad m = 0, 1, 2, \dots, \min\{n, M\}. \quad (3.8)$$

Hier haben wir benutzt, dass die Anzahl schwarze Kugeln, welche gezogen werden können, einerseits beschränkt ist durch  $n$  und andererseits auch beschränkt ist durch  $M$ .

**Definition 3.34 (hypergeometrische Verteilung)** Eine Zufallsgröße  $X$  mit der Wahrscheinlichkeitsfunktion (3.8) heißt **hypergeometrisch verteilt** mit den Parametern  $n, N, M$ , kurz

$$X \sim H(n, N, M).$$

Die Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung ist nachfolgend für  $n = 100$ ,  $N = 300$  und verschiedene Werte von  $M$  abgebildet. Im Spezialfall von  $M = N/2 = 150$  ist sie achsensymmetrisch um ihren Maximalwert angeordnet, während sie ansonsten unsymmetrisch ist.



**Berechnung von  $\mathbb{E}(X)$  im Falle  $X \sim H(n, N, M)$ :** Es gilt gemäß Definition des Erwartungswerts

$$\mathbb{E}(X) = \sum_{m=0}^{\min\{n, M\}} m \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}}.$$

Ersetzen der Binomialkoeffizienten liefert

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{m=1}^{\min\{n,M\}} m \cdot \frac{M!}{m!(M-m)!} \cdot \frac{(N-M)!}{(n-m)!((N-M)-(n-m))!} \cdot \frac{n!(N-n)!}{N!} \\
&= n \cdot \frac{M}{N} \sum_{m=1}^{\min\{n,M\}} \frac{(M-1)!}{(m-1)!(M-1-(m-1))!} \\
&\quad \cdot \frac{((N-1)-(M-1))!}{(((n-1)-(m-1))!(((N-1)-(M-1))-((n-1)-(m-1))))!} \\
&\quad \cdot \frac{(n-1)!((N-1)-(n-1))}{(N-1)!} \\
&= n \cdot \frac{M}{N} \sum_{m=1}^{\min\{n,M\}} \frac{\binom{M-1}{m-1} \binom{(N-1)-(M-1)}{(n-1)-(m-1)}}{\binom{N-1}{n-1}}.
\end{aligned}$$

Hierin erfüllt der letzte Term

$$\sum_{m=1}^{\min\{n,M\}} \frac{\binom{M-1}{m-1} \binom{(N-1)-(M-1)}{(n-1)-(m-1)}}{\binom{N-1}{n-1}} \stackrel{j:=m-1}{=} \sum_{j=0}^{\min\{n-1,M-1\}} \frac{\binom{M-1}{j} \binom{(N-1)-(M-1)}{(n-1)-j}}{\binom{N-1}{n-1}} = 1,$$

da dies der Aufsummation der Wahrscheinlichkeitsfunktion einer Zufallsgröße  $Y \sim H(n-1, N-1, M-1)$  entspricht. Damit erhalten wir schließlich

$$\mathbb{E}(X) = n \cdot \frac{M}{N}.$$

**Berechnung von  $\mathbb{V}(X)$  im Falle  $X \sim H(n, N, M)$ :** Ähnlich wie beim Erwartungswert ergibt sich

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{m=0}^{\min\{n,M\}} m^2 \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}} \\
&= \sum_{m=0}^{\min\{n,M\}} m(m-1) \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}} + \mathbb{E}(X) \\
&= n(n-1) \cdot \frac{M(M-1)}{N(N-1)} \underbrace{\sum_{m=2}^{\min\{n,M\}} \frac{\binom{M-2}{m-2} \binom{(N-2)-(M-2)}{(n-2)-(m-2)}}{\binom{N-2}{n-2}}}_{=1} + \mathbb{E}(X) \\
&= n(n-1) \cdot \frac{M(M-1)}{N(N-1)} + n \cdot \frac{M}{N}.
\end{aligned}$$

Hieraus folgt nach kleinerer Rechnung das Ergebnis:

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\
&= n(n-1) \cdot \frac{M(M-1)}{N(N-1)} + n \cdot \frac{M}{N} - n^2 \cdot \frac{M^2}{N^2} \\
&= \frac{nM(M-N)(n-N)}{N^2(N-1)} \\
&= n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}.
\end{aligned}$$

## 4. Stetige Verteilungen

### 4.1 Dichtefunktion

**Definition 4.1** (stetige Zufallsgröße) Eine Zufallsgröße  $X$  heißt **stetig**, wenn eine reelle integrierbare Funktion  $f(x)$  existiert, so dass

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) \, dx$$

gilt. Dabei heißt  $f(x)$  die **Dichtefunktion** der Zufallsgröße  $X$ .

#### Eigenschaften der Dichtefunktion:

1. Die Dichtefunktion ist nichtnegativ, dies bedeutet

$$f(x) \geq 0 \quad \text{für alle } x \in \mathbb{R}.$$

2. Es gilt

$$\mathbb{P}(-\infty < X < \infty) = \mathbb{P}(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x) \, dx = 1.$$

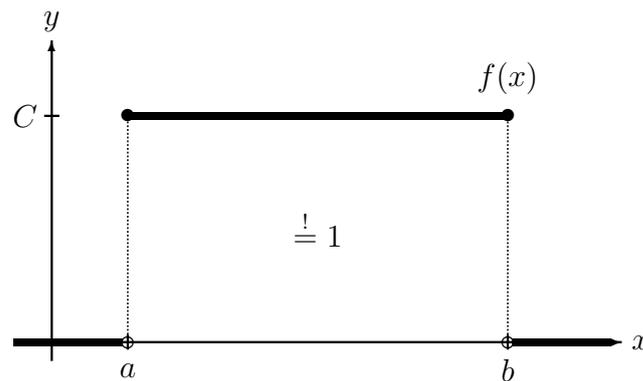
Das Integral ist dabei im Sinne von Riemann oder Lebesgue zu verstehen. Als Funktionen  $f(x)$  treten vorzugsweise stetige oder stückweise stetige Funktionen auf, die gegebenenfalls auch schwache Polstellen besitzen dürfen. Wegen

$$\mathbb{P}(a \leq X \leq a) = \mathbb{P}(X = a) = \int_a^a f(x) \, dx = 0$$

ist die Wahrscheinlichkeit, dass  $X$  genau einen festen Wert annimmt, immer gleich Null.

**Beispiel 4.2** (Gleichverteilung) Bei der Gleichverteilung auf dem Intervall  $[a, b]$  sind die Werte von  $X$  gleichwahrscheinlich über das Intervall  $[a, b]$  verteilt. Außerhalb des Intervalls  $[a, b]$  kann  $X$  keine Werte annehmen. Damit ergibt sich für die Dichtefunktion

$$f(x) = \begin{cases} C, & x \in [a, b], \\ 0, & \text{sonst.} \end{cases}$$



Die Normierungseigenschaft der Dichtefunktion impliziert

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_a^b C dx = C(b-a),$$

das heißt, der konkrete Wert der Konstanten ist

$$C = \frac{1}{b-a}.$$

△

## 4.2 Erwartungswert und Varianz

Wir wollen den Erwartungswert einer stetigen Zufallsgröße  $X$  mittels Approximation einführen. Dazu definieren wir für  $n \in \mathbb{N}$  die diskrete Zufallsgröße  $X_n$  mit den Werten  $k/n$  und der Wahrscheinlichkeitsfunktion

$$\mathbb{P}(X_n = k/n) = \mathbb{P}\left(\frac{k}{n} \leq X < \frac{k+1}{n}\right), \quad k \in \mathbb{Z}.$$

Insbesondere ist dann

$$X_n \leq X \leq X_n + \frac{1}{n}$$

und folglich

$$|X_n - X_m| \leq |X_n - X| + |X - X_m| \leq \frac{1}{n} + \frac{1}{m}.$$

Existiert  $\mathbb{E}(X_n)$  für ein  $n$ , so existiert wegen

$$\begin{aligned} |\mathbb{E}(X_m)| &\leq |\mathbb{E}(X_m) - \mathbb{E}(X_n)| + |\mathbb{E}(X_n)| \\ &\leq \mathbb{E}(|X_m - X_n|) + |\mathbb{E}(X_n)| \\ &\leq \frac{1}{n} + \frac{1}{m} + |\mathbb{E}(X_n)| \end{aligned}$$

auch  $\mathbb{E}(X_m)$  für beliebiges  $m \in \mathbb{N}$ . Speziell folgt aus

$$|\mathbb{E}(X_n) - \mathbb{E}(X_m)| \leq \frac{1}{n} + \frac{1}{m},$$

dass  $\mathbb{E}(X_n)$  eine Cauchy-Folge ist. Wir sagen, dass  $\mathbb{E}(X)$  existiert, falls  $\mathbb{E}(X_1)$  existiert, und setzen

$$\mathbb{E}(X) := \lim_{n \rightarrow \infty} \mathbb{E}(X_n).$$

**Satz 4.3 (Erwartungswert)** Es sei  $X$  eine stetige Zufallsgröße mit Dichtefunktion  $f$  und  $g : \mathbb{R} \rightarrow \mathbb{R}$  eine stetige Funktion. Dann existiert  $\mathbb{E}(g(X))$  genau dann, wenn

$$I := \int_{-\infty}^{\infty} |g(x)|f(x) dx < \infty,$$

und in diesem Fall gilt

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

Insbesondere ist

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

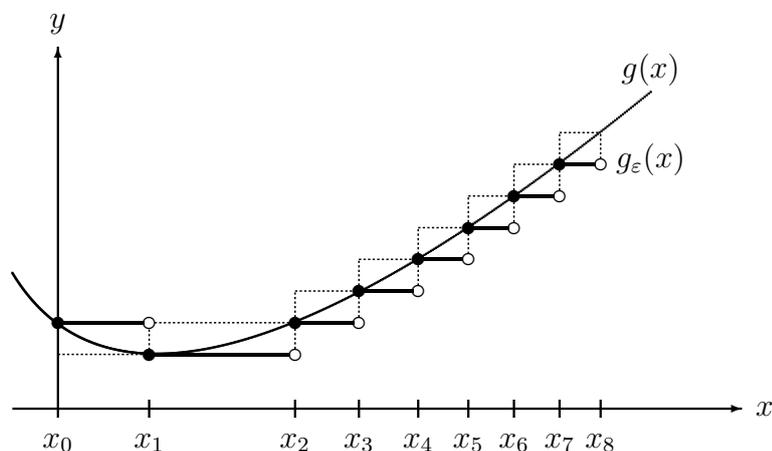
falls  $\int_{-\infty}^{\infty} |x|f(x) dx < \infty$ .

*Beweis.* Aufgrund der Stetigkeit von  $g$  existiert zu jedem  $\varepsilon > 0$  eine strikt monoton wachsende Folge  $\{x_n\}_{n \in \mathbb{Z}}$  mit  $x_n \rightarrow -\infty$  für  $n \rightarrow -\infty$  und  $x_n \rightarrow \infty$  für  $n \rightarrow \infty$ , so dass

$$|g(x) - g(x_n)| < \varepsilon \quad \text{für } x_n \leq x \leq x_{n+1}.$$

Setzen wir  $g_\varepsilon(x) := g(x_n)$  für alle  $x_n \leq x < x_{n+1}$  und  $n \in \mathbb{Z}$ , so gilt

$$|g_\varepsilon(x) - g(x)| < \varepsilon \quad \text{für alle } x \in \mathbb{R}. \quad (4.1)$$



Aus (4.1) folgern wir

$$\mathbb{E}(g_\varepsilon(X)) = \sum_{n=-\infty}^{\infty} g(x_n)\mathbb{P}(x_n \leq X < x_{n+1}) = \sum_{n=-\infty}^{\infty} g(x_n) \int_{x_n}^{x_{n+1}} f(x) dx.$$

Hierin konvergiert die letzte Summe genau dann absolut, wenn  $I$  endlich ist, denn

$$\begin{aligned} \sum_{n=-\infty}^{\infty} |g(x_n)| \int_{x_n}^{x_{n+1}} f(x) dx &\leq \int_{-\infty}^{\infty} |g(x)|f(x) dx + \sum_{n=-\infty}^{\infty} \int_{x_n}^{x_{n+1}} \underbrace{|g(x_n) - g(x)|}_{< \varepsilon} f(x) dx \\ &< I + \varepsilon \end{aligned}$$

und

$$\begin{aligned} \sum_{n=-\infty}^{\infty} |g(x_n)| \int_{x_n}^{x_{n+1}} f(x) dx &\geq \int_{-\infty}^{\infty} |g(x)| f(x) dx - \sum_{n=-\infty}^{\infty} \int_{x_n}^{x_{n+1}} \underbrace{|g(x_n) - g(x)|}_{< \varepsilon} f(x) dx \\ &> I - \varepsilon. \end{aligned}$$

Insbesondere ist wegen (4.1)

$$\begin{aligned} &\left| \mathbb{E}(g(X)) - \int_{-\infty}^{\infty} g(x) f(x) dx \right| \\ &\leq \underbrace{|\mathbb{E}(g(X)) - \mathbb{E}(g_\varepsilon(X))|}_{=|\mathbb{E}(g(X) - g_\varepsilon(X))|} + \left| \mathbb{E}(g_\varepsilon(X)) - \int_{-\infty}^{\infty} g(x) f(x) dx \right| < 2\varepsilon, \end{aligned}$$

woraus die Behauptung folgt.  $\square$

**Definition 4.4 (Varianz)** Sei  $X$  eine stetige Zufallsgröße mit Dichtefunktion  $f$  und

$$\int_{-\infty}^{\infty} x^2 f(x) dx < \infty.$$

Dann ist die **Varianz** von  $X$  definiert als

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

**Beispiel 4.5 (Gleichverteilung (Fortsetzung von Beispiel 4.2))** Die Dichtefunktion der Gleichverteilung lautet

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

Für den Erwartungswert folgt daher

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Weiter ergibt sich wegen

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} \frac{x^2}{b-a} dx = \int_a^b \frac{x^2}{b-a} dx = \frac{x^3}{3(b-a)} \Big|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

für die Varianz

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\ &= \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} \\ &= \frac{4a^2 + 4ab + 4b^2 - 3(a^2 + 2ab + b^2)}{12} \\ &= \frac{a^2 - 2ab + b^2}{12} \\ &= \frac{(a-b)^2}{12}. \end{aligned}$$

△

**Satz 4.6 (Tschebyscheffsche Ungleichung)** Auch für die stetigen Zufallsgrößen gilt die Tschebyscheffsche Ungleichung

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}, \quad \varepsilon > 0.$$

*Beweis.* Analog zum Beweis von Satz 3.21 folgt die Behauptung durch Umstellen von

$$\begin{aligned} \mathbb{V}(X) &= \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx \\ &\geq \int_{|x - \mathbb{E}(X)| \geq \varepsilon} \underbrace{(x - \mathbb{E}(X))^2}_{\geq \varepsilon^2} f(x) dx \\ &\geq \varepsilon^2 \int_{|x - \mathbb{E}(X)| \geq \varepsilon} f(x) dx \\ &= \varepsilon^2 \mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon). \end{aligned}$$

□

### 4.3 Verteilungsfunktion

**Definition 4.7 (Verteilungsfunktion)** Die reelle Funktion

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t) dt$$

heißt **Verteilungsfunktion** der stetigen Zufallsgröße  $X$ .

**Eigenschaften der Verteilungsfunktion:**

1. Die Verteilungsfunktion besitzt die beiden Asymptoten

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

2. Die Wahrscheinlichkeit dafür, dass  $a \leq X \leq b$  gilt, ist gegeben gemäß

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx = F(b) - F(a).$$

3.  $F(x)$  ist monoton wachsend (nicht notwendigerweise streng), das heißt

$$x < y \Rightarrow F(x) \leq F(y).$$

4.  $F(x)$  ist eine *stetige* Funktion für alle  $x \in \mathbb{R}$ . Dies bedeutet, dass wir im Gegensatz zur Verteilungsfunktion einer diskreten Zufallsgröße links- und rechtsseitige Stetigkeit haben.
5. Falls  $f(x)$  in  $y \in \mathbb{R}$  stetig ist, so ist  $F(x)$  in  $y$  differenzierbar und es gilt

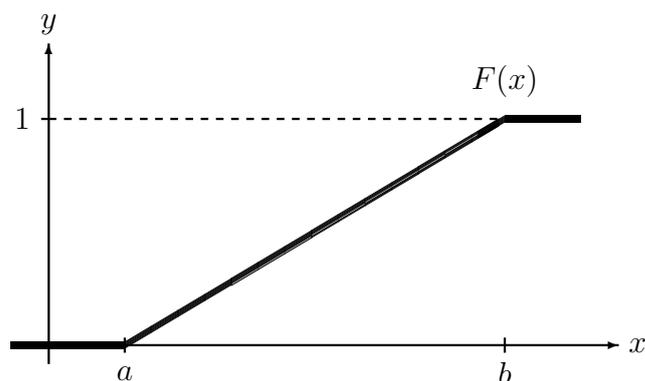
$$F'(y) = f(y).$$

**Beispiel 4.8** (Gleichverteilung (Fortsetzung von Beispiel 4.2)) Wir berechnen nun die Verteilungsfunktion für die Gleichverteilung. Sei  $x \in [a, b]$ , dann gilt

$$F(X) = \int_{-\infty}^x f(t) dt = \int_a^x \frac{1}{b-a} dt = \frac{t}{b-a} \Big|_a^x = \frac{x-a}{b-a}.$$

Folglich ergibt sich für die Verteilungsfunktion

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



△

## 4.4 Exponentialverteilung\*

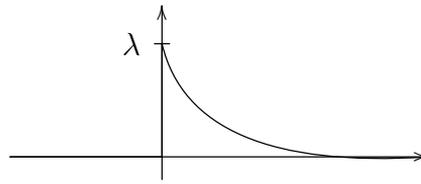
**Definition 4.9** (Exponentialverteilung) Eine stetige Zufallsgröße mit der Dichtefunktion

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

heißt **exponentialverteilt** mit dem Parameter  $\lambda > 0$ , kurz

$$X \sim \text{Ex}(\lambda).$$

Die Dichtefunktion  $f(x)$  besitzt folgendes Aussehen:



**Satz 4.10** Die Verteilungsfunktion einer exponentialverteilten Zufallsgröße lautet

$$F(X) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

*Beweis.* Im Falle  $x < 0$  ist nichts zu zeigen. Sei also  $x \geq 0$ , dann folgt

$$F(X) = \int_0^x \lambda \cdot e^{-\lambda t} dt = -e^{-\lambda t} \Big|_0^x = 1 - e^{-\lambda x}.$$

□

**Zusammenhang zwischen Exponential- und Poisson-Verteilung:** Wir betrachten wieder die Telefonzentrale, wobei

$\mu \hat{=}$  "Intensität",

$X_t \hat{=}$  "Anzahl der Anrufe im Zeitintervall der Länge  $t$ ",

$T \hat{=}$  "Zeitabstand zwischen 2 Anrufen"

bezeichne. Wir wissen bereits, dass

$$X_t \sim \pi_{\mu t}$$

gilt. Im nachfolgenden Satz wird gezeigt, dass  $T$  exponentialverteilt ist mit  $\mu$  als Parameter, das heißt

$$T \sim \text{Ex}(\mu).$$

**Satz 4.11** Die Zufallsgröße  $X_t$  zähle das Eintreten eines Ereignisses  $A$  innerhalb eines Zeitintervalls der Länge  $t$ . Die Zeit  $T$ , die nach dem Eintreten von  $A$  bis zum nächsten Eintreten verstreicht, ist exponentialverteilt

$$T \sim \text{Ex}(\mu),$$

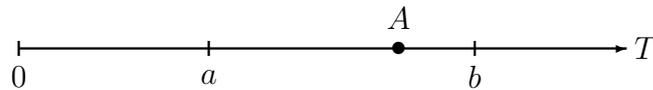
falls  $X_t$  Poisson-verteilt ist mit der Intensität  $\mu$ , das heißt

$$X_t \sim \pi_{\mu t}.$$

*Beweis.* Wir wollen zeigen, dass

$$\mathbb{P}(a \leq T \leq b) = \int_a^b \mu \cdot e^{-\mu t} dt$$

gilt. Hierzu beachte man, dass zum Zeitpunkt  $a$  das Ereignis  $A$  noch nicht, zum Zeitpunkt  $b$  jedoch *mindestens* einmal eingetreten ist (vergleiche Skizze).



Dies bedeutet,

$$X_a = 0, \quad \text{und} \quad X_b \geq 1.$$

Aus  $[0, a] \subseteq [0, b]$  folgt

$$\begin{aligned} \mathbb{P}(T \in [a, b]) &= \mathbb{P}(a \leq T \leq b) = \mathbb{P}(X_a = 0) - \mathbb{P}(X_b = 0) = e^{-\mu a} - e^{-\mu b} \\ &= \int_a^b \mu e^{-\mu t} dt. \end{aligned}$$

□

**Beispiel 4.12 (Telefonzentrale)** In einer Telefonzentrale kommen im Mittel pro Stunde 20 Anrufe an. Gesucht ist die Wahrscheinlichkeit dafür, dass 3 bis 6 Minuten zwischen zwei Anrufen vergehen. Legt man Minuten als verbindliche Zeiteinheit fest, so folgt aus

$$\mathbb{E}(X_{60}) = \mu \cdot 60 \stackrel{!}{=} 20,$$

dass  $\mu = \frac{1}{3}$  gilt.  $T$  sei die Zufallsgröße, welche die Zeit zwischen 2 Anrufen misst. Dann gilt

$$\begin{aligned} \mathbb{P}(3 \leq T \leq 6) &= F(6) - F(3) \\ &= (1 - e^{-\mu \cdot 6}) - (1 - e^{-\mu \cdot 3}) \\ &= -e^{-2} + e^{-1} \\ &= 0.2325. \end{aligned}$$

△

**Erwartungswert und Streuung der Exponentialverteilung:** Gemäß Definition des Erwartungswertes gilt

$$\mathbb{E}(X) = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx.$$

Substituieren wir  $t = \lambda x$ , so folgt

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{\lambda} \int_0^{\infty} \underset{u}{t} \cdot \underset{v'}{e^{-t}} dt \\ &= \frac{1}{\lambda} \left\{ \underset{u}{t} \cdot \underset{v}{(-e^{-t})} \Big|_0^{\infty} - \int_0^{\infty} \underset{u'}{1} \cdot \underset{v}{(-e^{-t})} dt \right\} \\ &= \frac{1}{\lambda} \left\{ -0 + 0 - e^{-t} \Big|_0^{\infty} \right\} \\ &= \frac{1}{\lambda} \{0 + 1\} = \frac{1}{\lambda}. \end{aligned}$$

In ähnlicher Weise folgt

$$\begin{aligned}\mathbb{E}(X^2) &= \int_0^\infty x^2 \cdot \lambda e^{-\lambda x} dx \\ &= \frac{1}{\lambda^2} \int_0^\infty \underset{u}{t^2} \underset{v'}{e^{-t}} dt \\ &= \frac{1}{\lambda^2} \left\{ \underset{u}{t^2} \cdot \underset{v}{(-e^{-t})} \Big|_0^\infty - \int_0^\infty \underset{u'}{2t} \cdot \underset{v}{(-e^{-t})} dt \right\} \\ &= \frac{1}{\lambda} \{-0 + 0 + 2\} = \frac{2}{\lambda^2},\end{aligned}$$

womit sich

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

ergibt.

**Exponentialverteilung als Lebensdauer-Verteilung:** Wartezeiten, Reparaturzeiten und die Lebensdauer von Bauelementen können als exponentialverteilt angenommen werden. Allerdings ist dabei zu beachten, dass *keine* Alterungseffekte modelliert werden können, wie folgende Überlegung zeigt:

Es sei  $X \sim \text{Ex}(\lambda)$  die Lebensdauer eines Bauelements. Dann gilt

$$\begin{aligned}\mathbb{P}(X \leq x + y \mid X \geq x) &= \frac{\mathbb{P}(x \leq X \leq x + y)}{\mathbb{P}(X \geq x)} \\ &= \frac{F(x + y) - F(x)}{1 - F(x)} \\ &= \frac{(1 - e^{-\lambda(x+y)}) - (1 - e^{-\lambda x})}{1 - (1 - e^{-\lambda x})} \\ &= \frac{e^{-\lambda x} - e^{-\lambda(x+y)}}{e^{-\lambda x}} \\ &= \frac{e^{-\lambda x} - e^{-\lambda x} e^{-\lambda y}}{e^{-\lambda x}} \\ &= 1 - e^{-\lambda y} = F(y) = \mathbb{P}(X \leq y).\end{aligned}$$

Dies bedeutet, dass die Wahrscheinlichkeit dafür, dass ein intaktes Bauteil in den nächsten  $y$  Zeiteinheiten kaputtgeht, unabhängig von dessen Alter  $X$  immer gleich ist.

## 4.5 Normalverteilung

Die *Normalverteilung*, die auch als *Gaußsche Verteilung* bezeichnet wird, ist die wohl bekannteste stetige Verteilung. Ihre Dichtefunktion ist die allseits beliebte *Gaußsche Glockenkurve*.

**Definition 4.13 (Normalverteilung)** Eine stetige Zufallsgröße heißt **normalverteilt** mit den Parametern  $\mu$  und  $\sigma^2 > 0$ , wenn ihre Dichtefunktion der Gleichung

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

genügt. Wir schreiben dann auch kurz

$$X \sim N(\mu, \sigma^2).$$

Die Verteilungsfunktion der  $N(\mu, \sigma^2)$ -Normalverteilung

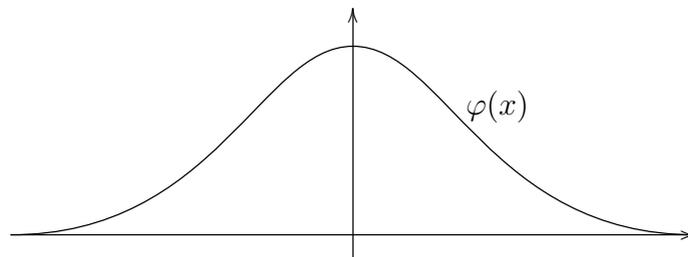
$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

ist im allgemeinen nicht geschlossen auswertbar. Daher nutzt man Techniken der *Standardisierung*, um wenigstens im Spezialfall die tabellierten Werte der Verteilungsfunktion nutzen zu können.

**Definition 4.14** (Standardisiert normalverteilt) Im Falle  $X \sim N(0, 1)$  bezeichnet man die Zufallsgröße  $X$  als **standardisiert normalverteilt**.

Für die Dichtefunktion einer standardisiert normalverteilten Zufallsgröße  $X$  gilt

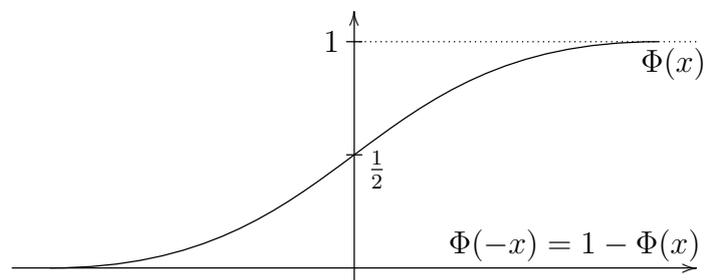
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty.$$



Die zugehörige Verteilungsfunktion

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt$$

ist für  $x \geq 0$  tabelliert.



Für die Berechnung der Werte der Verteilungsfunktion für negative Argumente nutzt man den aus der Symmetrie der Verteilung resultierenden Zusammenhang

$$\Phi(-x) = 1 - \Phi(x).$$

**Satz 4.15** Die standardisiert normalverteilte Zufallsgröße  $X \sim N(0, 1)$  besitzt den Erwartungswert  $\mathbb{E}(X) = 0$  und die Varianz  $\mathbb{V}(X) = 1$ .

+	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Tabelle 4.1: Tabelle zur Normalverteilung.

*Beweis.* Wegen  $x\varphi(x) = -\varphi'(x)$  folgt

$$\int_{-\infty}^{\infty} |x|\varphi(x) dx = 2 \lim_{C \rightarrow \infty} \int_0^C \varphi'(x) dx = 2 \lim_{C \rightarrow \infty} (\varphi(0) - \varphi(C)) = 2\varphi(0) < \infty$$

und daher ist (die Funktion  $x\varphi(x)$  ist punktsymmetrisch bezüglich 0)

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\varphi(x) dx = 0.$$

Weiter ergibt sich mit partieller Integration

$$\mathbb{V}(X) = \mathbb{E}(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot x e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \left[ \underbrace{-x e^{-\frac{x^2}{2}}}_{=0} \Big|_{-\infty}^{\infty} + \underbrace{\int_{-\infty}^{\infty} 1 \cdot e^{-\frac{x^2}{2}} dx}_{=\sqrt{2\pi}} \right] = 1.$$

Das letzte Integral rechnet man dabei wie folgt aus:

$$\left( \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right)^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}} dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy.$$

Verwenden wir nun Polarkoordinaten

$$x = r \cos \varphi, \quad y = r \sin \varphi,$$

so ergibt sich schließlich die gesuchte Identität

$$\left( \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right)^2 = \int_0^{2\pi} \int_0^{\infty} r \cdot e^{-\frac{r^2}{2}} dr d\varphi = \int_0^{2\pi} -e^{-\frac{r^2}{2}} \Big|_0^{\infty} d\varphi = \int_0^{2\pi} 1 d\varphi = 2\pi.$$

□

**Bemerkung 4.16** Die lineare Transformation

$$Y = \frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}}$$

einer Zufallsgröße  $X$  heißt *Standardisierung* von  $X$ . Dabei gilt

$$\mathbb{E}(Y) = \mathbb{E} \left( \frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}} \right) = \frac{1}{\sqrt{\mathbb{V}(X)}} (\mathbb{E}(X) - \mathbb{E}(X)) = 0$$

und

$$\mathbb{V}(Y) = \mathbb{V} \left( \frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}} \right) = \frac{1}{\mathbb{V}(X)} \mathbb{V}(X) = 1.$$

△

**Satz 4.17** Falls  $X \sim N(\mu, \sigma^2)$ , so gilt  $(X - \mathbb{E}(X))/\sqrt{\mathbb{V}(X)} \sim N(0, 1)$ . Insbesondere gilt  $\mathbb{E}(X) = \mu$  und  $\mathbb{V}(X) = \sigma^2$ .

*Beweis.* Sei  $Y \sim N(0, 1)$ , dann folgt für  $X := \sigma Y + \mu$  dass

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq \sigma Y + \mu \leq b) = \mathbb{P}\left(\frac{a - \mu}{\sigma} \leq Y \leq \frac{b - \mu}{\sigma}\right) = \int_{\frac{a - \mu}{\sigma}}^{\frac{b - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Mit der Substitution

$$x := \sigma y + \mu$$

ergibt sich hieraus

$$\mathbb{P}(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx,$$

dies bedeutet,  $X \sim N(\mu, \sigma^2)$ . Schließlich erhalten wir

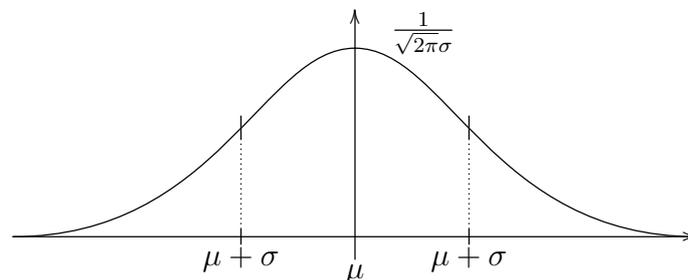
$$\mathbb{E}(X) = \mathbb{E}(\sigma Y + \mu) = \sigma \mathbb{E}(Y) + \mu = \mu$$

und

$$\mathbb{V}(X) = \mathbb{V}(\sigma Y + \mu) = \sigma^2 \mathbb{V}(Y) = \sigma^2.$$

□

**Interpretation der Parameter  $\mu$  und  $\sigma^2$ :** Die Dichtefunktion der Normalverteilung ist eine Glockenkurve, die durch den Lageparameter  $\mu$  und den Formparameter  $\sigma^2 > 0$  charakterisiert wird.



Der Parameter  $\mu$  gibt die Lage der Symmetrieachse an, während  $\sigma^2$  die Breite der Glocke bestimmt. Die Glocke hat ihr Maximum in  $\mu$  mit dem Wert  $\frac{1}{\sqrt{2\pi}\sigma}$  und besitzt in  $\mu \pm \sigma$  je einen Wendepunkt. Ist  $\sigma^2$  groß, so erhalten wir eine breitgezogene Glockenkurve, für kleines  $\sigma^2$  ergibt sich hingegen eine nadelförmige Glockenkurve.

**Wahrscheinlichkeitsberechnung bei normalverteilten Zufallsgrößen:** Die praktischen Berechnungen zur Normalverteilung erfordern in der Regel die Bestimmung von Wahrscheinlichkeiten des Typs  $\mathbb{P}(a \leq X \leq b)$  für  $X \sim N(\mu, \sigma^2)$ . Durch Ausnutzung des Standardisierungsgedankens führt man dies auf die Berechnung einer Differenz zweier Werte der Verteilungsfunktion  $\Phi$  der Standardnormalverteilung zurück:

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}\left(\frac{a - \mu}{\sigma} \leq \underbrace{\frac{X - \mu}{\sigma}}_{=: Y \sim N(0,1)} \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Als normalverteilt können Zufallsgrößen angesehen werden, die durch Überlagerung einer großen Anzahl von Einflüssen entstehen, wobei jede Einflussgröße nur einen im Verhältnis zur Gesamtsumme unbedeutenden Beitrag liefert. Beispiele normalverteilter Zufallsgrößen sind:

- zufällige Beobachtungs- oder Messfehler,
- zufällige Abweichungen vom Nennmaß bei der Fertigung von Werkstücken,
- Effekte beim Prozess der Brownschen Molekularbewegung.

**Beispiel 4.18 (Fertigungstoleranzen)** Ein Werkstück besitzt die gewünschte Qualität, wenn die Abweichung eines bestimmten Maßes vom entsprechendem Nennmaß dem Betrage nach nicht größer als  $3.6mm$  ist. Der Herstellungsprozess sei so beschaffen, dass dieses Maß als normalverteilte Zufallsgröße angesehen werden kann, wobei der Erwartungswert mit dem Nennmaß übereinstimmt. Weiter sei  $\sigma = 3mm$  bekannt. Wieviel Prozent der Werkstücke einer Serie werden durchschnittlich mit gewünschter Qualität produziert?

Es sei

$$X \hat{=} \text{“zufällige Abweichung vom Nennmaß”},$$

dann gilt

$$X \sim N(0, 9).$$

Wegen

$$\begin{aligned} \mathbb{P}(|X| \leq 3.6) &= P\left(\underbrace{\frac{|X|}{3}}_{\sim N(0,1)} \leq \frac{3.6}{3}\right) = \Phi(1.2) - \Phi(-1.2) = \Phi(1.2) - (1 - \Phi(1.2)) \\ &= 2\Phi(1.2) - 1 = 0.88493 \cdot 2 - 1 = 0.76983 \end{aligned}$$

genügen durchschnittlich etwa 77% aller Werkstücke den Qualitätsansprüchen.  $\triangle$

## 5. Simulationsverfahren

### 5.1 Pseudozufallszahlen

Voraussetzung für die stochastische Simulation sind sogenannte gleichverteilte Pseudozufallszahlen, die von Pseudozufallszahlengeneratoren erzeugt werden. Unter Pseudozufallszahlen versteht man Zahlenfolgen, die vom Computer mittels gewissen Algorithmen berechnet werden. Sie sollen sich wie "echte" Zufallszahlen verhalten. Eine oftmals verwendete Klasse solcher Algorithmen bilden die *linearen Kongruenzgeneratoren*.

Man wählt nichtnegative ganze Zahlen  $m$  (den Modul),  $a$  (den Multiplikator),  $b$  (das Inkrement) und  $z_0$  (den Startwert) mit  $z_0 \leq m - 1$ , und berechnet rekursiv

$$z_{j+1} := (az_j + b) \pmod{m}, \quad j = 0, 1, \dots \quad (5.1)$$

Dabei bedeutet das Rechnen *modulo*  $m$ , dass der beim Teilen durch  $m$  übrig bleibende kleinste nichtnegative Rest der Zahl  $az_j + b$  gebildet wird. Demnach gilt  $0 \leq z_j < m$ . Durch die Normierungsvorschrift

$$x_j := \frac{z_j}{m}, \quad j = 0, 1, \dots,$$

liefert das Schema (5.1) eine Folge  $x_0, x_1, \dots$  im Einheitsintervall  $[0, 1]$ . Bei einigen gebräuchlichen linearen Kongruenzgeneratoren sind die verschiedenen Konstanten wie folgt gewählt:

Name	Modul $m$	Multiplikator $a$	Inkrement $b$
Texas Instruments (TI 59)	199017	24298	99991
Turbo Pascal	$2^{32}$	33797	1
ANSI-C	$2^{31}$	1103515245	12345
Minimal Standard	$2^{31} - 1$	16807	0
RANDU	$2^{31}$	65539	0

**Beispiel 5.1** Als Zahlenbeispiel betrachten wir den Fall  $m = 100$ ,  $a = 18$ ,  $b = 11$  und  $z_0 = 50$ . Hier gilt

$$\begin{aligned} z_1 &= (18 \cdot 50 + 11) \pmod{100} = 911 \pmod{100} = 11, \\ z_2 &= (18 \cdot 11 + 11) \pmod{100} = 209 \pmod{100} = 9, \\ z_3 &= (18 \cdot 9 + 11) \pmod{100} = 173 \pmod{100} = 73, \\ z_4 &= (18 \cdot 73 + 11) \pmod{100} = 1325 \pmod{100} = 25, \\ z_5 &= (18 \cdot 25 + 11) \pmod{100} = 461 \pmod{100} = 61, \end{aligned}$$

dies bedeutet, es werden die Pseudozufallszahlen

$$x_0 = 0.5, \quad x_1 = 0.11, \quad x_2 = 0.09, \quad x_3 = 0.73, \quad x_4 = 0.25, \quad x_5 = 0.61$$

erzeugt. △

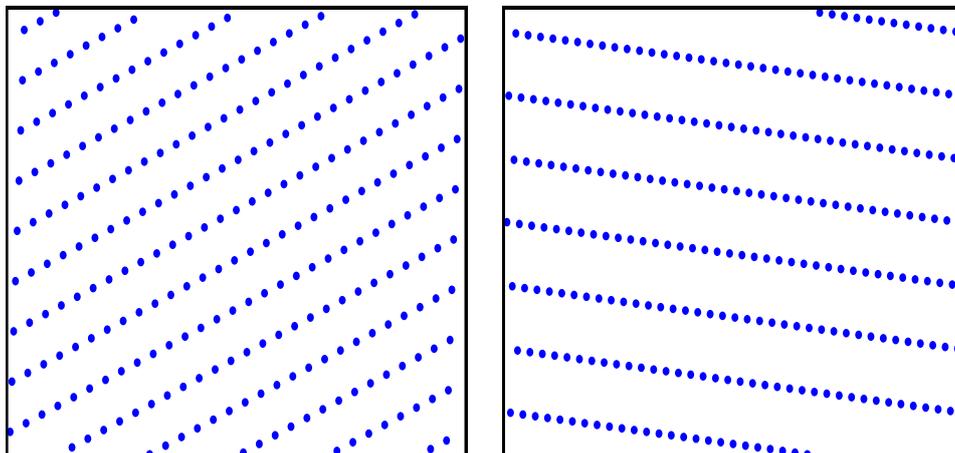
Weil stets gilt  $z_j \in \{0, 1, \dots, m-1\}$ , kann jeder lineare Kongruenzgenerator mit Modul  $m$  höchstens  $m$  verschiedene Zufallszahlen erzeugen. Daher wählt man  $m$  sehr groß, nämlich  $m = 2^k$  auf Binärcomputern, wenn  $k$  die maximale Wortlänge ist. Im obigen Zahlenbeispiel ist  $z_6 = 9 = z_2$  (bitte nachrechnen!), so dass nur sechs verschiedene Zahlen auftreten und der Generator schon nach zwei Schritten in die Periode 9, 73, 25, 61 der Länge vier läuft. Die maximal mögliche Periodenlänge  $m$  ist für  $b \geq 1$  genau dann sichergestellt, wenn folgende Bedingungen erfüllt sind:

- $b$  ist teilerfremd zu  $m$ .
- Jede Primzahl, die  $m$  teilt, teilt auch  $a - 1$ .
- Ist  $m$  durch 4 teilbar, so muss auch  $a - 1$  durch 4 teilbar sein.

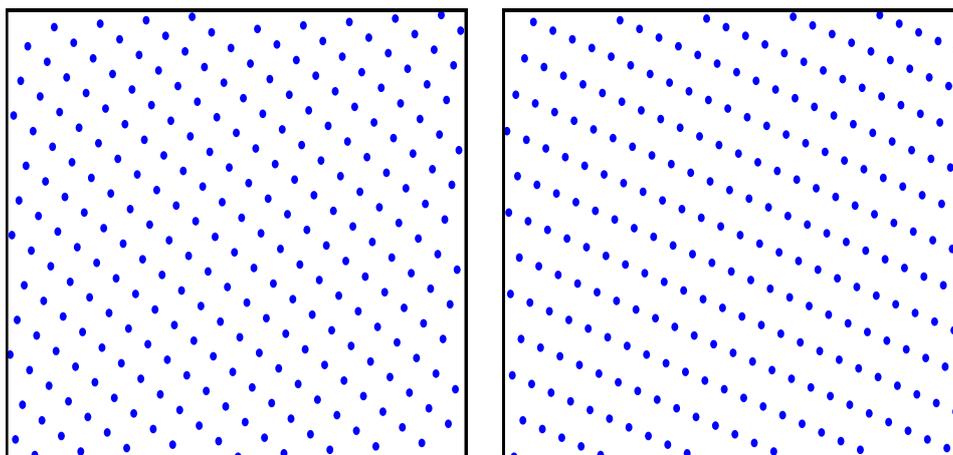
Dass ein linearer Kongruenzgenerator maximale Periodenlänge besitzt, bedeutet allerdings nur, dass alle Zahlen  $j/m$ ,  $0 \leq j < m$ , nach  $(m-1)$ -maligem Aufruf der Iterationsvorschrift (5.1) genau einmal aufgetreten sind. Die obigen drei Bedingungen sagen jedoch noch nichts über die statistische Qualität der erzeugten Zufallszahlen aus. So besitzt etwa das lineare Kongruenzschema  $z_{j+1} = (z_j + 1) \bmod m$  die maximale Periodenlänge  $m$ ; die hierdurch generierte Folge wird man jedoch kaum als zufällig erzeugt ansehen. Um die Aussicht auf die Vermeidung derart pathologischer Fälle zu vergrößern, empfiehlt es sich, den Multiplikator  $a$  weder zu klein noch zu groß zu wählen<sup>1)</sup>.

Ein grosser Nachteil linearer Kongruenzgeneratoren ist deren Gitterstruktur. Dies bedeutet, dass für jedes  $n \geq 2$  die  $n$ -dimensionalen Vektoren  $[x_i, x_{i+1}, \dots, x_{i+n-1}]^T$ ,  $i \geq 0$ , auf einem regelmäßigen Gitter im  $n$ -dimensionalen Einheitswürfel  $[0, 1]^n$  liegen. Der Abstand zwischen nachfolgenden Zufallszahlen ist also nicht willkürlich, was dem Zufallsgedanken offensichtlich widerspricht.

**Beispiel 5.2** Für den linearen Kongruenzgenerator mit  $m = 256$ ,  $a = 29$ ,  $b = 13$  und  $z_0 = 1$  zeigen die nachfolgenden Bilder, dass die 256 Pseudozufalls-Paare  $(z_j, z_{j+1})$ ,  $(z_j, z_{j+2})$ ,  $(z_j, z_{j+3})$  und  $(z_j, z_{j+4})$  offensichtlich allesamt auf Geraden liegen:



<sup>1)</sup>Es wird empfohlen, den Faktor  $a$  so zu wählen, dass er der Ungleichung  $\sqrt{m} < a < m - \sqrt{m}$  genügt.



△

Zur theoretischen Begründung dieser Gitterstruktur betrachten wir für festes  $n \geq 2$  die Spaltenvektoren  $\mathbf{z}_j = [z_j, z_{j+1}, \dots, z_{j+n-1}]^T$ ,  $0 \leq j < m$ . Durch vollständige Induktion nach  $j$  ergibt sich aus (5.1) zunächst

$$z_{j+k} - z_j = (a^j(z_k - z_0)) \pmod{m}, \quad j, k \geq 0.$$

Dies impliziert

$$\mathbf{z}_j - \mathbf{z}_0 = \begin{bmatrix} z_j - z_0 \\ a(z_j - z_0) \\ \vdots \\ a^{n-1}(z_j - z_0) \end{bmatrix} \pmod{m} = (z_j - z_0) \begin{bmatrix} 1 \\ a \\ \vdots \\ a^{n-1} \end{bmatrix} \pmod{m},$$

wobei die Kongruenz modulo  $m$  komponentenweise zu verstehen ist. Nach Definition der Kongruenzrelation schließen wir daher

$$\mathbf{z}_j - \mathbf{z}_0 = (z_j - z_0) \begin{bmatrix} 1 \\ a \\ \vdots \\ a^{n-1} \end{bmatrix} + m \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{n-1} \end{bmatrix}$$

für geeignete ganze Zahlen  $s_1, s_2, \dots, s_n$ . Folglich können die Differenzen  $\mathbf{z}_j - \mathbf{z}_0$  als ganzzahlige Linearkombinationen der Vektoren

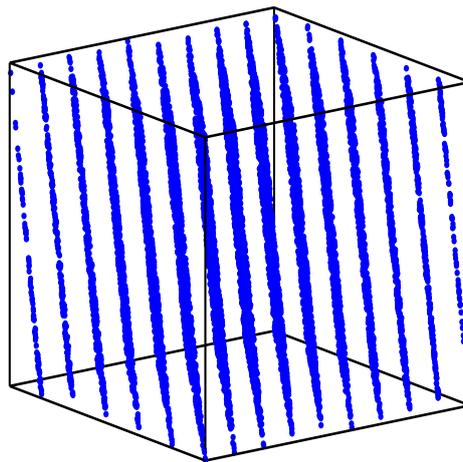
$$\begin{bmatrix} 1 \\ a \\ \vdots \\ a^{n-1} \end{bmatrix}, \begin{bmatrix} m \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ m \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ m \end{bmatrix}$$

dargestellt werden. Da sich der Vektor  $[m, 0, \dots, 0]^T$  als (ganzzahlige) Linearkombination der anderen Vektoren schreiben lässt, erkennt man, dass alle Differenzen  $\mathbf{z}_j - \mathbf{z}_0$  auf dem Gitter

$$\mathcal{G}_m = \left\{ \mathbf{z} = s_1 \begin{bmatrix} 1 \\ a \\ \vdots \\ a^{n-1} \end{bmatrix} + s_2 \begin{bmatrix} 0 \\ m \\ \vdots \\ 0 \end{bmatrix} + \dots + s_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ m \end{bmatrix} : s_1, \dots, s_n \in \mathbb{Z} \right\}$$

liegen. Die Punkte  $[x_j, x_{j+1}, \dots, x_{j+n-1}]^T$  liegen daher aufgrund der Normierungsvorschrift alle auf einem Gitter, das sich aus  $\mathcal{G}_m$  durch Verschiebung um  $\mathbf{z}_0$  und Skalierung mit dem Faktor  $1/m$  ergibt.

**Beispiel 5.3** Bei dem von IBM Anfang der sechziger Jahre eingeführten und fast 10 Jahre lang weitverbreiteten linearen Kongruenzgenerator RANDU ist  $a = 65539, b = 0, m = 2^{31}$ . Die Periodenlänge ist zwar mit  $2^{29}$  fast maximal, aber alle  $2^{29}$  Tripel  $[x_i, x_{i+1}, x_{i+2}]^T$  liegen auf nur 15 parallelen Ebenen, wie dies aus der nachfolgenden Abbildung ersichtlich ist:



△

Ziel bei der Konstruktion eines linearen Kongruenzgenerators ist demnach, die Zahlen  $a$  und  $b$  so zu wählen, dass die Gitterstruktur hinreichend fein ist. Prinzipiell kann die Gitterstruktur auch durch einen nichtlinearen Kongruenzgenerator vermieden werden, also der Rekursion  $z_{j+1} = g(z_j)$  mit einer nichtlinearen Funktion  $g : \{0, 1, \dots, m-1\} \rightarrow \{0, 1, \dots, m-1\}$ , etwa  $g(z) = (az^2 + bz + c) \bmod m$ . Derartige Generatoren vermeiden zwar die Gitterstruktur der  $n$ -Tupel, aber der Rechenaufwand wird gross. Zudem hat man immer noch höchstens  $m$  verschiedene  $n$ -Tupel. Daher haben sich diese Generatoren nicht wirklich durchgesetzt.

In den letzten Jahren hat sich der Mersenne-Twister als Zufallszahlengenerator der Wahl etabliert. Der Algorithmus gehört zur Klasse der *twisted generalized feedback shift register*. Er besitzt eine extrem lange Periode ( $2^{19937} - 1 \approx 4.3 \cdot 10^{6001}$ ), erzeugt sehr gute Gleichverteilungen (nachgewiesen bis zur Dimension 623) und ist schnell. Er wird in allen modernen Programmbibliotheken wie beispielsweise der GNU Scientific Library eingesetzt.

## 5.2 Simulation beliebiger Verteilungen

### 5.2.1 Diskrete Verteilungen

Gegeben sei ein Experiment, das mit Wahrscheinlichkeit  $p_j$  den Ausgang  $j \in \{1, \dots, n\}$  besitzt, wobei  $p_1 + \dots + p_n = 1$  gelte. Wir erzeugen eine gleichverteilte Pseudozufallszahl  $x \in [0, 1]$  und stellen fest, in welchem der disjunkten Intervalle  $[0, p_1), [p_1, p_1 + p_2), \dots, [p_1 + p_2 + \dots + p_{n-1}, 1]$  sie liegt. Das Experiment besitzt den Ausgang  $j$ , wenn  $x$  in das Intervall mit dem rechten Endpunkt  $p_1 + \dots + p_j$  fällt. Weil für eine auf  $[0, 1]$  gleichverteilte

Zufallsgröße  $X$  gerade gilt

$$\mathbb{P}(p_1 + \dots + p_{j-1} \leq X < p_1 + \dots + p_j) = p_j,$$

kann auf diese Weise die zugrundeliegende diskrete Verteilung simuliert werden.

## 5.2.2 Verwerfungsmethode

Gegeben sei eine stetige Verteilung auf  $[0, 1]$  mit Verteilungsfunktion  $F(x)$  und beschränkter Dichtefunktion  $f(x)$ . Wir bilden ein Tupel  $(x, y)$  aus zwei auf  $[0, 1]$  gleichverteilten Pseudozufallszahlen. Falls

$$y \leq g(x) := \frac{f(x)}{\max_{z \in [0,1]} f(z)} \leq 1$$

ist, dann wird  $x$  angenommen. Ansonsten wird  $x$  verworfen und ein neues Tupel gezogen.

### Algorithmus 5.4 (Verwerfungsmethode)

**input:** Dichtefunktion  $f : [0, 1] \rightarrow [0, M]$

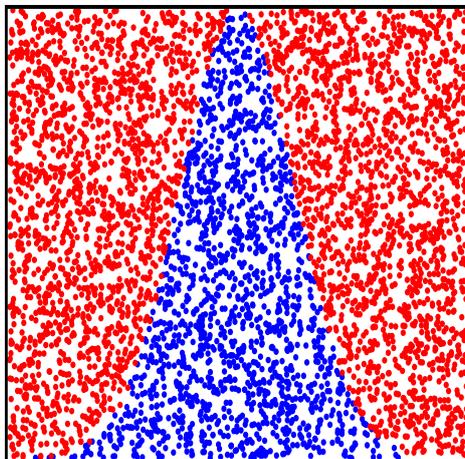
**output:** Zufallszahl  $x$  gemäss der vorgegebenen Verteilung

- ① ziehe zwei auf  $[0, 1]$  gleichverteilte Zufallszahlen  $(x, y)$
- ② falls gilt

$$My \leq f(x),$$

dann gebe  $x$  aus, ansonsten gehe nach ①.

**Beispiel 5.5** Simulieren wir beispielsweise die Normalverteilung  $N(0.5, 16)$  eingeschränkt auf das Intervall  $[0, 1]$ , dann ergibt sich für 5000 Punktepaare das folgende Bild, wobei akzeptierte Paare blau und verworfene Paare rot gekennzeichnet sind:



**Satz 5.6** Die Verwerfungsmethode produziert Zufallszahlen gemäß der durch die Verteilungsfunktion  $F(x)$  gegebenen Verteilung.

*Beweis.* Wir müssen nachweisen, dass die Verwerfungsmethode wirklich Zufallszahlen der gesuchten Verteilung  $F$  generiert. Dazu muss gezeigt werden, dass für zwei auf  $[0, 1]$  gleichverteilte Zufallsgrößen  $X, Y$  gilt

$$\mathbb{P}(X \leq a | Y \leq g(X)) = F(a) \quad \text{für alle } a \in [0, 1].$$

Gemäß der Definition der bedingten Wahrscheinlichkeit folgt

$$\mathbb{P}(X \leq a | Y \leq g(X)) = \frac{\mathbb{P}((X \leq a) \cap (Y \leq g(X)))}{\mathbb{P}(Y \leq g(X))}. \quad (5.2)$$

Um den Nenner in (5.2) auszurechnen, beachten wir die Identität

$$\mathbb{P}(Y \leq g(X) | X = x) = \mathbb{P}(Y \leq g(x)) = \int_0^{g(x)} dy = g(x).$$

Folglich gilt

$$\begin{aligned} \mathbb{P}(Y \leq g(X)) &= \int_0^1 \mathbb{P}(Y \leq g(X) | X = x) dx \\ &= \int_0^1 g(x) dx \\ &= \frac{1}{\max_{z \in [0,1]} f(z)}. \end{aligned}$$

Analog kann man den Zähler in (5.2) ausrechnen:

$$\begin{aligned} \mathbb{P}((X \leq a) \cap (Y \leq g(X))) &= \int_0^a \mathbb{P}(Y \leq g(X) | X = x) dx \\ &= \int_0^a g(x) dx \\ &= \frac{F(a)}{\max_{z \in [0,1]} f(z)}. \end{aligned}$$

Damit ist der Beweis vollständig erbracht. □

### 5.2.3 Inversionsmethode

Es sei  $F(x) := \mathbb{P}(X \leq x)$  eine Verteilungsfunktion, wobei  $X$  eine Zufallsgröße und  $\mathbb{P}$  die mit  $F$  verbundene Wahrscheinlichkeit ist. Die Inversionsmethode beruht auf dem folgenden Resultat.

**Satz 5.7** Es seien  $Y$  eine auf  $[0, 1]$  gleichverteilte Zufallsgröße und  $F$  eine stetige und streng monoton wachsende Verteilungsfunktion. Dann besitzt  $X := F^{-1}(Y)$  die Verteilungsfunktion  $F$ .

*Beweis.* Da  $Y$  auf  $[0, 1]$  gleichverteilt ist, gilt

$$\mathbb{P}(Y \leq y) = y \quad \text{für alle } y \in [0, 1].$$

Damit folgt

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(Y) \leq x) = \mathbb{P}(Y \leq F(x)) = F(x).$$

□

Dieser Satz kann angewandt werden, um Zufallszahlen zu erzeugen, die die Verteilungsfunktion  $F$  besitzen. Allerdings muss in jedem Schritt eine nichtlineare Gleichung gelöst werden, was numerisch beispielsweise mit dem Newton-Verfahren durchgeführt werden kann.

### Algorithmus 5.8 (Inversionsverfahren)

**input:** streng monoton wachsenden Verteilungsfunktion  $F(x)$

**output:** Zufallszahl  $x$  gemäß der vorgegebenen Verteilung

① ziehe eine auf  $[0, 1]$  gleichverteilte Zufallszahl  $y$

② löse die nichtlineare Gleichung  $y = F(x)$  nach  $x$  auf

Natürlich kann man versuchen, die inverse Verteilungsfunktion direkt zu approximieren. Im Falle einer standardisiert normalverteilten Zufallsgröße führt dies auf den *Beasley-Springer-Moro-Algorithmus*: Im Intervall  $0.5 \leq y \leq 0.92$  gilt

$$\Phi^{-1}(y) \approx \frac{\sum_{k=0}^3 a_k (y - 0.5)^{2k+1}}{1 + \sum_{k=0}^3 b_k (y - 0.5)^{2k+2}}$$

und für  $y > 0.92$

$$\Phi^{-1}(y) \approx \sum_{k=0}^8 c_k \left( \log(-\log(1-y)) \right)^k,$$

wobei die Konstanten gemäß

$$\begin{aligned} a_0 &= 2.50662823884, & b_0 &= -8.47351093090, \\ a_1 &= -18.61500062529, & b_1 &= 23.08336743743, \\ a_2 &= 41.39119773534, & b_2 &= -21.06224101826, \\ a_3 &= -25.44106049637, & b_3 &= 3.13082909833, \end{aligned}$$

und

$$\begin{aligned} c_0 &= 0.3374754822726147, & c_5 &= 0.0003951896511919, \\ c_1 &= 0.9761690190917186, & c_6 &= 0.0000321767881768, \\ c_2 &= 0.1607979714918209, & c_7 &= 0.0000002888167364, \\ c_3 &= 0.0276438810333863, & c_8 &= 0.0000003960315187, \\ c_4 &= 0.0038405729373609, & & \end{aligned}$$

zu wählen sind. Im Intervall  $0 \leq y < 0.5$  kann man schließlich die Symmetrie der Standardnormalverteilung benutzen.

## 5.3 Monte-Carlo-Verfahren

Die stetige Zufallsgrösse  $X$  besitze die Dichtefunktion  $f : (a, b) \rightarrow \mathbb{R}$ , wobei hier  $a = -\infty$  oder  $b = \infty$  zugelassen sind. Für eine stetige Funktion  $h$  sei der Erwartungswert  $\mathbb{E}(h(X))$  zu berechnen. Gemäß Definition des Erwartungswerts entspricht dies der Auswertung des Integrals

$$S = \int_a^b h(x)f(x) dx.$$

Eine naheliegende Idee ist die Approximation dieses Integrals durch den Mittelwert einer Sequenz von unabhängigen Pseudozufallszahlen  $x_i$  zur durch die Dichtefunktion  $f$  gegebenen Verteilung:

$$s_n = \frac{1}{n} \sum_{i=1}^n h(x_i). \quad (5.3)$$

Wir betrachten demnach den *Monte-Carlo-Schätzer*

$$S_n = \frac{1}{n} \sum_{i=1}^n h(X_i).$$

Dieser erfüllt

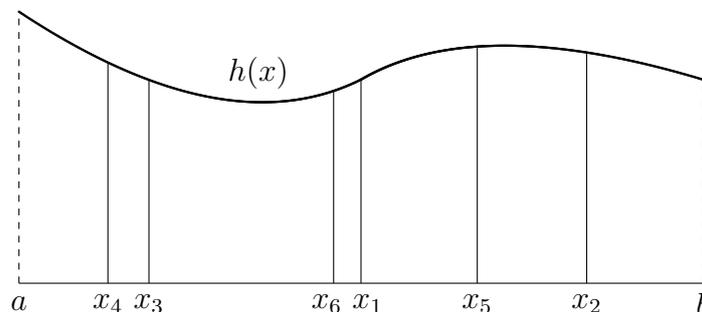
$$\mathbb{E}(S_n) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(h(X_i)) = S,$$

das heisst, im Mittel liefert  $S_n$  den gewünschten Wert  $S$ . Für die Varianz des Schätzers gilt

$$\mathbb{V}(S_n) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(h(X_i)) = \frac{1}{n} \int_a^b (h(x) - S)^2 f(x) dx. \quad (5.4)$$

Dies bedeutet, für  $n \rightarrow \infty$  wird die Schätzung von  $S = \mathbb{E}(h(X))$  immer genauer.

Die Monte-Carlo-Approximation (5.3) ist genaugenommen eine *Monte-Carlo-Quadratur* zur Berechnung des Integrals  $S$ . Dabei werden zufällig Stützstellen ausgewählt und das Integral durch den Mittelwert approximiert.



Natürlich kann der Fehler beliebig groß sein. Jedoch wird mit wachsendem  $n$  die mittlere Abweichung  $\mathbb{E}(|S - S_n|)$  klein.

**Satz 5.9** Sei

$$v_h := \int_a^b (h(x) - S)^2 f(x) dx.$$

Dann ist die mittlere Abweichung der Fehler beschränkt durch

$$\mathbb{E}(|S - S_n|) \leq \sqrt{\frac{v_h}{n}}.$$

*Beweis.* Die Behauptung ergibt sich sofort aus

$$\mathbb{E}^2(|S - S_n|) \leq \mathbb{E}((S - S_n)^2) = \mathbb{V}(S_n)$$

und (5.4). □

Da  $\sqrt{v_h}$  eine Konstante ist, die nur vom Integranden  $h$  abhängt, ist  $1/\sqrt{n}$  die Konvergenzrate. Mit anderen Worten, der Erwartungswert des Fehlers konvergiert wie  $\mathcal{O}(1/\sqrt{n})$ .

## 6. Beschreibende Statistik

Die beschreibende Statistik befasst sich mit der Auswertung von Daten. Ziel ist es, die Daten geordnet und verdichtet darzustellen, um sie überschaubar zu machen. Bei einer Datenerhebung werden ein oder mehrere *Merkmale* (Beobachtungsmerkmale) untersucht. Die beobachteten Werte heissen *Merkmalsausprägungen*. Bei Befragungen sind die Merkmalsausprägungen etwa die Antwortmöglichkeiten, die der Befragte angeben kann.

**Beispiel 6.1** Verschiedene Merkmale und zugehörige Ausprägungen sind in nachfolgender Tabelle zusammengestellt:

Grundgesamtheit	Merkmal	Ausprägung
Bevölkerung	Gewicht	$x$ kg mit $x \in \mathbb{R}_{\geq 0}$
	Geschlecht	m/w
	Gesundheitszustand	gut/schlecht
Kandidaten einer Parlamentswahl	Stimmenanzahl	$n \in \mathbb{N}_0$
Bolzen	Durchmesser	$d$ mm mit $d \in \mathbb{R}_{\geq 0}$

△

Die Menge der unbearbeiteten *Beobachtungswerte* heisst *Urliste*. Dieses Zahlenmaterial soll nun so geordnet und gegebenenfalls verdichtet werden, dass es überschaubar wird. Ein erster Schritt besteht darin, die Beobachtungswerte der Urliste  $\{x_1, x_2, \dots, x_n\}$  der Gösse zu ordnen. Dies führt auf die *Variationsreihe*  $\{x_1^*, x_2^*, \dots, x_n^*\}$  mit  $x_1^* \leq x_2^* \leq \dots \leq x_n^*$ , welche die *Variationsbreite*  $R = x_n^* - x_1^*$  besitzt.

**Beispiel 6.2** Wir beobachten für 20 produzierte Bolzen die Durchmesser:

183 181 183 180 182 182 185 182 184 179  
182 184 180 181 179 180 182 180 181 183

Die Variationsreihe zu dieser Urliste lautet:

$d_{\min} = 179$  179 180 180 180 180 181 181 181 182  
182 182 182 182 183 183 183 184 184 185 =  $d_{\max}$

Die Variationsbreite  $R = d_{\max} - d_{\min}$  der Beobachtungswerte ist 6.

△

**Häufigkeitstabelle:** Sehr häufig tritt wie im letzten Beispiel der Fall ein, dass in der Urliste Merkmalswerte mehrmals auftreten. Für das Ordnen des Materials ist es daher günstig, eine Häufigkeitstabelle zu erstellen.

**Beispiel 6.3** Für Beispiel 6.2 ergibt sich folgende Häufigkeitstabelle:

$m$	$d_m$	absolute Häufigkeit		relative Häufigkeit	relative Summenhäufigkeit
		Strichliste	$h_m$	$H_m$ in %	$\sum_m H_m$ in %
1	179		2	10	10
2	180		4	20	30
3	181		3	15	45
4	182		5	25	70
5	183		3	15	85
6	184		2	10	95
7	185		1	5	100
		$n = 20$		100	

Die Häufigkeitstabelle zeigt beispielsweise, dass der Wert 182 am häufigsten vorkommt.  $\triangle$

**Klasseneinteilung:** Enthält die Urliste eine grosse Anzahl unterschiedlicher Beobachtungswerte, so können wir das Material weiter verdichten, indem wir die vorliegenden Beobachtungswerte in *Klassen* einteilen. Diese werden durch ihre *obere* und *untere Klassengrenze*, durch ihre *Klassenbreite* und ihre *Klassenmitte* charakterisiert.

**Beispiel 6.4** Vorgelegt seien nun nicht wie in Beispiel 6.2 nur 20, sondern 100 Beobachtungswerte für den Durchmesser der Bolzen (in  $mm$ ):

179 187 183 185 165 170 176 181 191 177  
 193 185 192 181 173 165 182 188 177 175  
 177 169 175 178 180 179 186 181 198 183  
 198 173 185 177 186 186 158 167 166 184  
 195 177 188 185 169 184 172 195 180 196  
 176 186 179 160 170 186 179 169 174 182  
 170 190 174 198 169 182 165 173 174 184  
 189 183 176 188 188 181 186 184 170 188  
 173 174 182 177 179 186 190 178 183 178  
 182 184 174 183 186 181 197 174 173 163

Weil  $d_{\min} = 158$  und  $d_{\max} = 198$  gilt, folgt eine Variationsbreite der Daten von 40. Diese teilen wir in 9 Klassen der Klassenbreite 5 ein und erhalten die folgende Häufigkeitstabelle für die Klassen:

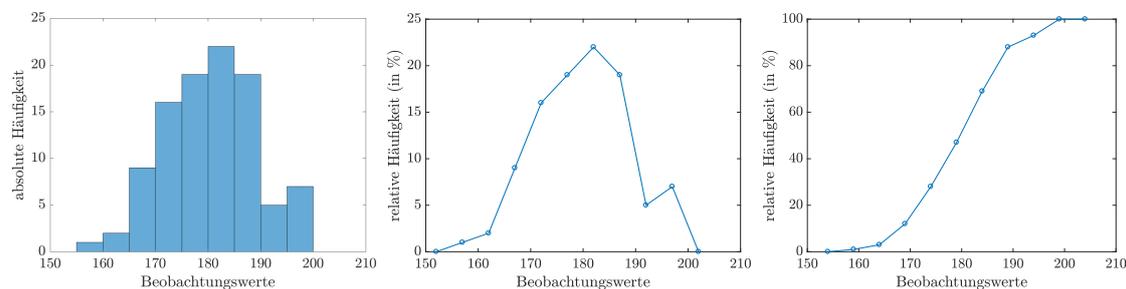
$m$	Klassengrenzen	Klassenmitte	absolute Häufigkeit	relative Häufigkeit	relative Summenhäufigkeit
			$h_m$	$H_m$ in %	$\sum_m H_m$ in %
1	155–159	157	1	1	1
2	160–164	162	2	2	3
3	165–169	167	9	9	12
4	170–174	172	16	16	28
5	175–180	177	19	19	47
6	180–184	182	22	22	69
7	185–189	187	19	19	88
8	190–194	192	5	5	93
9	195–199	197	7	7	100
			$n = 100$	100	

△

### Visualisierung von Daten:

- Von einem *Histogramm* sprechen wir dann, wenn über jeder Klasse ein Rechteck mit einer der absoluten oder relativen Häufigkeit der jeweiligen Klasse entsprechenden Fläche errichtet wird. Ist die Klassenbreite variabel, so ist die Höhe des Histogramms folglich nicht proportional zur Häufigkeit der Klasse.
- Alternativ zum Histogramm kann man auch die absolute oder relative Häufigkeit gegen die Klassenmittelpunkte plotten und diese Punkte durch einen Polygonzug verbinden. Auf diese Weise erhalten wir ein *Häufigkeitspolygon*, das offensichtlich die zugrundeliegende Dichtefunktion angenähert.
- Beim *Summenpolygon* wird die relative Summenhäufigkeit gegen die obere Klassengrenze eingezeichnet und mit einem Polygonzug verbunden. Dieser Polygonzug stellt dann eine Approximation der Verteilungsfunktion dar.

**Beispiel 6.5** Die verschiedenen Visualisierungstechniken führen im Fall der Daten aus Beispiel 6.4 auf die folgenden graphischen Darstellungen:



△

**Statistische Mittelwertmaße:** Durch den Mittelwert der Beobachtungswerte erhalten wir Aufschluss über die Lage des Zentrums dieser Beobachtungswerte. Für ihn sind folgende statistische Maßzahlen gebräuchlich

- Das *arithmetische Mittel* ist definiert durch

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i.$$

- Der *Median* teilt die Beobachtungswerte in zwei gleichgroße Mengen. In der einen Menge sind alle Beobachtungswerte kleiner oder gleich dem Median, während in der anderen Menge sind alle Beobachtungswerte größer oder gleich dem Median. Um ihn zu bestimmen, sortieren wir die Beobachtungswerte  $x_i$ ,  $i = 1, 2, \dots, n$ , der Größe nach, um die Variationsreihe  $x_1^* \leq x_2^* \leq \dots \leq x_n^*$  zu erhalten. Der Median berechnet sich dann gemäß

$$\tilde{x}_n = \begin{cases} \frac{1}{2}(x_{n/2}^* + x_{n/2+1}^*), & \text{falls } n \text{ gerade,} \\ x_{(n+1)/2}^*, & \text{falls } n \text{ ungerade.} \end{cases}$$

- Der *Modalwert*  $D$  ist der am häufigsten vorkommende Beobachtungswert. Der Modalwert ist im allgemeinen nicht eindeutig.

**Beispiel 6.6** Im Fall der Variationsreihe

179 179 180 180 180 180 181 181 181 182  
182 182 182 182 183 183 183 184 184 185

ist der Mittelwert  $\bar{x}_{20} = 181.65$ , der Median  $\tilde{x}_{20} = \frac{1}{2}(x_{10}^* + x_{11}^*) = 182$  und der Modalwert  $D = 182$ . △

**Geometrisches Mittel:** Ein weiteres Mittelwertmaß ist das *geometrische Mittel*. Bei der Untersuchung von Wachstumsprozessen ist das geometrische Mittel von grosser Bedeutung. Liegt eine Menge von positiven Beobachtungswerten  $x_i$ ,  $i = 1, 2, \dots, n$ , vor, dann berechnet sich das geometrische Mittel gemäss

$$G = \sqrt[n]{\prod_{i=1}^n x_i}.$$

Seine Berechnung erfolgt zweckmässig auf logarithmischen Wege:

$$\log G = \frac{1}{n} \sum_{i=1}^n \log x_i.$$

**Beispiel 6.7** Ein Guthaben  $G_0$  wird im ersten Jahr mit 2%, im zweiten Jahr mit 1.5% und im dritten Jahr mit 3% verzinst. Welcher über konstante Zinssatz  $p$  hätte nach drei Jahren das gleiche Kapital  $G_3$  ergeben?

Es gilt

$$G_3 = (1 + 0.02)(1 + 0.015)(1 + 0.03)G_0 \stackrel{!}{=} (1 + p)^3 G_0.$$

Wir haben also

$$1 + p = \sqrt[3]{(1 + 0.02)(1 + 0.015)(1 + 0.03)} = 1.0216,$$

dies bedeutet, der durchschnittliche Zinssatz beträgt 2.16%. △

**Streumaß:** Häufigkeitsverteilungen können sich trotz gleicher Mittelwertmaße erheblich unterscheiden, wenn deren Werte unterschiedlich um den Mittelwert streuen. Zur Charakterisierung dieses Sachverhalts dienen die sogenannten Streumaß.

- Die *empirische Varianz* ist definiert als

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Dass der Nenner wirklich  $n - 1$  und nicht  $n$  lauten muss, werden wir später einsehen. Bilden wir die (positive) Wurzel der empirischen Varianz, so erhalten wir die *empirische Standardabweichung*.

- Die *Variationsbreite* haben wir schon kennengelernt. Sie ist die Differenz aus dem größten Beobachtungswert und dem kleinsten Beobachtungswert:

$$R = x_{\max} - x_{\min} = x_n^* - x_1^*.$$

- Der *empirische Variationskoeffizient* ist ein normiertes Streumaß mit dem Vorteil, dass dieses Streumaß relativ zum arithmetischen Mittel gebildet wird:

$$v_n = \frac{s_n}{\bar{x}_n} \cdot 100 \%$$

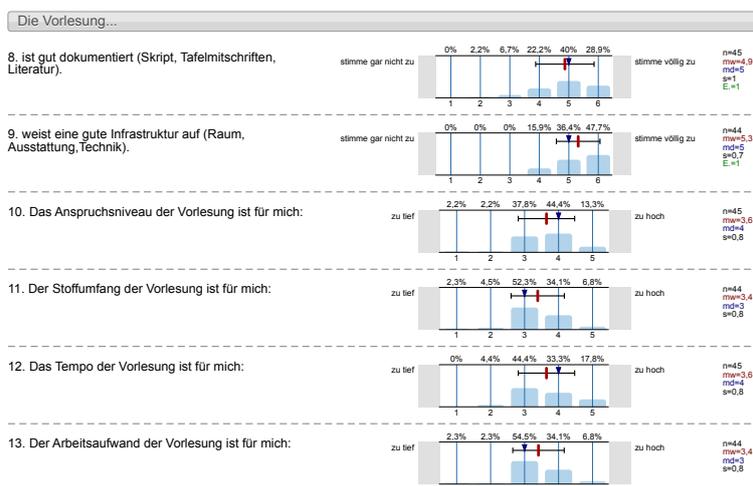
Man überlegt sich leicht, dass er der empirische Variationskoeffizient nur sinnvoll für Messreihen mit ausschließlich positiven oder ausschließlich negativen Beobachtungswerten ist.

**Beispiel 6.8** Im Fall der Variationsreihe

179 179 180 180 180 180 181 181 181 182  
182 182 182 182 183 183 183 184 184 185

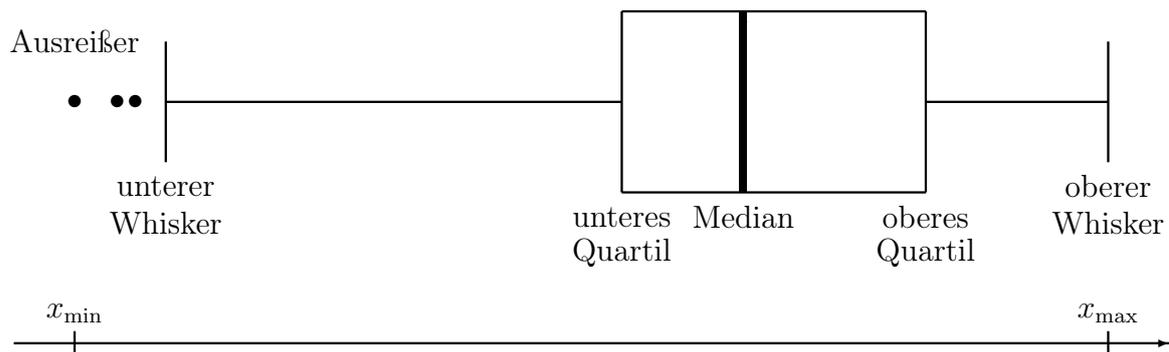
ist die empirische Varianz  $s_{20}^2 = 2.8711$ , die Variationsbreite  $R = 6$  und der empirische Variationskoeffizient  $v_n = 0.9328 \%$ .  $\triangle$

**Visualisierung statistischer Maßzahlen** Ein mögliche graphische Darstellung der statistischen Maßzahlen findet sich in der nachfolgenden Graphik. Im Histogramm der Beobachtungswerte sind neben dem arithmetischen Mittel (roter Strich / mw), der Median (blaues Dreieck / md) auch die empirische Standardabweichung (schwarzes Intervall / s) eingetragen:



Eine weitere populäre Darstellung sind *Boxplots*. Dabei entspricht die Box dem Bereich, in dem die mittleren 50 % der Daten liegen. Sie wird also durch das obere und das untere *Quartil* begrenzt. Links vom unteren Quartil liegen demnach 25 % der Daten. Rechts vom oberen Quartil liegen ebenfalls 25 % der Daten. Der Median wird als durchgehender Strich in der Box eingezeichnet. Durch die *Antennen*, üblicherweise *Whisker* genannt, werden die ausserhalb der Box liegenden Werte dargestellt. Die Länge der Antennen ist allerdings nicht einheitlich. Gebräuchlich sind folgende Definitionen:

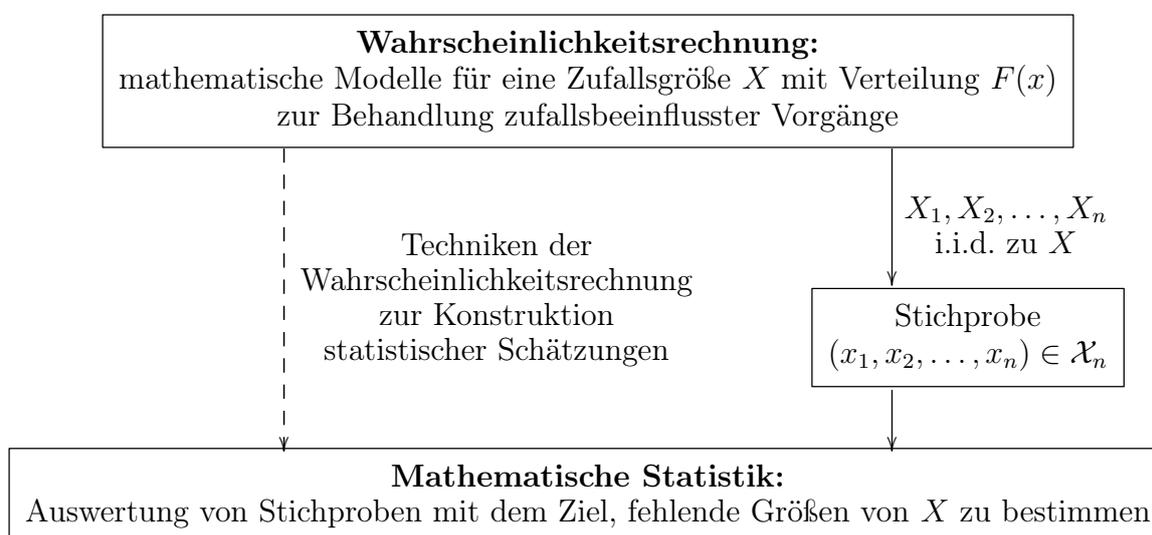
- Die Länge der Antenne ist auf das maximal 1.5-fache der Boxlänge beschränkt. Außerhalb der Antennen liegende Daten heißen *Ausreißer*. Beobachtungswerte, welche einen Abstand zur Box von mehr als dem dreifachen der Boxlänge haben, werden extrem genannt. Bei allen anderen handelt es sich um milde Ausreißer.
- Die untere Antenne ist das 2.5%-Quantil und die obere das 97.5%-Quantil. In dieser Darstellung gibt es also ab einem bestimmten Stichprobenumfang immer einzeln dargestellte Punkte, welche dann nicht automatisch als Ausreißer interpretiert werden sollten.
- Die Antennen reichen bis zum grössten bzw. kleinsten Wert aus den Daten. Hier gibt es keine Ausreisser.



## 7. Mathematische Statistik

### 7.1 Grundbegriffe

#### Abgrenzung Wahrscheinlichkeitsrechnung und Statistik:



**Grundgesamtheit:** Die Grundgesamtheit umfasst die Zufallsgröße  $X$  mit Verteilungsgesetz  $F_{\vartheta}(x) = \mathbb{P}(X \leq x)$ , wobei  $\vartheta \in \Theta$  für einen oder mehrere noch zu bestimmende Parameter steht.

**Stichprobe vom Umfang  $n$ :** Eine Stichprobe vom Umfang  $n$  besteht aus  $n$  Realisierungen der Zufallsgröße  $X$  ohne gegenseitige Beeinflussung der Ergebnisse. Dabei liegt

- abstrakt der Stichprobe ein  $n$ -dimensionaler Zufallsvektor  $(X_1, X_2, \dots, X_n)$  von  $n$  unabhängig und identisch verteilten Zufallsgrößen  $X_1, X_2, \dots, X_n$  zugrunde, während
- die konkrete Realisierung ein Vektor  $(x_1, x_2, \dots, x_n) \in \mathcal{X}_n$  ist, der die Realisierungen  $x_i$  von  $X_i$  enthält.

Wir sagen auch, die Zufallsgrößen  $X_1, X_2, \dots, X_n$  seien vom Typ i.i.d. (kurz für *independent and identically distributed*).

**Parameterraum  $\Theta$ :** Der Parameterraum  $\Theta$  enthält die Menge aller möglichen Parameterwerte  $\vartheta$  des Verteilungsgesetzes. Der Stichprobe liegt ein fester, aber bis dahin unbekannter Parameter  $\vartheta \in \Theta$  zugrunde. Dieser feste Wert  $\vartheta$  ist dann aus der Stichprobe zu *schätzen*. Beispielsweise könnte  $\vartheta = (\mu, \sigma^2) \in \Theta \in \mathbb{R} \times \mathbb{R}^+$  im Falle  $X \sim N(\mu, \sigma^2)$  sein.

**Stichprobenraum  $\mathcal{X}_n$ :** Der Stichprobenraum  $\mathcal{X}_n$  besteht aus der Menge aller möglichen Stichprobenwerte, zum Beispiel  $(x_1, x_2, \dots, x_n) \in \mathcal{X}_n = \mathbb{R}^n$ , falls  $X \sim N(\mu, \sigma^2)$  ist.

**Stichprobenfunktion:** Die Stichprobenfunktion  $T_n : \mathcal{X}_n \rightarrow \Theta$  ist eine Abbildung, die jedem Stichprobenwert  $(x_1, x_2, \dots, x_n) \in \mathcal{X}_n$  einen Parameterwert  $T_n(x_1, x_2, \dots, x_n) \in \Theta$  zuordnet. Sie wird zum Schätzen von Parametern oder Bereichen beziehungsweise zum Testen von Hypothesen verwendet. Für unsere theoretischen Untersuchungen müssen wir die der Stichprobenfunktion entsprechende Zufallsgröße  $T_n(X_1, X_2, \dots, X_n)$  betrachten.

**Beispiel 7.1** Eine Firma stellt Sicherungen her. Bei der Produktion wird von einer Ausschussrate  $p \in (0, 1)$  ausgegangen. Die Zufallsgröße  $X$  erfülle

$$X = \begin{cases} 1, & \text{betrachtete Sicherung ist defekt,} \\ 0, & \text{betrachtete Sicherung funktioniert,} \end{cases}$$

wobei  $\mathbb{P}(X = 1) = p$  und  $\mathbb{P}(X = 0) = 1 - p$ . Es gilt  $X \sim \text{Bin}(1, p)$  für den unbekannt Parameter  $\vartheta = p \in (0, 1) = \Theta$ . Die Stichprobe  $(X_1, X_2, \dots, X_n)$  vom Umfang  $n$  besitzt die Realisierung

$$(x_1, x_2, \dots, x_n) \in \mathcal{X}_n = \{(x_1, x_2, \dots, x_n) : x_i = 0 \vee x_i = 1\}.$$

Die Stichprobenfunktion  $\sum_{i=1}^n X_i \sim \text{Bin}(n, p)$  gibt die Zahl der defekten Sicherungen an. Die relative Häufigkeit  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  ist eine Stichprobenfunktion, die offensichtlich die Ausschusswahrscheinlichkeit  $p$  schätzt.  $\triangle$

**Punktschätzung:** Hier wird ein Schätzwert  $\hat{\vartheta}$  aus der Stichprobe mit “möglichst guten” Eigenschaften bestimmt.

**Bereichsschätzung:** Zu einer vorgegebenen Irrtumswahrscheinlichkeit  $\alpha$  (übliche Werte sind etwa 1%, 5% oder 10%) wird ein *Konfidenz-* oder *Vertrauensintervall*

$$[U(x_1, x_2, \dots, x_n), O(x_1, x_2, \dots, x_n)]$$

gesucht, so dass  $\vartheta$  mit der Wahrscheinlichkeit  $1 - \alpha$  in diesem Intervall liegt.

**Tests:** Ausgehend von einer *Hypothese*  $H_0$  über den zu schätzenden Parameter (z.B.  $H_0 : p \leq p_0$ ), wird mit Hilfe einer Testgröße getestet, ob die Hypothese  $H_0$  abgelehnt werden muss. Bei Ablehnung wird eine *Alternativhypothese*  $H_1$  (hier  $H_1 : p > p_0$ ) angenommen. Grundlage ist dabei eine Irrtumswahrscheinlichkeit  $\alpha$ .

## 7.2 Punktschätzungen

Auf Grundlage der Schätzfunktion  $T_n = T_n(X_1, X_2, \dots, X_n)$  soll ein möglichst guter Schätzwert  $\hat{\vartheta} = T_n(X_1, X_2, \dots, X_n)$  zum unbekannt Parameter  $\vartheta \in \Theta$  der Grundgesamtheit gefunden werden. Dabei ist die Schätzfunktion eine spezielle Stichprobenfunktion mit Werten im Parameterraum  $\Theta$ .

**Bemerkung 7.2** Häufig wird anstelle  $\vartheta$  selbst eine reelle Funktion  $\tau(\vartheta)$  geschätzt. Beispielsweise könnte die Varianz  $np(1-p)$  einer Binomialverteilung zu schätzen sein. Im Falle der Normalverteilung kann vom zweidimensionalen Parameter  $\vartheta = (\mu, \sigma^2)$  nur der Erwartungswert  $\tau(\vartheta) = \mu$  zu schätzen sein.  $\triangle$

**Definition 7.3** Eine Schätzfunktion  $T_n(X_1, X_2, \dots, X_n)$  für die Funktion  $\tau(\vartheta)$  heißt **erwartungstreu**, wenn für jeden Parameterwert  $\vartheta \in \Theta$  gilt

$$\mathbb{E}_\vartheta(T_n(X_1, X_2, \dots, X_n)) = \tau(\vartheta).$$

Die Forderung bedeutet, dass der Schätzer im Mittel richtig schätzt.

**Satz 7.4** Existieren in einer Grundgesamtheit  $X$  sowohl  $\mathbb{E}_\vartheta(X)$ , als auch  $\mathbb{V}_\vartheta(X)$ , so gilt:

1. Für  $\tau(\vartheta) = \mathbb{E}_\vartheta(X)$  ist

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{bzw. konkret} \quad \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

eine erwartungstreue Schätzfunktion.

2. Für  $\tau(\vartheta) = \mathbb{V}_\vartheta(X)$  ist

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad \text{bzw. konkret} \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

eine erwartungstreue Schätzfunktion.

3. Ist  $\mu = \mathbb{E}_\vartheta(X)$  bekannt, so ist

$$V_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad \text{bzw. konkret} \quad v_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

eine erwartungstreue Schätzfunktion für  $\tau(\vartheta) = \mathbb{V}_\vartheta(X)$ .

*Beweis.* 1. Es gilt

$$\mathbb{E}_\vartheta(\bar{X}_n) = \mathbb{E}_\vartheta\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\vartheta(X_i) = \frac{1}{n} \cdot n \cdot \mathbb{E}_\vartheta(X) = \mathbb{E}_\vartheta(X) = \tau(\vartheta).$$

2. Mit

$$X_i - \bar{X}_n = -\frac{1}{n}X_1 - \dots - \frac{1}{n}X_{i-1} + \underbrace{\left(1 - \frac{1}{n}\right)}_{=\frac{n-1}{n}}X_i - \frac{1}{n}X_{i+1} - \dots - \frac{1}{n}X_n$$

folgt

$$\begin{aligned} \mathbb{E}_\vartheta([X_i - \bar{X}_n]^2) - \underbrace{\mathbb{E}_\vartheta^2(X_i - \bar{X}_n)}_{=0} &= \mathbb{V}_\vartheta(X_i - \bar{X}_n) \\ &= \frac{n-1}{n^2} \mathbb{V}_\vartheta(X) + \frac{(n-1)^2}{n^2} \mathbb{V}_\vartheta(X) \\ &= \frac{n(n-1)}{n^2} \mathbb{V}_\vartheta(X). \end{aligned}$$

Damit ergibt sich

$$\begin{aligned} \mathbb{E}_\vartheta\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2\right) &= \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}_\vartheta([X_i - \bar{X}_n]^2) \\ &= \frac{1}{n-1} \cdot n \cdot \frac{n \cdot (n-1)}{n^2} \mathbb{V}_\vartheta(X) \\ &= \mathbb{V}_\vartheta(X) = \tau(\vartheta). \end{aligned}$$

3. Es gilt

$$\mathbb{E}_\vartheta\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2\right) = \frac{1}{n} \cdot n \cdot \left\{ \mathbb{E}_\vartheta([X - \mu]^2) - \underbrace{\mathbb{E}_\vartheta^2(X - \mu)}_{=0} \right\} = \mathbb{V}_\vartheta(X) = \tau(\vartheta).$$

□

**Definition 7.5** Eine Schätzfunktion  $T_n(X_1, X_2, \dots, X_n)$  für die Funktion  $\tau(\vartheta)$  des unbekanntes Parameters  $\vartheta$  der Grundgesamtheit heißt **konsistent**, wenn für alle  $\vartheta \in \Theta$  und beliebiges  $\varepsilon > 0$  gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}_\vartheta(|T_n(X_1, X_2, \dots, X_n) - \tau(\vartheta)| \geq \varepsilon) = 0.$$

**Satz 7.6** Die Schätzfunktionen  $\bar{X}_n$  für  $\mathbb{E}_\vartheta(X)$  und  $S_n^2$  beziehungsweise  $V_n^2$  bei bekanntem Erwartungswert für  $\mathbb{V}_\vartheta(X)$  sind konsistent, wobei im Fall der Streuung  $\mathbb{E}_\vartheta(X^4) < \infty$ , das heißt, die Existenz der vierten Momente, vorauszusetzen ist.

*Beweis.* Die Konsistenz des Mittelwertes  $\bar{X}_n$  folgt sofort aus dem schwachen Gesetz der großen Zahlen. Wegen  $\mathbb{E}_\vartheta(X_i^4) < \infty$  sind die Varianzen der Zufallsgrößen  $(X_i - \mu)^2$  beschränkt. Da

$$\mathbb{E}_\vartheta([X_i - \mu]^2) = \mathbb{V}_\vartheta(X_i) = \mathbb{V}_\vartheta(X),$$

folgt aus dem schwachen Gesetz der großen Zahlen:

$$\lim_{n \rightarrow \infty} \mathbb{P}_\vartheta(|V_n^2 - \mathbb{V}_\vartheta(X)| \geq \varepsilon) = \lim_{n \rightarrow \infty} \mathbb{P}_\vartheta\left(\left|\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \mathbb{V}_\vartheta(X)\right| \geq \varepsilon\right) = 0.$$

Für  $S_n^2$  zeigt man die Behauptung anhand der Zerlegung

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n [(X_i - \mu) - (\bar{X}_n - \mu)]^2.$$

□

**Bemerkung 7.7** Für  $X \sim N(\mu, \sigma^2)$  ist übrigens  $\mathbb{E}(X^4) < \infty$  immer erfüllt. △

**Definition 7.8** Es seien  $T_n^{(1)}(X_1, X_2, \dots, X_n)$  und  $T_n^{(2)}(X_1, X_2, \dots, X_n)$  erwartungstreue Schätzfunktionen für  $\tau(\vartheta)$ . Wenn gilt

$$\mathbb{V}(T_n^{(1)}(X_1, X_2, \dots, X_n) - \tau(\vartheta)) \leq \mathbb{V}(T_n^{(2)}(X_1, X_2, \dots, X_n) - \tau(\vartheta)),$$

so heißt die Schätzfunktion  $T_n^{(1)}(X_1, X_2, \dots, X_n)$  **wirksamer** als die Schätzfunktion  $T_n^{(2)}(X_1, X_2, \dots, X_n)$ . Diejenige Schätzfunktion, die unter allen erwartungstreuen Schätzfunktionen die kleinste Streuung besitzt, heißt die **wirksamste Schätzfunktion** für  $\tau(\vartheta)$ .

**Bemerkung 7.9** Für  $X \sim N(\mu, \sigma^2)$  ist  $\bar{X}_n$  die wirksamste Schätzfunktion für  $\mathbb{E}_\vartheta(X)$ . △

## 7.3 Maximum-Likelihood-Methode

Die Maximum-Likelihood-Methode ist ein wichtiges Verfahren, um Punktschätzungen mit möglichst vielen der zuletzt genannten “guten” Eigenschaften zu konstruieren. Dazu sei  $X$  eine Grundgesamtheit von bekanntem Verteilungstyp. Die Parameter  $\vartheta \in \Theta$  dieser Verteilung seien unbekannt und sollen unter Verwendung einer Stichprobe  $X_1, X_2, \dots, X_n$  i.i.d. zu  $X$  vom Umfang  $n$  geschätzt werden.

**Definition 7.10** Ist  $x_1, x_2, \dots, x_n$  eine aus der Grundgesamtheit  $X$  gezogene Stichprobe vom Umfang  $n$  und ist  $X$  eine diskrete Zufallsgröße mit den Einzelwahrscheinlichkeiten  $\mathbb{P}_\vartheta(X = x_i)$ ,  $i = 1, 2, \dots, n$ , wobei der Parameter  $\vartheta \in \Theta$  unbekannt ist, dann wird die Funktion

$$L(\vartheta) = \prod_{i=1}^n \mathbb{P}_\vartheta(X = x_i)$$

als **Likelihood-Funktion** bezeichnet. Ist  $X$  hingegen eine stetige Zufallsgröße mit Dichtefunktion  $f_\vartheta(t)$ , so lautet die Likelihood-Funktion

$$L(\vartheta) = \prod_{i=1}^n f_\vartheta(x_i).$$

Das Prinzip der Maximum-Likelihood-Methode, entwickelt von R.A. Fisher im Jahr 1921, besteht nun darin, als Punktschätzwert  $\hat{\vartheta}$  für den unbekannt Parameter  $\vartheta \in \Theta$  denjenigen zu ermitteln, für den die Likelihood-Funktion das Maximum annimmt.

**Beispiel 7.11** Der Parameter  $\vartheta = p \in (0, 1) = \Theta$  einer binomialverteilten Zufallsgröße  $X \sim \text{Bin}(1, p)$  sei zu bestimmen. Ist in einer Stichprobe vom Umfang  $n$   $\ell$ -mal  $X = 1$  und  $(n - \ell)$ -mal  $X = 0$  aufgetreten, so gilt

$$L(p) = p^\ell (1 - p)^{n - \ell}.$$

Zur Maximierung einer Likelihood-Funktion ist es oftmals günstiger, deren Logarithmus zu betrachten, das heißt, in unserem Fall die Aufgabe

$$\tilde{L}(p) = \log L(p) = \ell \cdot \log p + (n - \ell) \log(1 - p) \rightarrow \max_{p \in (0, 1)}.$$

Die notwendige Optimalitätsbedingung lautet

$$\tilde{L}'(p) = \ell \cdot \frac{1}{p} - (n - \ell) \cdot \frac{1}{1 - p} = \frac{\ell \cdot (1 - p) + (\ell - n) \cdot p}{p \cdot (1 - p)} \stackrel{!}{=} 0,$$

dies bedeutet  $\hat{p} = \ell/n$ . Die Maximum-Likelihood-Methode liefert in diesem Fall das bekannte  $\bar{x}_n$ , unsere erwartungstreue und konsistente Schätzung.  $\triangle$

**Beispiel 7.12** Die Parameter  $\vartheta = (\mu, \sigma^2)$  einer normalverteilten Grundgesamtheit  $X \sim N(\mu, \sigma^2)$  seien aus einer Stichprobe  $x_1, x_2, \dots, x_n$  mittels Maximum-Likelihood-Methode zu bestimmen. Fassen wir die folgenden  $\sigma^2$  als Symbol auf, so gilt

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

Um das Maximum  $(\hat{\mu}, \hat{\sigma}^2)$  zu bestimmen, gehen wir wieder von

$$\tilde{L}(\mu, \sigma^2) = \log L(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \rightarrow \max_{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+}$$

aus. Differentiation nach  $\mu$  liefert

$$\frac{\partial \tilde{L}}{\partial \mu}(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{\sum x_i}{\sigma^2} - \frac{n \cdot \mu}{\sigma^2} \stackrel{!}{=} 0,$$

das heißt, wir erhalten die bekannte Schätzung  $\hat{\mu} = \bar{x}_n$ . Ferner folgt aus Differentiation nach  $\sigma^2$

$$\frac{\partial \tilde{L}}{\partial \sigma^2}(\mu, \sigma^2) = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \stackrel{!}{=} 0.$$

Indem wir hier das bereits bekannte Optimum  $\hat{\mu} = \bar{x}_n$  einsetzen, ergibt sich

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{n-1}{n} s_n^2.$$

Demnach ist die Maximum-Likelihood-Schätzung für  $\sigma^2$  im Gegensatz zu  $s_n^2$  *nicht* erwartungstreu.  $\triangle$

**Eigenschaften von Maximum-Likelihood-Schätzungen:**

1. Alle Maximum-Likelihood-Schätzer sind konsistent.
2. Existiert eine wirksamste Schätzfunktion, so wird sie durch den Maximum-Likelihood-Schätzer angenommen.
3. Maximum-Likelihood-Schätzer sind asymptotisch normalverteilt mit Mittelwert  $\vartheta$ .

## 7.4 Verteilungen wichtiger Stichproben

### 7.4.1 Binomialverteilte Grundgesamtheit

Für die Grundgesamtheit  $X \sim \text{Bin}(1, p)$  und  $\vartheta = p \in (0, 1) = \Theta$  sei  $X_i = \{0, 1\}$  i.i.d. zu  $X$ , das heißt

$$\mathbb{P}(X_i = k) = \begin{cases} p, & k = 1, \\ 1 - p, & k = 0. \end{cases}$$

Dann erfüllt die Schätzfunktion  $T_n^{(0)} = \sum_{i=1}^n X_i$

$$T_n^{(0)} = \sum_{i=1}^n X_i \sim \text{Bin}(n, p).$$

Für großes  $n$  (Faustregel:  $np(1-p) > 9$ ) ist  $T_n^{(0)}$  näherungsweise normalverteilt:

$$T_n^{(0)} \approx N(np, np(1-p)).$$

Dies motiviert die Schätzfunktion

$$T_n^{(1)} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n \approx N\left(p, \frac{p \cdot (1-p)}{n}\right).$$

Standardisierung von  $T_n^{(1)}$  liefert

$$T_n^{(2)} = \frac{\bar{X}_n - p}{\sqrt{p \cdot (1-p)}} \sqrt{n} \approx N(0, 1).$$

### 7.4.2 Normalverteilte Grundgesamtheit

Für die normalverteilte Grundgesamtheit  $X \sim N(\mu, \sigma^2)$  gilt für den Mittelwert  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \sigma^2/n)$  und nach Standardisierung

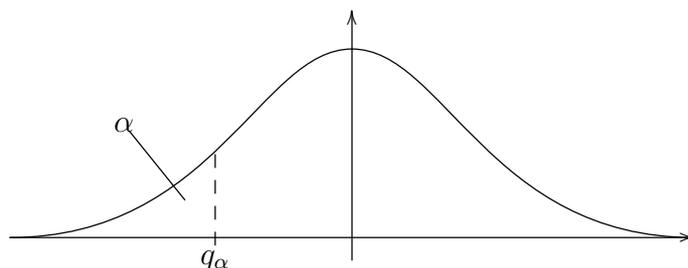
$$T_n^{(3)} = \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \sim N(0, 1).$$

**Definition 7.13** Es sei  $X$  eine stetige Zufallsgröße mit Dichtefunktion  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$ . Dann heißt die Zahl  $q_\alpha$  das  $\alpha$ -**Quantil** zu  $f$ , falls

$$\mathbb{P}(X \leq q_\alpha) = \int_{-\infty}^{q_\alpha} f(x) dx = \alpha.$$

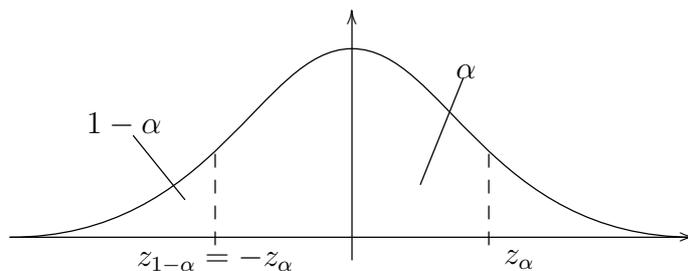
Analog ist  $q_{1-\alpha}$  das  $(1 - \alpha)$ -**Quantil** zu  $f$ , falls

$$\mathbb{P}(X \leq q_{1-\alpha}) = \int_{-\infty}^{q_{1-\alpha}} f(x) dx = 1 - \alpha.$$



**Beispiel 7.14** Das  $q_\alpha$ -Quantil der Standardnormalverteilung  $N(0, 1)$  bekam aufgrund seiner großen Bedeutung den Namen  $z_\alpha$ . Es gilt beispielsweise:

$\alpha$	90%	95%	99%
$z_\alpha$	1.282	1.645	2.326



△

**Definition 7.15** Die Zufallsgröße  $X$  besitzt eine  $\chi^2$ -**Verteilung mit  $n$  Freiheitsgraden**, kurz  $X \sim \chi_n^2$ , wenn  $X$  die Dichtefunktion

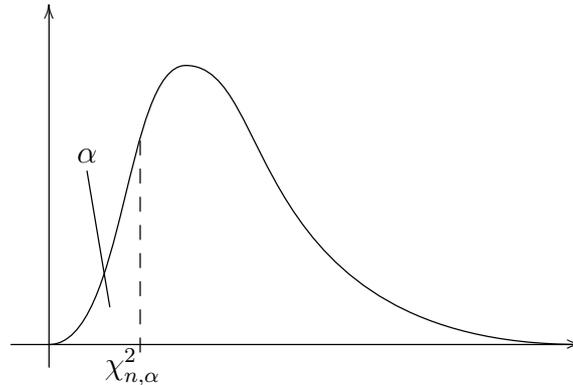
$$f_n(x) = \begin{cases} 0, & x \leq 0, \\ \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \end{cases}$$

besitzt. Dabei ist

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

die Gammafunktion.

**Bemerkung 7.16** Die Gammafunktion erfüllt  $\Gamma(n+1) = n!$  für alle  $n \in \mathbb{N}$ .  $\triangle$



**Satz 7.17** Für  $X \sim N(\mu, \sigma^2)$  und  $X_1, X_2, \dots, X_n$  i.i.d. zu  $X$  gelten die folgenden Aussagen:

1. Die Schätzfunktion  $T_n^{(4)}$  erfüllt

$$T_n^{(4)} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{n}{\sigma^2} V_n^2 \sim \chi_n^2,$$

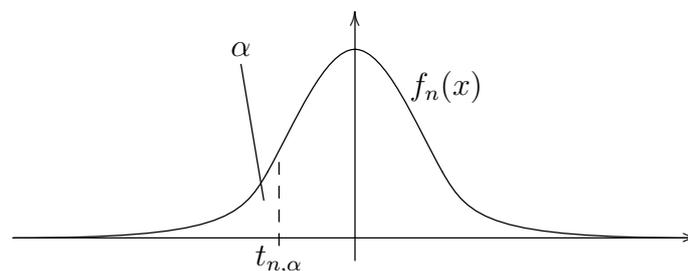
falls  $\mu$  bekannt ist.

2. Die Schätzfunktion  $T_n^{(5)}$  erfüllt

$$T_n^{(5)} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2.$$

**Definition 7.18** Die Zufallsgröße  $X$  besitzt eine **Student-Verteilung** oder **t-Verteilung** mit  $n$  Freiheitsgraden, kurz  $X \sim t_n$ , wenn  $X$  die folgende Dichtefunktion besitzt:

$$f_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} (1+x^2)^{-\frac{n+1}{2}}$$



**Satz 7.19** Für  $X \sim N(\mu, \sigma^2)$  und  $X_1, X_2, \dots, X_n$  i.i.d. zu  $X$  gilt

$$T_n^{(6)} = \frac{\bar{X}_n - \mu}{S_n} \sqrt{n} \sim t_{n-1}.$$

**Bemerkung 7.20** Bei der Berechnung von Konfidenzintervallen und Tests spielt

- die  $t$ -Verteilung eine wichtige Rolle, wenn die Größe des Mittelwertes  $\mu = \mathbb{E}(X)$  einer normalverteilten Zufallsgröße ermittelt werden soll und
- die  $\chi^2$ -Verteilung eine wichtige Rolle, wenn die Größe  $\sigma^2 = \mathbb{V}(X)$  einer normalverteilten Zufallsgröße bestimmt werden soll.

△

## 7.5 Konfidenzintervalle

**Ziel:** Wir suchen anhand einer Stichprobe  $(X_1, X_2, \dots, X_n)$  zur Grundgesamtheit  $X$  mit der Verteilung  $F_\vartheta(x)$  ein möglichst kleines Intervall

$$I = [U(X_1, X_2, \dots, X_n), (X_1, X_2, \dots, X_n)],$$

in welchem mit der Wahrscheinlichkeit  $1 - \alpha$  der unbekannt Parameter  $\vartheta \in \Theta$  tatsächlich liegt, das heißt

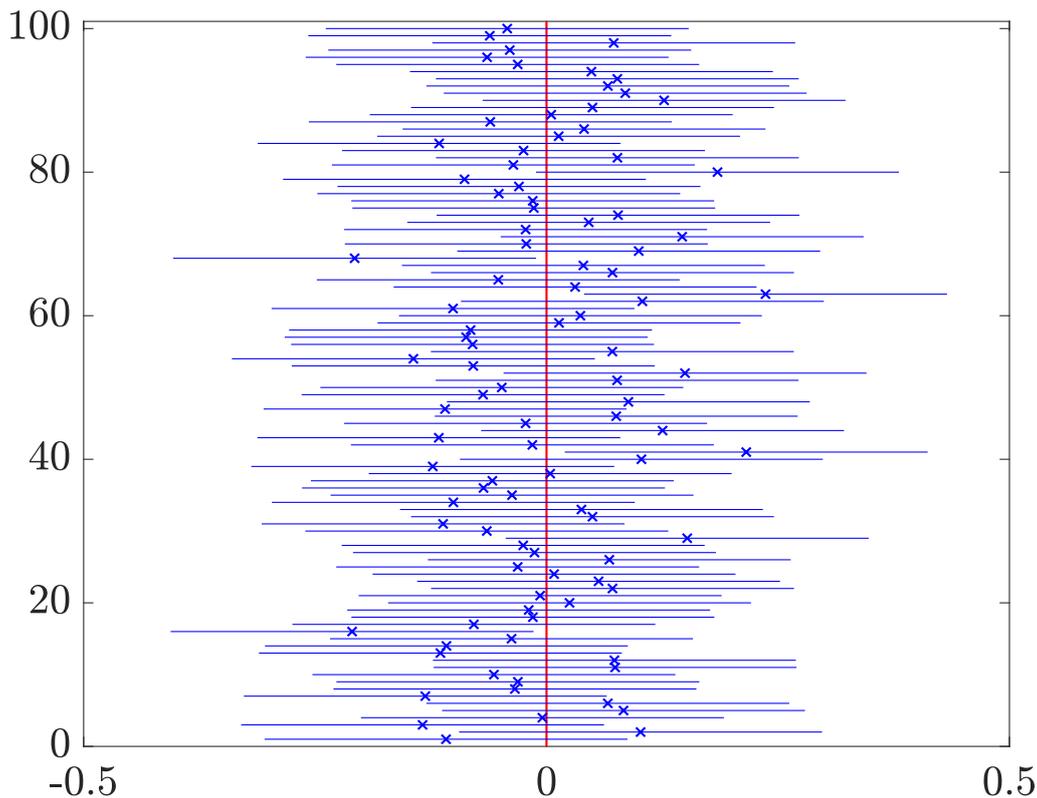
$$\mathbb{P}(U(X_1, X_2, \dots, X_n) \leq \vartheta \leq O(X_1, X_2, \dots, X_n)) \geq 1 - \alpha.$$

Dabei sind  $U : \mathcal{X}_n \rightarrow \Theta$  und  $O : \mathcal{X}_n \rightarrow \Theta$  zwei Schätzfunktionen. Die Größe  $1 - \alpha$  heißt *Konfidenzniveau* und  $\alpha$  die *Irrtumswahrscheinlichkeit*, welche stets vorzugeben ist.

**Beispiel 7.21** (Illustration von Konfidenzintervallen) Wir betrachten eine Zufallsgröße  $X$ , welche  $N(\mu, \sigma^2)$  verteilt ist. Die Gleichung

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{X - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) = \Phi(z_{1-\frac{\alpha}{2}}) - \Phi(-z_{1-\frac{\alpha}{2}}) = 2\Phi(z_{1-\frac{\alpha}{2}}) - 1 = 1 - \alpha$$

impliziert, dass jede Realisierung  $x$  von  $X$  mit der Wahrscheinlichkeit  $1 - \alpha$  im Intervall  $[\mu - z_{1-\frac{\alpha}{2}}\sigma, \mu + z_{1-\frac{\alpha}{2}}\sigma]$  liegt. Dies bedeutet, es gilt mit Wahrscheinlichkeit  $1 - \alpha$  für jede Realisierung  $x$  die Ungleichung  $|x - \mu| \leq z_{1-\frac{\alpha}{2}}\sigma$ . Mit der Wahrscheinlichkeit  $1 - \alpha$  demnach liegt der Erwartungswert  $\mu$  im *Konfidenzintervall*  $[x - z_{1-\frac{\alpha}{2}}\sigma, x + z_{1-\frac{\alpha}{2}}\sigma]$ . Im Fall  $X \sim N(0, 1/10)$  finden sich in nachfolgender Abbildung für  $n = 100$  Realisierungen  $x$  der Zufallsgröße  $X$  (blaue Kreuzchen) die zugehörigen Konfidenzintervalle (blaue Linien) zum Konfidenzniveau  $1 - \alpha = 95\%$ :



Im Mittel beinhalten offensichtlich nur  $\alpha \cdot n$  Intervalle der Länge  $2z_{1-\alpha/2}$  nicht den echten Erwartungswert  $\mu = 0$  (rote Linie). Im konkreten Beispiel sind es 4.  $\triangle$

### 7.5.1 Konfidenzintervalle bei binomialverteilter Grundgesamtheit

Sei  $X \sim \text{Bin}(1, p)$  und  $\vartheta = p \in (0, 1) = \Theta$ . Für die Stichprobe  $(X_1, \dots, X_n)$  gilt  $\sum_{i=1}^n X_i \sim \text{Bin}(n, p)$ , das heißt

$$\bar{X}_n \approx N\left(p, \frac{p(1-p)}{n}\right).$$

1. Mit dem  $(1 - \frac{\alpha}{2})$ -Quantil  $z_{1-\frac{\alpha}{2}}$  erhalten wir

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \underbrace{\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n}}_{=T_n^{(2)} \approx N(0,1)} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

2. Das Umrechnen von

$$-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n} \leq z_{1-\frac{\alpha}{2}}$$

in ein Konfidenzintervall ergibt eine quadratische Ungleichung für  $p$ . Auflösen ergibt

$$U(X_1, X_2, \dots, X_n) = \frac{n}{n + z_{1-\frac{\alpha}{2}}^2} \left[ \bar{X}_n + \frac{z_{1-\frac{\alpha}{2}}^2}{2n} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n} + \left(\frac{z_{1-\frac{\alpha}{2}}}{2n}\right)^2} \right]$$

und

$$O(X_1, X_2, \dots, X_n) = \frac{n}{n + z_{1-\frac{\alpha}{2}}^2} \left[ \bar{X}_n + \frac{z_{1-\frac{\alpha}{2}}^2}{2n} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n} + \left(\frac{z_{1-\frac{\alpha}{2}}}{2n}\right)^2} \right]$$

*Beweis.* Wir haben

$$|\bar{X}_n - p| \sqrt{n} \leq z_{1-\frac{\alpha}{2}} \sqrt{p(1-p)}$$

nach  $p$  aufzulösen. Quadrieren liefert

$$0 \geq p^2(z_{1-\frac{\alpha}{2}}^2 + n) - p(2n\bar{X}_n + z_{1-\frac{\alpha}{2}}^2) + \bar{X}_n^2 n$$

das heißt

$$\begin{aligned} O/U &= \frac{1}{2(z_{1-\frac{\alpha}{2}}^2 + n)} \left[ 2n\bar{X}_n + z_{1-\frac{\alpha}{2}}^2 \pm \sqrt{(2n\bar{X}_n + z_{1-\frac{\alpha}{2}}^2)^2 - 4(z_{1-\frac{\alpha}{2}}^2 + n)\bar{X}_n^2 n} \right] \\ &= \frac{1}{2(z_{1-\frac{\alpha}{2}}^2 + n)} \left[ 2n\bar{X}_n + z_{1-\frac{\alpha}{2}}^2 \pm \sqrt{z_{1-\frac{\alpha}{2}}^4 + 4n\bar{X}_n z_{1-\frac{\alpha}{2}}^2 - 4nz_{1-\frac{\alpha}{2}}^2 \bar{X}_n^2} \right] \\ &= \frac{1}{2(z_{1-\frac{\alpha}{2}}^2 + n)} \left[ 2n\bar{X}_n + z_{1-\frac{\alpha}{2}}^2 \pm 2nz_{1-\frac{\alpha}{2}} \sqrt{\frac{z_{1-\frac{\alpha}{2}}^2}{4n^2} + \frac{\bar{X}_n(1-\bar{X}_n)}{n}} \right]. \end{aligned}$$

□

**Beispiel 7.22** Anhand einer Stichprobe von  $n = 100$  Sicherungen soll die Ausschussrate  $p$  bestimmt werden. Es werden  $c = 2$  defekte Sicherungen gefunden, das heißt,  $\hat{p} = \bar{x}_n = \frac{2}{100} = 0.02$ . Gesucht ist ein Konfidenzintervall zum Konfidenzniveau  $1 - \alpha = 95\%$ . Mit  $z_{97.5\%} = 1.96$  ergibt sich

$$I = [0.0055, 0.0700].$$

Hat man hingegen  $n = 1000$  Sicherungen geprüft und  $c = 20$  defekte Sicherungen gefunden, es ändert sich also nur das  $n$ , dann folgt

$$I = [0.0130, 0.0304].$$

△

## 7.5.2 Konfidenzintervalle bei normalverteilter Grundgesamtheit

Wir betrachten für eine normalverteilte Grundgesamtheit  $X \sim N(\mu, \sigma^2)$  die Stichprobe  $(X_1, X_2, \dots, X_n)$  mit  $X_i$  i.i.d. zu  $X$ .

**Konfidenzintervalle für  $\mu$  bei bekanntem  $\sigma^2$ :** Für  $T_n^{(3)}$  aus Abschnitt 7.4.2 folgt

$$T_n^{(3)} = \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \sim N(0, 1).$$

Für vorgegebenes  $\alpha$  erhalten wir die Gleichung

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Umrechnung von

$$-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \leq z_{1-\frac{\alpha}{2}}$$

in ein Konfidenzintervall für  $\mu$  liefert

$$U(X_1, \dots, X_n) = \bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = O(X_1, \dots, X_n).$$

**Konfidenzintervalle für  $\mu$  bei unbekanntem  $\sigma^2$ :** Aus Satz 7.19 folgt

$$T_n^{(6)} = \frac{\bar{X}_n - \mu}{S_n} \sqrt{n} \sim t_{n-1}.$$

Unter Verwendung der  $t_{n-1, 1-\frac{\alpha}{2}}$ -Quantile folgt aus der Identität

$$\mathbb{P}\left(-t_{n-1, 1-\frac{\alpha}{2}} \leq \frac{\bar{X}_n - \mu}{S_n} \sqrt{n} \leq t_{n-1, 1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

durch Auflösen nach  $\mu$ , dass

$$U(X_1, X_2, \dots, X_n) = \bar{X}_n - t_{n-1, 1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{n-1, 1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} = O(X_1, X_2, \dots, X_n).$$

**Bemerkung 7.23** Natürlich kann man auch nur einseitige Konfidenzintervalle bestimmen. Dann muss für  $U$  bzw.  $O$  das entsprechende  $(1 - \alpha)$ -Quantil anstelle des  $(1 - \alpha/2)$ -Quantils eingesetzt werden, während für  $O$  bzw.  $U$   $+\infty$  bzw.  $-\infty$  zu nehmen ist.  $\triangle$

**Konfidenzintervalle für  $\sigma^2$  bei bekanntem  $\mu$ :** Für  $T_n^{(4)}$  gilt gemäß Satz 7.17

$$T_n^{(4)} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{n}{\sigma^2} V_n^2 \sim \chi_n^2.$$

Ausgehend von

$$\mathbb{P}\left(\chi_{n, \frac{\alpha}{2}}^2 \leq \frac{1}{\sigma^2} n V_n^2 \leq \chi_{n, 1-\frac{\alpha}{2}}^2\right) = 1 - \alpha$$

ergibt sich durch Umformen

$$U(X_1, \dots, X_n) = \frac{n V_n^2}{\chi_{n, 1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{n V_n^2}{\chi_{n, \frac{\alpha}{2}}^2} = O(X_1, \dots, X_n).$$

**Konfidenzintervalle für  $\sigma^2$  bei unbekanntem  $\mu$ :** Satz 7.17 besagt, dass

$$T_n^{(5)} = \frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2.$$

gilt. Die Forderung

$$\mathbb{P}(\chi_{n-1, \frac{\alpha}{2}}^2 \leq \frac{1}{\sigma^2} (n-1) S_n^2 \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2) = 1 - \alpha$$

impliziert durch Umformen, dass

$$U(X_1, \dots, X_n) = \frac{(n-1) S_n^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{(n-1) S_n^2}{\chi_{n-1, \frac{\alpha}{2}}^2} = O(X_1, \dots, X_n).$$

**Bemerkung 7.24** Da die Varianz von unten immer durch 0 beschränkt ist, sind gerade im Fall der Varianzschätzung einseitige Intervalle gebräuchlich. Als untere Schranke wird dann 0, als obere das passende  $O(X_1, \dots, X_n)$  gewählt, wobei die Quantile  $\chi_{n, \frac{\alpha}{2}}^2$  bzw.  $\chi_{n-1, \frac{\alpha}{2}}^2$  durch  $\chi_{n, \alpha}^2$  bzw.  $\chi_{n-1, \alpha}^2$  ersetzt werden müssen.  $\triangle$

**Zusammenstellung der wichtigen Konfidenzintervalle:**

Verteilung der Grundgesamtheit	zu schätzender Parameter	Intervallgrenzen $U(X_1, X_2, \dots, X_n)/O(X_1, X_2, \dots, X_n)$
Bin(1, $p$ )	$p$ ( $n$ groß)	$\frac{n}{n + z_{1-\frac{\alpha}{2}}^2} \left[ \bar{X}_n + \frac{z_{1-\frac{\alpha}{2}}^2}{2n} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n} + \left(\frac{z_{1-\frac{\alpha}{2}}}{2n}\right)^2} \right]$
$N(\mu, \sigma^2)$	$\mu$ ( $\sigma$ bekannt)	$\bar{X}_n \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
	$\mu$ ( $\sigma$ geschätzt)	$\bar{X}_n \pm t_{n-1, 1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}$
	$\sigma^2$ ( $\mu$ bekannt)	$\left[ \frac{nV_n^2}{\chi_{n, 1-\frac{\alpha}{2}}^2}, \frac{nV_n^2}{\chi_{n, \frac{\alpha}{2}}^2} \right]$
	$\sigma^2$ ( $\mu$ geschätzt)	$\left[ \frac{(n-1)S_n^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{(n-1)S_n^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right]$

## 7.6 Tests

### 7.6.1 Einführung

Ziel eines Tests ist es, anhand von Daten einer Stichprobe eine Annahme über die Größe eines Parameters  $\vartheta \in \Theta$  (Parametertest) oder über das Verhalten einer unbekanntem Verteilung (parameterfreier Test) zu überprüfen. Eine solche Annahme nennen wir *Hypothese*  $H_0$ .  $H_0$  wird auch als *Nullhypothese* bezeichnet. Diese Hypothese wird *abgelehnt*, wenn sie mit den Daten in signifikanter Weise unverträglich ist. Deshalb werden solche Tests *Signifikanztests* genannt. Falls die Daten mit  $H_0$  hingegen hinreichend verträglich sind, so bestehen gegen  $H_0$  keine Einwände. Es gibt also zwei Ausgänge von Tests:

1. gegen  $H_0$  bestehen keine Einwände, das heißt,  $H_0$  wird angenommen,
2.  $H_0$  ist unverträglich mit den Daten, das heißt,  $H_0$  wird abgelehnt.

Falls  $H_0$  abzulehnen ist, so nehmen wir eine *Alternativhypothese*  $H_1$  an, die das Komplement zu  $H_0$  darstellt.

**Beispiel 7.25 (Grundidee der Entscheidungsfindung)** Bei der Produktion von Sicherungen sei die Ausschussrate  $p$  unbekannt. Als Nullhypothese wählen wir  $H_0 : p \leq p_0$ . Für die Zufallsgröße  $X$ , definiert durch

$$X = \begin{cases} 1, & \text{Sicherung ist defekt,} \\ 0, & \text{Sicherung ist okay,} \end{cases}$$

gilt  $\mathbb{P}(X = 1) = p$  und  $\mathbb{P}(X = 0) = 1 - p$ , das heißt  $X \sim \text{Bin}(1, p)$ . Ferner seien  $X_1, X_2, \dots, X_n$  i.i.d. zu  $X$ .

Um eine Regel zur Entscheidung zu formulieren, ob  $H_0$  akzeptiert oder abgelehnt werden soll, betrachten wir die Testgröße  $T_n^{(0)} = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$ , die die Anzahl der defekten Sicherungen unter den  $n$  betrachteten zählt.

Konkret seien nun  $n = 100$ ,  $H_0 : p \leq 0.01$  und  $\sum_{i=1}^{100} x_i = c := 2$ . Die vorgegebene Irrtumswahrscheinlichkeit sei  $\alpha = 5\%$ . Falls  $\mathbb{P}_{H_0}(T_n^{(0)} = \sum_{i=1}^n X_i \geq c) \leq \alpha$  ist, so lehnen

wir  $H_0$  ab und nehmen die Alternativhypothese  $H_1 : p > 0.01$  an. Andernfalls nehmen wir  $H_0$  an.

Es treten bei diesem Vorgehen folgende Fehler auf, deren Größen gegenläufig sind:

- *Fehler 1. Art:*  $H_0$  ist gültig, wird aber abgelehnt. Die Größe dieses Fehlers ist  $\alpha$ , also vorgegeben.
- *Fehler 2. Art:*  $H_0$  ist falsch, wird aber trotzdem angenommen. Die Größe dieses Fehlers ist im allgemeinen nicht bekannt.

Im Beispiel ergibt sich bei Gültigkeit von  $H_0$

$$\begin{aligned} \mathbb{P}_{H_0}(T_n^{(0)} \geq c = 2) &= 1 - \mathbb{P}_{H_0}(T_n^{(0)} < 2) \\ &= 1 - \mathbb{P}_{H_0}(T_n^{(0)} = 0) - \mathbb{P}_{H_0}(T_n^{(0)} = 1) \\ &\leq 1 - \binom{100}{0} \cdot 0.01^0 \cdot 0.99^{100} - \binom{100}{1} \cdot 0.01^1 \cdot 0.99^{99} \\ &= 0.264238. \end{aligned}$$

Dies bedeutet, dass  $H_0$  bei  $c = 2$  defekten Sicherungen akzeptiert werden kann. Bei  $c = 4$  fällt die entsprechende Wahrscheinlichkeit unter 5% und man lehnt  $H_0$  ab.  $\triangle$

## 7.6.2 Allgemeines Testschema

Wir betrachten zu der Grundgesamtheit  $X$  mit der Verteilung  $F_{\vartheta}(X)$ , wobei  $\vartheta \in \Theta$  ein unbekannter Parameter ist, die Stichprobe  $(X_1, X_2, \dots, X_n)$  mit  $X_i$  i.i.d. zu  $X$ .

1. Aufstellen einer Hypothese  $H_0$  über den Parameter  $\vartheta \in \Theta$  und der entsprechenden Alternativhypothese, wobei  $\vartheta_0 \in \Theta$  gegeben ist:

$$\begin{array}{ll} H_0 : \vartheta = \vartheta_0 & H_1 : \vartheta \neq \vartheta_0 \\ H_0 : \vartheta \leq \vartheta_0 & H_1 : \vartheta > \vartheta_0 \\ H_0 : \vartheta \geq \vartheta_0 & H_1 : \vartheta < \vartheta_0 \end{array}$$

2. Vergabe der Irrtumswahrscheinlichkeit  $\alpha$ , das heißt, der Wahrscheinlichkeit dafür, dass  $H_0$  abgelehnt wird, obwohl  $H_0$  gilt, also die Größe des Fehlers 1. Art.
3. Finden einer Stichprobenfunktion (Testfunktion)  $T_n = T_n(X_1, X_2, \dots, X_n)$ , deren Verteilung bekannt ist, falls  $H_0$  gilt.
4. Bestimmung eines *kritischen Bereiches*  $K$ , so dass  $\mathbb{P}_{H_0}(T_n \in K) \leq \alpha$ . Bei Gültigkeit von  $H_0$  darf der Wert der Testfunktion höchstens mit der Irrtumswahrscheinlichkeit im kritischen Bereich liegen.
5. Entscheidung:
  - $T_n \in K \Rightarrow H_0$  wird abgelehnt, stattdessen wird  $H_1$  angenommen. Die Irrtumswahrscheinlichkeit ist  $\alpha$ .
  - $T_n \notin K \Rightarrow$  gegen  $H_0$  bestehen keine Einwände. Die Irrtumswahrscheinlichkeit ist unbekannt.

## 7.6.3 Tests bei normalverteilter Grundgesamtheit

**Hypothesen über den Mittelwert  $\mu$  bei bekanntem  $\sigma^2$ :** Seien  $\mu_0$  und  $\alpha$  vorgegeben und seien Null- und entsprechende Alternativhypothese durch einen der folgenden Fälle

erfasst:

$$H_0 : \begin{cases} \mu = \mu_0 \\ \mu \leq \mu_0 \\ \mu \geq \mu_0 \end{cases} \quad H_1 : \begin{cases} \mu \neq \mu_0 \\ \mu > \mu_0 \\ \mu < \mu_0 \end{cases}$$

Die Testfunktion ist bekannt:

$$T_n^{(3)}(X_1, X_2, \dots, X_n) = \frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n} \sim N(0, 1).$$

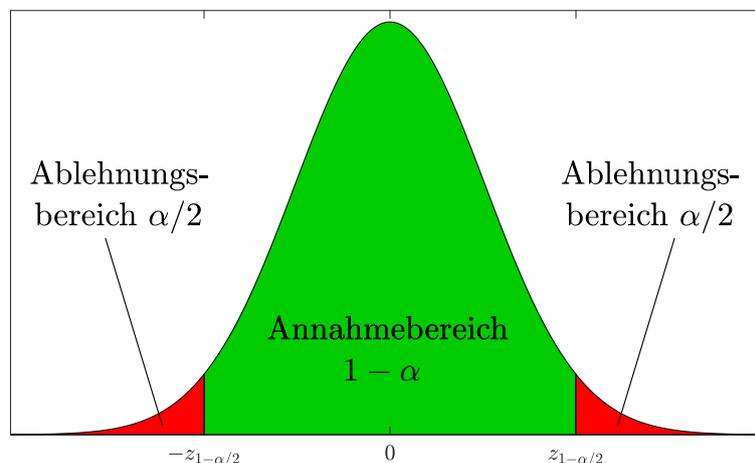
Als kritischer Bereich ergibt sich

$$K = \begin{cases} \{x : |x| > z_{1-\frac{\alpha}{2}}\} \\ \{x : x > z_{1-\alpha}\} \\ \{x : x < -z_{1-\alpha}\} \end{cases}$$

da gilt

$$\begin{aligned} \mathbb{P}_{H_0}(T_n \in K) &= \left\{ \begin{array}{l} 1 - \mathbb{P}\left(\left|\frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n}\right| \leq z_{1-\frac{\alpha}{2}}\right) \\ \mathbb{P}\left(\frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n} > z_{1-\alpha}\right) \\ \mathbb{P}\left(\frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n} < -z_{1-\alpha}\right) \end{array} \right\} = \left\{ \begin{array}{l} 1 - [\Phi(z_{1-\frac{\alpha}{2}}) - \Phi(-z_{1-\frac{\alpha}{2}})] \\ 1 - \Phi(z_{1-\alpha}) \\ \Phi(-z_{1-\alpha}) \end{array} \right\} \\ &= \left\{ \begin{array}{l} 2 - 2\Phi(z_{1-\frac{\alpha}{2}}) \\ 1 - \Phi(z_{1-\alpha}) \\ 1 - \Phi(z_{1-\alpha}) \end{array} \right\} = \alpha, \end{aligned}$$

vergleiche auch die nachfolgende Visualisierung für den ersten der drei Fälle:



**Hypothesen über den Mittelwert  $\mu$  bei unbekanntem  $\sigma^2$ :** Seien  $\mu_0$  und  $\alpha$  vorgegeben und Null- und entsprechend Alternativhypothese durch einen der drei oben genannten Fälle erfasst. Mit der Testfunktion

$$T_n^{(6)}(X_1, X_2, \dots, X_n) = \frac{\bar{X}_n - \mu_0}{S_n} \sqrt{n} \sim t_{n-1}$$

ergibt sich dann analog zu oben als kritischer Bereich

$$K = \begin{cases} \{x : |x| > t_{n-1, 1-\frac{\alpha}{2}}\} \\ \{x : x > t_{n-1, 1-\alpha}\} \\ \{x : x < -t_{n-1, 1-\alpha}\} \end{cases}$$

**Hypothesen über die Varianz  $\sigma^2$  bei bekanntem  $\mu$ :** Null- und Alternativhypothese seien wie folgt gegeben:

$$H_0 : \begin{cases} \sigma^2 = \sigma_0^2 \\ \sigma^2 \leq \sigma_0^2 \\ \sigma^2 \geq \sigma_0^2 \end{cases} \quad H_1 : \begin{cases} \sigma^2 \neq \sigma_0^2 \\ \sigma^2 < \sigma_0^2 \\ \sigma^2 > \sigma_0^2 \end{cases}$$

Mit Hilfe der Testfunktion

$$T_n^{(4)}(X_1, X_2, \dots, X_n) = \frac{n \cdot V^2}{\sigma_0^2} \sim \chi_n^2$$

ergibt sich vermittels einem zu oben analogen Vorgehen der kritische Bereich

$$K = \begin{cases} \{x : x < \chi_{n, \frac{\alpha}{2}}^2 \vee x > \chi_{n, 1-\frac{\alpha}{2}}^2\} \\ \{x : x > \chi_{n, 1-\alpha}^2\} \\ \{x : x < \chi_{n, \alpha}^2\} \end{cases}$$

**Hypothesen über die Varianz  $\sigma^2$  bei unbekanntem  $\mu$ :** Null- und Alternativhypothese seien wie folgt gegeben:

$$H_0 : \begin{cases} \sigma^2 = \sigma_0^2 \\ \sigma^2 \leq \sigma_0^2 \\ \sigma^2 \geq \sigma_0^2 \end{cases} \quad H_1 : \begin{cases} \sigma^2 \neq \sigma_0^2 \\ \sigma^2 > \sigma_0^2 \\ \sigma^2 < \sigma_0^2 \end{cases}$$

Mit der Testfunktion

$$T_n^{(5)}(X_1, X_2, \dots, X_n) = \frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi_{n-1}^2$$

ergibt sich der kritische Bereich

$$K = \begin{cases} \{x : x < \chi_{n-1, \frac{\alpha}{2}}^2 \vee x > \chi_{n-1, 1-\frac{\alpha}{2}}^2\} \\ \{x : x > \chi_{n-1, 1-\alpha}^2\} \\ \{x : x < \chi_{n-1, \alpha}^2\} \end{cases}$$

**Beispiel 7.26** Eine Referenzstrecke von  $1\text{km}$  Länge werde mit einem Messgerät gemessen. Das Messgerät liefert folgende Ergebnisse in  $m$ :

998.0, 1001.0, 1003.0, 1000.5, 999.0, 997.5, 1000.0, 999.5, 996.0, 998.5.

Es gilt also

$$\bar{x}_n = 999.3\text{ m}, \quad s_n^2 = 3.9\text{ m}^2, \quad s_n = 1.975\text{ m}.$$

1. Um zu testen, ob das Gerät im Mittel die korrekte Entfernung wiedergibt, wählen wir  $H_0 : \mu = \mu_0 = 1000\text{ m}$  mit  $H_1 : \mu \neq \mu_0$ . Für  $\alpha = 0.05$  ergibt sich nach dem allgemeinen Testschema:

$$T_n^{(4)} = \frac{\bar{x}_n - \mu_0}{s_n} \sqrt{n} = \frac{999.3 - 1000}{1.975} \sqrt{10} = -1.12.$$

Der kritische Bereich ist

$$K = \{x : |x| > t_{9,0.975} = 2.262\} = (-\infty, -2.262) \cup (2.262, +\infty).$$

Da  $T_n \notin K$  gibt es gegen  $H_0$  keine Einwände, dies bedeutet, das Messgerät liefert im Mittel die korrekte Entfernung. Beachte, dass die Irrtumswahrscheinlichkeit unbekannt ist (Fehler 2. Art).

2. Um zu überprüfen, ob die Behauptung  $\sigma^2 \leq 2m^2$  mit den Daten vereinbar ist, wählen wir  $H_0 : \sigma^2 \leq \sigma_0^2 = 2m^2$  und  $H_1 : \sigma^2 > \sigma_0^2$ . Das Vorgehen gemäß dem allgemeinen Testschema liefert für  $\alpha = 0.05$

$$T_n^{(6)} = (n-1) \frac{s_n^2}{\sigma_0^2} = 9 \cdot \frac{3.9 m^2}{2 m^2} = 17.55$$

und

$$K = \{x : x < \chi_{9,0.95}^2 = 16.919\} = (16.919, \infty).$$

Da  $T_n \in K$ , können wir die Behauptung, dass die Varianz des Messgeräts kleiner ist als  $2m^2$ , mit der Irrtumswahrscheinlichkeit  $\alpha$  ablehnen (dies bedeutet, das Messgerät ist schlechter).

△

### Zusammenstellung wichtiger Tests:

Verteilung der Grundgesamtheit	$H_0$	$H_1$	Testfunktion $T_n(X_1, X_2, \dots, X_n)$	kritischer Bereich $K$
$N(\mu, \sigma^2)$ $\sigma$ bekannt	$\mu = \mu_0$	$\mu \neq \mu_0$	$T_n = \frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n}$	$ T_n  > z_{1-\frac{\alpha}{2}}$
	$\mu \leq \mu_0$	$\mu > \mu_0$		$T_n > z_{1-\alpha}$
	$\mu \geq \mu_0$	$\mu < \mu_0$		$T_n < -z_{1-\alpha}$
$N(\mu, \sigma^2)$ $\sigma$ geschätzt	$\mu = \mu_0$	$\mu \neq \mu_0$	$T_n = \frac{\bar{X}_n - \mu_0}{S_n} \sqrt{n}$	$ T_n  > t_{n-1, 1-\frac{\alpha}{2}}$
	$\mu \leq \mu_0$	$\mu > \mu_0$		$T_n > t_{n-1, 1-\alpha}$
	$\mu \geq \mu_0$	$\mu < \mu_0$		$T_n < -t_{n-1, 1-\alpha}$
$N(\mu, \sigma^2)$ $\mu$ bekannt	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$T_n = n \frac{V_n^2}{\sigma_0^2}$	$T_n < \chi_{n, \frac{\alpha}{2}}^2 \vee T_n > \chi_{n, 1-\frac{\alpha}{2}}^2$
	$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$		$T_n > \chi_{n, 1-\alpha}^2$
	$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$		$T_n < \chi_{n, \alpha}^2$
$N(\mu, \sigma^2)$ $\mu$ geschätzt	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$T_n = (n-1) \frac{S_n^2}{\sigma_0^2}$	$T_n < \chi_{n-1, \frac{\alpha}{2}}^2 \vee T_n > \chi_{n-1, 1-\frac{\alpha}{2}}^2$
	$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$		$T_n > \chi_{n-1, 1-\alpha}^2$
	$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$		$T_n < \chi_{n-1, \alpha}^2$
Bin(1, $p$ ) $n$ groß	$p = p_0$	$p \neq p_0$	$T_n = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}$	$ T_n  > z_{1-\frac{\alpha}{2}}$
	$p \leq p_0$	$p > p_0$		$T_n > z_{1-\alpha}$
	$p \geq p_0$	$p < p_0$		$T_n < -z_{1-\alpha}$

### 7.6.4 $\chi^2$ -Test

Der  $\chi^2$ -Test ist ein parameterfreier Test. Sein Ziel ist es zu testen, ob eine Zufallsgröße  $X$  mit einer vorgegebenen Verteilung übereinstimmt. Dazu sei eine Stichprobe

$X_1, X_2, \dots, X_n$  i.i.d. zu  $X$  vom Umfang  $n$  gegeben. Weiter betrachten wir die Hypothese

$$H_0 : F_X(t) = F_0(t), \quad H_1 : F_X(t) \neq F_0(t)$$

für eine vorgegebene Verteilungsfunktion  $F_0(t)$ . Unabhängig davon, ob  $X$  eine diskrete oder stetige Zufallsgröße ist, teilen wir die reelle Achse in zueinander disjunkte Intervalle auf

$$\mathbb{R} = I_1 \cup I_2 \cup \dots \cup I_r = (-\infty, a_1] \cup (a_1, a_2] \cup \dots \cup (a_{r-2}, a_{r-1}] \cup (a_{r-1}, \infty).$$

Bezeichnet  $y_i$  jeweils die Anzahl der Stichprobenelemente im Intervall  $I_i$ , so gilt  $\sum_{i=1}^r y_i = n$ . Seien ferner  $y_i^0$  die "Idealzahlen" der Stichprobenwerte im Intervall  $I_i$ , die für die gegebene Verteilung gelten, also

$$y_i^0 = n \cdot \mathbb{P}_{H_0}(X \in I_i) = n \cdot (F_0(a_i) - \lim_{\varepsilon \searrow 0} F_0(a_{i-1} + \varepsilon))$$

mit  $a_0 := -\infty$  und  $a_r := +\infty$ . Um hinreichende Genauigkeit des Tests sicherzustellen, gelte  $y_i^0 \geq 5$  für alle  $i = 1, 2, \dots, r$ . Dies kann stets durch ein Vergrößern der Stichprobe oder ein Vergrößern des zugrundeliegenden Intervalls erzielt werden.

Wir betrachten die folgende Testfunktion

$$T_n(X_1, X_2, \dots, X_n) = \sum_{i=1}^r \frac{(y_i - y_i^0)^2}{y_i^0} \sim \chi_{r-m-1}^2,$$

wobei  $m$  die Anzahl der unbekannt Parameter in der Verteilungsfunktion  $F_0(t)$  angibt. Wir lehnen die Hypothese  $H_0$  ab, falls  $T_n$  zu groß wird, das heißt, bei gegebener Irrtumswahrscheinlichkeit  $\alpha > 0$  gilt

$$K = \{x : x > \chi_{r-m-1, 1-\alpha}^2\}.$$

**Beispiel 7.27 (Radioaktiver Zerfall)** Rutherford und Geiger veröffentlichten 1910 das folgende Versuchsergebnis. In 2608 Zeitintervallen von je 7.5 Sekunden Länge wurden die emittierten  $\alpha$ -Teilchen einer radioaktiven Substanz gezählt:

Teilchen pro 7.5 Sek.	Abs. Häufigkeit	Theor. Häufigkeit
0	57	54.8
1	203	211.2
2	383	406.8
3	525	524.2
4	532	508.6
5	408	393.8
6	273	253.0
7	139	140.8
8	45	67.8
9	27	28.7
$\geq 10$	16	18.3

Mit der Irrtumswahrscheinlichkeit  $\alpha = 5\%$  soll geprüft werden, ob die entsprechende Grundgesamtheit  $X$  ( $X$  ist die Anzahl der innerhalb von 7.5 Sekunden emittierten  $\alpha$ -Teilchen) durch die Poisson-Verteilung beschrieben werden kann. Die Nullhypothese  $H_0$  lautet also

$$H_0 : \mathbb{P}(X = m) = \frac{\lambda_0^m}{m!} e^{-\lambda_0}, \quad m = 0, 1, 2, \dots$$

Wir schätzen den Mittelwert  $\lambda_0$  durch die entsprechende Punktschätzung

$$\bar{x}_n = 3.87.$$

Damit ergibt sich für die theoretischen Häufigkeiten

$$y_m^0 = n \cdot \mathbb{P}(X = m) = 2608 \cdot \frac{3.87^m}{m!} e^{-3.87}, \quad m = 0, 1, 2, \dots$$

Es folgt  $T_n = 13.05$  und

$$K = \{x : x > \chi_{11-1-1,0.95}^2 = 16.9\} = (16.9, \infty).$$

Folglich wird die Nullhypothese wegen  $T_n \notin K$  nicht abgelehnt.  $\triangle$

**Beispiel 7.28 (Normalverteilung)** Ein Nennmaß  $X$  wird anhand einer Stichprobe überprüft. Um die Nullhypothese

$$H_0 : F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

zu zeigen, wählen wir eine Stichprobe vom Umfang  $n = 150$ , welche

$$\bar{x}_n = 40.48, \quad s_n = 5.71,$$

und

$i$	$I_i$	$y_i$	$y_i^0$
1	0 – 30.5	5	6.03
2	30.5 – 33.5	13	10.59
3	33.5 – 36.5	23	19.81
4	36.5 – 39.5	22	28.35
5	39.5 – 42.5	29	30.94
6	42.5 – 45.5	29	25.81
7	45.5 – 48.5	16	16.44
8	$\geq 48.5$	13	12.01
		$\sum = 150$	

ergibt. Es folgt

$$T_n = \sum_{i=1}^8 \frac{(y_i - y_i^0)^2}{y_i^0} = 3.27$$

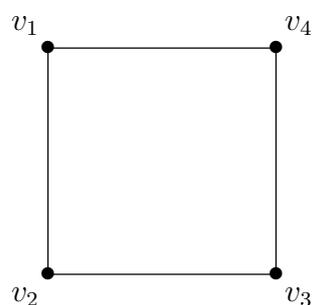
und mit  $\alpha = 0.10$

$$K = \{x : x > \chi_{8-2-1,0.9}^2 = 9.24\} = (9.24, \infty),$$

also  $T_n \notin K$ . Damit erscheint die Annahme der Normalverteilung im Sinne der Stichprobe vernünftig.  $\triangle$

## 8. Markov-Ketten

**Beispiel 8.1** (Simple Random Walk) Wir betrachten einen “Random Walker” in einer sehr kleinen Stadt, die nur vier Straßen und vier Straßenecken besitzt:



Zum Zeitpunkt 0 steht der Random Walker an der Ecke  $v_1$ . Zum Zeitpunkt 1 wirft er eine Münze und geht nach  $v_2$  bzw.  $v_4$ , je nachdem ob er Zahl oder Kopf geworfen hat. Zum Zeitpunkt 2 wirft er wieder eine Münze, um zu entscheiden zu welcher nächstgelegenen Straßenecke er nun geht. Diese Prozedur wird für die Zeitschritte 3, 4, ... iteriert.

Für jedes  $n \in \mathbb{N}_0$  sei

$$X_n \hat{=} \text{“Straßenecke zum Zeitpunkt } n\text{”}.$$

Dann gilt

$$\mathbb{P}(X_0 = v_1) = 1$$

und

$$\mathbb{P}(X_1 = v_2) = \frac{1}{2}, \quad \mathbb{P}(X_1 = v_4) = \frac{1}{2}.$$

Um die Wahrscheinlichkeitsverteilung für  $X_n$  mit  $n \geq 2$  zu berechnen, ist es vorteilhaft, bedingte Wahrscheinlichkeiten zu betrachten. Angenommen, der Random Walker steht zum Zeitpunkt  $n$  an der Ecke  $v_2$ . Dann gilt

$$\mathbb{P}(X_{n+1} = v_1 | X_n = v_2) = \frac{1}{2}, \quad \mathbb{P}(X_{n+1} = v_3 | X_n = v_2) = \frac{1}{2}.$$

Diese Wahrscheinlichkeiten ändern sich nicht, wenn wir die gesamte Vergangenheit in Betracht ziehen:

$$\mathbb{P}(X_{n+1} = v_1 | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = v_2) = \frac{1}{2}$$

und

$$\mathbb{P}(X_{n+1} = v_3 | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = v_2) = \frac{1}{2}$$

für jede Wahl möglicher  $i_0, i_1, \dots, i_{n-1}$ . Insbesondere ist die Wahrscheinlichkeit  $P(X_{n+1} = v_i | X_n = v_j)$  unabhängig vom Zeitpunkt  $n$ . △

**Definition 8.2** Es sei  $S$  eine endliche oder abzählbar unendliche Menge. Dann bildet die Folge  $X_0, X_1, \dots$  von Zufallsgrößen eine **Markov-Kette**, falls die **Markov-Eigenschaft** gilt: Für alle  $n \in \mathbb{N}_0$  und für  $s_0, s_1, \dots, s_{n+1} \in S$  mit

$$\mathbb{P}(X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) > 0$$

gilt

$$\mathbb{P}(X_{n+1} = s_{n+1} | X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) = \mathbb{P}(X_{n+1} = s_{n+1} | X_n = s_n).$$

Die Menge  $S$  wird **Zustandsraum** genannt.

**Definition 8.3** Es sei  $\mathbf{P} = [p_{i,j}]_{i,j=1}^k$  eine  $(k \times k)$ -Matrix. Eine Markov-Kette  $(X_0, X_1, \dots)$  mit endlichem Zustandsraum  $S = \{s_1, s_2, \dots, s_k\}$  heißt **homogene Markov-Kette** mit **Übergangsmatrix**  $\mathbf{P}$ , falls für alle  $n \in \mathbb{N}_0$  und  $i, j \in \{1, 2, \dots, k\}$  gilt

$$\mathbb{P}(X_{n+1} = s_i | X_n = s_j) = p_{i,j}.$$

Die Einträge von  $\mathbf{P}$  heißen **Übergangswahrscheinlichkeiten**.

**Bemerkung 8.4** Die Übergangsmatrix ist eine *stochastische Matrix*, das heißt, es gilt

$$p_{i,j} \geq 0 \quad \text{für alle } i, j \in \{1, 2, \dots, k\}$$

und

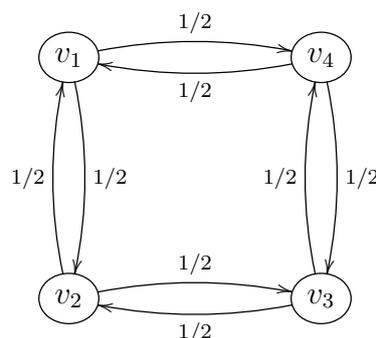
$$\sum_{i=1}^k p_{i,j} = 1 \quad \text{für alle } j \in \{1, 2, \dots, k\}.$$

△

Für unser Einführungsbeispiel ergibt sich beispielsweise die Übergangsmatrix

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}.$$

Neben der Übergangsmatrix ist der *Übergangsgraph* eine weitere sehr hilfreiche Darstellung von Markov-Ketten. Im Falle des Simple Random Walks erhalten wir:



Die Wahrscheinlichkeitsverteilung von  $X_n$  wollen wir nun als Spaltenvektor auffassen, also

$$\boldsymbol{\mu}^{(n)} = \begin{bmatrix} \mu_1^{(n)} \\ \mu_2^{(n)} \\ \vdots \\ \mu_k^{(n)} \end{bmatrix} = \begin{bmatrix} \mathbb{P}(X_n = s_1) \\ \mathbb{P}(X_n = s_2) \\ \vdots \\ \mathbb{P}(X_n = s_k) \end{bmatrix}.$$

Die Wahrscheinlichkeitsverteilung  $\boldsymbol{\mu}^{(0)}$  von  $X_0$  heißt auch *Startverteilung*.

**Satz 8.5** Für die homogene Markov-Kette  $(X_0, X_1, \dots)$  mit endlichem Zustandsraum  $S = \{s_1, s_2, \dots, s_k\}$ , Startverteilung  $\boldsymbol{\mu}^{(0)}$  und Übergangsmatrix  $\mathbf{P}$  gilt für jedes  $n \in \mathbb{N}$

$$\boldsymbol{\mu}^{(n)} = \mathbf{P}^n \boldsymbol{\mu}^{(0)}.$$

*Beweis.* Im Fall  $n = 1$  gilt für alle  $i \in \{1, 2, \dots, k\}$

$$\begin{aligned} \mu_i^{(1)} &= \mathbb{P}(X_1 = s_i) = \sum_{j=1}^k \mathbb{P}(X_0 = s_j, X_1 = s_i) \\ &= \sum_{j=1}^k \underbrace{\mathbb{P}(X_0 = s_j)}_{=\mu_j^{(0)}} \cdot \underbrace{\mathbb{P}(X_1 = s_i \mid X_0 = s_j)}_{=p_{i,j}} \\ &= \sum_{j=1}^k p_{i,j} \mu_j^{(0)} = [\mathbf{P} \boldsymbol{\mu}^{(0)}]_i. \end{aligned}$$

Mit vollständiger Induktion folgt der Schritt  $n \mapsto n + 1$ :

$$\begin{aligned} \mu_i^{(n+1)} &= \mathbb{P}(X_{n+1} = s_i) = \sum_{j=1}^k \mathbb{P}(X_n = s_j, X_{n+1} = s_i) \\ &= \sum_{j=1}^k \underbrace{\mathbb{P}(X_n = s_j)}_{=\mu_j^{(n)}} \cdot \underbrace{\mathbb{P}(X_{n+1} = s_i \mid X_n = s_j)}_{=p_{i,j}} \\ &= \sum_{j=1}^k p_{i,j} \mu_j^{(n)} = [\mathbf{P} \boldsymbol{\mu}^{(n)}]_i \stackrel{\text{Induktions-}}{\text{annahme}} [\mathbf{P}^{n+1} \boldsymbol{\mu}^{(0)}]_i. \end{aligned}$$

□

**Beispiel 8.6 (Basler Wettermodell)** Es wird gelegentlich behauptet, dass die beste Wettervorhersage diejenige ist, bei der das Wetter morgen dem heutigen entspricht. Ist diese Behauptung richtig, so ist es naheliegend, das Wetter als Markov-Kette zu modellieren. Wir nehmen einfachheitshalber an, dass es nur die Zustände  $s_1 = \text{“Regen”}$  und  $s_2 = \text{“Sonne”}$  gibt. Vorausgesetzt unsere Wettervorhersage ist mit der Wahrscheinlichkeit  $p \in [0, 1]$  richtig, so folgt die Übergangsmatrix

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

△

**Beispiel 8.7 (Londoner Wettermodell)** Wir haben im letzten Beispiel eine perfekte Symmetrie zwischen “Regen” und “Sonne”. Dies mag in Basel richtig sein, sicherlich jedoch nicht in London. Dort dürfte die Übergangsmatrix

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

mit  $q < p$  realistischer sein.

△

**Definition 8.8** Es sei  $\mathbf{P}^{(1)} = [p_{i,j}^{(1)}]_{i,j=1}^k$ ,  $\mathbf{P}^{(2)} = [p_{i,j}^{(2)}]_{i,j=1}^k, \dots$  eine Folge von stochastischen  $(k \times k)$ -Matrizen. Eine Markov-Kette  $(X_0, X_1, \dots)$  mit endlichem Zustandsraum  $S = \{s_1, s_2, \dots, s_k\}$  heißt **inhomogene Markov-Kette** mit Übergangsmatrizen  $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots$ , falls für alle  $n \in \mathbb{N}$  und  $i, j \in \{1, 2, \dots, k\}$  gilt

$$\mathbb{P}(X_{n+1} = s_i \mid X_n = s_j) = p_{i,j}^{(n+1)}.$$

**Beispiel 8.9 (verfeinertes Basler Wettermodell)** Wir verfeinern unser Basler Wettermodell, indem wir zusätzlich zwischen “Winter” und “Sommer” unterscheiden. Hierzu erweitern wir den Zustandsraum um den Wert  $s_3 = \text{“Schnee”}$  und unterscheiden zwischen Sommer (Mai–September) und Winter (Oktober–April). Realistisch wäre beispielsweise die inhomogene Markov-Kette mit

$$\mathbf{P}_{\text{Sommer}} = \begin{bmatrix} 0.75 & 0.25 & 0.5 \\ 0.25 & 0.75 & 0.5 \\ 0 & 0 & 0 \end{bmatrix}$$

und

$$\mathbf{P}_{\text{Winter}} = \begin{bmatrix} 0.5 & 0.15 & 0.2 \\ 0.3 & 0.7 & 0.3 \\ 0.2 & 0.15 & 0.5 \end{bmatrix}.$$

Im Sommer entspricht das verfeinerte Modell dem alten Modell im Falle  $p = 0.75$  außer eventuellem Tauwetter am 1. Mai.

△

**Satz 8.10** Für die inhomogene Markov-Kette  $(X_0, X_1, \dots)$  mit endlichem Zustandsraum  $S = \{s_1, s_2, \dots, s_k\}$ , Startverteilung  $\boldsymbol{\mu}^{(0)}$  und Übergangsmatrizen  $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots$  gilt für jedes  $n \in \mathbb{N}$

$$\boldsymbol{\mu}^{(n)} = \mathbf{P}^{(n)} \mathbf{P}^{(n-1)} \cdot \dots \cdot \mathbf{P}^{(1)} \boldsymbol{\mu}^{(0)}.$$

*Beweis.* Die Aussage wird ähnlich zum Beweis von Satz 8.5 gezeigt.

□

# Index

- $\chi^2$ -Test, 90
- $\chi^2$ -Verteilung, 80
- $\sigma$ -Algebra, 10
- Übergangs
  - graph, 94
  - matrix, 94
  - wahrscheinlichkeit, 94
- absolute Häufigkeit, 10
- Algorithmus
  - Beasley-Springer-Moro-, 64
  - Inversionsverfahren, 64
  - Verwerfungsmethode, 62
- Alternativhypothese, 74
- Antenne, 72
- Ausreißer, 72
- Baumdiagramm, 20
- Bayessche Formel, 21
- Beasley-Springer-Moro-Algorithmus, 64
- bedingte Wahrscheinlichkeit, 17
- Beobachtungswerte, 67
- Bereichsschätzung, 74
- Binomialverteilung, 38
- Boxplot, 72
- Dichtefunktion, 44
  - der Exponentialverteilung, 49
  - der Gleichverteilung, 47
  - der Normalverteilung, 52
- Elementarereignis, 8
- Ereignis
  - Elementar-, 8
  - Komplementär-, 8
  - sicheres, 8
  - unmögliches, 8
  - unvereinbares, 9
  - zufälliges, 7
- Ereignisalgebra, 10
- erwartungstreu, 75
- Erwartungswert, 30
  - Funktion einer Zufallsgröße, 32, 46
- Exponentialverteilung, 49
- Fehler
  - 1. Art, 87
  - 2. Art, 87
- Gammafunktion, 80
- Gaußsche
  - Glockenkurve, 52
  - Verteilung, 52
- Geburtstagsparadoxon, 15
- geometrische Verteilung, 28
- Gleichverteilung, 44
- Graph
  - Übergangs-, 94
- Grundgesamtheit, 73
- Häufigkeit
  - absolute, 10
  - relative, 10
- Häufigkeitspolygon, 69
- Histogramm, 69
- homogene Markov-Kette, 94
- Hypergeometrische Verteilung, 42
- Hypothese, 74, 86
- inhomogene Markov-Kette, 96
- Intensität der Poisson-Verteilung, 40, 41
- Inversionsverfahren, 64
- Irrtumswahrscheinlichkeit, 82
- Klasse, 68
- Klassen
  - breite, 68
  - einteilung, 68
  - grenze
    - obere –, 68
    - untere –, 68
  - mitte, 68

- Kombination, 13
- Komplementärereignis, 8
- Konfidenzintervall, 74, 82
- Konfidenzniveau, 82
- konsistent, 76
- kritischer Bereich, 87
- Laplace-Modell, 12
- Likelihood-Funktion, 77
- linearer Kongruenzgenerator, 58
- Lotto (Urnenmodell), 16
- Markov-Eigenschaft, 94
- Markov-Kette, 94
  - homogene, 94
  - inhomogene -, 96
- Matrix
  - Übergangs-, 94
- Median, 70
- Merkmal, 67
- Merkmalsausprägung, 67
- Mittel
  - arithmetisches -, 69
  - geometrisches -, 70
- Mittelwert, 30
- Mittelwertmaß, 69
- Modalwert, 70
- Monte-Carlo-
  - Quadratur, 65
  - Schätzer, 65
  - Verfahren, 65
- Multiplikationsregel, 19
  - erweiterte, 20
- Normalverteilung, 52
  - Standard-, 53
- Nullhypothese, 86
- Parallelschaltung, 23
- Parameterraum, 73
- Pfadregel, 21
- Poisson-Verteilung, 39, 40
- Poissonsche Annahmen, 39
- Produkt
  - maß, 26
  - von Wahrscheinlichkeitsräumen, 25
- Pseudozufallszahlen, 58
- Punktschätzung, 74
  - Monte-Carlo-, 65
- Quantil, 80
- Quartil, 72
- Random Walk, 93
- relative Häufigkeit, 10
  - Eigenschaften, 11
  - schwache Konvergenz, 37
- Satz
  - von Bayes, 21
  - von der totalen Wahrscheinlichkeit, 21
- Schätzer
  - Monte-Carlo-, 65
- Schätzfunktion
  - wirksamste -, 77
- schwaches Gesetz der großen Zahlen, 36
- Serienschaltung, 23
- sicheres Ereignis, 8
- Signifikanztest, 86
- Standardabweichung, 32
  - empirische -, 71
- Standardnormalverteilung, 53
- Startverteilung, 95
- Stichprobe, 73
- Stichprobenfunktion, 74
- Stichprobenraum, 74
- stochastische Matrix, 94
- Streumaß, 71
- Streuung, 32, 47
- Student-Verteilung, 81
- Summenpolygon, 69
- t-Verteilung, 81
- Test (statistischer-), 74
- totale Wahrscheinlichkeit, 21
- Tschebyscheffsche Ungleichung, 34, 48
  - unabhängig, 23, 24, 35
  - Ungleichung
    - von Tschebyscheff, 34, 48
  - unmögliches Ereignis, 8
  - unvereinbares Ereignis, 9
- Urliste, 67
- Varianz, 32, 47
  - empirische -, 71
- Variation, 13
- Variations
  - breite, 67

- reihe, 67
- Variationsbreite, 71
- Variationskoeffizient
  - empirischer –, 71
- Versuch, 6
- Verteilung
  - Binomial-, 38
  - Exponential-, 49
  - Gaußsche, 52
  - geometrische, 28
  - hypergeometrische, 42
  - Normal-, 52
  - Poisson-, 39, 40
  - Standardnormal-, 53
- Verteilungsfunktion, 29, 48
  - der Exponentialverteilung, 50
  - der Gleichverteilung, 49
  - der Normalverteilung, 53
  - Eigenschaften, 29, 48
- Vertrauensintervall, 74
- Verwerfungsmethode, 62
  
- Wahrscheinlichkeit, 11
  - bedingte, 17
  - totale, 21
- Wahrscheinlichkeits
  - funktion, 28
  - raum, 6, 13
- Whisker, 72
- wirksam, 77
  
- Zerlegung
  - von  $\Omega$ , 12
- Zufalls
  - situation, 6
  - variable, 27
- Zufallsgröße
  - Definition, 27
  - diskrete, 27
  - stetige, 44
- Zustandsraum, 94