

Studentisches Skript zur Vorlesung

# Mathematik für Ingenieure C4

gehalten von Dr. Wigand Rathmann  
im Sommersemester 2020

---

Verfasst von Frederik Hennig

mit Unterstützung von Franziska Bumiller

**Datum:** 5. Juli 2023  
**Kontakt:** [frederik.hennig@fau.de](mailto:frederik.hennig@fau.de)  
**Git:** [gitlab.cs.fau.de/da15siwa/mathingc4script](https://gitlab.cs.fau.de/da15siwa/mathingc4script)

## Disclaimer

Dieses Dokument ist eine studentische Vorlesungsmitschrift und stellt somit *kein* offizielles Skript der Lehrveranstaltung „Mathematik für Ingenieure C4“ oder irgendeiner anderen Lehrveranstaltung dar. Es erhebt keinen Anspruch auf Richtigkeit oder Vollständigkeit.

## Vorwort

Dieses Skript entstand während des Sommersemesters 2020 als Mitschrift zur Vorlesung „Mathematik für Ingenieure C4“, gehalten von Dr. Wigand Rathmann, in erster Linie aus meinem Wunsch nach einer strukturierten und mathematisch rigorosen Zusammenfassung des Stoffes. Ich habe mich bemüht, möglichst viele Aussagen zu begründen und zu beweisen; nicht nur der Vollständigkeit, sondern vor allem des Verständnisses wegen.

Leider enthält diese Mitschrift noch einige Lücken und bestimmt noch einige Fehler. Ich hoffe dennoch, dass sie zukünftigen Studenten der Mathematik C4 zumindest ein wenig weiterzuhelfen vermag. Über Feedback, Korrekturen und Ergänzungen würde ich mich natürlich freuen. Wer möchte, kann sich gerne per Git an diesem Werk beteiligen.

## Referenzen

Die folgende Literatur wurde zur Erstellung dieses Skripts herangezogen:

- Rathmann; Script zur Vorlesung „Mathematik für Ingenieure C4“, Sommersemester 2020
- Meintrup, Schäffler; Stochastik. Theorie und Anwendungen.
- Toutenburg, Schomaker, Wißmann, Heumann; Arbeitsbuch zur deskriptiven und induktiven Statistik.
- Die deutsch- und englischsprachige Wikipedia

# Inhaltsverzeichnis

# 1 Einleitung

## 2 Grundlagen der beschreibenden Statistik

In der Statistik werden Beobachtungen aus Messungen **Datensatz** oder Urliste genannt, ist diese nicht vollständig, wird sie **Stichprobe** genannt.

Diese erhobenen Daten kann man in 3 Unterkategorien einteilen:

- nominal (Daten ohne Ordnung) z.B. männlich / weiblich,
- ordinal (Daten mit Ordnung) z.B. Gewicht,
- metrisch/rational (Ausprägungen, welche mittels Zahlen beschrieben werden kann) z.B. Bewertungsscalen in Fragebögen.

Für uns sind im Folgenden vor allem ordinale Datensätze von Bedeutung.

**Statistik** ist die Komprimierung/Darstellung eines Datensatzes.

# 3 Wahrscheinlichkeitsmodelle

## 3.1 Grundlegende Definitionen

### Definition 3.1 (Ergebnisse und Ereignisse)

- Die **Ergebnismenge**  $\Omega$  ist eine nichtleere Menge von Ergebnissen  $\omega$ . Sie beschreibt die möglichen Ausgänge eines Zufallsexperiments.
- Eine Teilmenge  $A \subseteq \Omega$  heißt **Ereignis**. Man sagt, „ein Ereignis  $A$  tritt ein“, falls ein Ergebnis  $\omega \in A$  beobachtet wird. Die Menge aller betrachteten Ereignisse heißt **Ereignissystem**  $\mathcal{A}$ .

Manche Ereignisse verdienen besondere Namen:

- $A = \emptyset$  ist das *unmögliche Ereignis*. Jede Durchführung eines Zufallsexperiments liefert ein Ergebnis aus  $\Omega$ . Die leere Menge als Ereignis würde jedoch alle Durchführungen beschreiben, in denen *kein* Ergebnis auftritt.
- $A = \Omega$  ist dementsprechend das *sichere Ereignis*. Es beinhaltet alle möglichen Ergebnisse und tritt daher immer ein.
- $A = \{\omega\}$ ,  $\omega \in \Omega$  ist ein *Elementarereignis*.

Die Zusammenhänge von Ergebnissen und Ereignissen lassen sich leicht an einem Beispiel illustrieren:

**Beispiel 3.2** Ein sechsseitiger Würfel wird geworfen. Die möglichen Ausgänge dieses Zufallsexperiments sind die Augenzahlen 1 bis 6. Eine sinnvolle Ergebnismenge ist also

$$\Omega_{\text{Würfel}} = \{1, 2, 3, 4, 5, 6\}.$$

Auf dieser Menge lassen sich diverse Ereignisse beschreiben, beispielsweise:

1. „Die Augenzahl ist durch drei teilbar“:  $A_1 = \{3, 6\}$
2. „Die Augenzahl ist gerade“:  $A_2 = \{2, 4, 6\}$

Ereignisse sind Mengen. Sie lassen sich also mit den üblichen mengentheoretischen Definitionen und Schreibweisen beschreiben. Einige von diesen möchten wir in diesem Kontext erneut einführen:

### Definition 3.3

Sei  $\Omega$  eine Ergebnismenge.

- Die Menge aller Teilmengen von  $\Omega$  heißt **Potenzmenge**  $\mathcal{P}(\Omega)$  und ist definiert als

$$\mathcal{P}(\Omega) := \{A \mid A \subseteq \Omega\}.$$

- Das **Komplement** eines Ereignisses  $A \subseteq \Omega$  ist definiert als

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

und wird **Gegenereignis** genannt. Es beinhaltet genau die Ergebnisse, die in  $A$  *nicht* enthalten sind. Anders ausgedrückt:  $A^c$  tritt genau dann ein, wenn  $A$  nicht eintritt.

- Zwei Ereignisse  $A_1, A_2$  heißen **disjunkt**, wenn ihr Schnitt leer ist:

$$A_1 \cap A_2 = \emptyset.$$

Eine Familie von Ereignissen  $A_1, \dots, A_n$ ,  $n \in \mathbb{N} \cup \{\infty\}$  heißt **paarweise disjunkt**, wenn für je zwei verschiedene Ereignisse  $A_i, A_j$ ,  $i \neq j$  gilt:  $A_i$  und  $A_j$  sind disjunkt.

- Für eine paarweise disjunkte Familie von Ereignissen  $A_1, \dots, A_n$  ist die **disjunkte Vereinigung** definiert als

$$A_i + A_j := A_i \cup A_j \quad (i \neq j)$$

beziehungsweise

$$\sum_{i=1}^n A_i := \bigcup_{i=1}^n A_i.$$

### 3.2 Ereignissysteme

Um Aussagen über ein Zufallsexperiment machen zu können, möchten wir nicht nur Ereignisse einzeln betrachten, sondern sie zu einem *Ereignissystem* zusammenfassen.

#### Definition 3.4 (Abgeschlossenes Mengensystem)

Sei  $\Omega$  eine beliebige Menge.

- Ein System  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  von Teilmengen von  $\Omega$  heißt **abgeschlossenes Mengensystem** oder  **$\sigma$ -Algebra** über  $\Omega$ , wenn gilt:
  1.  $\Omega \in \mathcal{A}$  (Die Grundmenge ist im System enthalten)
  2.  $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$  (Zu jeder Menge aus  $\mathcal{A}$  ist auch ihr Komplement in  $\mathcal{A}$  enthalten)
  3. Für jede Auswahl von abzählbar vielen Mengen  $A_1, \dots, A_n \in \mathcal{A}$ ,  $n \in \mathbb{N} \cup \{\infty\}$  ist auch ihre Vereinigung in  $\mathcal{A}$  enthalten:

$$A_1, \dots, A_n \in \mathcal{A} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{A}$$

- Ist  $\Omega$  eine Ergebnismenge, so heißt ein abgeschlossenes Mengensystem  $\mathcal{A}$  von Ereignissen  $A_i \subseteq \Omega$  **Ereignissystem** über  $\Omega$ .

**Beispiel 3.5** Gegeben sei eine Ergebnismenge  $\Omega$  mit  $\Omega = \{0, 1\}^2$  und  $A = \{(0, 1), (1, 0)\}$ . Ein abgeschlossenes Mengensystem  $\mathcal{A}$  mit A wäre

$$\mathcal{A} = \{\Omega, \emptyset, A, A^c\}.$$

Nimmt man nun noch  $B = \{(1, 1)\}$  hinzu, ändert sich entsprechend auch das abgeschlossene Mengensystem  $\mathcal{A}$ . Zuerst muss nun wieder das Komplement von B ( $B^c = \{(0, 0), (0, 1), (1, 0)\}$ ) Nun kann noch entsprechend die Vereinigung von A und B gebildet werden ( $C = \{(0, 1), (1, 0), (1, 1)\}$ ). Auch von C muss nun entsprechend das Komplement ergänzt werden. Eben dieses Komplement kann nun ebenfalls wieder vereinigt werden usw. Das Ereignissystem ist dann

$$\mathcal{A} = \{\Omega, \emptyset, A, A^c, B, B^c, C, C^c\}.$$

*TBD: Borel-Algebra*

### 3.3 Wahrscheinlichkeitsmaße und Zufallsvariablen

#### Definition 3.6 (Messraum)

Sei  $\Omega$  eine beliebige Menge und  $\mathcal{A}$  eine  $\sigma$ -Algebra auf  $\Omega$ . Dann heißt das Tupel  $(\Omega, \mathcal{A})$  **Messraum** oder **messbarer Raum**. Ist  $\Omega$  eine stochastische Ergebnismenge, dann heißt der Messraum auch **Ereignisraum**. Eine Menge  $A \in \mathcal{A}$  heißt **messbare Menge**.

#### Definition 3.7 (Messbare Funktion)

Sei eine Funktion  $f : \Omega_1 \rightarrow \Omega_2$  zwischen zwei Messräumen  $(\Omega_1, \mathcal{A}_1)$  und  $(\Omega_2, \mathcal{A}_2)$  gegeben.

- Für eine Menge  $A_2 \in \mathcal{A}_2$  heißt

$$f^{-1}(A_2) = \{\omega \in \Omega_1 | f(\omega) \in A_2\}$$

das **Urbild** von  $A_2$  in  $f$ .

- $f$  heißt **messbar**, wenn das Urbild jeder Menge  $A_2 \in \mathcal{A}_2$  unter  $f$  ein Element aus  $\mathcal{A}_1$  ist, also:

$$\forall A_2 \in \mathcal{A}_2 : f^{-1}(A_2) \in \mathcal{A}_1.$$

**Definition 3.8 (Maß, Maßraum)**

Sei  $(\Omega, \mathcal{A})$  ein Messraum. Eine Abbildung  $\mu : \mathcal{A} \rightarrow [0, \infty) \cup \{\infty\}$  heißt **Maß** auf  $\mathcal{A}$ , wenn sie die folgenden Eigenschaften hat:

1.  $\mu(\emptyset) = 0$
2.  $\sigma$ -Additivität: Für jede Familie  $A_1, \dots, A_n$ ,  $n \in \mathbb{N} \cup \{\infty\}$  paarweise disjunkter Mengen aus  $\mathcal{A}$  gilt:

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

Das Tupel  $(\Omega, \mathcal{A}, \mu)$  heißt **Maßraum**.

**Definition 3.9 (Wahrscheinlichkeitsmaß, Wahrscheinlichkeitsraum)**

Sei  $(\Omega, \mathcal{A})$  ein Ereignisraum. Eine Abbildung  $P : \mathcal{A} \rightarrow [0, 1]$  heißt **Wahrscheinlichkeitsmaß**, wenn  $P$  ein Maß ist und zusätzlich die Normiertheitseigenschaft

$$P(\Omega) = 1$$

erfüllt.

Der Maßraum  $(\Omega, \mathcal{A}, P)$  heißt **Wahrscheinlichkeitsraum**.

Wir kennen nun die notwendigen Voraussetzungen, um den Begriff der *Zufallsgröße* einzuführen. Eine Zufallsgröße  $X$  ordnet jedem möglichen Ergebnis  $\omega \in \Omega$  eines Zufallsexperiments einen Wert  $X(\omega)$  zu. Diese Werte bilden eine eigene Ergebnismenge  $\Omega'$ , auf welchem sich natürlich auch ein Ereignissystem  $\mathcal{A}'$  bilden lässt. Zunächst die Definition:

**Definition 3.10 (Zufallsvariable)**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(\Omega', \mathcal{A}')$  ein Messraum. Eine messbare Funktion  $X : \Omega \rightarrow \Omega'$  heißt **Zufallsgröße** oder **Zufallsvariable**.

Für eine Zufallsvariable  $X$  führen wir für das Urbild einer Menge  $A' \in \mathcal{A}'$  in  $X$  die Schreibweise

$$\{X \in A'\} := X^{-1}(A')$$

ein und nennen eine Teilmenge  $A \in \mathcal{A}$  dieser Form *beschreibbar* durch  $X$ .

**Beispiel 3.11** Beispielhaft können wir das zweimalige Werfen eines sechsseitigen Würfels betrachten. Die Ergebnismenge ist

$$\Omega = \{(a_1, a_2) | a_1, a_2 = 1, \dots, 6\}$$

und wir können nun eine Zufallsvariable definieren, die jedem Ergebnis die Summe der beiden Augenzahlen zuordnet:

$$X : \Omega \rightarrow \{2, \dots, 12\}, X((a_1, a_2)) = a_1 + a_2$$

Die Ziel-Ergebnismenge ist hier  $\Omega' = \{2, \dots, 13\}$ . Betrachten wir nun ein Ereignis  $A' \in \mathcal{P}(\Omega')$ , beispielsweise „Augensumme gleich 7“:

$$A' = \{s \in \Omega' | s = 7\}$$

Nun wird deutlich, wieso wir fordern, dass die Zufallsvariable  $X$  *messbar* sein soll. Die Messbarkeit ermöglicht uns, die Wahrscheinlichkeit des Ereignisses  $A'$  zu bestimmen, indem wir das auf der ursprünglichen Ergebnismenge  $\Omega$  definierte Wahrscheinlichkeitsmaß  $P$  nutzen. Denn: über die Messbarkeit von  $X$  ist gefordert, dass die durch  $A'$  beschriebene Menge  $A \subset \Omega$ :

$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

im Ereignissystem  $\mathcal{A}$  von  $\Omega$  enthalten ist. Wir sehen:  $P(A) = \frac{1}{6}$ . Setzen wir  $P'(A')$  auf denselben Wert, so können wir auf diesem Weg direkt ein Wahrscheinlichkeitsmaß  $P'$  auf  $\Omega'$  definieren.

Um die Messbarkeit einer Zufallsvariable sicherzustellen, wird als Ereignissystem auf endlichen oder abzählbaren Wahrscheinlichkeitsräumen in der Regel die Potenzmenge genutzt. In überabzählbaren Wahrscheinlichkeitsräumen ist dies jedoch nicht ohne weiteres möglich.

### 3.4 Eigenschaften von Wahrscheinlichkeiten

Wir haben nun die formelle Definition des Wahrscheinlichkeitsmaßes kennengelernt, welches jedem Ergebnis eines Zufallsexperiments und auch jedem Ereignis einen Zahlenwert zuzuordnen vermag. Doch wie ist dieses Konzept vereinbar mit der Vorstellung, die wir von Wahrscheinlichkeiten aus dem Alltag mitbringen?

Wahrscheinlichkeiten sind ein abstraktes Konzept, welches wir nutzen, um Vorgänge der realen Welt in Zahlen zu fassen. Wir nutzen sie, um das Element des Zufalls greifbar zu machen, um Unsicherheit quantifizieren zu können und um Ungewissheit über den Ausgang realer Geschehnisse zum Ausdruck zu bringen. Doch die Systeme, die wir so beschreiben, scheitern nicht um die Zahlen, die wir ihnen zuweisen. Diese Zahlen, das Wahrscheinlichkeitsmaß im realen Sinne, sind nur eine Abstraktion, um komplexe Prozesse besser greifbar zu machen.

Ein Beispiel: Es ist wohl jedem intuitiv klar, dass die Wahrscheinlichkeit, bei einem fairen Münzwurf das Ergebnis „Kopf“ zu erhalten, bei 50% liegen muss. Doch der Zufall, der das Ergebnis dieses Vorgangs bestimmt, richtet sich nicht nach dieser Zahl. Der Vorgang ist nicht einmal wirklich zufällig. Er hängt allein von einer Reihe von Umwelteinflüssen ab: der oben liegenden Seite vor dem Wurf, der exakten Kraftausübung des Werfenden, die Flugdauer, das Gewicht der Münze, und so weiter. Mit hinreichend genauen Informationen über diese Faktoren könnte man den Ausgang des Experiments Münzwurf exakt vorausbestimmen.

Wenn hier also überhaupt kein Zufall im Spiel ist, was bedeuten dann die fünfzig Prozent? Diese Zahl stellt eine starke Abstraktion der komplizierten Realität dar: Sie ist eine statistische Aussage darüber, was man, ausgehend von früheren Erfahrungen und beschränktem Wissen über das System, für das Ergebnis erwarten kann.

#### 3.4.1 Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit

TBD

#### 3.4.2 Rechnen mit Wahrscheinlichkeiten

Die Definition des Wahrscheinlichkeitsmaßes legt bereits einige seiner Eigenschaften fest. Daraus ableiten lassen sich eine Reihe nützlicher Rechenregeln.

##### Lemma 3.12 (Rechenregeln für Wahrscheinlichkeitsmaße)

1. Wahrscheinlichkeit des Gegenereignisses:

$$P(A^c) = 1 - P(A)$$

2. Wahrscheinlichkeit der Differenzmenge:

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

3. Wahrscheinlichkeit der (nicht disjunkten) Vereinigung:

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

4. Subadditivität:

$$P(A \cup B) \leq P(A) + P(B)$$

5. Monotonie:

$$A \subset B \Rightarrow P(A) \leq P(B)$$

6. Stetigkeit von Unten:

$$A_1 \subset A_2 \subset \dots \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i)$$

7. Stetigkeit von Oben:

$$A_1 \supset A_2 \supset \dots \Rightarrow P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i)$$

BEWEIS. Zur Übung.

□

### 3.4.3 Bedingte Wahrscheinlichkeiten und Stochastische Unabhängigkeit

In der Praxis kommt es häufig vor, dass Ereignisse voneinander abhängig sind. Das Wissen darüber, ob ein bestimmtes Ereignis  $A$  eingetreten ist, verändert die Wahrscheinlichkeit dafür, dass auch das Ereignis  $B$  eingetreten ist. Ein einfaches Beispiel: Ohne Vorwissen ist die Wahrscheinlichkeit, mit einem fairen Würfel eine 6 zu würfeln, gerade  $\frac{1}{6}$ . Ist allerdings bekannt, dass eine gerade Zahl gewürfelt wurde, dann steigt die Wahrscheinlichkeit der Zahl 6 auf  $\frac{1}{3}$ . Man sagt: Das Ereignis  $A :=$  „Eine 6 werfen“ ist vom Ereignis  $B :=$  „Eine gerade Zahl werfen“ abhängig, die Wahrscheinlichkeit von  $A$  wird durch das Wissen über das Eintreten von Ereignis  $B$  bedingt.

#### Definition 3.13 (Bedingte Wahrscheinlichkeit)

Seien  $A, B \subset \Omega$  zwei Ereignisse und  $P(B) > 0$ . Dann heißt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3.1)$$

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

#### Lemma 3.14 (Rechenregeln für bedingte Wahrscheinlichkeiten)

1. Wahrscheinlichkeit des Schnittes:

$$P(A \cap B) = P(B) \cdot P(A|B) \quad (3.2)$$

2. Verkettungsregel:

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B) \quad (3.3)$$

3. Formel der totalen Wahrscheinlichkeit: Sei  $B_1, \dots, B_n$ ,  $n \in \mathbb{N}$  eine abzählbare Zerlegung von  $\Omega$ . Dann gilt für ein Ereignis  $A \subset \Omega$ :

$$P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(B_i) \cdot P(A|B_i) \quad (3.4)$$

BEWEIS.

1. Durch Umstellen von (??).

2. Durch wiederholtes Anwenden von (??).

3. Es ist

$$\Omega = \sum_{i=1}^n B_i$$

eine disjunkte Zerlegung von  $\Omega$ , und nach dem Distributivgesetz für Mengen ist

$$A = A \cap \Omega = \sum_{i=1}^n A \cap B_i.$$

eine ebenfalls disjunkte Zerlegung von  $A$ . Die Formel (??) ergibt sich nun aus der  $\sigma$ -Additivität des Wahrscheinlichkeitsmaßes und (??).

□

### Satz 3.15 (Satz von Bayes)

Seien  $A, B \subset \Omega$  zwei Ereignisse. Dann gilt:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}. \quad (3.5)$$

BEWEIS. Wir stellen (??) um:

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

□

Natürlich kann es passieren, dass das Wissen über das Eintreten eines Ereignisses  $A$  keinerlei Schlüsse auf ein zweites Ereignis  $B$  zulässt, und umgekehrt. Die beiden Ereignisse sind dann unabhängig voneinander.

### Definition 3.16

Zwei Ereignisse  $A, B \subset \Omega$  heißen **stochastisch unabhängig**, wenn gilt:

$$P(A \cap B) = P(A) \cdot P(B) \quad (3.6)$$

Direkt aus der Definition folgt für zwei stochastisch unabhängige Ereignisse  $A$  und  $B$ :

$$P(A|B) = P(A) \text{ sowie } P(B|A) = P(B).$$

## 4 Wahrscheinlichkeitsmaße

Wir kennen bereits den Begriff des *Wahrscheinlichkeitsmaßes*. Zur Erinnerung: Ein Wahrscheinlichkeitsmaß ist eine Abbildung, die jedem *Ereignis* einen Wert zwischen 0 und 1 zuordnet. Wir nennen diesen Wert die *Wahrscheinlichkeit* des Ereignisses. In diesem Kapitel werden wir Funktionen definieren, um Wahrscheinlichkeitsmaße auf verschiedenen Grundmengen besser darstellen und analysieren zu können. Diese Funktionen nennen wir *Dichten*. Je nachdem, ob die Grundmenge abzählbar oder überabzählbar ist, kann man auf verschiedenen Wegen Wahrscheinlichkeitsmaße und Dichten eindeutig einander zuordnen.

### 4.1 Diskrete Wahrscheinlichkeitsmaße und Zähldichten

#### 4.1.1 Eigenschaften von Zähldichten

Zunächst befassen wir uns mit Wahrscheinlichkeitsmaßen für endliche oder abzählbar unendliche Ergebnismengen. In einem abzählbaren Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  kann  $P$  allein durch das Wissen über die Wahrscheinlichkeiten der Elementarereignisse  $\omega \in \Omega$  eindeutig definiert werden. Die Zuordnung dieser elementaren Wahrscheinlichkeiten können wir als eine Funktion betrachten:

**Definition 4.1 (Zähldichte)**

Sei  $\Omega$  eine abzählbare Ergebnismenge. Eine Funktion  $f : \Omega \rightarrow [0, 1]$  mit der Eigenschaft

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

heißt **Zähldichte** oder **Wahrscheinlichkeitsfunktion** auf  $\Omega$ .

**Satz 4.2**

Sei  $(\Omega, \mathcal{A})$  ein abzählbarer Ereignisraum.

- Ist  $f : \Omega \rightarrow [0, 1]$  eine Zähldichte, dann ist die Abbildung  $P_f : \mathcal{A} \rightarrow [0, 1]$  mit

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

ein Wahrscheinlichkeitsmaß auf  $\mathcal{A}$ .

- Ist umgekehrt  $P : \mathcal{A} \rightarrow [0, 1]$  ein Wahrscheinlichkeitsmaß auf  $\mathcal{A}$ , dann ist  $f_P : \Omega \rightarrow [0, 1]$  mit

$$f(\omega) = P(\{\omega\})$$

eine Zähldichte.

- Die so gegebene Zuordnung zwischen Wahrscheinlichkeitsmaßen und Zähldichten ist eindeutig.

BEWEIS. TBD.

□

#### 4.1.2 Diskrete Verteilungen auf endlichen Mengen

Im Folgenden wollen wir einige praktisch relevante diskrete Wahrscheinlichkeitsverteilungen und deren Zähldichten betrachten. Für verschiedene Ergebnismengen  $\Omega$  soll im Folgenden stets die  $\sigma$ -Algebra  $\mathcal{A} = \mathcal{P}(\Omega)$  verwendet werden.

**Die Bernoulli-Verteilung** Ein Zufallsexperiment mit einer zweielementigen Ergebnismenge  $\Omega = \{\omega_1, \omega_2\}$  heißt *Bernoulli-Experiment*. Die beiden möglichen Ausgänge werden in der Regel als *Erfolg* und *Misserfolg* bezeichnet. Auf dem dazugehörigen Ereignissystem lässt sich eine sehr einfache Klasse von Wahrscheinlichkeitsmaßen definieren:

**Definition 4.3 (Bernoulli-Verteilung)**

Ein Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit  $\Omega = \{\omega_1, \omega_2\}$  und

$$P(\omega_1) = p, \quad P(\omega_2) = 1 - p \quad (p \in [0, 1])$$

heißt **Bernoulli-Modell**.  $P$  heißt **Bernoulli-Verteilung** zum Parameter  $p$ , kurz  $B(p)$ .

Bernoulli-Modelle kommen überall dort zum Einsatz, wo nur zwischen zwei Fällen unterschieden wird. In diesem Kapitel werden einige weitere Wahrscheinlichkeitsverteilungen vorgestellt, welche verschiedene Aspekte der wiederholten Ausführung eines Bernoulli-Experiments beschreiben. Dazu gehört die erwartete Anzahl von Erfolgen bei  $n$  Wiederholungen, oder auch die zu erwartende Wartezeit bis zum ersten Erfolg.

**Die Laplace-Verteilung** Die Laplace-Verteilung beschreibt Zufallsexperimente mit endlich vielen, gleich wahrscheinlichen Ausgängen. Sie ist für die Modellierung von fairen Würfeln, Münzwürfen und ähnlichem geeignet.

**Definition 4.4 (Laplace-Verteilung)**

Das auf einem Ereignisraum  $(\Omega, \mathcal{A})$  mit einer endlichen Ergebnismenge  $\Omega = \{\omega_1, \dots, \omega_N\}$  durch die Zähldichte  $f$  mit

$$f(\omega_i) = \frac{1}{N} \quad (i = 1, \dots, N)$$

definierte Wahrscheinlichkeitsmaß  $P$  heißt **Laplace-Verteilung** über  $\Omega$ .

**Die Binomialverteilung** Als Motivation für die Binomialverteilung soll folgendes Beispiel dienen:

**Beispiel 4.5** Es wird der gleichzeitige Wurf von 10 Würfeln betrachtet. Nun möchte man die Wahrscheinlichkeit für das Ereignis berechnen, das bei genau 6 der Würfel die Augenzahl 1 oben liegt. Hierbei gibt der Binomialkoeffizient die Anzahl der möglichen geordneten Kombinationen an und dies wird mit dem Eintreten eines dieser Ereignisse multipliziert.

**Definition 4.6 (Binomialverteilung)**

Sei  $(\Omega, \mathcal{A})$  ein endlicher Wahrscheinlichkeitsraum mit  $\Omega = \{0, 1, \dots, n\}$ . Seien außerdem  $p \in [0, 1]$ . Die Zähldichte  $f : \Omega \rightarrow [0, 1]$  mit

$$f(k) = b_{n,p}(k) := \binom{n}{k} p^k (1-p)^{n-k}$$

heißt **Binomial-Zähldichte**. Das durch  $f$  auf  $\Omega$  definierte Wahrscheinlichkeitsmaß  $B(n, p)$  heißt **Binomialverteilung**.

Unter anderem beschreibt die Binomialverteilung das  $n$ -malige „Ziehen mit Zurücklegen ohne Beachtung der Reihenfolge“ aus einer Urne, welche Kugeln in zwei Farben enthält. Haben die Kugeln die Farben  $a$  und  $b$  mit jeweils den relativen Häufigkeiten  $p_a$  und  $p_b$  ( $p_a + p_b = 1$ ), dann wird die Frage nach der Wahrscheinlichkeit, bei  $n$  Zügen  $k$  mal die Farbe  $a$  zu ziehen, durch

$$B_{n,p_a}(k)$$

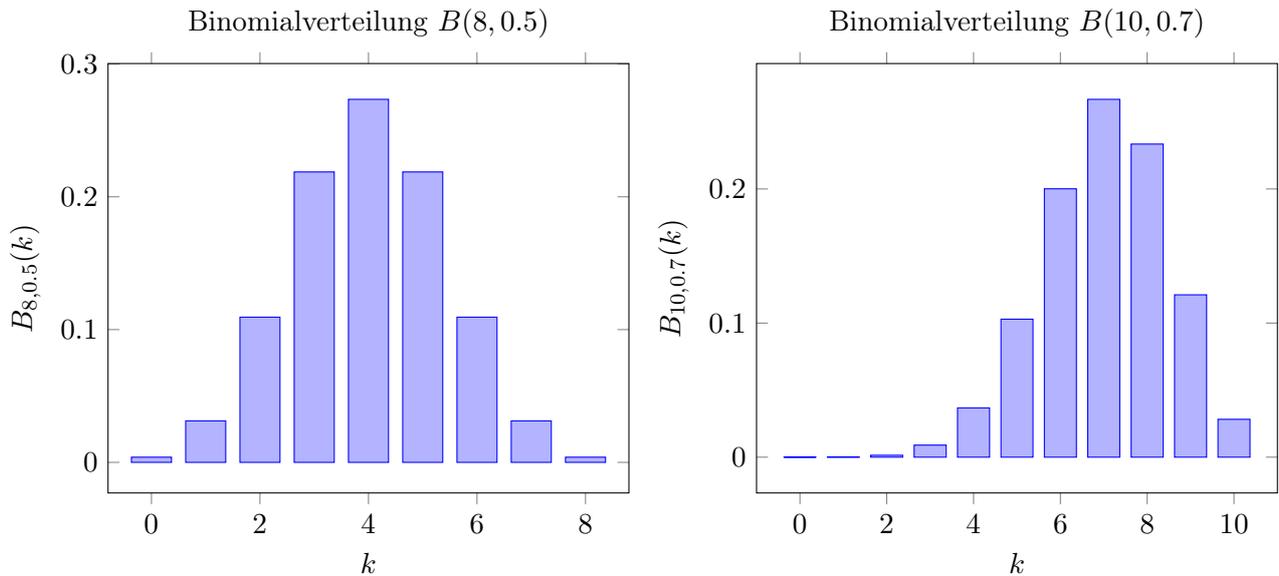


Abbildung 4.1: Binomialverteilungen

beschrieben.

Allgemein gesprochen ist „Farbe  $a$ “ hier ein *Erfolg*, und „Farbe  $b$ “ ist ein *Misserfolg*. Der Parameter  $p$  ist die Wahrscheinlichkeit eines Erfolgs bei jedem Versuch. Die Formel kommt nun durch die Frage nach der Wahrscheinlichkeit von genau  $k$  Erfolgen bei  $n$  Versuchen zu Stande. Das Produkt  $p^k(1-p)^{n-k}$  beschreibt die Wahrscheinlichkeit, zunächst  $k$  Erfolge ( $p^k$ ) und anschließend  $n-k$  Misserfolge ( $(1-p)^{n-k}$ ) zu erzielen. Der Binomialkoeffizient trägt anschließend alle möglichen Reihenfolgen für diese Erfolge und Misserfolge bei. In diesem Sinne beschreibt die Binomialverteilung die zu erwartenden Ergebnisse eines wiederholt durchgeführten Bernoulli-Experiments.

#### 4.1.3 Diskrete Verteilungen auf den natürlichen Zahlen

**Die geometrische Verteilung** Mit der geometrischen Verteilung wollen wir erstmals eine geläufige Wahrscheinlichkeitsverteilung auf den natürlichen Zahlen einführen. Die geometrische Verteilung beschreibt ein Wahrscheinlichkeitsmaß auf  $\mathbb{N}_0$ , welches für  $n \rightarrow \infty$  immer weiter fallende, aber nie verschwindende Wahrscheinlichkeiten liefert.

##### Definition 4.7 (Geometrische Verteilung)

Sei  $(\Omega, \mathcal{A})$  ein Ereignisraum mit  $\Omega = \mathbb{N}_0$ , und sei  $q \in (0, 1)$ . Die Zähldichte  $G_q : \mathbb{N}_0 \rightarrow [0, 1]$  mit

$$G_q(k) := (1-q)q^k$$

heißt **geometrische Zähldichte** und definiert das Wahrscheinlichkeitsmaß der **geometrischen Verteilung**.

Die geometrische Verteilung beschreibt die beliebig häufige Durchführung eines Bernoulli-Experiments, wobei bei jedem Versuch die Wahrscheinlichkeit des *Erfolges*  $1-q$  beträgt. Der Wert  $G(k)$  beschreibt dann die Wahrscheinlichkeit, dass der Erfolg beim  $k+1$ -ten Versuch zum ersten Mal eintritt. Der Teil  $q^k$  des Terms beschreibt die  $k$  vorangegangenen Misserfolge, und  $(1-q)$  das Eintreten des Erfolges.

**Beispiel 4.8** Eine Glühbirne habe bei jedem Einschalten eine Wahrscheinlichkeit von 1 %, durchzubrennen. Dazu gehört auch, dass sie das Einschalten mit 99%-iger Wahrscheinlichkeit unbeschadet übersteht. Nun ist

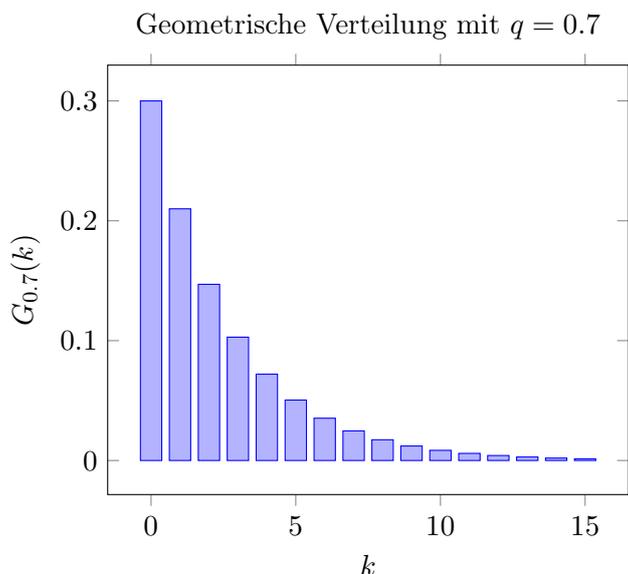


Abbildung 4.2: Geometrische Verteilung

beispielsweise die Wahrscheinlichkeit, dass sie beim achten Mal durchbrennt, gegeben durch

$$\underbrace{0.99 \cdot 0.99 \cdot \dots \cdot 0.99}_{7 \text{ mal überlebt}} \cdot \underbrace{0.01}_{1 \text{ mal durchgebrannt}}$$

Das Leben dieser Glühbirne wird also durch die geometrische Verteilung  $G_{0.99}$  beschrieben.

Die Formel der geometrischen Verteilung lässt sich aus der geometrischen Reihe ableiten. Für  $q \in (0, 1)$  ist

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \tag{4.1}$$

Diese Formel lässt sich per Multiplikation mit  $(1 - q)$  umstellen zu

$$\sum_{k=0}^{\infty} (1-q)q^k = 1.$$

Nun ist es naheliegend, für eine Zähldichte auf  $\mathbb{N}_0$  die einzelnen Summanden zu wählen. Damit ist die erforderliche Normierungseigenschaft bereits erfüllt.

In der Praxis tritt die geometrische Verteilung meist in abgewandelter Form auf. Es ist etwas unintuitiv, dass in der ursprünglichen Definition der Parameter  $q$  nicht die Wahrscheinlichkeit auf *Erfolg*, sondern die eines *Misserfolges* beschreibt. Außerdem beinhaltet die Urform der geometrischen Verteilung einen „nullten“ Versuch. Um reale Szenarien besser modellieren zu können, führen wir also zwei leicht abgewandelte Formen der geometrischen Verteilung ein:

**Definition 4.9 (Varianten der geometrischen Verteilung)**

Sei  $p \in [0, 1]$  die Erfolgswahrscheinlichkeit eines wiederholt durchgeführten Bernoulli-Experiments.

- Die geometrische Verteilung  $\text{Geo}^+(p)$  ist definiert durch die Zähldichte

$$\text{geo}^+(p; k) := p \cdot (1-p)^{k-1}$$

und beschreibt die Wahrscheinlichkeit, dass der Erfolg beim  $k$ -ten Versuch erstmals eintritt.

- Die geometrische Verteilung  $\text{Geo}^0(p)$  ist definiert durch die Zähldichte

$$\text{geo}^0(p; k) := p \cdot (1 - p)^k$$

und beschreibt die Wahrscheinlichkeit, dass der Erfolg beim  $k + 1$ -ten Versuch erstmals eintritt, bzw. dass genau  $k$  Misserfolge vor dem ersten Erfolg eintreten.

**Die negative Binomialverteilung** Eine Verallgemeinerung der geometrischen Verteilung ist die negative Binomialverteilung. Sie beschreibt ebenso die wiederholte Durchführung eines Bernoulli-Experiments, liefert allerdings die Wahrscheinlichkeiten für die nötige Anzahl an Versuchen, um  $r > 0$  Erfolge zu erhalten.

**Definition 4.10 (Negative Binomialverteilung)**

Sei  $p \in [0, 1]$  die Erfolgswahrscheinlichkeit eines wiederholt durchgeführten Bernoulli-Experiments, und sei  $r \in \mathbb{N}$  die erwünschte Anzahl an Erfolgen. Die **negative Binomialverteilung** beschreibt die Anzahl der notwendigen Versuche für  $r$  Erfolge.

- Die negative Binomialverteilung  $\text{Nb}^+(r, p)$  ist definiert durch die Zähldichte  $\text{nb}^+ : \{r, r + 1, \dots\} \rightarrow [0, 1]$ ,

$$\text{nb}^+(r, p; k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

und beschreibt die Wahrscheinlichkeit, dass der  $r$ -te Erfolg beim  $k$ -ten Versuch eintritt.

- Die negative Binomialverteilung  $\text{Nb}^0(r, p)$  ist definiert durch die Zähldichte  $\text{nb}^0 : \mathbb{N}_0 \rightarrow [0, 1]$ ,

$$\text{nb}^0(r, p; k) = \binom{k+r-1}{r-1} p^r (1-p)^k$$

und beschreibt die Wahrscheinlichkeit, dass vor dem  $r$ -ten Erfolg insgesamt  $k$  Misserfolge eintreten.

Die Dichte der negativen Binomialverteilung ist ein Hybrid aus der geometrischen und der Binomialverteilung. Der Teilterm  $p^r (1-p)^{k-r}$  errechnet die Wahrscheinlichkeit, bei  $k$  Versuchen  $r$  Erfolge und damit gleichzeitig  $k - r$  Misserfolge zu erzielen; allerdings ohne Beachtung der Reihenfolge. Der letzte Erfolg soll stets der  $k$ -te Versuch sein. Für die vorherigen  $r - 1$  Erfolge gibt es also  $\binom{k-1}{r-1}$  Möglichkeiten für die Zeitpunkte, zu denen sie unter den ersten  $k - 1$  Versuchen auftreten.

**Die Poisson-Verteilung** Die Poisson-Verteilung ist der Binomialverteilung konzeptionell sehr ähnlich. Sie dient der Modellierung von Vorgängen, in denen ein bestimmtes Ereignis mit einer bekannten mittleren Häufigkeit eintritt. Die einzelnen Auftritte müssen dabei voneinander unabhängig sein. Die Mittlung geschieht über ein beliebiges, vom zu modellierenden Sachverhalt abhängiges Zeitintervall oder räumliches Gebiet.

**Definition 4.11 (Poisson-Verteilung)**

Die von dem Parameter  $\lambda \in \mathbb{R}, \lambda > 0$  abhängige Zähldichte  $\pi_\lambda : \mathbb{N}_0 \rightarrow [0, 1]$  definiert durch

$$\pi_\lambda(k) = \exp(-\lambda) \cdot \frac{\lambda^k}{k!}$$

heißt **Poisson-Dichte** und definiert das Wahrscheinlichkeitsmaß der **Poisson-Verteilung**  $\pi(\lambda)$ .

Anschaulich gesprochen beschreibt der Parameter  $\lambda$  die zu erwartende Anzahl von Ereignissen. Er beschreibt außerdem sowohl Mittelwert als auch Streuung der Verteilung, welche ihr Maximum bei  $\lambda$  selbst hat und ähnlich der Binomialverteilung nach außen abflacht.

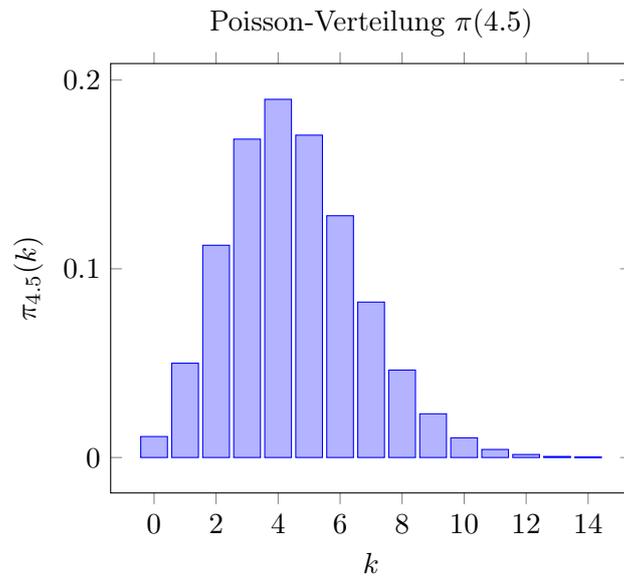


Abbildung 4.3: Poisson-Verteilung

## 4.2 Wahrscheinlichkeitsverteilungen auf den reellen Zahlen

Im Folgenden soll die überabzählbare Menge der reellen Zahlen  $\mathbb{R}$  als Ergebnismenge mit der Borel- $\sigma$ -Algebra  $\mathbb{B}$  als Ereignissystem betrachtet werden.

Ist die Ergebnismenge überabzählbar groß, so können wir die Wahrscheinlichkeiten von Ereignissen nicht mehr als Summe der Wahrscheinlichkeiten ihrer Elementarereignisse darstellen. In  $\mathbb{R}$  wäre ein Elementarereignis die Menge einer einzelnen reellen Zahl  $\{x\}, x \in \mathbb{R}$ . Zur Erinnerung: Die Borel- $\sigma$ -Algebra wird erzeugt durch halboffene Intervalle  $(a, b] \subset \mathbb{R}$ . Können wir  $\{x\}$  durch ein solches Intervall darstellen? Leider nicht:

- Das Intervall  $(x, x]$  ist die leere Menge, sie kann also  $x$  nicht enthalten.
- Doch jedes Intervall  $(x - \varepsilon, x]$  enthält bereits unendlich viele Elemente! Es kann also kein Elementarereignis beschreiben.

Die Borel- $\sigma$ -Algebra also nicht für die Darstellung von Elementarereignissen geeignet. Doch vielleicht finden wir eine andere Algebra, mit der es funktioniert? Man kann sich leicht davon überzeugen, dass das nicht möglich ist. Dies hängt mit der Überabzählbarkeit der reellen Zahlen zusammen.

### 4.2.1 Eigenschaften stetiger Wahrscheinlichkeitsmaße

Wir können also ein Wahrscheinlichkeitsmaß für die reellen Zahlen nicht über die Wahrscheinlichkeiten der Elementarereignisse definieren. Auch das Konzept der Zähldichte kann hier keine Anwendung mehr finden. Stattdessen definieren wir auf den reellen Zahlen eine andere Art von Dichte:

#### Definition 4.12 (Riemann-Dichte)

Eine Riemann-integrierbare Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit den Eigenschaften

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

und

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

heißt **Riemann-Dichte** oder **stetige Dichte** über  $\mathbb{R}$ .

Die Art und Weise, wie wir Wahrscheinlichkeiten von Ereignissen durch eine Zähldichte bestimmt haben, lässt sich auf Riemann-Dichten übertragen. Das, was bei den Zähldichten eine Summation über die Elementarereignisse war, wird nun zu einem Integral.

**Lemma 4.13**

Sei  $A \in \mathbb{B}$ . Dann existiert eine disjunkte Familie von Borel-Intervallen  $(a_i, b_i]$ , sodass

$$A := \sum_{i=1}^{\infty} (a_i, b_i].$$

BEWEIS. Per Definition der Borel- $\sigma$ -Algebra ist  $A \in \mathbb{B}$  eine Vereinigung abzählbar vieler halboffener Intervalle:

$$A = \bigcup_{i=1}^n (x_i, y_i] \quad n \in \mathbb{N} \cup \{\infty\}$$

Wir zeigen per vollständiger Induktion, dass sich  $A$  auch als disjunkte Vereinigung darstellen lässt.

- $n = 0$ :  $A_0 = \emptyset = (x, x]$ .
- $n \rightarrow n + 1$ : Sei  $A_{n+1} = (x_{n+1}, y_{n+1}] \cup \bigcup_{i=1}^n (x_i, y_i]$ . Per Induktionsvoraussetzung ist

$$A_n := \bigcup_{i=1}^n (x_i, y_i] = \sum_{i=1}^k (a_i, b_i].$$

Wir zerlegen nun das Intervall  $(x_{n+1}, y_{n+1}]$  in seine Schnittintervalle mit  $A_n$  und zu  $A_n$  komplementären Intervalle. Die Teilintervalle dieser Zerlegung sind disjunkt:

$$(x_{n+1}, y_{n+1}] = \sum_{j=k+1}^h (a_j, b_j] + (b_j, a_{j+1}]$$

wobei  $(a_j, b_j] \not\subseteq A_n$  und  $(b_j, a_{j+1}] \subseteq A_n$  sind. Nun können wir  $A_{n+1}$  als disjunkte Vereinigung von Borel-Intervallen angeben:

$$A_{n+1} = \sum_{i=1}^k (a_i, b_i] + \sum_{j=k+1}^h (a_j, b_j] = \sum_{i=1}^h (a_i, b_i]$$

□

**Satz 4.14**

Sei  $f : \mathbb{R} \rightarrow \mathbb{R}$  eine Riemann-Dichte, und die Abbildung  $p : \mathcal{G}_1 \rightarrow [0, 1]$  definiert durch

$$p((a, b]) = \int_a^b f(x) dx.$$

Sei außerdem  $A \in \mathbb{B}$  gegeben als disjunkte Vereinigung

$$A := \sum_{i=1}^{\infty} (a_i, b_i].$$

Dann ist die Abbildung  $P : \mathbb{B} \rightarrow [0, 1]$  mit

$$P(A) = \sum_{i=1}^{\infty} p((a_i, b_i])$$

ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathbb{B})$ .

BEWEIS.  $P$  muss die Eigenschaften eines Wahrscheinlichkeitsmaßes erfüllen:

- Maß der leeren Menge:

$$P(\emptyset) = P((a, a]) = \int_a^a f(x)dx = 0$$

- Normiertheit:

$$P(\mathbb{R}) = \int_{-\infty}^{\infty} f(x)dx = 1$$

- $\sigma$ -Additivität:

$$P\left(\sum_{i=1}^{\infty} (a_i, b_i]\right) = \sum_{i=1}^{\infty} p((a_i, b_i]) = \sum_{i=1}^{\infty} P((a_i, b_i])$$

□

Auf den reellen Zahlen werden die Wahrscheinlichkeiten von Ereignissen also durch Integration ermittelt. Die Wahrscheinlichkeit eines Elementarereignisses ist dabei paradoxerweise Null. Anders als bei den Zähldichten stellt der Funktionswert einer Riemann-Dichtefunktion an einer Stelle  $x$  *nicht* die Wahrscheinlichkeit dieses Elementarereignisses dar.

**Beispiel 4.15** Insert Beispiel Here

Gelegentlich stoßen wir auf Aufgabenstellungen, in denen nach der Wahrscheinlichkeit geschlossener, offener oder rechtsseitig halboffener Intervalle bzw. ihrer Vereinigungen gefragt ist. Dies sind nach unserer Definition zwar keine Borel-Mengen, wodurch der Satz ?? streng genommen nicht anwendbar ist. Das soll uns allerdings nicht stören.

**Korollar 4.16**

Das Ergebnis des Satzes ?? ist auf alle Arten von Intervallen anwendbar.

BEWEIS. Mit angepasstem Definitionsbereich für  $p$  ist

$$p((a, b)) = p([a, b]) = p((a, b]) = p([a, b)) = \int_a^b f(x)dx.$$

□

Um die Wahrscheinlichkeitsmaße von Riemann-Dichten auszuwerten, müssen wir die Stammfunktion der Dichte bestimmen. Diese Stammfunktion wird *Verteilungsfunktion* heißen, und wir können sie unabhängig von der Riemann-Dichte definieren:

**Definition 4.17 (Verteilungsfunktion)**

Sei  $P$  ein Wahrscheinlichkeitsmaß über  $(\mathbb{R}, \mathbb{B})$ . Die Abbildung  $F : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$F(x) := P((-\infty, x])$$

heißt **Verteilungsfunktion** von  $P$ .

Ist die Verteilungsfunktion bekannt, so lassen sich Wahrscheinlichkeiten von Ereignissen einfach berechnen:

**Lemma 4.18**

Ist  $P$  ein Wahrscheinlichkeitsmaß über  $(\mathbb{R}, \mathbb{B})$  mit Verteilungsfunktion  $F$ , so gilt:

$$P((a, b]) = F(b) - F(a), \quad a, b \in \mathbb{R}, a \leq b$$

BEWEIS. Unter Ausnutzung der  $\sigma$ -Additivität von  $P$  ergibt sich:

$$\begin{aligned} F(b) - F(a) &= P((-\infty, b]) - P((-\infty, a]) \\ &= P((-\infty, a]) + P((a, b]) - P((-\infty, a]) \\ &= P((a, b]) \end{aligned}$$

□

Im speziellen Fall eines durch eine Riemann-Dichte definierten Wahrscheinlichkeitsmaßes ist auch die Verteilungsfunktion leicht aufzustellen.

**Satz 4.19 (Verteilungsfunktion einer Riemann-Dichte)**

Sei  $f : \mathbb{R} \rightarrow \mathbb{R}$  eine Riemann-Dichte und  $P$  das durch  $f$  auf  $(\mathbb{R}, \mathbb{B})$  definierte Wahrscheinlichkeitsmaß. Dann gilt für die Verteilungsfunktion  $F$  von  $P$ :  $F$  ist Stammfunktion von  $f$  und

$$F(x) = \int_{-\infty}^x f(t) dt.$$

BEWEIS.

$$F(x) = P((-\infty, x]) = \int_{-\infty}^x f(t) dt$$

□

Verteilungsfunktionen lassen sich durch drei Eigenschaften charakterisieren:

**Satz 4.20 (Charakterisierung der Verteilungsfunktion)**

Eine Funktion  $F : \mathbb{R} \rightarrow [0, 1]$  ist genau dann eine Verteilungsfunktion, wenn sie die folgenden Eigenschaften erfüllt:

- $F$  ist monoton steigend, d.h. für alle  $x, y \in \mathbb{R}, x < y$  gilt:  $F(x) \leq F(y)$ .
- $F$  ist rechtsseitig stetig, d.h. für jeden Punkt  $x_0 \in \mathbb{R}$  gilt für den rechtsseitigen Grenzwert

$$\lim_{x \searrow x_0} F(x) = F(x_0)$$

- $F$  ist normiert, d.h.

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

BEWEIS. **Hinrichtung:** Sei  $F$  eine Verteilungsfunktion. Dann gilt:

- Monotonie:

$$F(x) = P((-\infty, x]) \leq P((-\infty, x]) + P((x, y]) = P((-\infty, y]) = F(y)$$

- Rechtsseitige Stetigkeit:

$$\begin{aligned} \lim_{x \searrow x_0} F(x) &= \lim_{x \searrow x_0} P((-\infty, x]) \\ &= P((-\infty, x_0]) + \lim_{x \searrow x_0} P((x_0, x]) \\ &= P((-\infty, x_0]) + P((x_0, x_0]) \\ &= P((-\infty, x_0]) \\ &= F(x_0) \end{aligned}$$

- Normiertheit:

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} P((-\infty, x]) = P(\emptyset) = 0$$

und

$$\lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} P((-\infty, x]) = P(\mathbb{R}) = 1.$$

**Rückrichtung:** Sei  $F$  monoton steigend, rechtsseitig stetig und normiert. Wir zeigen: Es existiert ein Wahrscheinlichkeitsmaß  $P$ , sodass  $F$  die zu  $P$  gehörige Verteilungsfunktion ist.

TBD.

□

#### Satz 4.21 (Korrespondenzsatz)

Sei  $F$  eine Verteilungsfunktion. Dann ist die zu  $F$  gehörige Wahrscheinlichkeitsverteilung  $P_F$  durch

$$P_F((-\infty, x]) := F(x)$$

eindeutig bestimmt.

Sei umgekehrt  $P$  ein Wahrscheinlichkeitsmaß. Dann ist auch die zu  $P$  gehörige Verteilungsfunktion  $F_P$  durch

$$F_P(x) := P((-\infty, x])$$

eindeutig bestimmt. Die Zuordnung von Wahrscheinlichkeitsmaßen zu Verteilungsfunktionen ist also bijektiv.

BEWEIS. TBD.

□

### 4.2.2 Spezielle stetige Wahrscheinlichkeitsverteilungen

Im Folgenden sollen einige praktisch relevante stetige Wahrscheinlichkeitsmaße, ihren Dichten und Verteilungsfunktionen vorgestellt werden. Viele dieser Verteilungen sind nicht auf den gesamten reellen Zahlen definiert, sondern nur auf Teilintervallen derselben. Für nur abschnittsweise definierten Verteilungen kann allerdings in jedem Fall angenommen werden, dass der Wert der Dichte außerhalb des Definitionsbereiches gleich Null wird.

Die für eine Verteilung tatsächlich relevante Menge möchten wir charakterisieren:

#### Definition 4.22 (Träger)

Sei  $P$  ein Wahrscheinlichkeitsmaß über  $\mathbb{R}$  mit Dichte  $f$ . Dann heißt die Menge

$$T = \{x \in \mathbb{R} \mid f(x) \neq 0\}$$

**Träger** von  $P$ .

Wir definieren einige Verteilungen lediglich auf ihrem Träger.

In vielen der noch folgenden Abschnitte werden wir häufig Dichten über ganz  $\mathbb{R}$  integrieren. In der Regel beschränkt sich der Integrationsbereich jedoch auf den Träger der Dichte.

**Die Rechtecksverteilung** Eine sehr einfache stetige Verteilung ist die Rechtecksverteilung. Sie ist dadurch charakterisiert, dass sie jedem Wert eines Intervalls die gleiche Wahrscheinlichkeit zuweist. Zunächst die Definition:

**Definition 4.23 (Rechtecksverteilung)**

Das Wahrscheinlichkeitsmaß der **stetigen Gleichverteilung** oder **Rechtecksverteilung**  $\mathcal{R}(a, b)$  über dem Intervall  $(a, b) \in \mathbb{R}$  ist definiert durch die Riemann-Dichte

$$f_{\mathcal{R}}(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in (a, b) \\ 0 & \text{sonst.} \end{cases}$$

Ihre Verteilungsfunktion ist

$$F_{\mathcal{R}}(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b. \end{cases}$$

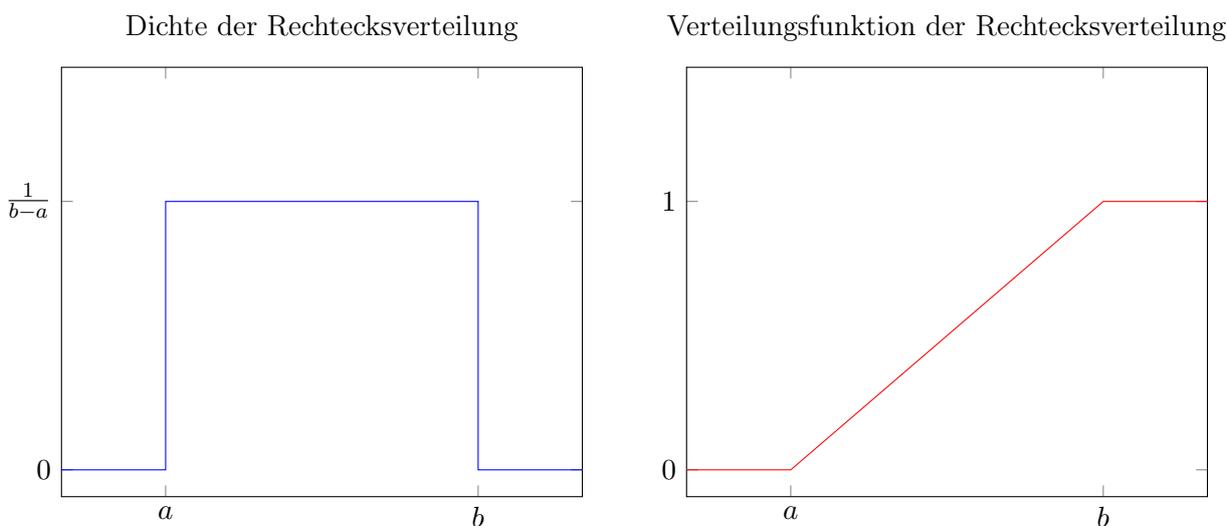


Abbildung 4.4: Rechtecksverteilung

Wie aus den Graphen von Dichte und Verteilungsfunktion (Abbildung ??) deutlich wird, sind die Wahrscheinlichkeiten hier gleichmäßig verteilt - jedes Intervall innerhalb von  $(a, b)$  hat eine Wahrscheinlichkeit proportional zu seiner Breite.

**Die Exponentialverteilung** Die Exponentialverteilung findet vor allem bei der Beschreibung von natürlichen Zerfallsprozessen Anwendung.

**Definition 4.24 (Exponentialverteilung)**

Das Wahrscheinlichkeitsmaß der **Exponentialverteilung**  $\text{Exp}_{\alpha}$  mit Parameter  $\alpha > 0$  ist definiert durch die Riemann-Dichte

$$f_{\alpha}(x) = \alpha \exp(-\alpha x) \cdot 1_{(0, \infty)} = \begin{cases} \alpha \exp(-\alpha x) & x > 0 \\ 0 & x \leq 0 \end{cases}$$

und die dazugehörige Verteilungsfunktion

$$F_{\alpha}(x) = \begin{cases} 1 - \exp(-\alpha x) & x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

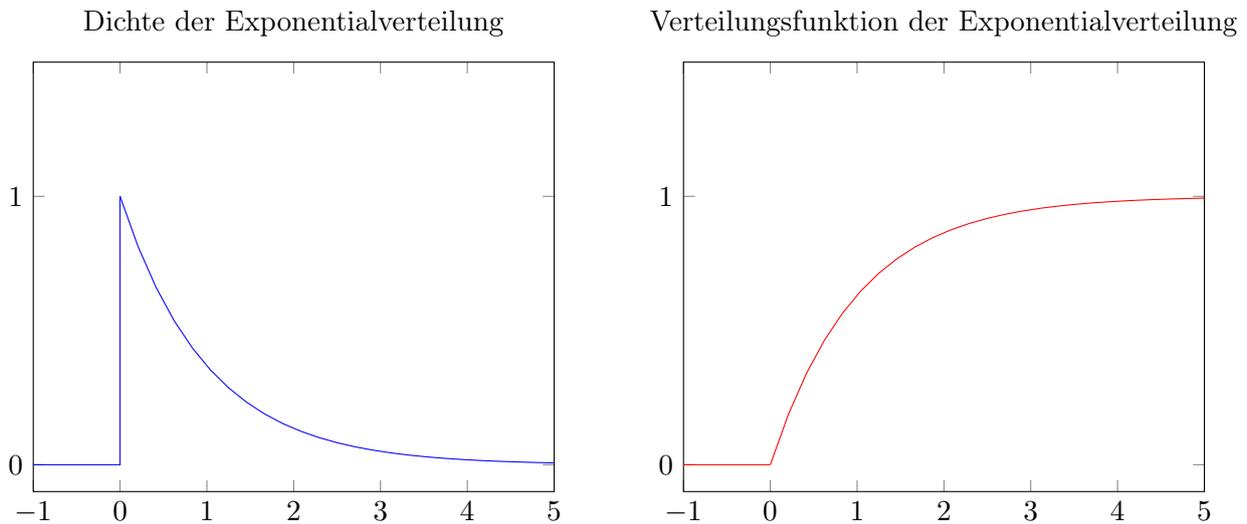


Abbildung 4.5: Exponentialverteilung

**Gamma- und Betaverteilung** Sowohl Beta- als auch Gammaverteilung sind parametrisierte Wahrscheinlichkeitsverteilungen und zeichnen sich durch eine hohe Flexibilität aus. Durch geeignete Wahl ihrer Parameter lassen sich viele verschiedene reale Sachverhalte modellieren. Außerdem eignen sie sich zur Abstraktion von spezielleren Verteilungen und beinhalten manche von diesen als Spezialfälle.

Beide Verteilungen stützen sich auf die Gamma-Funktion; diese ist eine Verallgemeinerung der Fakultät auf die komplexen Zahlen. Uns genügt es, die Gamma-Funktion auf den positiven reellen Zahlen zu betrachten.

**Definition 4.25 (Gamma-Funktion)**

Für positive reelle Zahlen ist die Gammafunktion  $\Gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  definiert durch

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \tag{4.2}$$

Bevor wir zu den beiden Verteilungen kommen, wollen wir einige Eigenschaften der Gammafunktion betrachten. Wie bereits angemerkt handelt es sich bei der Gammaverteilung um eine Verallgemeinerung der Fakultätsfunktion. Diese Tatsache führt direkt zur folgenden Eigenschaft.

**Lemma 4.26 (Funktionalgleichung der Gammafunktion)**

Für  $x \in \mathbb{R}^+$  gilt

$$x \cdot \Gamma(x) = \Gamma(x + 1). \tag{4.3}$$

BEWEIS. Per partieller Integration erhalten wir

$$\begin{aligned} \Gamma(x + 1) &= \int_0^\infty t^x e^{-t} dt \\ &= [-t^x e^{-t}]_0^\infty + \int_0^\infty x t^{x-1} e^{-t} dt \\ &= x \cdot \int_0^\infty t^{x-1} e^{-t} dt = x \cdot \Gamma(x), \end{aligned}$$

da

$$\lim_{s \rightarrow \infty} [-t^x e^{-t}]_0^s = \lim_{s \rightarrow \infty} (-s^x e^{-s}) = 0.$$

□

Mit dieser Erkenntnis folgt direkt die gewünschte Eigenschaft mit Bezug auf die Fakultät.

**Satz 4.27**

Für  $k \in \mathbb{N}_0$  gilt

$$\Gamma(k+1) = k!$$

BEWEIS. Per vollständiger Induktion:

- $k = 0$ :  $\Gamma(1) = \int_0^\infty e^{-t} dt = 1 = 0!$
- $k - 1 \rightarrow k$ :

$$\Gamma(k+1) \stackrel{\text{Lemma ??}}{=} k \cdot \Gamma(k) \stackrel{\text{I.V.}}{=} k \cdot (k-1)! = k!$$

□

Nun wollen wir die bereits beschriebenen Verteilungen einführen:

**Definition 4.28 (Gammaverteilung)**

Das Wahrscheinlichkeitsmaß der **Gammaverteilung**  $\Gamma_{\alpha, v}$  mit den Parametern  $\alpha, v > 0$  ist definiert durch die Riemann-Dichte

$$\gamma(x) := \begin{cases} \frac{\alpha^v}{\Gamma(v)} x^{v-1} e^{-\alpha x} & x > 0 \\ 0 & x \leq 0. \end{cases} \quad (4.4)$$

**Bemerkung 4.29** Für  $v = 1$  ergibt sich die Exponentialverteilung als Spezialfall der Gammaverteilung.

**Definition 4.30 (Betaverteilung)**

- Die Funktion  $B : (0, \infty)^2 \rightarrow (0, \infty)$  mit

$$B(p, q) := \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \quad (4.5)$$

heißt **Betafunktion**.

- Das Wahrscheinlichkeitsmaß der **Betaverteilung**  $\text{Beta}(p, q)$  mit den Parametern  $p, q > 0$  ist definiert durch die Riemann-Dichte

$$f(x) = \begin{cases} \frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1} & x \in (0, 1) \\ 0 & \text{sonst.} \end{cases} \quad (4.6)$$

**Die Normalverteilung** Eine sehr besondere Rolle in der Welt der Statistik nimmt die Normalverteilung, auch Gauß'sche Normalverteilung oder Gauß-Verteilung ein.

**Definition 4.31 (Normalverteilung)**

Für jeden Parameter  $\mu \in \mathbb{R}$  und  $\sigma > 0$  ist die **Normalverteilung** mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma$ , geschrieben  $\mathcal{N}(\mu, \sigma^2)$ , definiert durch die Riemann-Dichte

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

Die Normalverteilung  $\mathcal{N}(0, 1)$  mit Erwartungswert 0 und Standardabweichung 1 heißt **Standardnormalverteilung**.

Dichte der Standardnormalverteilung

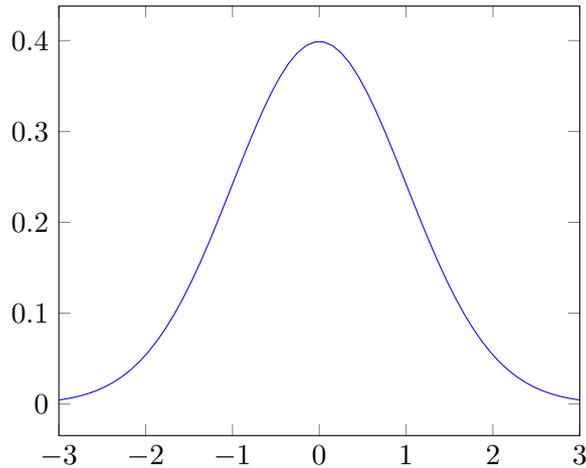


Abbildung 4.6: Standardnormalverteilung

Das Integral der Dichte der Normalverteilung lässt sich nicht in geschlossener Form berechnen, weswegen man zur Bestimmung der Werte der Verteilungsfunktion auf numerische Methoden und Tabellenwerke angewiesen ist. Die Verteilungsfunktion der Normalverteilung ist im Allgemeinen gegeben als

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt. \quad (4.7)$$

Für den Fall der Standard-Normalverteilung ergibt sich die Verteilungsfunktion

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt. \quad (4.8)$$

Die Berechnung der Verteilungsfunktion der allgemeinen Normalverteilung kann vereinfacht werden:

**Satz 4.32**

Sei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung  $\mathcal{N}(0, 1)$ . Dann ist die Verteilungsfunktion einer beliebigen Normalverteilung  $\mathcal{N}(\mu, \sigma^2)$  ( $\mu \in \mathbb{R}$ ,  $\sigma > 0$ ) gegeben als

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right). \quad (4.9)$$

BEWEIS. Zunächst stellen wir (??) um:

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

Wir substituieren nun im Integral  $z := \frac{t-\mu}{\sigma}$ , also  $dt = \sigma dz$ :

$$= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\frac{x-\mu}{\sigma}} \exp\left(-\frac{1}{2}z^2\right) dz = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

□

Diese Erkenntnis kommt vor allem bei der Arbeit mit Tabellenwerken zum Einsatz: Es genügt also, eine Tabelle für die Standardnormalverteilung zu haben, um Werte für jede beliebige Normalverteilung zu berechnen.

## 5 Mehrstufige Modelle

Bei der mehrfachen Durchführung eines Zufallsexperiments kann es sein, dass sich die Wahrscheinlichkeiten der möglichen Ereignisse zu jedem Versuch ändern, abhängig von den zuvor beobachteten Ereignissen. Beispielsweise ist beim Ziehen ohne Zurücklegen aus einer Urne die Wahrscheinlichkeit, beim zweiten Mal Ziehen eine bestimmte Farbe zu erhalten, anders als beim ersten Mal: Durch das Entfernen einer Kugel haben sich die relativen Häufigkeiten der Farben geändert.

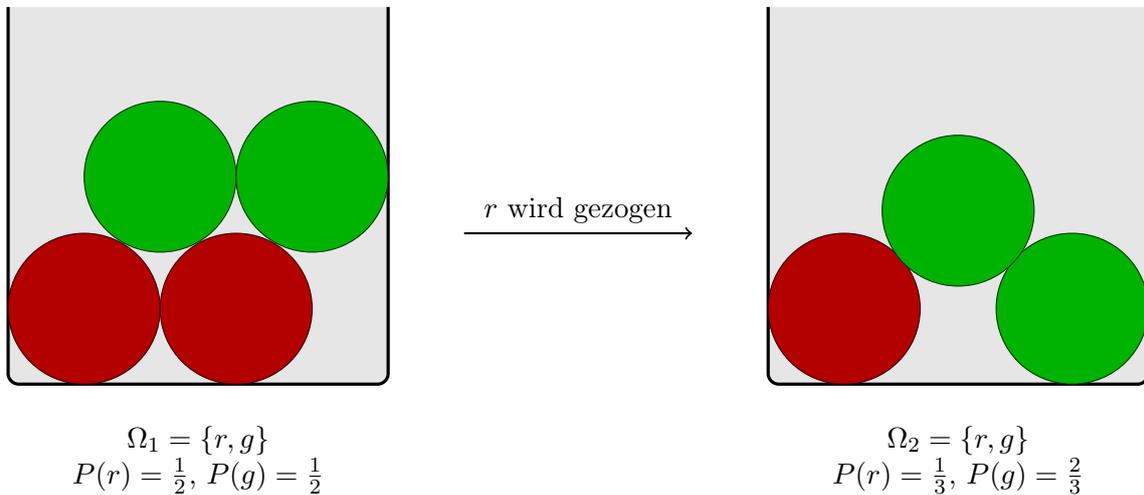


Abbildung 5.1: Ziehen ohne Zurücklegen ändert die Wahrscheinlichkeiten

Das wiederholte Ziehen ohne Zurücklegen aus einer Urne ist ein Beispiel für ein *Mehrstufiges Zufallsexperiment*. Es besteht aus mehreren Versuchen, genannt Stufen. Sowohl die Menge der möglichen Ergebnisse  $\Omega_i$  der  $i$ -ten Stufe als auch deren Wahrscheinlichkeiten hängen von den Ergebnissen der vorherigen Stufen ab. Abbildung ?? illustriert dies anschaulich. Nachdem in der ersten Stufe eine rote Kugel gezogen wurde, haben sich die Wahrscheinlichkeiten für den Versuch der zweiten Stufe verändert.

Im Folgenden wollen wir Mittel und Wege ergründen, derartige Zufallsexperimente mathematisch zu beschreiben. Wir entwickeln die sogenannten *gekoppelten Modelle* zur Beschreibung dieser Experimente und betrachten insbesondere die Wahrscheinlichkeitsdichten, die diesen Modellen zu Grunde liegen.

### 5.1 Gekoppelte Modelle und ihre Dichten

Zunächst wollen wir das mehrstufige Zufallsexperiment als formales Konstrukt darstellen.

#### Definition 5.1 (Mehrstufiges Zufallsexperiment)

Ein **mehrstufiges Zufallsexperiment** mit  $n$  Stufen wird beschrieben durch die Ergebnismenge

$$\Omega = \Omega_1 \times \cdots \times \Omega_n,$$

wobei  $\Omega_i$  die Menge der möglichen Ergebnisse in der  $i$ -ten Stufe ist.

Im Beispiel der Urne gilt  $\Omega_1 = \Omega_2 = \dots$ , die möglichen Ergebnisse ändern sich also von Stufe zu Stufe nicht (ihre Wahrscheinlichkeiten allerdings schon; werden alle Kugeln einer Farbe entnommen, so lässt sich dies durch  $P(c) = 0$  ausdrücken). Es gibt aber auch mehrstufige Zufallsexperimente, bei denen die Ergebnismengen der einzelnen Stufen völlig verschieden sind.

**Beispiel 5.2** TBD

Wir beschreiben ein mehrstufiges Zufallsexperiment, indem wir seinen einzelnen Stufen eigene Dichten sowie Wahrscheinlichkeitsmaße zuordnen. Da diese Dichten voneinander abhängen, werden sie und die dazugehörigen Wahrscheinlichkeitsmaße als *gekoppelt* bezeichnet.

**Definition 5.3 (Übergangsdichten, Kopplung)**

Sei  $\Omega = \Omega_1 \times \dots \times \Omega_n$  die Ergebnismenge eines mehrstufigen Zufallsexperiments.

- Die der  $i$ -ten Stufe vorausgehenden Beobachtungen

$$(\omega_1, \dots, \omega_{i-1}) \in \Omega_1 \times \dots \times \Omega_{i-1}$$

heißen **Vorgeschichte zur Stufe  $i$** .

- Die von der Vorgeschichte abhängige Familie von Wahrscheinlichkeitsdichten auf  $\Omega_i$

$$f_i^{i-1} : (\Omega_1 \times \dots \times \Omega_{i-1}) \times \Omega_i \rightarrow [0, \infty)$$

heißen **Übergangsdichten** von  $\Omega_1 \times \dots \times \Omega_{i-1}$  nach  $\Omega_i$ .

- Die das gesamte Zufallsexperiment beschreibende Dichte  $f : \Omega \rightarrow [0, \infty)$  ist definiert durch

$$f(\omega_1, \dots, \omega_n) = f_1(\omega_1) \cdot f_2^1(\omega_1; \omega_2) \dots f_n^{n-1}(\omega_1, \dots, \omega_{n-1}; \omega_n) \tag{5.1}$$

und heißt **Kopplung** von  $f_1, f_2^1, \dots, f_n^{n-1}$ . Sie wird geschrieben als

$$f := f_1 \otimes f_2^1 \otimes \dots \otimes f_n^{n-1}.$$

**Bemerkung 5.4** • Bei bekannter Vorgeschichte  $\omega_1, \dots, \omega_{i-1}$  ist die Wahrscheinlichkeitsdichte der  $i$ -ten Stufe gegeben als

$$\omega_i \mapsto f_i^{i-1}(\omega_1, \dots, \omega_{i-1}; \omega_i).$$

- Sind die einzelnen Übergangsdichten Zähldichten, so lässt sich die Tatsache, dass die durch (??) gegebene Funktion tatsächlich eine Dichte ist, beispielsweise durch vollständige Induktion über die Anzahl der Stufen zeigen.

**5.1.1 Kopplung stetiger Modelle**

Für Ausprägungsräume eines mehrstufigen Zufallsexperiments in den reellen Zahlen  $\Omega_1, \dots, \Omega_n \subseteq \mathbb{R}$  wollen wir die Kopplung von Riemann-Dichten noch etwas genauer betrachten. Es kann ohne weiteres angenommen werden, dass die Dichte der ersten Stufe  $f_1$  eine Riemann-Dichte ist. Genauso ist auch für eine feste Vorgeschichte  $x_1, \dots, x_{i-1}$  die Übergangsdichte

$$x_i \mapsto f_i^{i-1}(x_1, \dots, x_{i-1}; x_i)$$

eine Riemann-Dichte. Die durch (??) gegebene Kopplung ist o.B.d.A. eine Funktion  $f : \mathbb{R}^n \rightarrow [0, \infty)$  und beinhaltet *alle möglichen* auftretenden Vorgeschichten. Die Riemann-Integrierbarkeit von  $f$  ist im Allgemeinen nicht garantiert und muss nachgeprüft werden.

**5.2 Unabhängige Kopplung**

Der einfachste Fall mehrstufiger Zufallsexperimente ist der, bei dem die Wahrscheinlichkeitsdichten der einzelnen Stufen *nicht* von der Vorgeschichte abhängen. In diesem Fall spricht man von *unabhängiger Kopplung*.

**Definition 5.5 (Produktdichte)**

Hängen die Übergangsdichten  $f_1, \dots, f_n$  eines mehrstufigen Zufallsexperimentes nicht von den jeweiligen Vorgeschichten ab, so spricht man von der **unabhängigen Kopplung** der Stufen. Die Gesamtdichte ist dann gegeben durch das Produkt der Einzeldichten

$$f(\omega_1, \dots, \omega_n) = \prod_{i=1}^n f_i(\omega_i) \quad (5.2)$$

und heißt **Produktdichte**.

**5.2.1 Unabhängig gekoppelte diskrete Verteilungen**

**n-faches Laplace-Experiment** Mehrere voneinander unabhängige Laplace-Versuche mit nicht notwendigerweise gleichen Ergebnismengen  $\Omega_1, \dots, \Omega_n$  können als mehrstufiges Zufallsexperiment zu einem einzelnen Laplace-Versuch mit Ergebnismenge  $\Omega = \Omega_1 \times \dots \times \Omega_n$  zusammengefasst werden. Die Mächtigkeit dieser Menge ist gerade das Produkt der Mächtigkeiten der Einzelmengen. Die Formel der Produktdichte liefert dann eine Gesamt-Laplace-Verteilung mit der Dichte  $f : \Omega \rightarrow [0, 1]$ ,

$$f(\omega_1, \dots, \omega_n) = \frac{1}{|\Omega_1|} \cdot \frac{1}{|\Omega_2|} \cdots \frac{1}{|\Omega_n|} = \frac{1}{|\Omega|}.$$

**5.2.2 Die n-dimensionale Standardnormalverteilung**

## 6 Zufallsvariablen

Wir hatten den Begriff der Zufallsvariable bereits in Abschnitt ?? eingeführt (siehe Definition ??). Eine Zufallsvariable ist eine Funktion, die die Ergebnisse eines Zufallsexperimentes messbar auf Größen abbildet, welche diese Ergebnisse beschreiben sollen. Auf diese Weise können Zufallsvariablen genutzt werden, um die möglicherweise sehr feingranulare Ergebnismenge eines Experiments durch eine leichter zu handhabende Bildmenge zu abstrahieren. Der Wert einer Zufallsvariable beschreibt für gewöhnlich eine ganze Klasse von Ergebnissen des Urbildraumes.

Der Name „Zufallsvariable“ stammt daher, dass es sich um eine Größe handelt, deren Wert vom Zufall abhängt. Insbesondere wird der Wert der Zufallsvariable durch die Wahrscheinlichkeitsverteilung ihres Urbildraumes beeinflusst. Ist diese bekannt, so lässt sich damit auch eine Wahrscheinlichkeitsverteilung auf dem Bildraum definieren. Dieser wird dadurch ebenfalls zum Wahrscheinlichkeitsraum. So können wir einer Zufallsvariable auch direkt eine Dichte und eine Verteilungsfunktion zuordnen.

In der Praxis und für unsere Zwecke kann der Urbildraum einer Zufallsvariable meist ausgeblendet werden. Wir betrachten in der Regel lediglich den Bildraum und insbesondere die Dichte der Zufallsvariable. Von besonderem Interesse sind *reellwertige Zufallsvariablen*, also solche Zufallsvariablen, welche den Raum  $(\mathbb{R}, \mathbb{B})$  als Bildraum haben.

Im Folgenden wollen wir einige grundlegende Definitionen für die Arbeit mit Zufallsvariablen klären. Im Anschluss soll auf konkrete Verteilungen von diskreten und stetigen Zufallsgrößen eingegangen werden. Außerdem wollen wir Zufallsvariablen in mehreren Dimensionen untersuchen. Wir möchten sie miteinander koppeln, ihre stochastische Abhängigkeit untersuchen und sie durch arithmetische Operationen verknüpfen.

### 6.1 Bildverteilungen von Zufallsvariablen

Eine Zufallsvariable stattet ihren Bildraum, welcher per Definition nur ein Messraum ist, mit einem Wahrscheinlichkeitsmaß aus. Dadurch wird dieser zum Wahrscheinlichkeitsraum, und wir können ihn mit dem gesamten uns bekannten Arsenal der Wahrscheinlichkeitstheorie analysieren. Die Zufallsvariable abstrahiert ihren Urbildraum zu einem ihn beschreibenden *Modell*. Daher nennen wir den Bildraum auch *Bildmodell* der Zufallsvariable, und das auf ihm definierte Wahrscheinlichkeitsmaß *Bildmaß*:

#### Definition 6.1 (Bildmodell)

Sei  $(\Omega_1, \mathcal{A}_1, P)$  ein Wahrscheinlichkeitsraum,  $(\Omega_2, \mathcal{A}_2)$  ein Messraum und  $X : \Omega_1 \rightarrow \Omega_2$  eine Zufallsvariable. Das Wahrscheinlichkeitsmaß  $P_X : \mathcal{A}_2 \rightarrow [0, 1]$  definiert durch

$$P_X(A) = P(X^{-1}(A))$$

heißt **Bildmaß** beziehungsweise **Verteilung** von  $X$ . Der Wahrscheinlichkeitsraum  $(\Omega_2, \mathcal{A}_2, P_X)$  heißt **Bildmodell** von  $(\Omega, \mathcal{A}, P)$  unter  $X$ . Wir schreiben

$$X \sim P_X$$

für „ $X$  hat die Verteilung  $P_X$ “.

**Zufallsvariablen und Ereignisse** Wir können nun Ereignisse auf dem Urbild- sowie auf dem Bildraum einer Zufallsvariable durch den von der Zufallsvariable angenommenen Wert beschreiben. Eine geläufige Schreibweise zur Beschreibung von Ereignissen auf dem *Urbildraum* mit Zufallsvariablen ist dabei

$$\{X \in A\} = X^{-1}(A),$$

wobei die Menge  $A$  ein Ereignis auf dem *Bildraum* ist. Meist ist dieses Ereignis durch eine gewisse Forderung an  $A$  definiert. Beispielsweise schreibt man auch kurz:

$$\{X \stackrel{\geq}{\leq} k\} = \left\{ X \in \{y \in \Omega_2 \mid y \stackrel{\geq}{\leq} k\} \right\}.$$

Eine Gleichheitsforderung entspricht gerade einem Elementarereignis des Bildraumes.

Die Menge  $\{X \in A\}$  ist eine Teilmenge des Urbildraumes  $\Omega_1$ . Wir nennen diese Menge die durch  $A$  *beschriebene* Menge. Jede Teilmenge von  $\Omega_1$ , die auf solche Weise durch eine Forderung an die Zufallsvariable beschrieben werden kann, heißt *durch  $X$  beschreibbar*.

Wir betrachten die Familie aller Elementarereignisse auf dem Bildraum

$$\text{Elem}_X = \{\{k\} | k \in \Omega_2\}.$$

Durch die Beschreibbarkeitsrelation induziert diese Mengenfamilie eine disjunkte Überdeckung des Bildraumes.

Das durch  $X$  auf  $\Omega_2$  bestimmte Wahrscheinlichkeitsmaß ist durch Rückführung auf das Maß  $P$  von  $\Omega_1$  definiert. Um daher  $P_X(A)$  zu bestimmen, müssen wir aktuell noch die Urbildmenge von  $A$  berechnen. Um uns diesen Umweg zu ersparen, wollen wir  $P_X$  durch eine eigene Dichte und Verteilungsfunktion ausdrücken.

**Verteilungen von Zufallsvariablen** Häufig ist lediglich die Verteilung beziehungsweise die dazugehörige Dichte einer Zufallsvariable von Interesse. Diese genügt bereits, um die Zufallsvariable vollständig zu beschreiben. Je nach Beschaffenheit des Bildraumes können Zufallsvariablen unterschiedlich klassifiziert werden. Dementsprechend kommen auch alle bereits vorgestellten Verteilungen für sie in Frage.

- Ist  $\Omega_2$  endlich oder abzählbar, so ist  $X$  eine *diskrete* Zufallsvariable. Als Verteilung kommen alle in Abschnitt ?? aufgeführten Modelle in Frage.
- Ist  $\Omega_2 = \mathbb{R}$  oder eine Borel-Teilmenge von  $\mathbb{R}$ , dann kann  $X$  durch die in Abschnitt ?? dargestellten stetigen Verteilungen beschrieben werden und heißt *stetige* Zufallsvariable. (Es gibt natürlich auch nicht stetige Verteilungen auf  $\mathbb{R}$ , aber diese sollen hier nicht behandelt werden.)

**Verkettung** Ist  $\Omega_X$  der Bildraum einer Zufallsvariable  $X$ , so kann natürlich auch eine weitere Zufallsvariable  $Y : \Omega_X \rightarrow \Omega_Y$  definiert werden, die  $\Omega_X$  als Urbildraum hat. Die Wahrscheinlichkeitsverteilung der neuen Zufallsvariable  $Y$  kann - analog zu oben - durch Rückführung auf  $X$  berechnet werden. Ist  $P_X$  das Bildmaß von  $X$ , dann ist

$$P_Y(A) = P_X(Y^{-1}(A))$$

das zu  $Y$  gehörige Bildmaß. Auf diese Weise können Zufallsvariablen beliebig verkettet werden.

Ein Spezialfall dessen ist, dass  $X$  und  $Y$  stetige Zufallsvariablen sind, und  $Y$  aus  $X$  durch Anwendung einer stetigen Funktion hervorgeht, also  $Y = f(X)$  gilt. Kennen wir Dichte  $f_X$  und Verteilungsfunktion  $F_X$  des Bildmaßes  $P_X$  von  $X$ , dann können wir auch  $P_Y$  durch eine Verteilungsfunktion  $F_Y$  definieren. Diese Verteilungsfunktion geht direkt aus  $F_X$  hervor:

**Lemma 6.2 (Transformation einer Zufallsvariable)**

Sei  $X : \Omega \rightarrow \mathbb{R}$  eine reelle, stetige Zufallsvariable mit Dichte  $f_X$ , und  $g \in C^0(\mathbb{R}, \mathbb{R})$  eine stetige Abbildung. Dann ist  $Y := g(X) : \Omega \rightarrow \mathbb{R}$  ebenfalls eine Zufallsvariable mit der Verteilungsfunktion

$$F_Y(y) = \int_{M(y)} f_X(x) dx \tag{6.1}$$

mit

$$M(y) = \{x \in \mathbb{R} | g(x) \leq y\}.$$

BEWEIS. Ist  $P_Y$  das Wahrscheinlichkeitsmaß von  $Y$ , dann ist nach Satz ??

$$F_Y(y) = P_Y((-\infty, y]) = P_X(g^{-1}((-\infty, y])) = P_X(M(y)) = \int_{M(y)} f_X(x) dx.$$

□

Ist die Verteilungsfunktion bekannt, so kann die Dichte daraus durch Differentiation errechnet werden; bekanntlich ist ja

$$f_Y(t) = \frac{d F_Y}{d y}(t)$$

Ist  $g : \mathbb{R} \rightarrow \mathbb{R}$  stetig differenzierbar und bijektiv, so erhalten wir für  $Y = g(X)$  die Verteilungsfunktion

$$F_Y(y) = \int_{-\infty}^{g^{-1}(y)} f_X(x) dx \quad (6.2)$$

sowie durch Differentiation mit der Kettenregel die Dichte

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{d y} g^{-1}(y). \quad (6.3)$$

Insbesondere wollen wir nun explizite Formeln für Dichten und Verteilungen von affin transformierten und quadrierten Zufallsvariablen betrachten. Dabei soll der bereits bekannten Normalverteilung ?? besondere Aufmerksamkeit zukommen.

### Satz 6.3 (Affine Transformation von Zufallsvariablen)

Sei  $X$  eine reelle Zufallsvariable mit Dichte  $f_X$  und Verteilungsfunktion  $F_X$ , und sei  $Y = a + b \cdot X$  mit  $a \in \mathbb{R}, b > 0$ . Dann ist auch  $Y$  eine reelle Zufallsvariable mit der Dichte

$$f_Y(y) = \frac{1}{b} f_X\left(\frac{y-a}{b}\right) \quad (6.4)$$

und der Verteilungsfunktion

$$F_Y(y) = F_X\left(\frac{y-a}{b}\right). \quad (6.5)$$

BEWEIS. Es ist  $bx + a < y \Leftrightarrow x < \frac{y-a}{b}$ . Mit Lemma ?? ist somit

$$F_Y(y) = \int_{-\infty}^{\frac{y-a}{b}} f_X(x) dx = F_X\left(\frac{y-a}{b}\right).$$

Die Dichte  $f_Y$  ist die Ableitung der Verteilungsfunktion  $F_Y$ ; also gilt nach dem Hauptsatz der Differential- und Integralrechnung

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

□

**Verschobene Normalverteilung** Wir können nun den eben bewiesenen Satz ?? auf die Normalverteilung anwenden. Sei also  $X$  eine Standard-Normalverteilte Zufallsvariable und  $Y = \sigma X + \mu$  mit  $\mu \in \mathbb{R}, \sigma > 0$ . Durch obigen Satz erhalten wir dann als Verteilungsfunktion für  $Y$

$$F_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right).$$

Nach Satz ?? wissen wir: Das ist gerade die Verteilungsfunktion der Normalverteilung  $\mathcal{N}(\mu, \sigma^2)$ ! Zum Zeitpunkt seiner Einführung erschien dieser Satz wohl ziemlich aus der Luft gegriffen. Hier wird nun seine tiefgreifendere Bedeutung deutlich. Bereits verdeutlicht wurde, dass die Parameter  $\mu$  und  $\sigma$  jeweils Mittelwert und Streuung der Normalverteilung beschreiben. Tatsächlich handelt es sich hierbei um nichts weiter als eine Skalierung und Verschiebung der Standardnormalverteilung, beziehungsweise einer standardnormalverteilten Zufallsvariable.

## 6.2 Mehrdimensionale Zufallsvariablen

In Abschnitt ?? haben wir die mehrstufigen Modelle kennengelernt. Dabei handelte es sich um Zufallsexperimente mit einem mehrdimensionalen Ausprägungsraum  $\Omega = \Omega_1 \times \dots \times \Omega_n$ . Wir können in einer allgemeineren Betrachtung solche Zufallsexperimente auch durch *mehrdimensionale Zufallsvariablen* beschreiben.

### Definition 6.4 (Mehrdimensionale Zufallsvariable)

Eine Zufallsvariable  $X = (X_1, \dots, X_N) : \Omega \rightarrow \Omega_1 \times \dots \times \Omega_n$  heißt *n-dimensionale Zufallsvariable*.

Bei der Betrachtung mehrdimensionaler Zufallsvariablen treten im Grunde zwei Aufgabenstellungen auf:

- Bei gegebener *n*-dimensionaler Zufallsvariable  $X = (X_1, \dots, X_N)$  mit bekannter Gesamtverteilung  $P_X$ , bestimme die Wahrscheinlichkeitsverteilungen  $P_i$  der Komponenten  $X_i$
- Bei gegebenen eindimensionalen Zufallsvariablen  $X_1, \dots, X_N$  mit bekannten Verteilungen  $P_1, \dots, P_n$ , bestimme die Verteilung der mehrdimensionalen Zufallsvariable  $X = (X_1, \dots, X_n)$ .

Das zweite Problem entspricht gerade der Frage nach der Kopplung der Zufallsvariablen.

### 6.2.1 Randverteilungen

Das erste Problem lässt sich relativ einfach lösen. Der Prozess, eine einzelne Dimension aus einem *n*-dimensionalen Konstrukt zu extrahieren, wird gemeinhin als *Projektion* bezeichnet. Die Projektion einer mehrdimensionalen Zufallsvariable auf eine ihrer Komponenten ist natürlich auch wieder eine Zufallsvariable. Nun interessieren wir uns für die Wahrscheinlichkeitsverteilung, der diese Projektion unterliegt. Diese wollen wir *Randverteilung* nennen:

### Definition 6.5 (Randverteilung)

Sei  $\Omega_X = \Omega_1 \times \dots \times \Omega_n$ , und  $X = (X_1, \dots, X_n) : \Omega \rightarrow \Omega_X$  eine *n*-dimensionale Zufallsvariable. Die Zufallsvariable  $X_i$  ist die Projektion von  $X$  auf ihre *i*-te Komponente. Ihre Verteilung  $P_i$  heißt *i-te Randverteilung* von  $X$ .

Ist die Verteilung  $P_X$  von  $X$  bekannt, so lassen sich die Randverteilungen leicht berechnen. Die Intuition dafür lässt sich anschaulich am Beispiel einer diskreten, zweidimensionalen Zufallsvariable entwickeln.

### Satz 6.6 (Berechnung der Randverteilung)

Sei  $\Omega_X = \Omega_1 \times \dots \times \Omega_n$ , und  $X : \Omega \rightarrow \Omega_X$  eine *n*-dimensionale Zufallsvariable mit Verteilung  $P_X$ .

- Ist  $\Omega_X$  endlich oder abzählbar unendlich, und  $f$  die zu  $P$  gehörige Zähldichte, dann hat die *i*-te Randverteilung die Dichte

$$f_{X_i}(x_i) = \sum_{\omega_1 \in \Omega_1} \dots \sum_{\omega_{i-1} \in \Omega_{i-1}} \sum_{\omega_{i+1} \in \Omega_{i+1}} \dots \sum_{\omega_n \in \Omega_n} f(\omega_1, \dots, \omega_{i-1}, x_i, \omega_{i+1}, \omega_n),$$

welche auch *i*-te Randdichte genannt wird.

- Ist  $\Omega_X = \mathbb{R}^n$  und  $f$  die zu  $P$  gehörige Riemann-Dichte, dann hat die *i*-te Randverteilung die Riemann-Dichte

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} f(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

BEWEIS. TBD. □

Die Randverteilung entspricht, wie sie hier vorgestellt wurde, der Projektion einer mehrdimensionalen Zufallsvariable auf eine ihrer Koordinatenachsen. Das Konzept lässt sich verallgemeinern. Wir können eine  $n$ -dimensionale Zufallsvariable auf jede beliebige Auswahl ihrer Komponenten projizieren.

## 6.2.2 Kopplung und Stochastische Unabhängigkeit

Im letzten Abschnitt haben wir untersucht, wie man aus einer bekannten mehrdimensionalen Zufallsvariable ihre Komponenten als eigene Zufallsvariablen extrahiert. Nun möchten wir, mit Rückbezügen auf Abschnitt ??, die Vereinigung mehrerer eindimensionaler Zufallsvariablen zu einer mehrdimensionalen Zufallsvariable betrachten, und in diesem Zusammenhang die Begriffe der stochastischen Bedingtheit und Unabhängigkeit für Zufallsvariablen präzisieren.

### Definition 6.7 (Gemeinsame Verteilung)

Seien  $X_i : \Omega \rightarrow \Omega_i$ ,  $i = 1, \dots, n$  Zufallsvariablen. Dann ist auch  $X = (X_1, \dots, X_n) : \Omega \rightarrow (\Omega_1, \dots, \Omega_n)$  eine Zufallsvariable. Die Wahrscheinlichkeitsverteilung  $P_X$  von  $X$  heißt **Gemeinsame Verteilung** der  $X_i$ . Die Dichte  $f$  von  $P_X$  heißt **gemeinsame Dichte**.

**Bemerkung 6.8** Offensichtlicherweise sind die  $i$ -ten Randverteilungen der Zufallsvariable  $X$  gerade die Verteilungen  $P_i$  der  $X_i$ .

Die gemeinsame Verteilung mehrerer Zufallsvariablen beschreibt im wesentlichen ein mehrstufiges Modell, dessen  $i$ -te Stufe durch die Zufallsvariable  $X_i$  beschrieben wird. Die Zufallsvariable  $X$  können wir daher als Kopplung der  $X_i$  betrachten. Zur Analyse dieser Kopplung können wir die bereits aus Abschnitt ?? bekannten Begriffe heranziehen.

Eine Erkenntnis der vorherigen Betrachtungen war, dass die einzelnen Stufen eines Zufallsexperiments sich gegenseitig beeinflussen können. Je nachdem, welche Art von Experiment wir modellieren, kann dies durch einen tatsächlichen kausalen Zusammenhang geschehen. Oder aber die verschiedenen Zufallsvariablen abstrahieren verschiedene Aspekte eines einzelnen Experiments. In beiden Fällen lässt das Wissen über die von  $X_1, \dots, X_{i-1}$  angenommenen Werte Schlüsse auf den wahrscheinlichen Wert von  $X_i$  zu. Um dieses Phänomen zu beschreiben, haben wir die Übergangsdichten eingeführt. Dieser Begriff soll nun für Zufallsvariablen konkretisiert werden.

### Definition 6.9 (Bedingte Dichten)

Seien  $X_i : \Omega \rightarrow \Omega_i$ ,  $i = 1, \dots, n$  gekoppelte Zufallsvariablen. Die Übergangsdichten des durch die  $X_i$  beschriebenen gekoppelten Modells

$$f_i^{i-1}(x_1, \dots, x_{i-1}; x_i)$$

heißen **bedingte Dichten**, und die durch sie definierten Wahrscheinlichkeitsmaße **bedingte Verteilungen**.

Die Definition der bedingten Dichten ist eng verwandt mit der bereits bekannten Definition der bedingten Wahrscheinlichkeit. Im Fall von diskreten Zufallsvariablen sind sie sogar äquivalent. Sind die  $\Omega_i$  endlich oder abzählbar unendlich, so gilt für die bedingte Wahrscheinlichkeit

$$P(X_i = x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) = f_i^{i-1}(x_1, \dots, x_{i-1}; x_i).$$

Im Fall von reellen Zufallsvariablen ist dies natürlich nicht der Fall. Die Intuition für bedingte Wahrscheinlichkeiten trifft hier dennoch zu. Die Berechnung der bedingten Dichten erfolgt nach demselben Prinzip.

Die im folgenden Satz verwendeten *verallgemeinerten Randdichten* werden analog zu Satz ?? berechnet. Allerdings wird hier nicht mehr nur auf die  $i$ -te Koordinatenvariable projiziert, sondern auf den gesamten Unterraum der Koordinatenvariablen  $1, \dots, i$ . Für Riemann-Dichten gilt also

$$f_{X_1, \dots, X_i}(x_1, \dots, x_i) = \int_{\mathbb{R}^{n-i}} f(x_1, \dots, x_n) dx_{i+1} \dots dx_n.$$

Für Zähldichten ist das Integral durch eine Summe zu ersetzen.

**Satz 6.10 (Berechnung der bedingten Dichte)**

Die gemeinsame Verteilung  $P_X$  der Zufallsvariablen  $X_1, \dots, X_n$  ist ein gekoppeltes Modell mit den Übergangsdichten

$$f_i^{i-1}(x_1, \dots, x_{i-1}; x_i) = \frac{f_{X_1, \dots, X_i}(x_1, \dots, x_i)}{f_{X_1, \dots, X_{i-1}}(x_1, \dots, x_{i-1})}.$$

BEWEIS. TBD. □

Analog zur bereits bekannten Definition können wir auch hier die stochastische Unabhängigkeit prägen.

**Definition 6.11 (Stochastische Unabhängigkeit von Zufallsvariablen)**

Die Zufallsvariablen  $X_i, i = 1, \dots, n$  heißen **stochastisch unabhängig**, wenn ihre gemeinsame Dichte  $f$  das Produkt der Randdichten  $f_{X_i}$  ist:

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i).$$

Andernfalls heißen sie **stochastisch abhängig**.

### 6.3 Summenverteilung und Faltung

Abschließend wollen wir uns noch der Summation mehrerer Zufallsvariablen widmen. Wir betrachten also für zwei Zufallsvariablen  $X$  und  $Y$  die Dichte und Verteilung ihrer Summe  $X + Y$ . Zunächst soll der allgemeine Fall möglicherweise stochastisch abhängiger Zufallsgrößen betrachtet werden. Anschließend diskutieren wir den Spezialfall der Faltung bei stochastischer Unabhängigkeit. Zuletzt möchten wir das Verhalten einiger spezieller diskreter und stetiger Verteilungen unter der Faltung genauer untersuchen.

#### 6.3.1 Die Summenverteilung

Um eine allgemeine Formel für die Summenverteilung genannte Verteilung von  $X + Y$  herzuleiten, stellen wir zunächst ein paar Vorüberlegungen an. Wir beschränken uns zunächst auf zwei ganzzahlige Zufallsvariablen  $X, Y : \Omega \rightarrow \mathbb{Z}$  mit bekannten Verteilungen  $P_X$  und  $P_Y$ .

Wir stellen uns die Frage: Für eine Zahl  $z \in \mathbb{Z}$ , wie groß ist die Wahrscheinlichkeit dafür, dass  $X + Y = z$  ist? Indem wir diese Frage für alle  $z$  beantworten, finden wir die Verteilung  $P_{X+Y}$ .

Wir fassen dafür die Zufallsvariablen zusammen zu einer mehrdimensionalen Variable  $(X, Y)$  und formulieren unsere Fragestellung als Ereignis:

$$S(z) = \{(x, y) \in \mathbb{Z}^2 | x + y = z\}.$$

Gesucht ist nun das Maß  $P(S(z))$ . Ein Ansatz zur Bestimmung dieser Verteilung wäre: Wenn der Wert  $x$  der Zufallsvariable  $X$  bekannt ist, welchen Wert  $y$  muss  $Y$  annehmen, damit  $x + y = z$  wird? Dies können wir mithilfe der Formel der totalen Wahrscheinlichkeit (??) ausdrücken, indem wir  $\mathbb{Z}^2$  nach  $x$  partitionieren:

$$P(S(z)) = \sum_{x \in \mathbb{Z}} P(X = x) \cdot P(S(z) | X = x). \tag{6.6}$$

Die bedingte Wahrscheinlichkeit  $P(S(z)|X = x)$  ist einfach bestimmt: Damit  $x + y = z$  ist, muss logischerweise  $y = z - x$  sein. Es existiert also genau ein  $y$ , für welches das Ereignis  $S(z)$  unter der Bedingung  $X = x$  eintritt, und es ist

$$P(S(z)|X = x) = P(Y = z - x).$$

Diese Erkenntnis möchten wir nun, unter Verwendung der gemeinsamen Zähl- bzw. Riemann-Dichte  $f_{(X,Y)}$ , als Satz formulieren.

**Satz 6.12 (Summenverteilung)**

Seien  $X, Y : \Omega \rightarrow M$  Zufallsvariablen mit gemeinsamer Dichte  $f_{(X,Y)}$ .

1. Sind  $X, Y$  ganzzahlig ( $M \subseteq \mathbb{Z}$ ), so besitzt die Zufallsvariable  $X + Y$  die Dichte  $f_{X+Y} : \mathbb{Z} \rightarrow [0, 1]$ ,

$$f_{X+Y}(z) = \sum_{x \in \mathbb{Z}} f_{(X,Y)}(x, z - x). \tag{6.7}$$

2. Sind  $X, Y$  reell ( $M \subseteq \mathbb{R}$ ) und  $f_{(X,Y)}$  eine Riemann-Dichte, dann hat die Zufallsvariable  $X + Y$  die Dichte  $f_{X+Y} : \mathbb{R} \rightarrow [0, 1]$ ,

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_{(X,Y)}(x, z - x) dx. \tag{6.8}$$

BEWEIS.

1. Siehe oben.
2. TBD.

□

**6.3.2 Faltung stochastisch unabhängiger Zufallsvariablen**

Sind die Zufallsvariablen  $X$  und  $Y$  *stochastisch unabhängig* und ihre gemeinsame Dichte somit das Produkt beider Dichten, handelt es sich bei der Summenverteilung um nichts anderes als die *Faltung* der beiden Variablen. Das Konzept der Faltung ist weit über die Stochastik hinaus relevant. Anschaulich bezeichnet die Faltung  $f * g$  einer Funktion  $f$  mit einer zweiten Funktion  $g$  die Bildung eines *gewichteten Mittelwertes* über die Werte von  $f$  an jedem Punkt  $x$ , wobei  $g$  die Gewichtung liefert.

Wir wollen hier nur die Faltung im eindimensionalen Raum betrachten.

**Definition 6.13 (Faltung)**

- Für zwei Funktionen  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  ist ihre **Faltung**  $f * g : \mathbb{R} \rightarrow \mathbb{R}$  definiert als

$$(f * g)(z) = \int_{-\infty}^{\infty} f(x)g(z - x) dx \tag{6.9}$$

- Für zwei diskrete Abbildungen  $F, G : \mathbb{Z} \rightarrow \mathbb{Z}$  ist ihre **Faltung**  $F * G : \mathbb{Z} \rightarrow \mathbb{Z}$  definiert als

$$(F * G)(z) = \sum_{x \in \mathbb{Z}} f(x)g(z - x) \tag{6.10}$$

Für stochastisch unabhängige Variablen  $X, Y$  können wir also aus dem Satz ?? folgern:

**Korollar 6.14 (Summenverteilung stochastisch unabhängiger Zufallsvariablen)**

Seien  $X, Y$  aus ?? stochastisch unabhängig, und  $f_X, f_Y$  ihre Dichten. Dann gilt für die Dichte  $f_{X+Y}$  der Zufallsvariable  $X + Y$

$$f_{X+Y} = f_X * f_Y.$$

BEWEIS. Für die gemeinsame Dichte gilt bei stochastischer Unabhängigkeit per Definition  $f_{(X,Y)} = f_X \cdot f_Y$ . Somit ist

$$(f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{\infty} f_{(X,Y)}(x, z-x) dx$$

□

**6.3.3 Faltung wichtiger Verteilungen**

Viele gebräuchliche, parametrisierte diskrete und stetige Wahrscheinlichkeitsverteilungen verändern unter der Faltung nicht ihre Gestalt, sondern lediglich ihre Parameter. Einige dieser Beziehungen möchten wir nun betrachten.

**Faltung der Binomialverteilung** Die Binomialverteilung  $B_{n,p}$  (??) beschreibt bekanntlich die  $n$ -fache Durchführung eines Bernoulli-Experimentes mit Erfolgswahrscheinlichkeit  $p$ . Wenn wir jede Durchführung des Bernoulli-Versuchs durch eine Zufallsvariable  $X_i : \Omega \rightarrow \{0, 1\}$  beschreiben, dann können wir die Binomialverteilung auch anders charakterisieren. Steht 0 für einen Misserfolg und 1 für einen Erfolg, so ist das Ereignis „ $k$  Erfolge“ gleichzusetzen mit  $\sum_{i=1}^n X_i = k$ . Da die Wahrscheinlichkeit dieses Ereignisses gerade der Wert  $B_{n,p}(k)$  ist, lässt sich die Binomialverteilung in natürlicher Weise als  $n$ -fache Faltung von  $n$  identisch Bernoulli-verteilten Zufallsvariablen schreiben:

$$B_{n,p} = \underbrace{B_p * \dots * B_p}_{n \text{ mal}}. \quad (6.11)$$

Sind  $X, Y$  zwei binomialverteilte Zufallsvariablen mit Verteilungen  $B_{n,p}$  und  $B_{m,p}$ , so ist damit auch ihre Summe wieder binomialverteilt, schließlich kann man sie als Faltung von  $m+n$  Bernoulli-Verteilungen schreiben. Es ist also

$$f_{X+Y} = B_{n,p} * B_{m,p} = B_{n+m,p}. \quad (6.12)$$

**Faltung der Poisson-Verteilung** Auch die Verteilung der Summe zweier stochastisch unabhängiger, Poisson-verteilter (??) Zufallsvariablen lässt sich sehr intuitiv berechnen. Sei  $X$   $\pi(\lambda)$ -verteilt und  $Y$   $\pi(\mu)$ -verteilt. Die Werte der Zufallsvariablen entsprechen dann jeweils der Beobachtung, dass ein gewisses Ereignis  $X$  beziehungsweise ein anderes Ereignis  $Y$  mal eingetreten ist.  $X + Y$  zählt dann praktisch die Häufigkeit beider Ereignisse. Da  $\lambda, \mu$  die zu erwartenden Häufigkeiten des jeweiligen Einzelereignisses beschreiben, ist es naheliegend, dass für das durch  $X + Y$  beschriebene kombinierte Ereignis die mittlere Häufigkeit  $\lambda + \mu$  erwartet werden kann. Tatsächlich ist die Verteilung von  $X + Y$  wieder eine Poisson-Verteilung, und wie erwartet addieren sich die Lambdas:

$$f_{X+Y} = \pi_\lambda * \pi_\mu = \pi_{\lambda+\mu} \quad (6.13)$$

Wir wollen das eben nachrechnen. Wir beginnen mit der Faltungsformel und erkennen, dass alle Summanden für  $k < 0$  und  $k > n$  wegfallen, da die Poisson-Dichte hier nicht definiert, also = 0 ist.

$$(\pi_\lambda * \pi_\mu)(n) = \sum_{k=0}^n \exp(-\lambda) \cdot \frac{\lambda^k}{k!} \cdot \exp(-\mu) \cdot \frac{\mu^{n-k}}{(n-k)!}$$

$$= \exp(-(\lambda + \mu)) \sum_{k=0}^n \frac{\lambda^k \mu^{n-k}}{k!(n-k)!}$$

An dieser Stelle sehen wir beinahe die Gestalt einer ausmultiplizierten binomischen Formel. Durch Erweitern mit  $n!$  können wir tatsächlich den Binomialkoeffizienten bilden und erhalten die gewünschte Form:

$$\begin{aligned} &= \exp(-(\lambda + \mu)) \sum_{k=0}^n \frac{n!}{k!(n-k)!} \frac{1}{n!} \lambda^k \mu^{n-k} \\ &= \exp(-(\lambda + \mu)) \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} \\ &= \exp(-(\lambda + \mu)) \frac{(\lambda + \mu)^n}{n!} = \pi_{\lambda+\mu}(n). \end{aligned}$$

**Faltung der Normalverteilung** Wir haben bereits gesehen, dass affine Transformationen normalverteilter (??) Zufallsvariablen wieder normalverteilt sind, allerdings mit anderem Mittelwert und Streuung. Es ist naheliegend, dass auch die Summe zweier stochastisch unabhängiger, normalverteilter Zufallsvariablen wieder normalverteilt ist. Auch die Kenngrößen  $\mu$  und  $\sigma$  der Summe ergeben sich auf natürliche Weise aus ihren Teilen.

Tatsächlich gilt für zwei normalverteilte Zufallsvariablen  $X$  mit  $\mathcal{N}(\mu, \sigma^2)$  und  $Y$  mit  $\mathcal{N}(\nu, \tau^2)$ , dass auch  $X + Y$  wieder normalverteilt ist mit  $\mathcal{N}(\mu + \nu, \sigma^2 + \tau^2)$ . Denn für die Dichten gilt:

$$\mathcal{N}(\mu, \sigma^2) * \mathcal{N}(\nu, \tau^2) = \mathcal{N}(\mu + \nu, \sigma^2 + \tau^2). \quad (6.14)$$

**Faltung der Gammaverteilung** Für die Faltung der Dichten zweier Gamma-Verteilungen mit dem selben Formparameter  $\alpha$  gilt die Formel

$$\Gamma_{\alpha, \nu} * \Gamma_{\alpha, \mu} = \Gamma_{\alpha, \nu + \mu}. \quad (6.15)$$

Für den Spezialfall der Exponentialverteilung ( $\nu = 1$ ) folgt entsprechend, dass die Faltung zweier Exponentialverteilungen mit gleichem Parameter  $\alpha$  gerade die Gammaverteilung  $\Gamma_{\alpha, 2}$  ergibt.

### 6.3.4 Zufällige Summen

Eine zufällige Summe ist eine Zufallsvariable, welche als Summe über eine *zufällige Anzahl* gleichartiger Zufallsvariablen definiert ist.

#### Definition 6.15 (Zufällige Summe)

Sei  $Y : \Omega \rightarrow \mathbb{N}_0$  eine positive ganzzahlige Zufallsvariable. Seien außerdem  $X_i, i = 1, 2, \dots$  identisch verteilte, reellwertige und untereinander sowie von  $Y$  stochastisch unabhängige Zufallsvariablen. Dann heißt die Zufallsvariable

$$S := \sum_{i=1}^Y X_i$$

**zufällige Summe.**

Unsere Erkenntnisse über die Anwendung der Faltung zur Bestimmung der Verteilung einer Summe von Zufallsvariablen können auf eine zufällige Summe nicht ohne weiteres angewendet werden. Allerdings können wir die Wahrscheinlichkeit  $P(S = x)$  mithilfe der Formel der totalen Wahrscheinlichkeit (??) schreiben als

$$P(S = k) = \sum_{j=0}^{\infty} P(Y = j) \cdot P(S = k | Y = j) \quad (6.16)$$

mit der bedingten Wahrscheinlichkeit

$$P(S = k|Y = j) = \sum_{i=1}^j X_i \quad (6.17)$$

welche wir - da die obere Summengrenze nun fest ist - mittels Faltung der Verteilung der  $X_i$  bestimmen können. Haben die  $X_i$  die Dichte  $f_X$ , so erhalten wir für die Summe (??) die bedingte Dichte

$$f_{(S|Y=j)} = \underbrace{(f_X * \cdots * f_X)}_{j \text{ mal}}. \quad (6.18)$$

Mit der Zähldichte  $f_Y$  von  $Y$  erhalten wir dann die Gesamtdichte  $f_S$  als

$$f_S(k) = \sum_{j=0}^{\infty} f_Y(j) f_{(S|Y=j)}(k). \quad (6.19)$$

Je nachdem welcher Verteilung  $Y$  gehorcht verschwinden häufig alle bis auf endlich viele der Summanden. Auch die Faltung lässt sich für viele gebräuchliche Verteilungen leicht induktiv berechnen.

## 7 Kenngrößen von Verteilungen

Bereits eingangs hatten wir in Abschnitt ?? verschiedene Kenngrößen für erhobene Daten eingeführt. Die dortigen Definitionen lassen sich auf Zufallsvariablen übertragen. Tatsächlich lassen sich statistisch erfasste Daten als eine Sammlung von Werten einer Zufallsvariablen ansehen, und die bekannten statistischen Kenngrößen als eine empirische Annäherung an die tatsächlichen Kenngrößen der ausgewerteten Zufallsvariablen.

Wir möchten zunächst die Begriffe Erwartungswert, Streuung und Varianz für Zufallsvariablen präzisieren und beispielhaft für einige bekannte Verteilungen bestimmen. Außerdem sollen Rechenregeln für diese Größen hergeleitet werden.

Im Folgenden soll allgemein eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  mit Dichte  $f_X$  und Verteilungsfunktion  $F_X$  betrachtet werden. Der Sonderfall einer diskreten Zufallsvariable ( $\Omega$  abzählbar) ist dabei mitinbegriffen. Definitionen und Sätze werden für reelle Zufallsvariablen eingeführt. In der Regel übertragen sich ihre Aussagen auf diskrete Zufallsvariablen.

Der Autor möchte seinen Ärger darüber zum Ausdruck bringen, dass dieser Abschnitt wesentlich kürzer hätte gefasst werden können, wäre der Begriff des Lebesgue-Integrals bekannt gewesen. Aber wir machen hier schließlich Ingenieursmathematik, also was soll's.

### 7.1 Der Erwartungswert einer Zufallsvariable

Als Erwartungswert einer Zufallsvariable wird jener Wert bezeichnet, welchen die Zufallsvariable ihrer Verteilung entsprechend im Mittel annimmt. Dabei muss die Variable diesen Wert nicht zwingend annehmen können. Beschreiben wir beispielsweise eine faire Münze als Bernoulli-verteilte Zufallsvariable mit  $p = 0.5$ , so ist ihr Erwartungswert gerade 0.5 - obwohl dies natürlich keine Ausprägung des Experiments „Münzwurf“ sein kann.

In der beschreibenden Statistik entspricht der Erwartungswert dem Mittelwert einer Datenreihe. Dieser war definiert als

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{k=1}^n x_k$$

über einen  $n$ -dimensionalen Datenvektor  $\mathbf{x}$ . Es erscheint intuitiv, diese Definition auf Zufallsvariablen zu übertragen. Dies gelingt leider nicht ohne weiteres:

- Ist  $\Omega$  endlich, so lässt sich obige Definition ohne weiteres anwenden.
- Ist  $\Omega$  jedoch unendlich, so wird aus der Summe entweder eine Reihe oder ein Integral. Deren Grenzwerte müssen jedoch nicht notwendigerweise existieren!

Im Folgenden stellen wir fest, wann der Erwartungswert tatsächlich existiert, und wie wir ihn berechnen können.

#### 7.1.1 Definition und Existenz

##### Definition 7.1 (Erwartungswert)

Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable mit der Dichte  $f_X$ .

- Ist  $\Omega = \mathbb{R}$  und  $f_X$  eine Riemann-Dichte, so ist der Erwartungswert  $\mathbb{E}(X)$  der Zufallsvariablen definiert durch

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot f_X(x) \, dx. \quad (7.1)$$

- Ist  $\Omega$  abzählbar und  $f_X$  eine Zähldichte mit Träger  $T$ , so ist der Erwartungswert  $\mathbb{E}(X)$  der Zufallsvariablen definiert durch

$$\mathbb{E}(X) = \sum_{k \in T} k \cdot f_X(k). \quad (7.2)$$

Diese Definition kann natürlich nur Anwendung finden, wenn das beteiligte Integral beziehungsweise die Summe auch existiert, also einen Wert  $c \in \mathbb{R} \cup \{\pm\infty\}$  annimmt.

Wir betrachten zunächst die Existenz des Erwartungswertes einer stetigen Zufallsvariable, also  $X$  mit Riemann-Dichte  $f_X$ . Um das uneigentliche Integral  $\int_{-\infty}^{\infty} x f_X(x) dx$  zu berechnen, müssen wir es sowieso in zwei Teilintegrale aufteilen. Dafür teilen wir auch  $X$  in einen Positiv- und einen Negativteil auf. Wir nennen

$$X^+ = \max(0, X)$$

*Positivteil* und

$$X^- = \max(0, -X)$$

*Negativteil* von  $X$ . Offensichtlich gilt:  $X = X^+ - X^-$ . Dann setzen wir

$$\mathbb{E}(X^+) = \int_0^{\infty} x \cdot f_X(x) dx = \lim_{s \rightarrow \infty} \int_0^s x \cdot f_X(x) dx$$

und analog

$$\mathbb{E}(X^-) = \int_{-\infty}^0 -x \cdot f_X(x) dx = \lim_{s \rightarrow -\infty} \int_s^0 -x \cdot f_X(x) dx.$$

Es sei festgehalten, dass diese beiden Integrale keine echten Erwartungswerte sind. Das selbe gilt für eine diskrete Zufallsvariable  $X$  mit Zähldichte  $f_X$  und Träger  $T$ . Hierbei setzen wir stattdessen

$$\mathbb{E}(X^+) = \sum_{k \in T, k \geq 0} k \cdot f_X(k), \quad \mathbb{E}(X^-) = \sum_{k \in T, k < 0} -k \cdot f_X(k).$$

Der Erwartungswert lässt sich nun schreiben als

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

Er existiert also genau dann, wenn sowohl  $\mathbb{E}(X^+)$  als auch  $\mathbb{E}(X^-)$  existieren und mindestens einer von beiden Ausdrücken einen endlichen Wert  $c < \infty$  annimmt. Sind beide Ausdrücke und somit auch  $\mathbb{E}(X)$  endlich, so heißt  $X$  *integrierbar*.

### 7.1.2 Rechnen mit Erwartungswerten

Die Verteilung einer Zufallsvariablen geht bekanntlich auf die Wahrscheinlichkeitsverteilung ihres Urbildraumes zurück. In der Definition des Erwartungswertes wird dieser per Summation über alle möglichen Ausprägungen von  $X$  bestimmt. Jede dieser Ausprägungen beschreibt gleichzeitig ein im allgemeinen nichtelementares Ereignis des Urbildraumes. Der Urbildraum wird durch die so beschriebenen Ereignisse disjunkt überdeckt.

Ist die Verteilung einer diskreten Zufallsvariable  $X$  unbekannt, aber ihre Urbildmaß  $P$  beziehungsweise dessen Dichte  $f_P$  bekannt, so können wir den Erwartungswert von  $X$  durch Rückführung auf das Urbildmaß bestimmen.

#### Lemma 7.2

Seien  $(\Omega_1, \mathcal{A}_1, P)$  ein abzählbarer Wahrscheinlichkeitsraum,  $\Omega_2 \subset \mathbb{R}$  ein abzählbarer Messraum und  $X : \Omega_1 \rightarrow \Omega_2$  eine diskrete Zufallsvariable. Dann gilt für den Erwartungswert von  $X$ :

$$\mathbb{E}(X) = \sum_{\omega \in \Omega_1} X(\omega) P(\omega) \tag{7.3}$$

falls diese Summe existiert.

BEWEIS. Ist  $P_X$  das Bildmaß von  $X$ , dann ist

$$f_X(k) = P_X(k) = P(X^{-1}(k)) = \sum_{\omega \in X^{-1}(k)} P(\{\omega\}).$$

Substituiere nun in (??) in jedem Summanden

$$k \cdot f_X(k) = \sum_{\omega \in X^{-1}(k)} k \cdot P(\{\omega\}) = \sum_{\omega \in X^{-1}(k)} X(\omega) \cdot P(\{\omega\}).$$

□

Auch der Erwartungswert einer transformierten Zufallsvariable  $Y = g(X)$  lässt sich, im diskreten wie auch im stetigen Fall, auf  $\mathbb{E}(X)$  zurückführen.

**Lemma 7.3 (Erwartungswert unter Transformation)**

Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable mit Dichte  $f_X$ , und  $g : \mathbb{R} \rightarrow \mathbb{R}$  eine integrierbare Abbildung.

- Ist  $\Omega = \mathbb{R}$  und  $f_X$  eine Riemann-Dichte, so ist

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \, dx \tag{7.4}$$

- Ist  $\Omega$  abzählbar,  $\Omega_X$  der Bildraum von  $X$  und  $f_X$  eine Zähldichte, so ist

$$\mathbb{E}(g(X)) = \sum_{k \in \Omega_X} g(k) \cdot f_X(k) \tag{7.5}$$

BEWEIS. Durch Rückführung der Dichte von  $g(X)$  auf  $f_X$ . □

Wir können auch zwei verschiedene Zufallsvariablen  $X$  und  $Y$  durch eine skalarwertige Abbildung  $h(X, Y)$  verknüpfen und den Erwartungswert der resultierenden Zufallsvariable bestimmen:

**Lemma 7.4**

Sei  $f_{(X,Y)}$  die gemeinsame Dichte zweier reeller Zufallsvariablen  $X$  und  $Y$ ,  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  und  $h \cdot f_{(X,Y)}$  integrierbar. Dann ist

$$\mathbb{E}(h(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f_{(X,Y)}(x, y) \, dy \, dx. \tag{7.6}$$

BEWEIS. TBD.

Ziemlich nützlich - auch im Beweis nachfolgender Aussagen - ist der Erwartungswert einer Einpunktverteilten Zufallsvariable.

**Lemma 7.5 (Erwartungswert der Einpunktverteilung)**

Sei  $X$  im Punkt  $a \in \mathbb{R}$  einpunkt-verteilt. Dann ist  $\mathbb{E}(X) = a$ .

BEWEIS.  $\sum_{k \in \{a\}} kP(k) = a \cdot 1 = a$ . □

Auch einige grundlegende algebraische Eigenschaften des Erwartungswertes möchten wir beleuchten:

**Lemma 7.6 (Eigenschaften des Erwartungswertes)**

Seien  $X, Y, X_1, \dots, X_N$  reellwertige Zufallsvariablen. Vorbehaltlich der Existenz beteiligter Ausdrücke gilt für die Erwartungswerte:

1. Monotonie:

$$X \leq Y \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y) \quad (7.7)$$

sowie

$$a \leq X \leq b \Rightarrow a \leq \mathbb{E}(X) \leq b \quad (a, b \in \mathbb{R}). \quad (7.8)$$

2. Linearität:

$$\mathbb{E}(a \cdot X + b) = a \cdot \mathbb{E}(X) + b \quad (a, b \in \mathbb{R}),$$

sowie, falls existent,

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

BEWEIS.

1. Sei  $T_X$  der Träger von  $X$  und  $T_Y$  der Träger von  $Y$ . Sei  $X < Y$ , also  $x < y$  für alle  $x \in T_X, y \in T_Y$ . Sei außerdem  $\tilde{x} = \sup(T_X), \tilde{y} = \inf(T_Y)$ . Insbesondere ist dann  $\tilde{x} \leq \tilde{y}$ . Es ergibt sich

$$\mathbb{E}(X) = \int_{T_X} x f_X(x) dx \leq \int_{T_X} \tilde{x} f_X(x) dx = \tilde{x} \leq \tilde{y} = \int_{T_Y} \tilde{y} f_Y(y) dY \leq \int_{T_Y} y f_Y(y) dY = \mathbb{E}(Y).$$

Die Aussage für Zähldichten und Kombinationen ergibt sich, wenn wir eines der Integrale (oder beide) durch Summen ersetzen. Insbesondere folgt auch (??), wenn man  $a$  und  $b$  als Erwartungswerte Einpunkt-verteilter Zufallsvariablen ansetzt.

2. Folgt direkt aus der Linearität der Summe beziehungsweise des Integrals. □

Natürlich verdient auch der Erwartungswert im Fall stochastischer Unabhängigkeit eine besondere Behandlung.

### Satz 7.7

Seien  $X, Y$  reelle, stochastisch unabhängige Zufallsvariablen mit endlichen Erwartungswerten. Dann ist

$$\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y). \quad (7.9)$$

BEWEIS. Mit  $h(x, y) = x \cdot y$  können wir Lemma ?? anwenden und erhalten, unter Ausnutzung der stochastischen Unabhängigkeit:

$$\begin{aligned} \mathbb{E}(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f_X(x) f_Y(y) dy dx \\ &= \int_{-\infty}^{\infty} x f_X(x) \cdot \left( \int_{-\infty}^{\infty} y f_Y(y) dy \right) dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx \cdot \int_{-\infty}^{\infty} y f_Y(y) dy = \mathbb{E}(X) \mathbb{E}(Y). \end{aligned}$$

Für diskrete Zufallsvariablen folgt die Aussage analog. □

Auch für bedingte Verteilungen kann der Erwartungswert berechnet werden.

### Definition 7.8 (Bedingter Erwartungswert)

Seien  $X, Y$  reelle Zufallsvariablen mit der bedingten Dichte  $f_{(X|Y)}$ . Dann heißt

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{(X|Y)}(y; x) dx$$

**bedingter Erwartungswert** von  $X$  unter der Bedingung  $Y$ .

**Lemma 7.9 (Formel des iterierten Erwartungswertes)**

Seien  $X, Y$  reelle Zufallsvariablen,  $f_Y$  die Dichte von  $Y$  und die bedingten Erwartungswerte  $\mathbb{E}(X|Y = y)$  bekannt. Dann gilt für den Erwartungswert von  $X$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} f_Y(y) \mathbb{E}(X|Y = y) \, dy. \quad (7.10)$$

BEWEIS. Es ist

$$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) f_{(X|Y)}(y; x) \, dy$$

und somit

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f_X(x) \, dx \\ &= \int_{-\infty}^{\infty} x \left( \int_{-\infty}^{\infty} f_Y(y) f_{(X|Y)}(y; x) \, dy \right) \, dx \\ &= \int_{-\infty}^{\infty} f_Y(y) \left( \int_{-\infty}^{\infty} x f_{(X|Y)}(y; x) \, dx \right) \, dy \\ &= \int_{-\infty}^{\infty} f_Y(y) \mathbb{E}(X|Y = y) \, dy. \end{aligned}$$

□

Für diskrete Zufallsvariablen gelten analog die Formeln für den bedingten und den iterierten Erwartungswert, mit Summen anstatt von Integralen. Für  $X : \Omega \rightarrow \Omega_X, Y : \Omega \rightarrow \Omega_Y$  diskret ist der bedingte Erwartungswert gegeben als

$$\mathbb{E}(X|Y = y) = \sum_{x \in \Omega_X} x \cdot P(X = x|Y = y), \quad (7.11)$$

und der iterierte Erwartungswert ist

$$\mathbb{E}(X) = \sum_{y \in \Omega_Y} P(Y = y) \mathbb{E}(X|Y = y). \quad (7.12)$$

Die Formeln für diskrete und stetige Zufallsvariablen lassen sich natürlich auch mischen. Der iterierte Erwartungswert findet unter anderem Anwendung bei der Berechnung des Erwartungswertes einer zufälligen Summe.

**Satz 7.10 (Erwartungswert zufälliger Summen)**

Sei  $Y$  eine Zufallsvariable mit Werten aus  $\mathbb{N}_0$ , sowie  $X_1, \dots, X_n$  identisch verteilt sowie stochastisch unabhängig untereinander und von  $Y$ . Sei  $S := \sum_{i=1}^Y X_i$  eine zufällige Summe. Dann ist

$$\mathbb{E}(S) = \mathbb{E}(Y) \cdot \mathbb{E}(X_i). \quad (7.13)$$

BEWEIS. Sei  $S_k := \sum_{i=1}^k X_i$  die  $k$ -te Teilsumme von  $S$ . Durch Iteration des Erwartungswertes (??) schreiben wir  $\mathbb{E}(S)$  als

$$\mathbb{E}(S) = \sum_{k=0}^{\infty} \mathbb{E}(S|Y = k) P(Y = k).$$

Wir substituieren  $\mathbb{E}(S|Y = k) = \mathbb{E}(S_k) = k \mathbb{E}(X_1)$  und erhalten

$$\mathbb{E}(S) = \mathbb{E}(X_1) \sum_{k=0}^{\infty} k P(Y = k) = \mathbb{E}(X_1) \mathbb{E}(Y).$$

□

## 7.2 Varianz und Standardabweichung

### Definition 7.11 (Varianz, Standardabweichung)

Sei  $X$  eine reelle Zufallsvariable mit endlichem Erwartungswert.

- Die **Varianz** von  $X$  ist definiert als

$$\text{Var}(X) := \mathbb{E} \left( (X - \mathbb{E}(X))^2 \right).$$

- Die **Standardabweichung** oder **Streuung** von  $X$  ist definiert als

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

Sowohl Varianz als auch Standardabweichung sind Maße für die mittlere Abweichung einer Zufallsvariable von ihrem Erwartungswert. Die Varianz wird häufig mit  $\sigma^2$  abgekürzt.

Setzen wir die Definition des Erwartungswertes ein und nutzen Lemma ??, so erhalten wir die Formel

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f_X(x) dx \quad (7.14)$$

für stetige beziehungsweise

$$\text{Var}(X) = \sum_{k \in T} (k - \mathbb{E}(X))^2 f_X(k) \quad (7.15)$$

für diskrete Zufallsvariablen. Durch Ausmultiplizieren und Anwendung der Linearität des Erwartungswertes erhalten wir außerdem die wesentlich einfachere Berechnungsvorschrift

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2. \quad (7.16)$$

Auch für die Varianz möchten wir einige Rechenregeln herleiten.

### Lemma 7.12 (Rechenregeln für die Varianz)

Für eine Zufallsvariable  $X$  und  $a, b \in \mathbb{R}$  gelten:

- $X$  ist genau dann konstant ( $X = a$ ), wenn die Varianz verschwindet:

$$\text{Var}(X) = 0 \Leftrightarrow P(X = a) = 1$$

- Die Varianz ist invariant unter Verschiebung, und ändert sich quadratisch unter Skalierung:

$$\text{Var}(aX + b) = a^2 \text{Var}(X) \quad \text{sowie} \quad \sigma(aX + b) = |a| \sigma(X).$$

BEWEIS. TBD.

### Satz 7.13

Seien  $X, Y$  stochastisch unabhängig. Dann ist

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

BEWEIS. Für  $X, Y$  stochastisch unabhängig gilt  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ . Mit (??) ergibt sich

$$\text{Var}(X + Y) = \mathbb{E}((X + Y)^2) - (\mathbb{E}(X + Y))^2$$

$$\begin{aligned}
&= \mathbb{E}(X^2 + 2XY + Y^2) - \mathbb{E}(X)^2 - 2\mathbb{E}(XY) - \mathbb{E}(Y)^2 \\
&= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - 2\mathbb{E}(XY) - \mathbb{E}(Y)^2 \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \\
&= \text{Var}(X) + \text{Var}(Y).
\end{aligned}$$

□

### 7.3 Kovarianz und Korrelation

#### Definition 7.14 (Kovarianz)

Für zwei reelle Zufallsvariablen  $X$  und  $Y$  mit endlichen Erwartungswerten ist die **Kovarianz** definiert als

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y)))$$

Auch für die Kovarianz können wir durch Ausmultiplizieren der Definition eine einfachere Darstellung erhalten:

$$\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y). \quad (7.17)$$

Direkt aus diesen Formeln lassen sich einige Rechenregeln herleiten.

#### Lemma 7.15 (Rechenregeln der Kovarianz)

Für Zufallsvariablen  $X, X_1, X_2, Y$  gelten, unter der Voraussetzung der Existenz aller Teilausdrücke, die Regeln:

1. Die Kovarianz ist symmetrisch und linear in beiden Argumenten:

$$\text{Cov}(X, Y) = \text{Cov}(Y, X) \quad (7.18)$$

und

$$\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y) \quad (7.19)$$

sowie

$$\text{Cov}(Y, X_1 + X_2) = \text{Cov}(Y, X_1) + \text{Cov}(Y, X_2). \quad (7.20)$$

2. Die Varianz einer Zufallsvariable ist ihre Kovarianz mit sich selbst:

$$\text{Cov}(X, X) = \text{Var}(X). \quad (7.21)$$

3. Die Kovarianz zweier stochastisch unabhängiger Zufallsvariablen verschwindet:

$$X, Y \text{ stoch. unabhängig} \Rightarrow \text{Cov}(X, Y) = 0. \quad (7.22)$$

BEWEIS.

1. Die Symmetrie folgt direkt aus der Definition aufgrund der Kommutativität der Multiplikation. Die Linearität in beiden Argumenten folgt ebenso durch Einsetzen der Definition und Ausmultiplizieren.
2. Gleichung (??) folgt durch

$$\text{Cov}(X, X) = \mathbb{E}((X - \mathbb{E}(X)) \cdot (X - \mathbb{E}(X))) = \mathbb{E}((X - \mathbb{E}(X))^2) = \text{Var}(X).$$

3. Für stochastisch unabhängige Zufallsvariablen folgt (??) aus (??) durch Anwendung von Satz ??.

□

Für einen Vektor  $\mathbf{Y} = (Y_1, \dots, Y_n)$  von  $n$  reellen Zufallsvariablen können wir jeweils paarweise die Kovarianzen der Komponenten bestimmen und erhalten so eine Matrix.

**Definition 7.16 (Kovarianzmatrix)**

Sei  $\mathbf{Y} = (Y_1, \dots, Y_n)$  ein  $n$ -dimensionaler reeller Zufallsvektor. Die Matrix  $\mathbf{K} \in \mathbb{R}^{n \times n}$  mit den Einträgen

$$k_{ij} := \text{Cov}(Y_i, Y_j) \quad (7.23)$$

heißt **Kovarianzmatrix** der Zufallsvariablen  $Y_1, \dots, Y_n$ .

Die Kovarianzmatrix ist symmetrisch positiv semidefinit. Sind die Zufallsvariablen  $Y_i$  linear unabhängig (siehe ??) und nicht Einpunkt-Verteilt, so ist  $\mathbf{K}$  auch positiv definit.

Als Maß für die (lineare) Abhängigkeit zweier Zufallsvariablen führen wir außerdem dem Korrelationskoeffizienten ein.

**Definition 7.17 (Korrelationskoeffizient)**

Für zwei reelle Zufallsvariablen  $X$  und  $Y$  mit endlichen Erwartungswerten heißt

$$\text{korr}(X, Y) = \rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

**Korrelationskoeffizient** von  $X$  und  $Y$ . Im Fall  $\sigma(X) = 0$  oder  $\sigma(Y) = 0$  ist er als  $\rho(X, Y) = 0$  festgelegt. Ist  $\rho(X, Y) = 0$ , so heißen  $X$  und  $Y$  **unkorreliert**.

Insbesondere sind stochastisch unabhängige Zufallsvariablen immer unkorreliert; die Umkehrung gilt allerdings nicht! Unkorrelierte Größen können somit immernoch voneinander abhängen.

Der Korrelationskoeffizient ist effektiv die durch die Varianzen normierte Kovarianz. Im Einklang damit gilt

$$-1 \leq \rho(X, Y) \leq 1. \quad (7.24)$$

**Satz 7.18 (Lineare Abhängigkeit)**

Die Zufallsvariablen  $X$  und  $Y$  sind genau dann linear abhängig, wenn  $\rho(X, Y) = \pm 1$  gilt. In diesem Fall existieren  $a, b \in \mathbb{R}$  mit  $\text{sgn}(a) = \text{sgn}(\rho(X, Y))$  sodass

$$Y = aX + b.$$

BEWEIS. TBD.

## 7.4 Gesetze der großen Zahlen

### 7.4.1 Zufall und Konvergenz

### 7.4.2 Abschätzungen von Wahrscheinlichkeiten

Eine wichtige und sehr praktische Abschätzung für die Wahrscheinlichkeiten einer Zufallsvariablen ist durch die Markov-Ungleichung, beziehungsweise ihre Instanzen gegeben.

**Satz 7.19 (Markow-Ungleichung)**

Sei  $X$  eine reelle Zufallsvariable,  $a \in \mathbb{R}$  eine Konstante und  $h : D \rightarrow [0, \infty]$  eine monoton wachsende Funktion auf  $D \subseteq \mathbb{R}$ . Dann gilt die Ungleichung

$$h(a)P(X \geq a) \leq \mathbb{E}(h(X)) \quad (7.25)$$

beziehungsweise für  $h(a) > 0$

$$P(X \geq a) \leq \frac{\mathbb{E}(h(X))}{h(a)} \quad (7.26)$$

BEWEIS. Für  $a \in \mathbb{R}$  betrachten wir das Ereignis  $A = \{X \geq a\}$  und seine Indikatorvariable

$$I_A := \begin{cases} 1 & X \geq a \\ 0 & \text{sonst} \end{cases}.$$

Dann ist

$$h(a)I_A \leq h(X),$$

denn: Ist  $X < a$ , so wird die linke Seite Null und die rechte Seite ist aufgrund der Forderungen an  $h$  nichtnegativ. Ist  $X \geq a$ , so bleibt auch  $h(X) \geq h(a)$  aufgrund der Monotonie von  $h$ . Anwendung des Erwartungswertes liefert

$$h(a)\mathbb{E}(I_A) \leq \mathbb{E}(h(X))$$

und mit der Substitution

$$\mathbb{E}(I_A) = 1 \cdot P(X \geq a) + 0 \cdot P(X < a)$$

erhalten wir

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

die gewünschte Ungleichung (??). Daraus ergibt sich, falls  $h(a)$  echt positiv ist, durch Division die Ungleichung (??).  $\square$

Aus der Markow-Ungleichung folgt unter anderem die Ungleichung von Tschebyscheff. Sie setzt die Wahrscheinlichkeit einer Abweichung vom Erwartungswert in Beziehung mit der Varianz einer Zufallsvariable. Die Aussage kann wie folgt in Worte gefasst werden: Umso größer die Varianz einer Zufallsvariable ist, umso größer ist auch die Wahrscheinlichkeit einer starken Abweichung vom Erwartungswert. Dies erscheint intuitiv.

**Satz 7.20 (Ungleichung von Tschebyscheff)**

Sei  $X$  eine reelle Zufallsvariable mit endlichem Erwartungswert  $\mu := \mathbb{E}(X)$  und endlicher Varianz  $\sigma^2 = \text{Var}(X)$ . Dann gilt für alle  $a > 0$  die Ungleichung

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2} \quad (7.27)$$

beziehungsweise für das Gegenereignis

$$P(|X - \mu| < a) \geq 1 - \frac{\sigma^2}{a^2} \quad (7.28)$$

BEWEIS. Mit der verschobenen Zufallsvariable  $Y := |X - \mu|$  und der Funktion

$$h : [0, \infty) \rightarrow [0, \infty), \quad h(x) = x^2$$

ergibt sich (??) aus der Markow-Ungleichung (??) als

$$P(Y \geq a) = P(|X - \mu| \geq a) \leq \frac{\mathbb{E}(|X - \mu|^2)}{a^2} = \frac{\sigma^2}{a^2}.$$

Die Darstellung (??) folgt daraus direkt durch Anwendung der Formel für das Gegenereignis

$$P(|X - \mu| < a) = 1 - P(|X - \mu| \geq a) \geq 1 - \frac{\sigma^2}{a^2}.$$

□

### 7.4.3 Die Gesetze der großen Zahlen und der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz ist eines der bedeutendsten Ergebnisse der Stochastik. Er liefert die Erklärung dafür, dass sich bei der Überlagerung vieler kleiner, gleichartiger und stochastisch unabhängiger Zufallsereignisse in der Regel eine Normalverteilung ergibt. Diese einzelnen Beobachtungen sollen hier als eine (unendliche) Folge von Zufallsvariablen  $X_1, X_2, \dots$  modelliert werden, welche alle stochastisch unabhängig sind und derselben Wahrscheinlichkeitsverteilung gehorchen. Wir sagen: Die  $X_i$  sind *stochastisch unabhängig identisch* verteilt (kurz *st. u. id.*, englisch *independent and identically distributed, i.i.d.*). Außerdem sollen ihr Erwartungswert  $\mu := \mathbb{E}(X)$  und ihre Varianz  $\sigma^2 := \text{Var}(X)$  endlich sein. Die Kumulation der Teilereignisse können wir als Summe über die Zufallsvariablen darstellen. Wir nennen dazu

$$S_n := \sum_{i=1}^n X_i \tag{7.29}$$

die  $n$ -te Teilsumme der Ereignisfolge. Diese Teilsumme hat den Erwartungswert  $\mathbb{E}(S_n) = n\mu$  und die Varianz  $\text{Var}(S_n) = n\sigma^2$ , wie sich aus den Rechenregeln für Erwartungswert und Varianz leicht ergibt:

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{E}(X_i) = n\mu \tag{7.30}$$

sowie

$$\text{Var}(S_n) = \text{Var}\left(\sum_{i=1}^n X_i\right) \stackrel{\text{st.u.}}{=} \sum_{i=1}^n \text{Var}(X_i) = n\sigma^2. \tag{7.31}$$

Damit können wir die Teilsummen normalisieren als

$$Z_n := \frac{S_n - n\mu}{\sqrt{n}\sigma}. \tag{7.32}$$

Der zentrale Grenzwertsatz besagt nun, dass für  $n \rightarrow \infty$  sich die Verteilung dieser normalisierten Teilsummen der Standardnormalverteilung annähert.

#### Satz 7.21 (Zentraler Grenzwertsatz)

Sei  $X_1, X_2, \dots$  eine Folge stochastisch unabhängiger, identisch verteilter reeller Zufallsvariablen mit  $\mu := \mathbb{E}(X_i)$  und  $\sigma^2 := \text{Var}(X_i)$  endlich, sowie  $S_n$  und  $Z_n$  definiert wie in (??) und (??). Dann konvergiert die Verteilung von  $Z_n$  für  $n \rightarrow \infty$  punktweise gegen die Standardnormalverteilung  $\mathcal{N}(0, 1)$ .

Mit anderen Worten konvergiert die Zufallsvariable  $Z_n$  *nach Verteilung* gegen eine standardnormalverteilte Zufallsvariable. Diese Konvergenz lässt sich auch durch die Verteilungsfunktion ausdrücken:

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z). \tag{7.33}$$

Auf den Beweis dieses Satzes soll hier verzichtet werden.

Der zentrale Grenzwertsatz kann genutzt werden, um Wahrscheinlichkeiten für endliche Stichproben zu approximieren.

## 8 Die Normalverteilung

Die Normalverteilung ist in diesem Skript in praktisch jedem Kapitel aufgetaucht. Sie verdient eine besondere Beachtung aufgrund ihrer vielfältigen Anwendbarkeit, ihrer Rolle in der Approximation anderer, insbesondere diskreter Verteilungen, und ihrer speziellen Eigenschaften in Bezug auf Transformationen und Kenngrößen.

Wir möchten in diesem Abschnitt zunächst einige bereits bekannte Erkenntnisse über die Normalverteilung zusammenfassen und miteinander in Relation setzen. Anschließend wollen wir die Zusammenhänge der Normalverteilung mit einigen anderen Verteilungen ergründen. Diese Überlegungen sollen in den zentralen Grenzwertsatz münden, welcher die herausragende Bedeutung dieser Verteilung begründet. Zuletzt möchten wir die Normalverteilung auf mehrdimensionale Zufallsvariablen verallgemeinern und für diese einige Rechenregeln herleiten.

### 8.1 Grundlegende Eigenschaften

Wir möchten zunächst alle bereits zuvor beschriebenen Eigenschaften der Normalverteilung zusammenfassen und ergänzen.

#### 8.1.1 Definition und Kenngrößen

Wie bereits zuvor vorgestellt, ist die stetige Normalverteilung  $\mathcal{N}(\mu, \sigma^2)$  (??) definiert durch die Riemann-Dichte

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (8.1)$$

Für gewöhnlich würden wir diese zur Bestimmung der Verteilungsfunktion integrieren. Im Falle der Normalverteilung jedoch ist dies nicht ohne weiteres möglich: Ihre Dichte ist nicht in geschlossener Form integrierbar. Anschaulich heißt das, dass wir ihre Verteilungsfunktion nicht ohne ein Integralzeichen hinschreiben können, und somit auch nicht einfach durch Werte einsetzen berechnen können.

Die Verteilungsfunktion der allgemeinen Normalverteilung ist naheliegenderweise

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt. \quad (8.2)$$

Eine Besonderheit der Normalverteilung ist, dass ihre Parameter  $\mu$  und  $\sigma^2$  direkt auch ihre wichtigsten Kenngrößen sind.

#### **Satz 8.1 (Erwartungswert und Standardabweichung der Normalverteilung)**

Sei  $X$  eine  $\mathcal{N}(\mu, \sigma^2)$ -verteilte reelle Zufallsvariable. Dann ist

- $\mathbb{E}(X) = \mu$  sowie
- $\text{Var}(X) = \sigma^2$  und somit  $\sigma(X) = \sigma$ .

BEWEIS. Durch Einsetzen und Ausrechnen.

- Erwartungswert:

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x \cdot \varphi(x) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \left[ \int_{-\infty}^{\infty} (x-\mu) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx + \int_{-\infty}^{\infty} \mu \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x - \mu) \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{y}{\sigma^2}\right) dy + \mu \quad (\text{Substitution: } y := \frac{(x - \mu)^2}{2}) \\
&= 0 + \mu = \mu
\end{aligned}$$

- Standardabweichung analog.

□

### 8.1.2 Die Standardnormalverteilung

Die spezielle Normalverteilung  $\mathcal{N}(0, 1)$  heißt *Standardnormalverteilung*. Ihre Dichte vereinfacht zu

$$\varphi_0(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (8.3)$$

Ihre Verteilungsfunktion lässt sich unter Verwendung der sogenannten Gauß'schen Fehlerfunktion schreiben.

#### Definition 8.2 (Fehlerfunktion)

Die Funktion  $\operatorname{erf} : \mathbb{R} \rightarrow \mathbb{R}$  ist definiert durch

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \quad (8.4)$$

und heißt **Fehlerfunktion** oder **Gauß'sche Fehlerfunktion**.

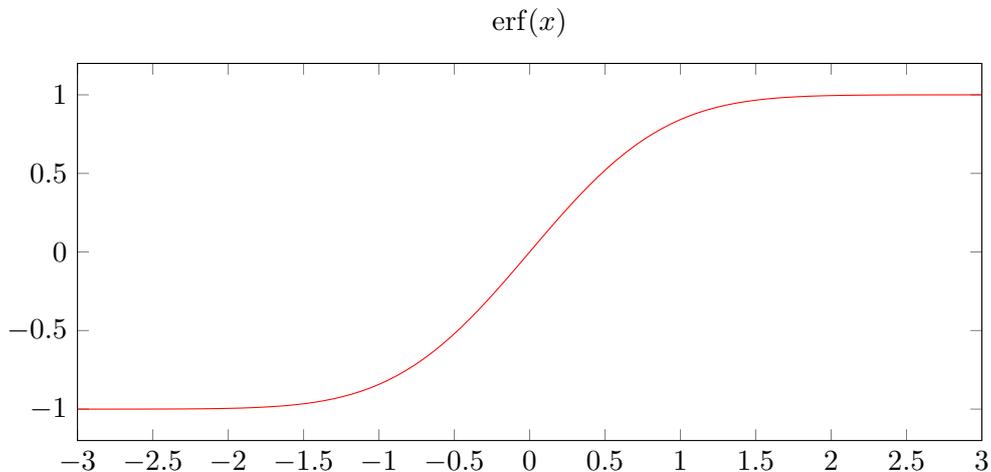


Abbildung 8.1: Verlauf der Fehlerfunktion

Damit ist die Verteilungsfunktion der Standardnormalverteilung

$$\Phi(x) = \frac{1}{2} \left( 1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt. \quad (8.5)$$

## 8.2 Beziehungen zu anderen Verteilungen

Aus Abschnitt ?? ist der Zentrale Grenzwertsatz bereits bekannt, der für sehr große Stichprobenmengen jede beliebige Verteilung gegen die Standardnormalverteilung konverieren lässt. Es gibt darüberhinaus diskrete Verteilungen, welche sich bei großem Stichprobenumfang explizit durch die Normalverteilung approximieren lassen.

### 8.2.1 Approximation der Binomialverteilung

Der **Satz von Moivre-Laplace** stellt die Verknüpfung zwischen der Binomialverteilung und der Normalverteilung her. Seine Kernaussage ist, dass sich für große Stichprobenmengen  $n \rightarrow \infty$  die Binomialverteilung  $B_{n,p}$  wie eine Normalverteilung verhält. Die zentrale, für uns relevante Folgerung aus dem Satz ist die Approximation der Binomial- durch die Normalverteilung:

#### Satz 8.3

Für  $n \rightarrow \infty$  kann die Binomialverteilung  $B_{n,p}$  durch die Normalverteilung  $\mathcal{N}(\mu, \sigma^2)$  approximiert werden mit

$$\mu := np \quad \text{und} \quad \sigma^2 = np(1-p).$$

Ist  $X$  eine  $B_{n,p}$ -verteilte Zufallsvariable, so lässt sich ihre Verteilung approximieren als

$$P(X \leq x) \approx \Phi\left(\frac{x - \mu}{\sigma}\right).$$

### 8.3 Die mehrdimensionale Normalverteilung

#### Definition 8.4 ( $n$ -dimensionale Normalverteilung)

Sei  $n \in \mathbb{N}$ ,  $\vec{a} \in \mathbb{R}^n$  sowie  $\mathbf{K} \in \mathbb{R}^{n \times n}$  symmetrisch positiv definit. Das durch die Riemann-Dichte

$$f(\vec{y}) = \frac{1}{\sqrt{\det \mathbf{K}}} \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \langle \vec{y} - \vec{a}, \mathbf{K}^{-1}(\vec{y} - \vec{a}) \rangle\right) \quad (8.6)$$

gegebene Wahrscheinlichkeitsmaß heißt  **$n$ -dimensionale Normalverteilung** und wird mit  $\mathcal{N}(\vec{a}, \mathbf{K})$  bezeichnet.

Die  $n$ -dimensionale Normalverteilung ist eine Verteilung für  $n$ -dimensionale Zufallsvektoren  $\vec{Y}$ . Dabei bezeichnet  $\vec{a}$  den *Erwartungsvektor*. Analog zur bekannten, eindimensionalen Normalverteilung speichert er die Erwartungswerte der einzelnen Komponenten von  $\vec{Y}$ . Die Matrix  $\mathbf{K}$  ist die *Kovarianzmatrix* der Zufallsvariablen  $Y_i$ .

Die  $n$ -dimensionale Normalverteilung ist in natürlicher Weise die gemeinsame Verteilung von  $n$  normalverteilten Zufallsvariablen.

#### Lemma 8.5 (Randdichten der Normalverteilung)

Sei  $\vec{Y} = (Y_1, \dots, Y_n)$  ein  $\mathcal{N}(\vec{a}, \mathbf{K})$ -verteilter Zufallsvektor. Dann hat die  $i$ -te Koordinatenvariable  $Y_i$  die Randverteilung  $\mathcal{N}(a_i, k_{ii})$ .

BEWEIS. TBD.

#### Satz 8.6 (Stochastische Unabhängigkeit normalverteilter Zufallsvariablen)

Sei  $\vec{Y} = (Y_1, \dots, Y_n)$   $\mathcal{N}(\vec{a}, \mathbf{K})$ -verteilt. Die  $Y_i$  sind genau dann stochastisch unabhängig, wenn  $\mathbf{K}$  eine Diagonalmatrix ist.

BEWEIS. Sei  $\mathbf{K} \in \mathbb{R}^{n \times n}$  gegeben als

$$\mathbf{K} := \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}.$$

Die Randdichten sind

$$f_i(y_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i - a_i)^2}{2\sigma_i^2}\right)$$

womit folgt

$$\begin{aligned} \prod_{i=1}^n f_i(y_i) &= \frac{1}{\sqrt{\sigma_1^2 \cdots \sigma_n^2}} \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - a_i)^2}{\sigma_i^2}\right) \\ &= \frac{1}{\sqrt{\det \mathbf{K}}} \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \langle (\vec{y} - \vec{a}), \mathbf{K}^{-1}(\vec{y} - \vec{a}) \rangle\right) \\ &= f(\vec{y}). \end{aligned}$$

□

Dieses Ergebnis ist besonders für die Normalverteilung: Es stellt nämlich eine Äquivalenz zwischen Unkorreliertheit und stochastischer Unabhängigkeit normalverteilter Zufallsvariablen her. Zur Erinnerung: Im Allgemeinen folgte aus stochastischer Unabhängigkeit zwar Unkorreliertheit; die Umkehrung war allerdings nicht gültig. Im Falle der Normalverteilung werden beide Begriffe aber tatsächlich gleichwertig.

Natürlich gibt es auch hier den Spezialfall der  $n$ -dimensionalen Standardnormalverteilung  $\mathcal{N}(0, \mathbf{I}_n)$ . Dabei ist  $\mathbf{I}_n$  die  $n$ -dimensionale Einheitsmatrix. Wie schon bei der eindimensionalen Variante geht jeder normalverteilte Zufallsvektor durch eine affin-lineare Transformation aus dem standardnormalverteilten Zufallsvektor  $\vec{X}$  hervor.

**Satz 8.7 (Affine Transformation normalverteilter Zufallsvektoren)**

Sei  $\vec{X} : \Omega \rightarrow \mathbb{R}^n$  ein  $\mathcal{N}(0, \mathbf{I}_n)$ -verteilter Zufallsvektor,  $\mathbf{A} \in \mathbb{R}^{n \times n}$  invertierbar sowie  $\vec{b} \in \mathbb{R}^n$ . Sei außerdem  $\vec{Y} : \Omega \rightarrow \mathbb{R}^n$  gegeben durch

$$\vec{Y} = \mathbf{A}\vec{X} + \vec{b}.$$

Genau dann ist  $\vec{Y} \mathcal{N}(\vec{b}, \mathbf{K})$ -verteilt mit der Kovarianzmatrix  $\mathbf{K} := \mathbf{A}\mathbf{A}^T$ .

BEWEIS. Die Komponenten  $X_i$  von  $\vec{X}$  sind alle  $\mathcal{N}(0, 1)$ -verteilt. Für die Komponenten  $Y_i$  von  $\vec{Y}$  gilt:

$$Y_i = \sum_{k=1}^n a_{ik} X_k + b_i \quad (8.7)$$

Wie wir wissen hat  $Y_i$  damit die Verteilung  $\mathcal{N}(b_i, \sigma^2)$  mit  $\sigma = \sum_{k=1}^n a_{ik}$ . Die Randverteilungen von  $\vec{Y}$  sind also Normalverteilungen; somit ist  $\vec{Y}$  ebenfalls normalverteilt. Der Erwartungsvektor ist  $\vec{b}$ . Für die Kovarianzmatrix gilt:

$$\begin{aligned} k_{ij} &= \text{Cov}(Y_i, Y_j) \\ &= \mathbb{E}((Y_i - \mathbb{E}(Y_i))(Y_j - \mathbb{E}(Y_j))) \\ &= \mathbb{E}\left(\left(\sum_{k=1}^n a_{ik} X_k\right) \left(\sum_{l=1}^n a_{jl} X_l\right)\right) \\ &= \sum_{k=1}^n a_{ik} a_{jk}. \end{aligned}$$

Denn: Es ist  $\mathbb{E}(X_k X_l) = 0$  für  $k \neq l$ , da  $X_k, X_l$  stochastisch unabhängig sind, und  $\mathbb{E}(X_k^2) = 1$ , da  $X_k^2$  Chi-Quadrat-verteilt ist. Damit ist  $\mathbf{K} = \mathbf{A}\mathbf{A}^T$ .

Umgekehrt kann jede positiv definite Matrix  $\mathbf{K}$  dargestellt werden als  $\mathbf{K} = \mathbf{A}\mathbf{A}^T$ , und mit  $\vec{b}$  dem Erwartungsvektor erhalten wir für ein gegebenes  $\vec{Y}$  die gewünschte Zerlegung. □

**Korollar 8.8**

Sei  $\vec{Y} \mathcal{N}(\vec{a}, \mathbf{K})$ -verteilt, und sei  $\vec{Z} := \mathbf{B}\vec{Y} + \vec{b}$  mit  $\mathbf{B} \in \mathbb{R}^{n \times n}$  invertierbar und  $\vec{b} \in \mathbb{R}^n$ .  
Dann ist  $\vec{Z} \mathcal{N}(\mathbf{B}\vec{a} + \vec{b}, \mathbf{BKB}^T)$ -verteilt.

BEWEIS. Expandiere

$$\vec{Z} = \mathbf{B}\vec{Y} + \vec{b} = \mathbf{B}(\mathbf{A}\vec{X} + \vec{a}) + \vec{b} = \mathbf{B}\mathbf{A}\vec{X} + \mathbf{B}\vec{a} + \vec{b}$$

und wende Satz ?? an. □

## 9 Stochastische Prozesse

Bisher hatten wir uns hauptsächlich mit einzelnen Zufallsvariablen und ihren Eigenschaften befasst, sowie einfache Abhängigkeiten und Kombinationen von Zufallsvariablen analysiert. In diesem Abschnitt widmen wir uns strukturierten Systemen von abhängigen Zufallsvariablen. Wir betrachten *stochastische Prozesse* im Allgemeinen und lernen, die Zustände eines sich zufällig in der Zeit entwickelnden Systems als Zufallsvariablen zu modellieren. Konkret betrachten wir *Markow-Ketten*: diskrete stochastische Prozesse, deren nächster Zustand allein vom aktuellen Zustand abhängt.

### 9.1 Grundlegendes zu stochastischen Prozessen

Ein stochastischer Prozess ist die Beschreibung eines Systems, welches seinen Zustand zufallsbedingt mit der Zeit verändert. Sowohl der Raum der möglichen Zustände als auch die betrachteten Zeiten können hierbei diskret (abzählbar) oder überabzählbar sein.

#### Definition 9.1 (Stochastischer Prozess)

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(Z, \mathcal{Z})$  ein Messraum, sowie  $T$  eine Indexmenge. Ein **stochastischer Prozess** ist eine Abbildung

$$X : \Omega \times T \rightarrow Z, \quad (\omega, t) \mapsto X_t(\omega),$$

sodass  $X_t : \omega \mapsto X_t(\omega)$  für alle  $t \in T$  eine messbare Abbildung, also eine Zufallsvariable ist. Der Raum  $Z$  heißt **Zustandsraum** des Prozesses.

Meist ist  $T = \mathbb{N}_0$  oder  $T = \mathbb{R}$ . Ein stochastischer Prozess ordnet also jedem Index beziehungsweise jeder Zeit  $t \in T$  eine Zufallsvariable zu, welche den Zustand des Systems zu diesem Zeitpunkt beschreibt. Wir werden uns hier nur mit zeitdiskreten stochastischen Prozessen ( $T = \mathbb{N}_0$ ) befassen. Im Wesentlichen handelt es sich hier um eine Verallgemeinerung der mehrstufigen Modelle auf unendlich viele Stufen, mit entsprechend unendlich vielen Übergangsdichten. Für ein mehrstufiges Experiment war es möglich, durch Kopplung eine Wahrscheinlichkeitsverteilung zu formulieren, welche das gesamte Experiment beschrieb. Es ist zunächst fraglich, ob dies auch für einen diskreten stochastischen Prozess möglich ist. Der **Satz von Ionescu-Tulcea** sichert die Existenz solcher unendlich gekoppelter Modelle.

### 9.2 Markow-Ketten

#### 9.2.1 Definitionen

Eine Markow-Kette ist ein ‚gedächtnisloser‘ stochastischer Prozess, bei dem der Folgezustand lediglich von dem aktuellen Zustand abhängt. Die einzelnen Stufen sind also Markow-gekoppelt. Wir betrachten hier lediglich Markow-Ketten mit einer endlichen Zustandsmenge  $Z$ . Wahrscheinlichkeitsverteilungen auf  $Z$  stellen wir daher als Vektoren dar:

#### Definition 9.2 (Wahrscheinlichkeitsvektor)

Ein Vektor  $\vec{v} \in \mathbb{R}^n$  heißt **Wahrscheinlichkeitsvektor** oder **stochastischer Vektor**, falls gilt:

1.  $v_i \in [0, 1]$  für alle  $i = 1, \dots, n$ ,
2.  $\sum_{i=1}^n v_i = 1$ .

Einen solchen stochastischen Vektor können wir auch als eine *Überlagerung* von Zuständen interpretieren: Jeder Eintrag gibt die Wahrscheinlichkeit dafür an, dass sich ein System in dem jeweiligen

Zustand befindet. In einem System mit 3 Zuständen  $Z = \{1, 2, 3\}$  beschreibt der Wahrscheinlichkeitsvektor

$$\vec{q}_0 := \begin{pmatrix} 0.3 \\ 0.5 \\ 0.2 \end{pmatrix}$$

eine Situation, in der sich das System mit Wahrscheinlichkeit 30% im Zustand 1 befindet, mit 50% im Zustand 2, und mit 20% im Zustand 3. Der tatsächliche Zustand des Systems ist allerdings unbekannt. Dennoch können wir, wenn wir die Wahrscheinlichkeiten der Zustandübergänge kennen, die zeitliche Fortentwicklung des Systems analysieren. Die Übergangswahrscheinlichkeiten können wir als eine Matrix formulieren.

**Definition 9.3 (Stochastische Matrix)**

Eine  $n \times n$ -Matrix  $\mathbf{P}$  heißt **stochastische Matrix** oder **Übergangsmatrix**, falls gilt:

1.  $p_{ij} \in [0, 1]$  für alle  $i, j = 1, \dots, n$ ,
2.  $\sum_{j=1}^n p_{ij} = 1$  für alle  $i = 1, \dots, n$ .

In einer stochastischen Matrix beschreibt der Eintrag der  $i$ -ten Zeile und  $j$ -ten Spalte die Wahrscheinlichkeiten, dass ein System, welches sich im Zustand  $i$  befindet, in den Zustand  $j$  übergeht. Zum Beispiel beschreibt die Matrix

$$\mathbf{P} := \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.2 & 0 & 0.8 \\ 0.7 & 0 & 0.3 \end{pmatrix}$$

ein System, welches von Zustand 1 mit 50%-iger Wahrscheinlichkeit in Zustand 1 bleibt, mit 40%-iger Wahrscheinlichkeit zu Zustand 2 und mit Wahrscheinlichkeit 10% in Zustand 3 übergeht. Außerdem kann das System nie in Zustand 2 verbleiben, oder von Zustand 3 in Zustand 2 übergehen. Wir können dieses System als einen *Übergangsgraphen* veranschaulichen, wie in Abbildung ??.

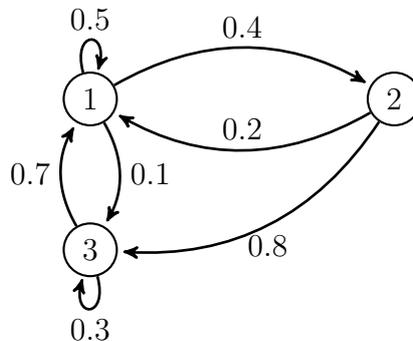


Abbildung 9.1: Übergangsgraph zur Matrix  $\mathbf{P}$

Wir können den aktuellen Zustand als Zufallsvariable  $Z_0$  und den Folgezustand als  $Z_1$  modellieren. Die Verteilung von  $Z_0$  ist dann gerade der Vektor  $\vec{q}_0$ , es ist

$$P(Z_0 = j) = q_{0,j}.$$

Die Matrix  $\mathbf{P}$  enthält nun alle Übergangsdichten von  $Z_0$  zu  $Z_1$  als bedingte Wahrscheinlichkeiten: Es ist

$$\mathbf{p}_{i,j} = P(Z_1 = j | Z_0 = i).$$

Die Verteilung des Folgezustandes nennen wir  $\vec{q}_1$ . Wir können sie mithilfe der Formel der totalen Wahrscheinlichkeit berechnen:

$$q_{1,j} = P(Z_1 = j) = \sum_{i=1}^n P(Z_0 = i)P(Z_1 = j|Z_0 = i) = \sum_{i=1}^n q_{0,i} \cdot p_{i,j}. \quad (9.1)$$

Wir erkennen: Dies ist lediglich die komponentenweise Formulierung des Matrixproduktes

$$\vec{q}_1 = \mathbf{P}^T \vec{q}_0 \Leftrightarrow \vec{q}_1^T = \vec{q}_0^T \mathbf{P}. \quad (9.2)$$

Nun haben wir alle Voraussetzungen geklärt, um die Markow-Kette formal zu definieren.

#### Definition 9.4 (Markow-Kette)

Sei  $Z := \{1, \dots, n\}$  eine Menge von Zuständen,  $q_0 : Z \rightarrow [0, 1]$  eine Zähldichte und  $\mathbf{P} \in \mathbb{R}^{n \times n}$  eine stochastische Matrix. Eine Folge  $(Z_k)_{k \in \mathbb{N}}$  von Zufallsvariablen  $Z_k : \Omega \rightarrow Z$  heißt (**homogene**) **Markow-Kette**, wenn gilt:

1.  $q_0$  ist die Verteilung von  $Z_0$ , also  $P(Z_0 = j) = q_0(j)$
2. Die  $Z_k$  sind Markow-gekoppelt und die Übergangsdichten sind durch die Matrix  $\mathbf{P}$  gegeben als

$$P(Z_{k+1} = j | Z_k = i) = \mathbf{p}_{i,j}.$$

Die Matrix  $\mathbf{P}$  wird als **Übergangsmatrix** bezeichnet, die Verteilung  $q_0$  als **Startverteilung**.

Die Markow-Kette selbst gibt die Menge an Zuständen und die möglichen Zustandsübergänge an, sowie deren Wahrscheinlichkeiten. Natürlich interessieren wir uns auch für konkrete Verläufe in einer Markow-Kette. Diese nennen wir Pfade. Ein Pfad ist eine konkrete Folge von Zuständen, wie sie eine Markow-Kette der Reihe nach durchlaufen kann.

#### Definition 9.5 (Pfad)

Sei  $(Z_k)_{k \in \mathbb{N}}$  eine Markow-Kette. Eine Folge  $(z_k)_{k \in \mathbb{N}}$  von Ausprägungen  $z_k$  der Zufallsvariablen  $Z_k$  heißt **Pfad** der Markow-Kette.

### 9.2.2 Zustände, Klassen und Übergänge

Im Folgenden meint  $q_k$  die Zähldichte der Zufallsvariable  $Z_k$  und  $\vec{q}_k$  den zugehörigen Zufallsvektor. Es gilt:  $\vec{q}_{k,i} := q_k(i)$ . Die Übergangsdichten von der  $k$ -ten zur  $k+1$ ten Stufe sind unabhängig von  $k$  und durch die Matrix  $\mathbf{P}$  vollständig definiert. Tatsächlich lassen sich mithilfe von  $\mathbf{P}$  sogar beliebig viele Schritte zusammenfassen.

#### Lemma 9.6 (Zustandsübergänge)

Sei  $(Z_k)_{k \in \mathbb{N}}$  eine Markow-Kette mit Übergangsmatrix  $\mathbf{P}^T$ . Seien  $k, m \in \mathbb{N}, k \leq m$  und  $q_k$  die Verteilung der Stufe  $Z_k$ . Dann hat die  $k+m$ -te Stufe  $Z_{k+m}$  die Verteilung  $q_{k+m}$  mit

$$\vec{q}_{k+m} = \mathbf{P}^m \cdot \vec{q}_k. \quad (9.3)$$

BEWEIS. Per vollständiger Induktion über  $m$ .

- $m = 0$ :  $\vec{q}_{k+0} = \vec{q}_k = \mathbf{P}^0 \cdot \vec{q}_k$ .
- Induktionsvoraussetzung: Sei  $m \in \mathbb{N}$  sodass (??) erfüllt ist.

- $m \rightarrow m + 1$ : Es ist  $\vec{q}_{k+m+1} = \mathbf{P} \cdot \vec{q}_{k+m}$ , da  $\mathbf{P}$  die Übergangsdichte liefert (vgl. (??), (??)). Dann folgt nach Induktionsvoraussetzung

$$\vec{q}_{k+m+1} = \mathbf{P} \cdot \vec{q}_{k+m} \stackrel{\text{I.V.}}{=} \mathbf{P} \cdot \mathbf{P}^m \cdot \vec{q}_k = \mathbf{P}^{m+1} \cdot \vec{q}_k.$$

□

Die Matrix  $\mathbf{P}^m$  heißt *m-Schritt-Übergangsmatrix*.

Wir möchten nun weiter die möglichen Pfade einer Markow-Kette untersuchen. Dabei interessiert uns zum Beispiel, welche Zustände von einem gewissen Ausgangszustand  $i$  überhaupt erreichbar sind, oder ob die Kette ‚kreisen‘ kann. Wir nutzen aus, dass eine Markow-Kette als ein gerichteter Graph mit gewichteten Kanten dargestellt werden kann, mit  $\mathbf{P}$  als Adjazenzmatrix. Wir werden nun an einige Konzepte der Graphentheorie erinnern und sie auf Markow-Ketten anwenden.

### Definition 9.7 (Pfade und Zusammenhang in Graphen)

Sei  $G = (V, E)$  ein gerichteter Graph.

- Ein **Pfad** in  $G$  ist ein Tupel  $(v_1, \dots, v_n) \in V^n$ , sodass  $(v_i, v_{i+1}) \in E$  für alle  $i = 1, \dots, n - 1$ .
- Der Graph  $G$  heißt **stark zusammenhängend**, falls für alle Knoten  $u, v \in V, u \neq v$  ein sowohl Pfad  $(u, \dots, v)$  als auch ein Pfad  $(v, \dots, u)$  in  $G$  existiert.
- Eine Knotenmenge  $U \subseteq V$  heißt **starke Zusammenhangskomponente** von  $G$ , wenn der durch  $U$  induzierte Teilgraph von  $G$  stark zusammenhängend ist. Eine starke Zusammenhangskomponente heißt **maximal**, wenn sie nicht durch Hinzunahme eines weiteren Knotens vergrößert werden kann.

### Lemma 9.8

Sei  $G = (V, E)$  ein gerichteter Graph. Sei die Relation  $\sim_G \subseteq V \times V$  gegeben, sodass  $u \sim_G v$  genau dann, wenn in  $G$  sowohl ein Pfad von  $u$  nach  $v$  als auch ein Pfad von  $v$  nach  $u$  existiert.

Dann ist  $\sim_G$  eine Äquivalenzrelation auf  $V$ , und ihre Äquivalenzklassen sind maximale starke Zusammenhangskomponenten in  $G$ .

BEWEIS.

1.  $\sim_G$  ist Äquivalenzrelation:

- $\sim_G$  ist reflexiv, da für jeden Knoten  $v \in V$  der triviale Pfad  $(v)$  in  $G$  existiert.
- $\sim_G$  ist symmetrisch: Aus  $u \sim_G v$  folgt  $v \sim_G u$  durch Vertauschen der Pfade.
- $\sim_G$  ist transitiv: Sei  $u \sim_G v$  und  $v \sim_G w$ . Dann folgt auch  $u \sim_G w$  durch Konkatenation der Pfade von  $u$  nach  $v$  und von  $v$  nach  $w$ .

Außerdem gilt: Ist  $u \sim_G w$  und es existiert ein Pfad  $p = (u, \dots, v, \dots, w)$ , dann ist auch  $u \sim_G v$ . Der Pfad von  $u$  nach  $v$  ist in  $p$  enthalten, und der Pfad von  $v$  nach  $u$  ergibt sich aus dem Restpfad von  $v$  nach  $w$  und dem Pfad von  $w$  nach  $u$ . Ebenso folgt  $v \sim_G w$ .

2.  $\sim_G$  induziert maximale starke Zusammenhangskomponenten auf  $G$ :

- Sei  $u \in V$  und  $U := [u]_{\sim_G}$  die Äquivalenzklasse von  $u$ . Sei  $G_U = (U, V_U)$  der von  $U$  induzierte Teilgraph von  $G$ .  $G_U$  enthält alle Pfade zwischen den Knoten in  $U$ , denn: Ein Pfad  $(a, \dots, b, \dots, c)$  von  $a \in U$  nach  $c \in U$  über  $b \notin U$  kann nicht existieren, da nach obiger Erkenntnis auch  $b \in U$  sein muss. Damit existieren in  $G_U$  also für alle Knoten  $a, c \in U$  Pfade hin und zurück;  $G_U$  ist also stark zusammenhängend.
- Gäbe es einen Knoten  $b \in V \setminus U$ , sodass der durch Hinzunahme von  $b$  zu  $G_U$  erzeugte Teilgraph von  $G$  immernoch stark zusammenhängend ist, dann wäre bereits  $u \sim_G b$  und somit  $b \in U$ .  $G_U$  ist also maximal.

□

Die Erkenntnis des vorausstehenden Lemmas ist grundlegend. Veranschaulicht bedeutet es, dass wir einen Graphen  $G$  vollständig und disjunkt in Teilgraphen zerlegen können, sodass man innerhalb der Teilgraphen von jedem Knoten zu jedem anderen kommt. Verlässt man jedoch einen dieser Teilgraphen, so kommt man nie wieder zu ihm zurück. Für Markow-Ketten ist dies eine entscheidende Erkenntnis: Durch Analyse der starken Zusammenhangskomponenten des Graphen einer Markow-Kette können wir die Zustände gruppieren, sowie mögliche Zyklen und Senken identifizieren.

### Satz 9.9 (Graphenzerlegung)

Für jeden gerichteten Graph  $G = (V, E)$  existiert eine eindeutige, disjunkte Zerlegung

$$V = \sum_{i=1}^k U_i \quad (9.4)$$

in seine maximalen starken Zusammenhangskomponenten  $U_i$ .

BEWEIS. Existenz folgt aus Lemma ??, da  $\sim_G$  eine disjunkte Zerlegung liefert. Diese ist eindeutig, denn: Gibt es eine zweite Zerlegung  $V = \sum_{j=1}^l W_j$  und einen Knoten  $v \in V$ ,  $v \in U_i$  und  $v \in W_j$ , dann gibt es für alle Knoten  $u \in U_i, w \in W_j$  einen Pfad von  $u$  nach  $w$  und zurück, jeweils über den Knoten  $v$ . Also ist auch  $U_i \cup W_j$  stark zusammenhängend. Da beide Mengen allerdings nach Voraussetzung maximal sind, ist  $U_i = W_j$ . □

Wir können die Zustände einer Markow-Kette also eindeutig in *Klassen* einteilen, indem wir den Graphen in seine maximalen starken Zusammenhangskomponenten zerlegen. Verlässt ein Pfad der Markow-Kette eine dieser Klassen, so kann er nie wieder in diese Klasse zurückkehren. Ist der Graph selbst stark zusammenhängend, so gehören alle Zustände zur selben Klasse, und wir nennen die Kette *irreduzibel*.

### Definition 9.10 (Irreduzibilität)

- Eine Matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  heißt **irreduzibel**, falls der durch  $\mathbf{A}$  als Adjazenzmatrix beschriebene gerichtete Graph stark zusammenhängend ist.
- Eine Markow-Kette  $(Z_k)_{k \in \mathbb{N}}$  mit Übergangsmatrix  $\mathbf{P}$  heißt **irreduzibel**, falls  $\mathbf{P}$  irreduzibel ist.

In einer irreduziblen Markow-Kette ist also sichergestellt, dass jeder Zustand immer wieder erreicht werden kann. Eine reduzible Markow-Kette hat ‚absorbierende‘ Zustandsklassen, in die sich ihre Pfade langfristig verzweigen.

## 9.2.3 Markow-Ketten im Gleichgewicht

### Definition 9.11 (Gleichgewichtsverteilung)

Eine Verteilung  $\pi$  auf  $Z$  heißt **Gleichgewichtsverteilung** einer Markow-Kette  $(Z_k)_{k \in \mathbb{N}}$ , wenn mit  $\pi$  als Startverteilung jede Stufe  $Z_k$  der Kette ebenfalls die Verteilung  $\pi$  hat. Die Markow-Kette ist dann im Gleichgewicht.

### Satz 9.12 (Charakterisierung der Gleichgewichtsverteilung)

Sei  $(Z_k)_{k \in \mathbb{N}}$  eine Markow-Kette mit Übergangsmatrix  $\mathbf{P}$ . Eine Verteilung  $\pi$  ist genau dann Gleichgewichtsverteilung von  $(Z_k)$ , wenn der Verteilungsvektor  $\vec{\pi}$  ein Linkseigenvektor von  $\mathbf{P}$  zum Eigenwert 1 ist, also

$$\vec{\pi}^T = \vec{\pi}^T \cdot \mathbf{P} \quad \text{beziehungsweise} \quad \vec{\pi} = \mathbf{P}^T \cdot \vec{\pi}. \quad (9.5)$$

BEWEIS. Sei  $\pi$  Gleichgewichtsverteilung von  $(Z_k)$ . Dann ist  $\pi$  Verteilung der Stufen  $Z_n$  und  $Z_{n+1}$  für  $n \in \mathbb{N}$ . Nach Lemma ?? ist somit  $\vec{\pi} = \mathbf{P}^T \cdot \vec{\pi}$ .  
Sei umgekehrt  $\vec{\pi} = \mathbf{P}^T \cdot \vec{\pi}$ , und  $\pi$  Startverteilung von  $(Z_k)$ . Nach Lemma ?? folgt dann für die Verteilung  $q_n$  der Stufe  $Z_n$ ,  $n \in \mathbb{N}$ :  $\vec{q}_n = (\mathbf{P}^T)^n \cdot \vec{\pi} = \vec{\pi}$ . Also ist  $\pi$  Gleichgewichtsverteilung.  $\square$

## 10 Schätzen von Parametern

Wir haben bereits diverse Wahrscheinlichkeitsverteilungen betrachtet, welche von einem oder mehreren Parametern abhängen. Anhand der Parameter konnten wir die Verteilungen analysieren und ihre Kenngrößen bestimmen. In der Praxis der mathematischen Statistik allerdings ist die Aufgabe meist umgekehrt. Aus Beobachtungsreihen und Stichproben möchte man zunächst Kenngrößen für die Grundgesamtheit extrapolieren, und anschließend die Grundgesamtheit durch eine Wahrscheinlichkeitsverteilung beschreiben. Die Entscheidung für eine Verteilung ist meist nicht schwer: So legt der zentrale Grenzwertsatz für viele Statistiken die Normalverteilung nahe. Auch die verschiedenen auf der Bernoulli-Verteilung basierenden Wahrscheinlichkeitsverteilungen eignen sich zur Beschreibung vieler realer Vorgänge. Die Parameter dieser Verteilungen ( $\mu$  und  $\sigma^2$  bei der Normalverteilung,  $p$  bei der Bernoulli-Verteilung, etc.) sind allerdings unbekannt. In diesem Abschnitt möchten wir Methoden ergründen, um diese Parameter anhand von Stichproben zu schätzen.

Zunächst betrachten wir die Berechnung eines Schätzwertes für einen beliebigen Parameter  $\theta$  aus einer Stichprobe, die sogenannte Punktschätzung. Als einen allgemeinen Ansatz zur Punktschätzung betrachten wir die Maximum-Likelihood-Methode. Zuletzt interessieren wir uns für die Güte unserer Schätzung, was uns zur Bestimmung der Konfidenzintervalle führt.

Die Begriffe *Grundgesamtheit* und *Stichprobe* aus der beschreibenden Statistik werden in diesem Abschnitt auf Zufallsvariablen erweitert. Wir beschreiben eine Grundgesamtheit durch eine Zufallsvariable  $X$ , deren Verteilung wir ergründen möchten. Dazu untersuchen wir Stichprobenvariablen  $\vec{X} = (X_1, \dots, X_n)$  als stochastisch unabhängige und zu  $X$  identisch verteilte Zufallsvariablen, deren Ausprägungen  $\vec{x} = (x_1, \dots, x_n)$  die eigentlichen Stichproben sind.

### 10.1 Punktschätzungen

Eine Punktschätzung ist ein Schätzverfahren, an dessen Ende ein fester Wert für den zu schätzenden Parameter  $\theta$  steht. Wir möchten Punktschätzverfahren entwickeln, die, wenn angewendet auf eine Stichprobe, einen Parameterschätzwert liefern, mit dem wir rechnen können. Dem Gesetz der großen Zahlen entsprechend soll ein solches Schätzverfahren bei wachsender Stichprobengröße zunehmend genauere Werte liefern. Die Güte eines Schätzverfahrens möchten wir daher analytisch untersuchen.

#### 10.1.1 Eigenschaften von Schätzfunktionen

Wir haben eine Stichprobe  $(x_1, \dots, x_n)$  von  $n$  Ausprägungen einer Zufallsvariable  $X$ , welche wir der Einfachheit wegen als Ausprägung eines Zufallsvektors  $(X_1, \dots, X_n)$  von stochastisch unabhängigen und zu  $X$  identisch verteilten Variablen  $X_i$  formulieren. Die Art der Verteilung ist bekannt, ihre Parameter allerdings nicht. Wir möchten nun anhand dieser Stichprobe einen Näherungswert für einen Parameter  $\theta$  der Verteilung schätzen. Den Schätzer stellen wir als Funktion  $\hat{\Theta}$  der Stichprobenvariablen  $(X_1, \dots, X_n)$  auf, er wird dadurch selbst auch zu einer Zufallsvariablen. Für die konkrete Stichprobe  $(x_1, \dots, x_n)$  erhalten wir so einen Schätzwert  $\hat{\theta}$  als Ausprägung der Zufallsvariable  $\hat{\Theta}$ .

##### Definition 10.1 (Schätzfunktion, Schätzer und Schätzwert)

Seien  $(X_1, \dots, X_n)$  Stichprobenvariablen einer Grundgesamtheit  $X$  mit der parameterabhängigen Dichte  $f(x; \theta)$ ,  $\theta \in \mathbb{R}$ . Sei außerdem  $(x_1, \dots, x_n)$  eine konkrete Stichprobe. Eine Funktion

$$h : \mathbb{R}^n \rightarrow \mathbb{R}$$

zur Schätzung des Parameters  $\theta$  heißt **Schätzfunktion**. Die Zufallsvariable

$$\hat{\Theta}_n = h(X_1, \dots, X_n)$$

heißt **Schätzer**, und die Ausprägung

$$\hat{\theta}_n = h(x_1, \dots, x_n)$$

**Schätzwert** für  $\theta$ .

Unser Ziel ist, für einen gesuchten Parameter  $\theta$  eine möglichst gute Schätzfunktion zu finden. Zur Analyse der Qualität der Schätzfunktion  $h$  betrachten wir die Zufallsvariable  $\hat{\Theta}$ . Als Zufallsvariable können wir für  $\hat{\Theta}$  unter anderem Erwartungswert und Varianz bestimmen. Für immer größere Stichproben ( $n \rightarrow \infty$ ) soll  $h$  immer genauere Schätzungen für  $\theta$  liefern. Mathematisch können wir also für die Verteilung von  $\hat{\Theta}$  fordern, dass sie für  $n \rightarrow \infty$  gegen die Einpunktverteilung konvergiert. Diese Eigenschaft bezeichnen wir als *Konsistenz*, und sie liefert uns ein Konvergenzkriterium.

**Definition 10.2 (Konsistenz)**

Eine Schätzfunktion  $h$  heißt **konsistent**, falls die Varianz für wachsende Stichprobengröße verschwindet, also

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\Theta}_n) = \lim_{n \rightarrow \infty} \text{Var}(h(X_1, \dots, X_N)) = 0.$$

**Satz 10.3**

Sei  $h$  eine konsistente Schätzfunktion für einen Parameter  $\theta$  der Verteilung einer reellen Zufallsvariable  $X$ . Seien  $(X_1, \dots, X_n)$  Stichprobenvariablen für  $X$  und  $\hat{\Theta} = h(X_1, \dots, X_n)$ . Dann gilt für  $\varepsilon > 0$  die Abschätzung

$$P(|\hat{\Theta}_n - \theta| \geq \varepsilon) \leq \frac{\text{Var}(\hat{\Theta}_n)}{\varepsilon^2} \tag{10.1}$$

sowie der Grenzwert

$$\lim_{n \rightarrow \infty} P(|\hat{\Theta}_n - \theta| \geq \varepsilon) = 0. \tag{10.2}$$

BEWEIS. Gleichung (??) ist lediglich eine Instanz der Tschebyscheff-Ungleichung (??). Aufgrund der Konsistenz von  $h$  konvergiert die rechte Seite für  $n \rightarrow \infty$  gegen Null. Da

$$0 \leq P(|\hat{\Theta}_n - \theta| \geq \varepsilon) \leq \frac{\text{Var}(\hat{\Theta}_n)}{\varepsilon^2}$$

konvergiert somit nach dem **Sandwichsatz** auch die rechte Seite gegen Null. □

Auch an den Erwartungswert von  $\hat{\Theta}_n$  können wir eine Forderung stellen: Es ist sinnvoll, dass er gerade dem unbekanntem Parameter  $\theta$  entspricht.

**Definition 10.4 (Erwartungstreue)**

Sei  $h$  eine Schätzfunktion mit Schätzer  $\hat{\Theta}_n$  für einen Parameter  $\theta$  einer Verteilung.

- Sowohl  $h$  als auch  $\hat{\Theta}_n$  heißen **erwartungstreu**, wenn für alle  $n \in \mathbb{N}$  gilt:

$$\mathbb{E}(\hat{\Theta}_n) = \theta. \tag{10.3}$$

- Sowohl  $h$  als auch  $\hat{\Theta}_n$  heißen **asymptotisch erwartungstreu**, wenn für wachsende Stichproben  $(X_1, \dots, X_n)$  gilt:

$$\lim_{n \rightarrow \infty} \mathbb{E}(\hat{\Theta}_n) = \theta. \tag{10.4}$$

Wir möchten uns nun mit der Schätzung der Parameter zweier wichtiger Verteilungen beschäftigen. Wir betrachten die Normalverteilung mit ihren Parametern  $\mu$  und  $\sigma^2$  sowie die Binomialverteilung mit dem Parameter  $q$ . Diese Parameter entsprechen jeweils einer markanten Kenngröße der Verteilungen. Es ist daher naheliegend, sie direkt aus den empirischen Pendanten dieser Kenngrößen zu schätzen.

### 10.1.2 Parameterschätzung der Normalverteilung

Die Normalverteilung  $\mathcal{N}(\mu, \sigma^2)$  ist parametrisiert durch ihren Erwartungswert  $\mu$  und ihre Varianz  $\sigma^2$ . Für eine gegebenen Stichprobe  $(x_1, \dots, x_n)$  einer normalverteilten Grundgesamtheit kennen wir bereits die entsprechenden Größen des arithmetischen Mittels  $\bar{x}$  und der empirischen Varianz  $s_x^2$ . Wir möchten nun ihre Eignung als Schätzfunktionen prüfen.

**Schätzung des Erwartungswertes durch das arithmetische Mittel** Es ist zu erwarten, dass sich Stichproben aus einer normalverteilten Grundgesamtheit entsprechend der Verteilung um den Erwartungswert gruppieren. Für immer größere Stichproben sollten ihre Mittelwerte dem Erwartungswert also immer näher kommen. Daher wählen wir für den Erwartungswert als Schätzfunktion das arithmetische Mittel:

$$h_\mu(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}. \quad (10.5)$$

Den durch  $h_\mu$  definierten Schätzer nennen wir  $\hat{M}$ . Das arithmetische Mittel ist, wie vermutet, erwartungstreu:

$$\mathbb{E}(\hat{M}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu. \quad (10.6)$$

Auch die Konsistenz dieses Schätzers lässt sich zeigen:

$$\text{Var}(\hat{M}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0. \quad (10.7)$$

Allem Anschein nach eignet sich das arithmetische Mittel also gut zur Schätzung von  $\mu$ . Im Abschnitt über die Maximum-Likelihood-Methode werden wir erkennen, dass es bezüglich des Likelihood-Kriteriums tatsächlich die beste Wahl ist.

**Schätzung der Varianz durch die empirische Varianz** Wie schon den Erwartungswert, so können wir auch die Varianz der Normalverteilung auf ihr empirisches Gegenstück zurückführen. Wir haben für Stichproben zwei verschiedene Abweichungsgrößen kennengelernt: Die statistische Varianz und die empirische Varianz. Als Schätzfunktion  $h_{\sigma^2}$  wählen wir die empirische Varianz:

$$h_{\sigma^2}(x_1, \dots, x_n) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = s_x^2 \quad (10.8)$$

### 10.1.3 Parameterschätzung der Binomialverteilung

Die Binomialverteilung  $B(n, p)$  ist die Verteilung der Anzahl der zu erwartenden Erfolge  $k$  eines  $n$  mal durchgeführten Bernoulli- $p$ -Experimentes. Bei einer hinreichend großen Zahl  $n$  von Versuchen ist zu erwarten, dass sich die relative Häufigkeit  $k/n$  der Erfolge der Erfolgswahrscheinlichkeit  $p$  annähert. Dementsprechend ist es naheliegend,  $p$  durch diese relative Häufigkeit zu schätzen. Für die binomialverteilte Zufallsvariable  $X$  betrachten wir daher die einzelnen Bernoulli-Experimente, also die  $B(p)$ -verteilten Zufallsvariablen  $X_i$ . Wir wählen für  $p$  die Schätzfunktion  $h_p$  als

$$h_p(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i = \frac{k}{n}, \quad (10.9)$$

wobei die Ausprägungen  $x_i$  der Versuche  $X_i$  lediglich die Werte 0 und 1 annehmen. Dementsprechend ist  $k = \sum_{i=1}^n x_i$  die Anzahl der Erfolge. Die Erwartungstreue des durch  $h_p$  gegebenen Schätzers  $\hat{P}$  folgt direkt:

$$\mathbb{E}(\hat{P}) = \frac{1}{n} \cdot np = p, \quad (10.10)$$

und auch die Konsistenz lässt sich leicht nachweisen:

$$\text{Var}(\hat{P}) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \xrightarrow{n \rightarrow \infty} 0. \quad (10.11)$$

## 10.2 Die Maximum-Likelihood-Methode

Für kompliziertere Verteilungen ist es meist schwer bis unmöglich, eine explizite Schätzfunktion anzugeben. Daher ist man an heuristischen Verfahren zum Finden eines Schätzwertes interessiert. In diesem Abschnitt werden wir als Beispiel für ein solches Verfahren die *Maximum-Likelihood-Methode* betrachten. Grundlage dieser Methode ist die Likelihood-Funktion, welche ein Maß für die Güte einer Schätzung liefert.

### Definition 10.5 (Likelihood-Funktion)

Sei  $g_\theta : \mathbb{R}^n \rightarrow \mathbb{R}$  eine durch  $\theta \in \mathbb{R}$  parametrisierte Dichtefunktion. Die Funktion  $L_x : \mathbb{R} \rightarrow \mathbb{R}$  definiert durch

$$L_x(\theta) = g_\theta(x)$$

heißt **Likelihood-Funktion** zu  $g_\theta$ .

Die Likelihood-Funktion dreht für die Dichte  $g_\theta$  praktisch den Spieß um. Sie liefert für eine gegebene Ausprägung  $x$  einer Zufallsvariablen eine Heuristik dafür, wie plausibel ein Parameter  $\theta$  zur Erklärung dieser Beobachtung ist. Für Stichprobenvariablen  $\vec{X} = (X_1, \dots, X_n)$  einer  $f_\theta$ -verteilten Grundgesamtheit erhalten wir aufgrund der stochastischen Unabhängigkeit die Gesamtdichte

$$g_\theta(\vec{x}) = \prod_{i=1}^n f_\theta(x_i). \quad (10.12)$$

Die Likelihood-Funktion der Stichprobe  $\vec{x} = (x_1, \dots, x_n)$  ist damit

$$L_{\vec{x}}(\theta) = \prod_{i=1}^n f_\theta(x_i). \quad (10.13)$$

Die durch die Likelihood-Funktion gelieferte Güteheuristik beruht auf der Erwartung, dass Ausprägungen  $x$  der Grundgesamtheit mit einem hohen Dichtewert  $f_\theta(x)$  wesentlich häufiger auftreten als solche mit einem niedrigen Dichtewert. Insgesamt wird also davon ausgegangen, dass die kumulierten Wahrscheinlichkeiten der Stichprobe so groß wie möglich sind. Gesucht wird nun der Schätzwert  $\hat{\theta}$ , für den diese kumulierten Wahrscheinlichkeiten tatsächlich maximal werden. Diese Überlegung führt zur Maximum-Likelihood-Methode.

### Definition 10.6 (Maximum-Likelihood-Schätzer)

Sei  $L_{\vec{x}}$  die Likelihood-Funktion der Stichprobe  $\vec{x} = (x_1, \dots, x_n)$  einer  $f_\theta$ -verteilten Grundgesamtheit. Die Schätzfunktion

$$h_L(\vec{x}) = \operatorname{argmax}_{\theta \in \mathbb{R}} L_{\vec{x}}(\theta)$$

heißt **Maximum-Likelihood-Schätzfunktion** und definiert den **Maximum-Likelihood-Schätzer**.

Für eine gegebene Stichprobe ist das Schätzen von  $\hat{\theta}$  mit der Maximum-Likelihood-Methode also ein eindimensionales Optimierungsproblem in  $\theta$ . Ist  $L$  in  $\theta$  einfach differenzierbar, so erhält man entsprechend

$$\frac{\partial L_{\vec{x}}}{\partial \theta} = 0 \quad (10.14)$$

als notwendiges Kriterium für einen Schätzwert.

Durch die Komposition der Likelihood-Funktion mit einer streng monoton steigenden Funktion wird das Ergebnis der Schätzung nicht verändert. Gerade, wenn  $f_\theta$  einen Exponentialterm beinhaltet, ist es häufig einfacher,  $L_{\vec{x}}$  vor der Lösung des Optimierungsproblems zu logarithmieren. Das Ergebnis wird nicht verfälscht, schließlich ist der Logarithmus streng monoton steigend. Auf diese Weise kommen wir zur logarithmierten Variante der Likelihood-Funktion.

**Definition 10.7 (Log-Likelihood-Funktion)**

Sei  $g_\theta : \mathbb{R}^n \rightarrow \mathbb{R}$  eine durch  $\theta \in \mathbb{R}$  parametrisierte Dichtefunktion. Die Funktion  $\mathcal{L}_x$  definiert durch

$$\mathcal{L}_x(\theta) = \ln(g_\theta(x))$$

heißt **Log-Likelihood-Funktion** zu  $g_\theta$ .

Für eine Stichprobe  $\vec{x}$  nimmt die Log-Likelihood-Funktion nach Vereinfachung des Logarithmus die Form

$$\ln(L_{\vec{x}}(\theta)) = \ln\left(\prod_{i=1}^n f_\theta(x_i)\right) = \sum_{i=1}^n \underbrace{\ln(f_\theta(x_i))}_{:=\ell_i(\theta)} = \sum_{i=1}^n \ell_i(\theta) \quad (10.15)$$

an.

### 10.3 Konfidenzschätzung

Bisher hatten wir uns mit der Schätzung eines konkreten Wertes für einen Parameter beschäftigt. Wir haben festgestellt, dass bei einem sinnvoll gewählten Schätzer die Genauigkeit der Schätzung mit größeren Stichproben zunimmt. Allerdings ist es oft nicht möglich, beliebig große Stichproben zu wählen. Daher interessieren wir uns für eine Abschätzung der Abweichung unserer Schätzung vom tatsächlichen Wert.

#### 10.3.1 Bestimmung von Konfidenzintervallen

Wir beschäftigen uns nun mit der Bestimmung sogenannter *Konfidenzintervalle*. Für einen Schätzer  $\hat{\Theta}$  eines Parameters  $\theta$  suchen wir ein Intervall  $(G_1, G_2)$  derart, dass

$$P(G_1 < \theta < G_2) = 1 - \alpha. \quad (10.16)$$

Dabei sind  $G_1 = G_1(X_1, \dots, X_n)$  und  $G_2 = G_2(X_1, \dots, X_n)$  Funktionen der Stichprobenvariablen und somit selbst auch Zufallsvariablen. In der Regel hängen sie über den Schätzer  $\hat{\Theta}$  von den Stichprobenvariablen ab. Für eine konkrete Stichprobe  $\vec{x} = (x_1, \dots, x_n)$  erhält man durch Einsetzen - analog zur Bestimmung des Schätzwertes - ein konkretes Konfidenzintervall  $(g_1, g_2) = (G_1(x_1, \dots, x_n), G_2(x_1, \dots, x_n))$ . Dieses ist nach Konstruktion der Stichprobe entsprechend gerade so groß, dass die Wahrscheinlichkeit dafür, dass der unbekannte Parameter  $\theta$  sich tatsächlich in dem Intervall  $(g_1, g_2)$  befindet, gerade  $1 - \alpha$  beträgt.

In Worten ist die Fragestellung also: Finde, in Abhängigkeit von der Stichprobe, ein Intervall  $(g_1, g_2)$ , in welchem sich der gesuchte Parameter mit hoher Wahrscheinlichkeit befindet.

**Definition 10.8 (Konfidenzintervall)**

Sei  $\hat{\Theta}$  ein Schätzer für einen Parameter  $\theta$  einer Verteilung und  $\vec{X}$  ein Vektor von Stichprobenvariablen. Seien  $g_1, g_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ . Das Zufallsintervall

$$(G_1 := g_1(\vec{X}), G_2 := g_2(\vec{X}))$$

heißt **Konfidenzintervall** für den Schätzer  $\hat{\Theta}$ , falls für  $\alpha \in [0, 1]$  gilt:

$$P(G_1 < \theta < G_2) = 1 - \alpha.$$

Die Zufallsvariablen  $G_1, G_2$  heißen **Konfidenzgrenzen**, die Wahrscheinlichkeit  $1 - \alpha$  ist das **Konfidenzniveau** und  $\alpha$  heißt **Irrtumswahrscheinlichkeit**.

In der Praxis interessiert man sich in der Regel für eine Umgebung der Schätzung  $\hat{\Theta}$ , mit von der Schätzung additiv oder multiplikativ abhängigen Grenzen. Eine häufige Wahl für die Grenzfunktionen ist

$$G_1 = \hat{\Theta} - \delta_1, \quad G_2 = \hat{\Theta} + \delta_2 \quad (10.17)$$

für zwei noch zu bestimmende Größen  $\delta_1, \delta_2 > 0$ . Diese obere bzw. untere Abweichung muss anschließend aus der Forderung

$$P(\hat{\Theta} - \delta_1 < \theta < \hat{\Theta} + \delta_2) = 1 - \alpha \quad (10.18)$$

bestimmt werden. Häufig ist vereinfachend  $\delta_1 = \delta_2$ , das Konfidenzintervall ist also um die Schätzung zentriert. Analog kann man die Grenzen auch in multiplikativer Abhängigkeit vom Schätzer als  $G_1 = \delta_1 \cdot \hat{\Theta}, G_2 = \delta_2 \cdot \hat{\Theta}$  setzen. Für Symmetrie bietet sich  $\delta_1 = 1/\delta_2$  an.

Anstatt einer beidseitigen Eingrenzung kann auch eine *einseitige Konfidenzschätzung* gewünscht sein. Man ist also lediglich an einer oberen oder unteren Grenze für den Parameter  $\theta$  interessiert. Ein Konfidenzintervall mit einer unendlichen Grenze, also

$$(-\infty, G_2) \quad \text{beziehungsweise} \quad (G_1, \infty), \quad (10.19)$$

heißt *einseitiges Konfidenzintervall*. Sind beide Grenzen endlich, ist es ein *zweiseitiges Konfidenzintervall*. Dementsprechend spricht man von einseitiger und zweiseitiger Konfidenzschätzung.

### 10.3.2 Konfidenzschätzung des Erwartungswertes der Normalverteilung

Als konkretes und äußerst relevantes Beispiel betrachten wir die Schätzung des Erwartungswertes  $\mu$  der Normalverteilung. Wir betrachten die Konfidenzschätzung für bekannte sowie unbekannte Varianz  $\sigma^2$ .

**Einseitige Konfidenzschätzung des Erwartungswertes bei bekannter Varianz** Wir betrachten eine normalverteilte Zufallsvariable  $X \sim \mathcal{N}(\mu, \sigma^2)$  deren Varianz  $\sigma^2$  bekannt ist. Das Ziel ist eine Konfidenzschätzung für den Erwartungswert  $\mu$ . Für die Stichprobenvariablen  $(X_1, \dots, X_n)$  wählen wir den bereits ergründeten Schätzer  $\hat{M} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . Zunächst interessieren wir uns für ein einseitiges Konfidenzintervall  $(-\infty, G_2)$  zum Konfidenzniveau  $1 - \alpha$ . Wir suchen  $\delta > 0$  sodass

$$P(\mu < \bar{X} + \delta) = 1 - \alpha. \quad (10.20)$$

Der Schätzer  $\bar{X}$  ist normalverteilt mit dem selben Erwartungswert  $\mu$  und einer für wachsendes  $n$  verschwindenden Varianz:  $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ . Wir können ihn also normalisieren als

$$Z := \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}. \quad (10.21)$$

Anstatt der Abweichung  $\delta$  suchen wir dementsprechend jetzt die normalisierte Abweichung  $\frac{\delta\sqrt{n}}{\sigma}$ . Wir stellen die Wahrscheinlichkeit (??) um zu

$$P\left(Z > -\frac{\delta\sqrt{n}}{\sigma}\right) = 1 - P\left(Z \leq -\frac{\delta\sqrt{n}}{\sigma}\right) = 1 - \Phi\left(-\frac{\delta\sqrt{n}}{\sigma}\right) = \Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) \stackrel{!}{=} 1 - \alpha. \quad (10.22)$$

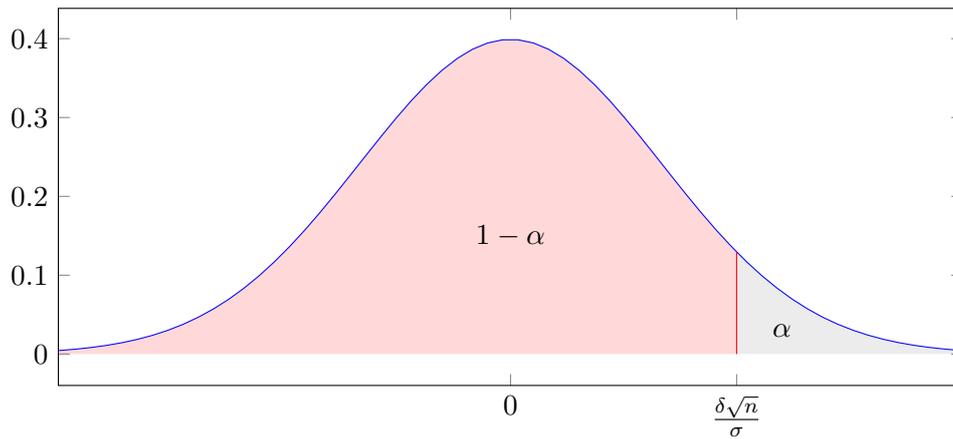


Abbildung 10.1: Das einseitige Konfidenzintervall in der Verteilung von  $Z$

Dies entspricht gerade dem  $1 - \alpha$ -Quantil  $z_{1-\alpha}$  der Standardnormalverteilung. Die Quantile der Normalverteilung lassen sich einem Tabellenwerk entnehmen. So finden wir also die einseitige Abweichung

$$\frac{\delta\sqrt{n}}{\sigma} = z_{1-\alpha} \Leftrightarrow \delta = \frac{\sigma}{\sqrt{n}} z_{1-\alpha}. \quad (10.23)$$

Wir erkennen: Die Größe des einseitigen Konfidenzintervalls hängt lediglich von der Stichprobengröße und der Standardabweichung ab. Erwartungsgemäß schrumpft das Konfidenzintervall mit wachsender Stichprobengröße, allerdings nur umgekehrt proportional zur Wurzel von  $n$ . Als Endergebnis erhalten wir das einseitige Konfidenzintervall

$$I_\mu = \left( -\infty, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right). \quad (10.24)$$

**Zweiseitige Konfidenzschätzung des Erwartungswertes bei bekannter Varianz** Zur Bestimmung eines zweiseitigen, symmetrischen Konfidenzintervalls  $(\bar{X} - \delta, \bar{X} + \delta)$  für den Erwartungswert  $\mu$  bei bekannter Varianz  $\sigma^2$  gehen wir analog vor. Die Forderung

$$P(\bar{X} - \delta < \mu < \bar{X} + \delta) = 1 - \alpha. \quad (10.25)$$

stellen wir um zu

$$\begin{aligned} & P(\bar{X} - \delta < \mu < \bar{X} + \delta) \\ &= P(-\delta < \mu - \bar{X} < \delta) \\ &= P(|\bar{X} - \mu| < \delta) \\ &= P\left(\left|\sqrt{n}\frac{\bar{X} - \mu}{\sigma}\right| < \frac{\delta\sqrt{n}}{\sigma}\right) \\ &= P\left(|Z| < \frac{\delta\sqrt{n}}{\sigma}\right) \stackrel{!}{=} 1 - \alpha. \end{aligned}$$

Wie hieraus  $\delta$  bestimmt werden kann ist noch nicht ganz klar. Abbildung ?? zeigt das gesuchte symmetrische Konfidenzintervall als Bereich unter der Normalverteilungsdichte. Der Bereich soll den Flächeninhalt  $1 - \alpha$  haben; mit der Symmetrie ist also auch der Flächeninhalt der beiden Randstreifen klar. Die normalisierte Abweichung  $\frac{\delta\sqrt{n}}{\sigma}$  ist gerade der Radius des roten Bereichs.

Durch die Abbildung wird deutlich: die obere Grenze des normalisierten Konfidenzintervalls ist gerade das  $1 - \frac{\alpha}{2}$ -Quantil  $z_{1-\frac{\alpha}{2}}$  der Standardnormalverteilung. Somit ist die Abweichung in beide Richtungen

$$\delta = \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \quad (10.26)$$

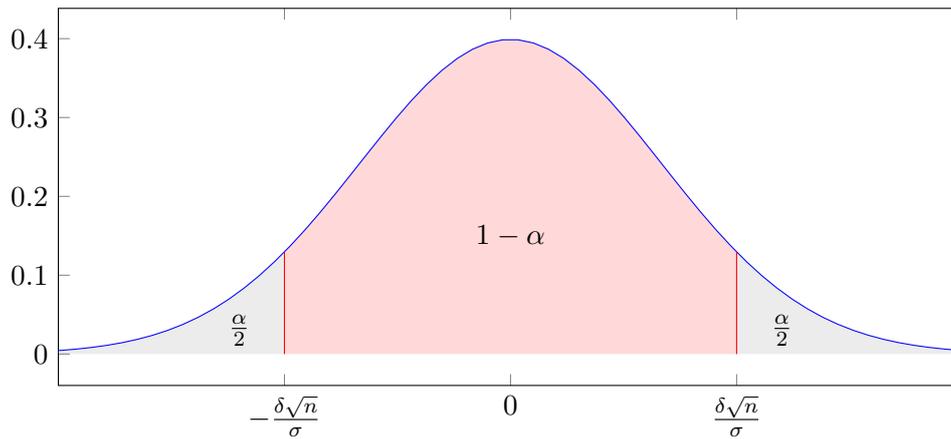


Abbildung 10.2: Das zweiseitige Konfidenzintervall in der Verteilung von  $Z$

und das zweiseitige Konfidenzintervall für den Erwartungswert ist

$$I_\mu = \left( \bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right). \quad (10.27)$$

**Schätzung des Erwartungswertes bei unbekannter Varianz** Nun betrachten wir eine normalverteilte Zufallsvariable  $X$  mit unbekannter Varianz, deren Erwartungswert wir abschätzen möchten. Für den Erwartungswert nutzen wir weiter das arithmetische Mittel  $\bar{X}$  als Schätzer. Die unbekannt Varianz wird durch die empirische Varianz  $S^2$  (??) geschätzt. Die Zufallsvariable  $\bar{X}$  hat somit Erwartungswert  $\mu$  und die geschätzte Varianz  $S^2/\sqrt{n}$ . Die normalisierte Zufallsvariable

$$Z := \sqrt{n} \frac{\bar{X} - \mu}{S} \quad (10.28)$$

ist nun **Student- $t$ -verteilt**.

Für eine einseitige Konfidenzschätzung zum Konfidenzniveau  $1 - \alpha$  suchen wir wieder die maximale Abweichung  $\delta_1$  für das Konfidenzintervall  $(-\infty, \bar{X} + \delta_1)$ . Normalisiert mit  $S/\sqrt{n}$  ergibt dies die Forderung

$$P\left(Z < \frac{\delta_1 \sqrt{n}}{S}\right) = 1 - \alpha \quad \Leftrightarrow \quad \delta_1 = \frac{S}{\sqrt{n}} \cdot t_{1-\alpha; n-1} \quad (10.29)$$

mit  $t_{1-\alpha; n-1}$  dem  $1 - \alpha$ -Quantil der Student- $t$ -Verteilung mit  $n - 1$  Freiheitsgraden.

Analog folgt für das zweiseitige Konfidenzintervall  $(\bar{X} - \delta_2, \bar{X} + \delta_2)$

$$P\left(|Z| < \frac{\delta_2 \sqrt{n}}{S}\right) = 1 - \alpha \quad \Leftrightarrow \quad \delta_2 = \frac{S}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}; n-1} \quad (10.30)$$

wobei wieder  $t_{1-\frac{\alpha}{2}; n-1}$  das  $1 - \frac{\alpha}{2}$ -Quantil der Student- $t$ -Verteilung mit  $n - 1$  Freiheitsgraden ist.

## 11 Hypothesentests

Im vorherigen Abschnitt haben wir uns mit der Abschätzung von Parametern anhand von Stichproben beschäftigt. In diesem Abschnitt gehen wir davon aus, dass wir bereits eine Annahme über den Wert eines Parameters haben, und diese Annahme auf Korrektheit prüfen möchten.

## A Integration im $\mathbb{R}^2$

Erläuterungen zu der Verallgemeinerung der Integration auf den  $\mathbb{R}^2$ . Diese Kenntnisse werden nun im Zusammenhang mit der Nutzung von 2 (oder mehr) Zufallsvariablen  $X$  und  $Y$  wichtig.

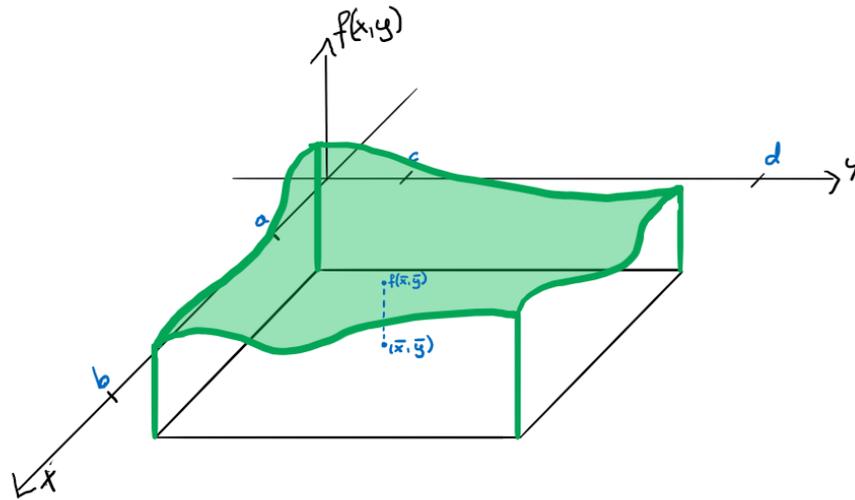


Abbildung A.1: Graphisch dargestellt

### A.1 Parameterintegral und Integralfunktion

#### Definition A.1 (Parameterintegral)

Eine Funktion  $F : I \rightarrow \mathbb{R}$  mit  $I \subset \mathbb{R}$  und der Gestalt

$$F(x) = \int_a^b f(t, x) dt$$

heißt Integralfunktion und das Integral wird Parameterintegral genannt.

Parameterintegrale gibt es als eigentliche und uneigentliche Parameterintegrale. Der Unterschied liegt darin, dass uneigentliche Parameterintegrale nur mit Tricks partiell differenzierbar sind.

### A.2 Projizierbarkeit und Umsetzung der Integration

Zur Umsetzung der Integration in 2 Dimensionen wollen wir uns zuerst nur ein Rechteck anschauen. Über dieses Rechteck lässt sich integrieren in dem zuerst das innere Integral auf geleitet wird und die Integralgrenzen eingesetzt werden. Dieser Vorgang wird dann auch für das außen liegende Integral wiederholt. somit integriert man nicht mehr über Streifen wie im  $\mathbb{R}$  sondern über kleine Rechtecke. Der Flächeninhalt dieser Rechtecke soll dann (ebenso wie bei den Streifen) gegen 0 gehen.

Im Folgenden wollen wir uns nun Projizierbarkeit anschauen. Diese soll dann genutzt werden um durch Substitution im Integral andere Mengen auf den Fall des Rechtecks zu reduzieren.

#### Definition A.2 (Projizierbarkeit)

$G$  sei eine nichtleere Teilmenge des  $\mathbb{R}^2$ .  $G$  heißt  $y$ -projizierbar wenn es auf einem Intervall  $[a, b]$  der  $x$ -Achse stetige Funktionen  $\underline{y}(x)$  und  $\bar{y}(x)$  gibt mit

$$\underline{y}(x) \leq \bar{y}(x) \forall x \in [a, b]$$

Dann gilt:

$$G = \{(x, y) \mid x \in [a, b], \underline{y}(x) \leq y \leq \bar{y}(x)\}$$

Analog ist die x-Projizierbarkeit definiert. Ist eine Menge sowohl x- als auch y- projizierbar, so heißt sie Standardmenge.

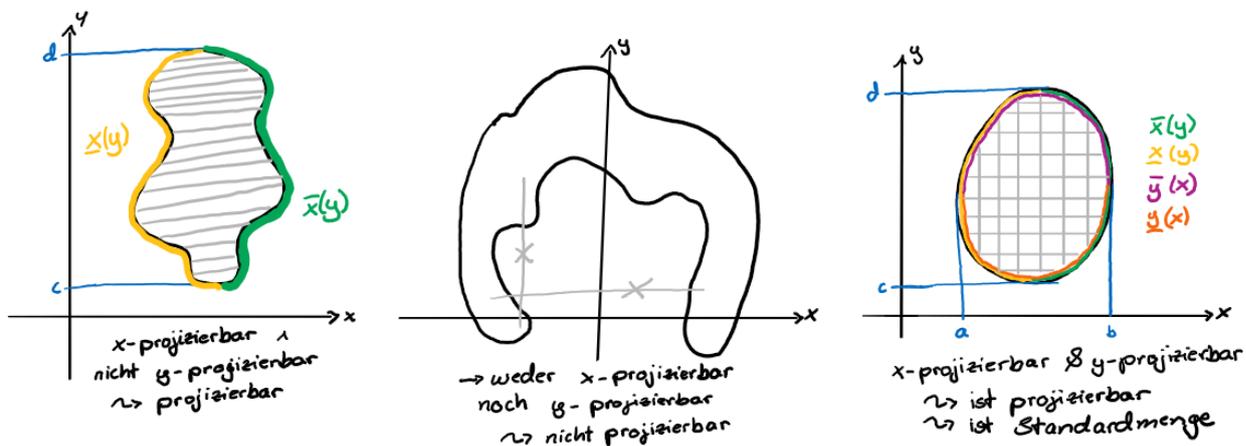


Abbildung A.2: Beispiele zur Projizierbarkeit

### Beispiel A.3

## A.3 Transformationsatz

Mit dem Transformationsatz wird nun über eine Menge  $G$  durch eine Menge  $H$  auf einer Funktion  $f(x, y)$  integriert. Dabei wird zuerst die Funktion  $T : H \rightarrow G$  definiert, die dann als Substitution verwendet wird.

# B Übersicht der Wahrscheinlichkeitsverteilungen

## B.1 Diskrete Verteilungen

Verteilung	Notation	Z-Dichte	EX	Var X	Faltung	Beispiel
Bernoulli	$B(p)$	$f^X(1) = p$	$p$	$p(1-p)$		(gezinkter) Münzwurf
Laplace	$L$	$f^X(a) = 1/N$	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$		fairer Würfel
Binomial	$B(n, p)$	$f(k) = \binom{n}{k} p^k q^{n-k}$	$np$	$np(1-p)$	$B(n+m, p) = B(n, p) * B(m, p)$	
negative Binomial	$Nb^+(r, p)$	$Nb^+(r, p; k) = \binom{k-1}{r-1} p^r q^{k-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$		
negative Binomial	$Nb^0(r, p)$	$Nb^0(r, p; k) = \binom{k+r-1}{r-1} p^r q^k$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$		
geometrische	$Geo^+(p)$	$geo^+(p, k) = pq^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$Nb^+(r, p) = Geo^+(p) * \dots * Geo^+(p)$	Warten mit Erfolg
geometrische	$Geo^0(p)$	$geo^0(p, k) = pq^k$	$\frac{1}{p}$	$\frac{1-p}{p^2}$		Warten ohne Erfolg
geometrische	$\pi(\lambda)$	$f(k) = (1-q)q^k$	$\lambda$	$\lambda$		
Poisson		$f(k) = \pi(\lambda; k) = e^{-\lambda} \frac{\lambda^k}{k!}$	$\lambda$	$\lambda$		
empirische		$f = \frac{1}{n} \sum_{i=1}^n 1_{x_i}(x)$				
hypergeometrische	$H(N, K, n)$	$h(N, N, n; k) = f^{Zn}(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$	$n \frac{K}{N}$	$n \frac{K(N-K)(N-n)}{N^2(N-1)}$		

## B.2 Stetige Verteilungen

Verteilung	Notation	R-Dichte	Verteilungsfunktion	EW	Var X	Faltung
Rechteck	$\mathcal{R}(a, b)$	$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b), \\ 0 & \text{sonst} \end{cases}$	$F(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x \leq b, \\ 1, & x > b. \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
Exponential	$Exp(\alpha)$	$f(x) = \alpha e^{-\alpha x} 1_{(0, \infty)}(x)$ , wenn $\alpha > 0$	$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\alpha x}, & x > 0 \end{cases}$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$Exp(\alpha) = \Gamma_{\alpha, 1}$
Normal	$\mathcal{N}(a, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, x \in \mathbb{R}$		$a$	$\sigma^2$	$N(a, \sigma^2) * N(b, \tau^2) = N(a+b, \sigma^2 + \tau^2)$
St. Normal	$\mathcal{N}(0, 1)$	$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}$		$0$	$1$	
Gamma	$\Gamma_{\alpha, v}(v) = \int_0^\infty x^{v-1} e^{-x} dx$	$\gamma(x) = \begin{cases} \frac{\alpha^v}{\Gamma(v)} x^{v-1} e^{-\alpha x}, & x > 0 \\ 0, & x \geq 0 \end{cases}$	$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$	$\frac{v}{\alpha}$	$\frac{v}{\alpha^2}$	$\Gamma_{\alpha, \mu} * \Gamma_{\alpha, v} = \Gamma_{\alpha, \mu+v}$
Beta( $\mu, v$ )	$be_{\mu, v}$	$be_{\mu, v} = \begin{cases} \frac{\Gamma(\mu+v)}{\Gamma(\mu)\Gamma(v)} x^{\mu-1} (1-x)^{v-1}, & 0 < x < 1 \\ 0, & x \leq 0 \end{cases}$		$\frac{\mu}{\mu+v}$		