

# STOCHASTIK FÜR INFORMATIK UND INGENIEURSWISSENSCHAFTEN

---

Noemi Kurt

INSTITUT FÜR MATHEMATIK  
TU BERLIN

Vorlesungsskript<sup>1</sup>

Vorläufige Version, unvollständig und fehlerhaft!

---

<sup>1</sup>© Noemi Kurt  
*Date:* 13. April 2016.

## INHALTSVERZEICHNIS

Einleitung	4
<b>Teil 1. Grundlagen der Wahrscheinlichkeitstheorie</b>	<b>6</b>
1. Wahrscheinlichkeitsräume und Zufallsvariablen	6
1.1. Wahrscheinlichkeiten und Ereignisse	6
1.2. Häufigkeiten und Wahrscheinlichkeiten	11
2. Bedingte Wahrscheinlichkeiten, Unabhängigkeit	13
2.1. Bedingte Wahrscheinlichkeiten	13
2.2. Mehrstufige Experimente und Formel von der Gesamtwahrscheinlichkeit	14
2.3. Unabhängigkeit von Ereignissen	17
2.4. Bayes'sche Umkehrformel	19
3. Diskrete Zufallsvariablen und Verteilungen	21
3.1. Zufallsvariablen, Verteilung, Verteilungsfunktion	21
3.2. Häufigkeiten, Histogramme und Verteilungen	26
3.3. Bernoulli-Verteilung, Binomialverteilung, geometrische Verteilung	26
3.4. Poisson-Verteilung	30
3.5. Zusammenfassung	31
3.6. Weitere diskrete Verteilungen und ihre Anwendungen	31
4. Gemeinsame Verteilung, Unabhängigkeit von Zufallsvariablen	32
4.1. Gemeinsame Verteilung, bedingte Verteilung	32
4.2. Unabhängigkeit von Zufallsvariablen	35
5. Kenngrößen für Zufallsvariablen	37
5.1. Erwartungswert	37
5.2. Anwendung: Laufzeit von Quicksort	40
5.3. Varianz	40
5.4. Markov- und Chebyshev-Ungleichung	42
5.5. Kovarianz und Korrelation	42
6. Zufallsvariablen mit Dichte	47
6.1. Allgemeine Zufallsvariablen, Verteilungsfunktion, Dichten	47
6.2. Exponentialverteilung	50
6.3. Normalverteilung	52
6.4. Weitere Beispiele	54
7. Markov-Ketten	55

7.1.	Markov-Ketten und stochastische Matrizen	55
7.2.	Eigenschaften von Markov-Ketten	60
7.3.	Invariante Verteilungen	61
7.4.	Anwendungen	64

## EINLEITUNG

Das Wort *Stochastik* bezeichnet die Lehre von der Gesetzmäßigkeit des Zufalls. Sie unterteilt sich grob in die *Wahrscheinlichkeitstheorie*, welche sich mit mathematischer Modellierung und Formalisierung beschäftigt, und die theoretische Grundlage für alle Anwendungen bildet, sowie die *Statistik*, welche die Analyse und Interpretation von Messwerten und Daten aus zufälligen Vorgängen mittels wahrscheinlichkeitstheoretischen Methoden zum Inhalt hat.

Der *Zufall* kann dabei entweder ein “echter” Zufall sein, oder ein lediglich angenommener Zufall, der zum Beispiel aus einem Mangel an Information herrühren kann, oder als Effekt vieler kleiner (einzeln nicht genau unkontrollierbarer) Ereignisse auftritt. Für die mathematische Modellierung ist es dabei meistens nicht relevant, woher die Zufälligkeit eines Prozesses genau stammt, so dass man die philosophische Frage über die Existenz von echtem Zufall umgehen kann, und gleichzeitig eine universelle Anwendbarkeit der Theorie erreicht wird. Beispiele für zufällig ablaufende Prozesse sind

- Würfeln, Lottoziehung
- Warteschlangenbildung, z.B. in der Verkehrsmodellierung
- Lagerbestände, z.B. im Versandhandel
- Stochastische Algorithmen (z.B. Googles PageRank)
- Bildung von Netzwerken
- Börsenkurse
- Ausbreitung von Epidemien
- Genetische Variabilität
- etc.

Im ersten Teil der Vorlesung werden wir uns mit der Wahrscheinlichkeitstheorie beschäftigen, da diese das Fundament bildet, auf dem Anwendungen aufbauen. Gleichzeitig liefert sie die Sprache, in die Probleme aus der Wirklichkeit übersetzt werden (müssen), um die Macht der Mathematik anwenden zu können. Dabei wird notgedrungen die Wirklichkeit idealisiert und formalisiert.

Für Anwendungen, welche insbesondere im zweiten und dritten Teil der Vorlesung wichtig werden, ist es deshalb wichtig, dass das mathematische Modell immer wieder mit der Wirklichkeit abgeglichen wird, und gegebenenfalls angepasst. Dies geschieht in vielen Fällen mit Hilfe der Statistik.

Wegen ihrer Universalität ist die Wahrscheinlichkeitstheorie sogar in manchen Bereichen nützlich, in denen a priori gar kein Zufall herrscht. Jedoch kann eine künstliche Zufälligkeit z.B. bei Suchalgorithmen enorm nützlich sein. Im dritten Teil der Vorlesung werden wir Algorithmen kennenlernen, welche auf wahrscheinlichkeitstheoretischen Methoden basieren.

Für diese Vorlesung wird die Kenntniss des Stoffes aus den Vorlesungen “Analysis I für Ingenieure” und “Lineare Algebra für Ingenieure” vorausgesetzt. Wichtig sind insbesondere

die gute Beherrschung der Grundbegriffe und Notationen was Mengen und Funktionen angeht, sowie das souveräne Arbeiten mit Integralen, Reihen, linearen Gleichungssystemen und Matrizen.

Jedes Kapitel der Vorlesung beginnt mit einem sehr knappen Ausblick auf seinen Inhalt, sowie einer Zusammenstellung der Lernziele des Kapitels und ggf. einem Hinweis auf besonders relevante Vorkenntnisse in diesem Kapitel. Die Kapitel bauen aufeinander auf.

## Teil 1. Grundlagen der Wahrscheinlichkeitstheorie

### 1. WAHRSCHEINLICHKEITSRÄUME UND ZUFALLSVARIABLEN

In diesem Kapitel lernen wir Wahrscheinlichkeitsräume kennen. In diesem Rahmen werden üblicherweise wahrscheinlichkeitstheoretische Modelle formuliert. Wie fast immer in der Mathematik werden wir uns erst einige Begriffe und Notationen aneignen müssen. Dies ist notwendig, um die Wirklichkeit in einer Art und Weise zu formalisieren und zu modellieren, welche die präzise und unmissverständliche mathematische Behandlung möglich macht. Wir müssen uns also sozusagen die Grundvokabeln der Stochastik aneignen.

Die **Lernziele** dieses Kapitel sind:

- Die Grundbegriffe und Standardnotation für Wahrscheinlichkeiten und Ereignisse kennen
- Geeignete Wahrscheinlichkeitsräume für einfache Zufallsexperimente angeben können
- Wahrscheinlichkeiten für Ereignisse in endlichen Wahrscheinlichkeitsräumen berechnen können.

Benötigte **Vorkenntnisse**: Grundbegriffe und Notation von Mengen und Funktionen wie sie z.B. aus der Analysis I bekannt sind.

#### 1.1. Wahrscheinlichkeiten und Ereignisse.

**Definition 1.1** (Zufallsexperiment). Ein *Zufallsexperiment* ist ein reproduzierbarer Vorgang mit zufälligem oder (a priori) unbestimmtem Ausgang.

Beispiele für Zufallsexperimente:

- Werfen einer Münze
- Lottoziehung
- Auswahl einer Stichprobe von  $k$  Bauteilen für einen Funktionstest, aus einer Gesamtzahl von  $n$  gleichartigen Teilen aus einer Produktion
- Zufälliges Surfen im Netz: Auswahl eines beliebigen Links auf einer bestimmten Website
- Zufällige Auswahl eines Punktes auf einer Zielscheibe, z.B. durch Werfen eines Dartspfeiles

Auf den ersten Blick mögen Experimente wie Münzwurf oder Dartspfeil uninteressant wirken, jedoch kann man kompliziertere zufällige Vorgänge oft auf solche einfachen Teilabläufe reduzieren, für die man dann interessante Größen einfach berechnen kann. Münzwürfe dienen z.B. zur Modellierung von zufälligen Entscheidungen, und Abfolgen von zufälligen Entscheidungen können zu geeigneten komplexen Algorithmen, z.B. zufällige Sortieralgorithmen, “zusammengebaut” werden. Wir werden im Verlauf der Vorlesung viele solche Beispiele kennenlernen.

**Definition 1.2** (Ergebnis, Ereignis). Ein möglicher Ausgang eines Zufallsexperiments ist ein *Ergebnis*. Die Menge aller möglichen Ergebnisse eines bestimmten Zufallsexperiment wird üblicherweise mit dem griechischen Buchstaben  $\Omega$  (Omega) bezeichnet. Für ein einzelnes Ergebnis schreiben wir entsprechend  $\omega \in \Omega$ .

Ein *Ereignis* ist eine Teilmenge von  $\Omega$ , in Mengenschreibweise  $A \subseteq \Omega$ . Ein Ereignis besteht also aus einem oder mehreren Ergebnissen.

**Beispiel 1.3.** Einfaches Würfeln ist ein Zufallsexperiment mit sechs möglichen Ergebnissen:

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

Beispiele für Ereignisse:

$$A = \{2, 4, 6\} = \{\omega \in \Omega : \omega \text{ ist eine gerade Zahl}\}$$

oder

$$B = \{3, 4, 5, 6\} = \{\omega \in \Omega : \omega \geq 3\}.$$

In Worten ausgedrückt beschreibt  $A$  das Ereignis, dass eine gewürfelte (und damit zufällige) Zahl gerade ist,  $B$  das Ereignis, dass eine gewürfelte Zahl mindestens den Wert 3 hat. Mit Hilfe von Mengenoperationen können neue Ereignisse beschrieben werden:

$$A \cap B = \{4, 6\}.$$

Dies ist das Ereignis dass  $A$  und  $B$  beide eintreten, das heißt, dass eine gewürfelte Zahl sowohl gerade als auch größer oder gleich 3 ist. Das Ereignis

$$A \cup B = \{2, 3, 4, 5, 6\}$$

bedeutet dass  $A$  oder  $B$  (oder beide) eintreten, d.h. dass die Zahl gerade oder mindestens 3 ist. Das Ereignis

$$A^c := \Omega \setminus A = \{1, 3, 5\}$$

ist das Komplement von Ereignis  $A$ , also das Ereignis, dass  $A$  nicht eintritt, bzw. dass die Zahl ungerade ist.  $A^c$  bezeichnet man auch als das *Gegenereignis* von  $A$ . Manchmal wird gleichbedeutend mit der Notation  $A^c$  auch  $\bar{A}$  geschrieben.

*Bemerkung.* Wie im obigen Beispiel klar werden sollte, kann man Ereignisse und (Teil-)Mengen eins zu eins in Verbindung miteinander setzen, und auch die Verknüpfung von Ereignissen in Mengenoperationen übersetzen, und umgekehrt. Die Tabelle stellt das noch einmal dar.

Tabelle (siehe Vorlesung)

**Beispiel 1.4** (Zufällige Auswahl eines Punktes aus einem Quadrat). siehe Vorlesung

Die beiden Beispiele haben einen großen quantitativen Unterschied: Im ersten Beispiel gibt es nur endlich viele mögliche Ergebnisse, nämlich genau 6. Im zweiten Beispiel gibt es unendlich (sogar überabzählbar) viele, denn ein Quadrat im  $\mathbb{R}^2$  besteht aus unendlich vielen einzelnen Punkten. Die Wahrscheinlichkeitstheorie ist in beiden Fällen anwendbar, und die Grundbegriffe sind für beide Fälle definiert. Viele Konzepte werden wir jedoch zuerst einmal an Hand von Beispielen mit endlich vielen möglichen Ergebnissen kennen

lernen.

Als nächstes wollen wir Ereignissen eine Wahrscheinlichkeit zuordnen. Dafür wollen wir uns erst einmal überlegen, was eine Wahrscheinlichkeit anschaulich bedeuten soll. Es hat sich eingebürgert, Wahrscheinlichkeiten mit Zahlen zwischen 0 und 1 anzugeben, wobei 1 bedeutet, dass ein entsprechendes Ereignis “sicher eintritt”, und 0 heißt dass das entsprechende Ereignis “sicher nicht” eintritt. Im Weiteren soll gelten, dass die Wahrscheinlichkeit für ein Ereignis, welches auf verschiedene, unterschiedliche Arten eintreten kann, sich als Summe der Wahrscheinlichkeiten für die unterschiedlichen Möglichkeiten berechnen lässt. Diese Eigenschaften hat der russische Mathematiker A. N. Kolmogorov (1903–1987) zu einer Definition (den sogenannten Axiomen von Kolmogorov) zusammengefasst, welche das Fundament der gesamten Wahrscheinlichkeitstheorie bilden.

**Definition 1.5** (Endlicher Wahrscheinlichkeitsraum). Ein *endlicher Wahrscheinlichkeitsraum* ist ein Paar bestehend aus einer endlichen *Ergebnismenge*  $\Omega$ , und einem *Wahrscheinlichkeitsmaß*  $\mathbb{P}$ , welches jedem Ereignis  $A \subseteq \Omega$  eine Zahl  $\mathbb{P}(A) \in [0, 1]$  zuordnet. Diese Zahl ist die *Wahrscheinlichkeit* des Ereignisses  $A$ .

Formal gesprochen ist ein Wahrscheinlichkeitsmaß eine Abbildung aus der Menge aller Teilmengen von  $\Omega$  nach  $[0, 1]$  mit den Eigenschaften

$$(P0) \quad 0 \leq \mathbb{P}(A) \leq 1 \text{ für alle } A \subseteq \Omega$$

$$(P1) \quad \mathbb{P}(\Omega) = 1$$

$$(P2) \quad \text{Für } A, B \subset \Omega \text{ mit } A \cap B = \emptyset \text{ gilt } \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Ein Wahrscheinlichkeitsmaß ordnet also jedem Ereignis eine Wahrscheinlichkeit zwischen 0 und 1 zu. Die Ergebnismenge  $\Omega$  hat die Wahrscheinlichkeit 1, was so verstanden werden kann, dass sie alle möglichen Ergebnisse enthält, also sicher eintritt. Wenn zwei Ereignisse  $A$  und  $B$  disjunkt sind, d.h. nicht gleichzeitig eintreten können, so berechnet sich die Wahrscheinlichkeit für das Eintreten von  $A$  oder  $B$  durch Addition der Einzelwahrscheinlichkeiten.

**Notation:** Für  $\omega \in \Omega$  schreiben wir oft  $\mathbb{P}(\omega)$  statt  $\mathbb{P}(\{\omega\})$ .

Es ist wichtig, zwischen Ergebnissen und Ereignissen sauber zu unterscheiden: Ergebnisse sind Elemente der Ergebnismenge  $\Omega$ , Ereignisse Teilmengen von  $\Omega$ .

**Beispiel 1.6.** [Fortsetzung von Beispiel 1.3] Beim Würfeln mit einem fairen Würfel sind alle Elemente von  $\Omega = \{1, \dots, 6\}$  gleich wahrscheinlich, also

$$\mathbb{P}(\{\omega\}) = \frac{1}{6} = \frac{1}{|\Omega|} \text{ für alle } \omega \in \Omega.$$

Damit können wir Wahrscheinlichkeiten von Ereignissen berechnen, z.B. von  $A = \{\omega \in \Omega : \omega \text{ ist gerade}\}$  :

$$\mathbb{P}(A) = \mathbb{P}(2) + \mathbb{P}(4) + \mathbb{P}(6) = 3 \cdot \frac{1}{6} = \frac{|A|}{|\Omega|} = \frac{1}{2}$$



Ebenso für  $B = \{\omega \in \Omega : \omega \geq 3\}$  und  $A \cup B = \{2, \dots, 6\}$  :

$$\mathbb{P}(B) = \frac{|B|}{|\Omega|} = \frac{4}{6} = \frac{2}{3},$$

und

$$\mathbb{P}(A \cup B) = \frac{|A \cup B|}{|\Omega|} = \frac{5}{6}.$$

Beachte:  $\mathbb{P}(A \cup B) \neq \mathbb{P}(A) + \mathbb{P}(B) = \frac{7}{6}$ . Dies ist kein Widerspruch zu (P2) in Definition 1.5, da  $A \cap B \neq \emptyset$ . Es gilt  $\mathbb{P}(A \cap B) = \frac{2}{6} = \frac{1}{3}$ . Daraus sehen wir

$$\frac{5}{6} = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \frac{3}{6} + \frac{4}{6} - \frac{2}{6}.$$

Aus der Definition 1.5 können weitere Rechenregeln hergeleitet werden.

**Satz 1.7** (Rechenregeln für Wahrscheinlichkeiten). *Sei  $(\Omega, \mathbb{P})$  ein Wahrscheinlichkeitsraum. Für  $A, B \subseteq \Omega$  gelten*

$$(P3) \quad \mathbb{P}(A^c) = 1 - \mathbb{P}(A),$$

$$(P4) \quad \mathbb{P}(\emptyset) = 0.$$

$$(P5) \quad \mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

$$(P6) \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

*Beweis.* Siehe Vorlesung. □

Im Beispiel 1.6 waren alle Ergebnisse gleich wahrscheinlich. Wahrscheinlichkeitsräume mit dieser Eigenschaft sind besonders wichtig.

**Definition 1.8** (Laplace-Raum). Ein *Laplace-Raum* ist ein endlicher Wahrscheinlichkeitsraum, in dem jedes Ergebnis  $\omega \in \Omega$  dieselbe Wahrscheinlichkeit hat.

**Satz 1.9** (Wahrscheinlichkeiten im Laplace-Raum). *In einem Laplace-Raum ist die Wahrscheinlichkeit für ein Ereignis  $A \subseteq \Omega$  gegeben durch*

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

Die Wahl eines geeigneten Wahrscheinlichkeitsraumes ist oft ein wichtiger Teil der mathematischen Beschreibung eines Problems aus der Wirklichkeit. Dabei gibt es üblicherweise nicht nur eine Möglichkeit. Oft ist es von Vorteil, einen Laplace-Raum zu wählen, da dann die Berechnung von Wahrscheinlichkeiten einfach wird.

**Beispiel 1.10** (Zweifacher Münzwurf). Betrachten wir folgendes einfache Zufallsexperiment: Eine faire Münze wird zweimal geworfen. Wir interessieren uns dafür, wie oft “Zahl” kommt. Was für einen Wahrscheinlichkeitsraum ist dafür geeignet? Dazu müssen wir die möglichen Ergebnisse des Zufallsexperiments “zwei Mal Münze werfen” formalisieren. Eine Möglichkeit besteht darin, ein Ergebnis als Paar  $\omega = (\omega_1, \omega_2)$  zu schreiben, wobei  $\omega_1$  das Ergebnis des ersten Wurfs bezeichnet, also  $\omega_1 = K$  falls der erste Wurf Kopf zeigt, und  $\omega_1 = Z$  falls der erste Wurf Zahl zeigt, und analog  $\omega_2 = K$  oder  $\omega_2 = Z$

falls der zweite Wurf Kopf bzw. Zahl zeigt. Der Ergebnisraum  $\Omega$  ist dann die Menge aller möglichen Ergebnisse, also

$$\Omega = \{(K, K), (K, Z), (Z, K), (Z, Z)\}.$$

Dabei bedeutet eben  $(K, K)$ , dass der erste Wurf Kopf und der zweite Wurf ebenfalls Kopf zeigt,  $(K, Z)$  dass erst Kopf und dann Zahl geworfen wurde, usw.

Das ist aber nicht die einzige Möglichkeit, die wir haben. Wenn wir uns nur dafür interessieren, wie oft Zahl geworfen wird, könnten wir auch sagen, dass wir den Ergebnisraum

$$\bar{\Omega} = \{0, 1, 2\}$$

betrachten, wobei 0 bedeutet, dass nie Zahl kommt (also zweimal Kopf), 1 dass einmal Kopf (und einmal Zahl) geworfen wurde, und 2 dass zweimal Kopf geworfen wurde.

Wir haben also mindestens zwei verschiedene Möglichkeiten, einen Ergebnisraum für dieses Experiment zu wählen. Um einen Wahrscheinlichkeitsraum zu haben, müssen wir in beiden Fällen noch das zugehörige Wahrscheinlichkeitsmaß angeben. Am einfachsten geht das, wenn wir uns das Zufallsexperiment als Baum aufzeichnen:

(Bild siehe Vorlesung)

Dabei stellen wir im ersten Fall fest, dass gilt

$$\mathbb{P}((K, K)) = \mathbb{P}((K, Z)) = \mathbb{P}((Z, K)) = \mathbb{P}((Z, Z)) = \frac{1}{4}.$$

Im zweiten Fall haben wir

$$\bar{\mathbb{P}}((0)) = \frac{1}{4}, \quad \bar{\mathbb{P}}((1)) = \frac{1}{2}, \quad \bar{\mathbb{P}}((2)) = \frac{1}{4}.$$

Somit liegt im ersten Fall ein Laplace-Raum vor, im zweiten aber nicht. Die Fragestellung “wie oft kommt Zahl” kann aber mit beiden Modellierungen untersucht werden.

**Beispiel 1.11** (Wiederholtes Zufallsexperiment). Der zweifache Münzwurf ist ein Beispiel für ein *wiederholtes* Zufallsexperiment. Dieses Prinzip können wir allgemein formulieren. Wir betrachten ein Zufallsexperiment mit  $k$  möglichen Ergebnissen, und wiederholen dieses  $n$  mal unter gleichbleibenden Bedingungen (wobei wir die Ergebnisse in der Reihenfolge ihres Auftretens notieren). Wir bezeichnen mit  $\Omega_n$  die Menge aller Ergebnisse des  $n$ -fach wiederholten Experiments. Dann gilt  $|\Omega_n| = k^n$ , d.h.  $\Omega_n$  hat  $k^n$  Elemente. Wenn alle Ergebnisse des (einmalig ausgeführten) Zufallsexperiments gleich wahrscheinlich sind, das heißt, wenn  $\Omega_1$  ein Laplace-Raum ist, dann ist auch  $\Omega_n$  ein Laplace-Raum.

**Beispiel 1.12** (Mehrfaches Würfeln). Wir würfeln zweimal mit einem fairen Würfel. Als Ergebnisraum wählen wir

$$\Omega_2 = \{(1, 1), (1, 2), (1, 3), \dots, (1, 6), (2, 1), (2, 2), (2, 3), \dots, (6, 5), (6, 6)\},$$

oder, in anderer Schreibweise

$$\Omega_2 = \{\omega = (\omega_1, \omega_2) : \omega_1 \in \{1, \dots, 6\}, \omega_2 \in \{1, \dots, 6\}\}.$$

Das heißt, ein Ergebnis wird als *Paar*  $(\omega_1, \omega_2)$  dargestellt, wobei der erste Eintrag das Ergebnis des ersten Wurfs, und der zweite Eintrag das Ergebnis des zweiten Wurfs ist.  $\Omega$  hat somit  $6^2 = 36$  Elemente, und da die Ergebnisse alle gleich wahrscheinlich sind, handelt

es sich um einen Laplace-Raum, also  $\mathbb{P}(\omega) = \frac{1}{36}$  für jedes  $\omega$  der Form  $(\omega_1, \omega_2) \in \Omega$ .  
Allgemeiner: Wenn wir  $n$  Mal würfeln, wählen wir als Ergebnisraum

$$\Omega_n = \{\omega = (\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \{1, \dots, 6\}, i = 1, \dots, n\}.$$

Auch das ist ein Laplace-Raum, und  $\Omega_n$  hat  $6^n$  Elemente.

(Bild siehe Vorlesung)

**1.2. Häufigkeiten und Wahrscheinlichkeiten.** Wir haben nun also eine Vorstellung von Wahrscheinlichkeiten entwickelt, und können auch bereits damit rechnen. Jedoch sind Wahrscheinlichkeiten, zumindest so, wie sie in diesem Skript eingeführt sind, ein rein mathematisches Konzept. In der Wirklichkeit können wir Wahrscheinlichkeiten nicht direkt beobachten, wir können sie nur mittels theoretischer Überlegungen berechnen. Werfen wir beispielsweise einen (nicht notwendigerweise fairen) Würfel, so sehen wir ein ganz bestimmtes Ergebnis, eine *Realisierung*, aber wir bekommen durch diese eine Beobachtung keine Information über zugrundeliegende Wahrscheinlichkeiten, dafür müssen wir die Form des Würfels kennen, und Annahmen über seine Symmetrie treffen, z.B. annehmen, dass er fair ist.

Wir können aber, wenn wir möchten, den Würfel nicht einmal, sondern sehr oft werfen, und uns die jeweiligen Ergebnisse notieren, also ein wiederholtes Zufallsexperiment ausführen. Statt einer Liste der verschiedenen Ergebnisse können wir die *Häufigkeiten* der jeweiligen Ergebnisse betrachten. Wenn wir den Würfel oft genug werfen, dann geben diese – empirisch ermittelten – Häufigkeiten Aufschluss über die Wahrscheinlichkeiten.

Konkret stellen wir uns vor, dass wir den Würfel  $n = 100$  mal werfen. Wir bezeichnen mit  $H_i$  die Häufigkeit, mit der dabei das Ergebnis  $i$  auftritt,  $i = 1, \dots, 6$ , d.h.  $H_1$  gibt an, wie viele unter den 100 Würfeln eine 1 ergeben habe,  $H_2$  die Anzahl der Zweien, usw. Sei  $p_i$  die (theoretische, unbekannte, nicht direkt messbare) Wahrscheinlichkeit, dass der Würfel den Wert  $i$  zeigt. Wir erwarten dann natürlich, dass

$$p_i \approx \frac{H_i}{n}$$

gilt, d.h. die *relative Häufigkeit*  $H_i/n$  ist eine *Näherung* der unbekannten Wahrscheinlichkeit  $p_i$ . Wir erwarten dass diese Näherung umso besser wird, je größer  $n$  ist.

Umgangssprachlich oder für praktische Zwecke sind viele “Wahrscheinlichkeiten” in diesem Sinne zu verstehen. Wird beispielsweise gesagt, dass es eine Wahrscheinlichkeit  $p$  gibt, an einer bestimmten Krankheit zu erkranken, so bedeutet das normalerweise, dass in einer (hinreichend großen) Menge von Personen die Krankheit mit einer relativen Häufigkeit  $p$  beobachtet wurde. Dies ist eine “statistische” Interpretation des Begriffs der Wahrscheinlichkeit, im Gegensatz zur axiomatischen, welche normalerweise in der Mathematik gebräuchlich ist. Entsprechend wird der Zusammenhang zwischen *beobachteten* Häufigkeiten und *theoretischen* Wahrscheinlichkeiten im Statistik-Teil dieser Vorlesung besonders wichtig sein.

**Beispiel 1.13** (Zufällige Zielscheibe; Monte-Carlo-Methode). Wir betrachten eine quadratische Zielscheibe, auf der eine komplizierte Figur eingezeichnet ist.

Bild siehe Vorlesung

Wir möchten wissen, mit welcher Wahrscheinlichkeit ein zufällig ausgewählter Punkt aus dem Quadrat im Innern der gezeichneten Figur liegt. Diese Wahrscheinlichkeit könnte durch Berechnung der Fläche der gezeichneten Figur bestimmt werden. Sollte die Intuition über relative Häufigkeiten stimmen, könnte die Wahrscheinlichkeit auch mit Hilfe eines Dartspeils bestimmt werden, mit folgendem Experiment: Ohne zu zielen wird sehr oft ein Darts-Pfeil auf das Quadrat geworfen, und gezählt, wie oft die Figur getroffen wird. Das Verhältnis der Treffer im Vergleich zur Gesamtzahl der geworfenen Pfeile sollte dann ungefähr dieser Wahrscheinlichkeit entsprechen.

Auch wenn man in Realität selten genau dieses Experiment ausführt, so wird die Grundidee davon, nämlich das wiederholte Ausführen eines Zufallsexperiments, in vielen praktischen Situationen zur näherungsweisen Bestimmung von Wahrscheinlichkeiten verwendet. Dies ist das Grundprinzip für sogenannte *Monte-Carlo-Methoden*, bzw. *Monte-Carlo-Simulation*.

**Beispiel 1.14** (Zufallszahlen mit R). R ist ein frei verfügbares Statistikprogramm, welches auf allen gängigen Betriebssystemen einfach zu bedienen ist. Mit dem Befehl `runif(n, a, b)` können Sie sich  $n$  zufällige Zahlen aus dem Intervall  $[a, b]$  angeben lassen, z.B. erhalten Sie mit `runif(20, -2, 13.7)` 20 zufällige Zahlen, welche alle zwischen  $-2$  und  $13.7$  liegen. Da jedes solche Intervall unendlich viele Zahlen enthält, werden Sie so gut wie nie eine Zahl doppelt bekommen (ein Computer kennt zwar nicht unendlich viele Zahlen, sondern wird auf eine gewisse Genauigkeit runden, aber dennoch wird nur mit sehr kleiner Wahrscheinlichkeit zweimal dieselbe Zahl ausgegeben werden).

Mit dem Befehl `sample(k:m, n, replace=T)` erhalten Sie  $n$  zufällige *natürliche* Zahlen zwischen  $k$  und  $m$ , wobei  $k < m$  natürliche Zahlen sein müssen. Beispielsweise erhalten Sie mit `sample(1:10, 100, replace=T)` 100 zufällige Zahlen, welche alle zwischen 1 und 10 liegen, wobei jede Zahl mit gleicher Wahrscheinlichkeit auftritt. Sie können sich solche Zufallszahlen ausgeben lassen, die jeweiligen Häufigkeiten bestimmten, und somit “überprüfen”, ob die Zahlen wirklich zufällig und gleich wahrscheinlich sind.

Mehr Hintergrund zu dieser Aufgabe werden wir in Kapitel 3 kennen lernen.

## 2. BEDINGTE WAHRSCHEINLICHKEITEN, UNABHÄNGIGKEIT

In diesem Kapitel werden wir etwas komplexere Zufallsexperimente betrachten. Insbesondere interessieren wir uns für Wahrscheinlichkeiten bei wiederholt ausgeführten oder mehrstufigen Zufallsexperimenten, und wir fragen uns, inwiefern partielle Informationen über das Ergebnis eines Zufallsexperiments die Wahrscheinlichkeiten beeinflussen. Dabei werden wir die zentralen Begriffe der *bedingten Wahrscheinlichkeit* und der *Unabhängigkeit von Ereignissen* kennen lernen, und die sogenannte *Bayes'sche Umkehrformel* herleiten und anwenden.

Die **Lernziele** dieses Kapitel sind:

- Bedingte Wahrscheinlichkeiten kennen und berechnen können
- Mehrstufige Experimente mathematisch modellieren und behandeln können
- Die Bayes'sche Umkehrformel anwenden können
- Die Unabhängigkeit von Ereignissen bestimmen können und die Implikationen davon kennen

### 2.1. Bedingte Wahrscheinlichkeiten.

**Beispiel 2.1** (Würfeln unter Zusatzinformation). Wir würfeln einmal mit einem fairen Würfel. Aus Kapitel 1 wissen wir, dass wir das mit Hilfe des Ergebnisraums  $\Omega = \{1, \dots, 6\}$  beschreiben können, und dass die Ergebnisse alle gleich wahrscheinlich sind, also

$$\mathbb{P}(1) = \mathbb{P}(2) = \dots = \mathbb{P}(6) = \frac{1}{6}.$$

Nun stellen wir uns vor, dass der Würfel verdeckt geworfen wird, dass wir also das Ergebnis nicht sehen können. Jemand darf sich nun das Ergebnis anschauen, und sagt uns, dass die gewürfelte Zahl gerade ist. Wir erhalten also eine Teilinformation über das Ergebnis. Wie verändern sich die Wahrscheinlichkeiten unter dieser Zusatzinformation? Wir schreiben  $\tilde{\mathbb{P}}$  für die Wahrscheinlichkeit unter Berücksichtigung dieser Zusatzinformation. Dann gelten

$$\tilde{\mathbb{P}}(2) = \tilde{\mathbb{P}}(4) = \tilde{\mathbb{P}}(6) = \frac{1}{3},$$

und

$$\tilde{\mathbb{P}}(1) = \tilde{\mathbb{P}}(3) = \tilde{\mathbb{P}}(5) = 0.$$

Denn ist die Zahl gerade kann sie weder 1 noch 3 noch 5 sein. Die übrigen Zahlen 2, 4 und 6 sind möglich. Da wir keine weiteren Informationen haben, sind diese alle gleich wahrscheinlich, und da Wahrscheinlichkeiten zu 1 aufsummieren müssen, müssen die Wahrscheinlichkeiten alle gleich  $1/3$  sein.

$\tilde{\mathbb{P}}$  aus dem obigen Beispiel ist eine sogenannte bedingte Wahrscheinlichkeit. Allgemein wird dies so definiert:

**Definition 2.2** (Bedingte Wahrscheinlichkeit). Sei  $(\Omega, \mathbb{P})$  ein endlicher Wahrscheinlichkeitsraum, und seien  $A, B \subseteq \Omega$  Ereignisse, mit  $\mathbb{P}(B) > 0$ . Die *bedingte Wahrscheinlichkeit für A gegeben B* ist definiert als

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

*Bemerkung.* Für jedes  $B \subseteq \Omega$  ist  $\mathbb{P}(\cdot | B) : \mathcal{P}(\Omega) \rightarrow [0, 1] : A \mapsto \mathbb{P}(A | B)$  ein Wahrscheinlichkeitsmaß (vgl. Kapitel 1).  $(\Omega, \mathbb{P}(\cdot | B))$  ist also ein Wahrscheinlichkeitsraum.

**Beispiel 2.3** (Würfeln unter Zusatzinformationen, Fortsetzung von Beispiel 2.1). Im Beispiel 2.1 vom Anfang können wir z.B.  $B = \{2, 4, 6\}$ , also das Ereignis “Ergebnis gerade” wählen. Dann gilt

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

Die Tatsache dass uns jemand mitteilt, dass das Ergebnis gerade ist, bedeutet, dass wir die Zusatzinformation erhalten, dass das Ereignis  $B$  eingetreten ist. Das heißt, dass wir uns für die *bedingte Wahrscheinlichkeit gegeben B* interessieren. Diese berechnen wir aus der Definition ???. Es gilt

$$\mathbb{P}(\{2\} \cap B) = \mathbb{P}(\{2\}) = \frac{1}{6},$$

also

$$\mathbb{P}(\{2\} | B) = \frac{\mathbb{P}(\{2\} \cap B)}{\mathbb{P}(B)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Andererseits gilt  $\mathbb{P}(\{1\} \cap B) = \mathbb{P}(\emptyset) = 0$ , also

$$\mathbb{P}(\{1\} | B) = \frac{\mathbb{P}(\{1\} \cap B)}{\mathbb{P}(B)} = \frac{0}{1/2} = 0.$$

Analog geht man für 3, 4, 5, 6 vor. Die Wahrscheinlichkeit  $\tilde{\mathbb{P}}$  aus dem Beispiel 2.1 war also genau die bedingte Wahrscheinlichkeit gegeben  $B$ , also  $\mathbb{P}(\cdot | B)$ .

**Beispiel 2.4** (Ausfallwahrscheinlichkeit, Sterbewahrscheinlichkeit). **siehe Vorlesung**

**Beispiel 2.5** (Urnenmodelle). **siehe Vorlesung**

## 2.2. Mehrstufige Experimente und Formel von der Gesamtwahrscheinlichkeit.

Wir haben nun schon mehrere Beispiele von mehrstufigen Experimenten gesehen, z.B. verschiedene Urnenmodelle. Dabei handelt es sich um Zufallsexperimente, welche in mehreren Schritten ablaufen. Sehr oft ist es sinnvoll, solche Experimente als *Wahrscheinlichkeitsbaum* darzustellen, wie wir es beim wiederholten Münzwurf 1.10 bereits getan haben. Ein “Baum” besteht dabei aus Kanten (oder Ästen) und Knoten, in denen sich die Äste aufspalten, und welche in Ebenen sortiert sind. Die erste Ebene besteht aus genau einem Knoten. Auf der letzten Ebene enden die Äste in “Blättern”. Die Bäume werden normalerweise von links nach rechts oder von oben nach unten gezeichnet:

**Bild siehe Vorlesung**

In den Baum trägt man nun die Ereignisse und Wahrscheinlichkeiten ein. Dabei geht man nach diesem Schema vor:

- Knoten des Baumes: Diese stellen die *Ergebnisse* der jeweiligen Stufe des Zufallsexperiment dar.
- Kanten (“Äste”) des Baumes: Hier tragen wir die *bedingte Wahrscheinlichkeit* des entsprechenden Ausgangs, gegeben das Ergebnis der vorherigen Stufe, ein.
- “Blätter” des Baumes: Endergebnis des Experiments entsprechend der Zwischenergebnisse die zu dem jeweiligen Blatt führen.

Bild siehe Vorlesung

**Beispiel 2.6** (Test auf Krankheit). Wir untersuchen eine Krankheit, von der bekannt ist, dass sie bei 2% der Bevölkerung auftritt. Es wird ein Bluttest entwickelt, der die Krankheit mit 99% Wahrscheinlichkeit erkennt, allerdings zeigt er auch bei 3% der gesunden Testpersonen falsch positiv an.

Diese Situation kann man als zweistufiges Zufallsexperiment auffassen: Wir bezeichnen mit  $K$  das Ereignis, dass eine zufällig ausgewählte Person aus der Bevölkerung krank ist, und mit  $P$  die Wahrscheinlichkeit, dass das Testergebnis positiv ist. Somit gilt also nach unseren Informationen  $\mathbb{P}(K) = 0.02$ . Außerdem ist die Wahrscheinlichkeit, dass eine Person gesund ist, gegeben durch  $\mathbb{P}(K^c) = 1 - \mathbb{P}(K) = 0.98$ .

Die Information, dass die Krankheit mit 99% Wahrscheinlichkeit erkannt wird, übersetzt sich nun in eine bedingte Wahrscheinlichkeit, nämlich

$$\mathbb{P}(P|K) = 0.99,$$

oder in Worten ausgedrückt, wenn eine Person krank ist, dann ist der Test mit einer Wahrscheinlichkeit von 0.99 positiv. Andererseits ist er mit 3% Wahrscheinlichkeit falsch positiv, also

$$\mathbb{P}(P|K^c) = 0.03.$$

Natürlich können wir auch wieder die Gegenwahrscheinlichkeiten berechnen:  $\mathbb{P}(P^c|K) = 0.01$ , und  $\mathbb{P}(P^c|K^c) = 0.97$ . (Übung: Überlegen Sie sich, was diese Wahrscheinlichkeiten in Worten bedeuten).

Somit können wir die Situation nun wieder als Baum darstellen:

Bild siehe Vorlesung

Wenn man die Formel für die bedingte Wahrscheinlichkeit aus Definition 2.2 umformt, erhält man die Gleichung

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

Da man im Wahrscheinlichkeitsbaum eines mehrstufigen Experiments die jeweiligen bedingten Wahrscheinlichkeiten auf die Äste schreibt, kann man die Definition der bedingten Wahrscheinlichkeit mit Hilfe der Baumdarstellung in folgende Rechenregel übersetzen:

**Satz 2.7** (Mehrstufiges Experiment: Multiplikationsregel). *In einem mehrstufigen Experiment berechnet sich die Wahrscheinlichkeit eines Ergebnisses durch Multiplikation der Wahrscheinlichkeiten entlang der Äste die zum Blatt mit diesem Ergebnis führen.*

*Beweis.* siehe Vorlesung

□

**Beispiel 2.8** (Test auf Krankheit, Fortsetzung von Beispiel 2.6). Im obigen Beispiel können wir also die Wahrscheinlichkeit für das Ereignis  $K \cap P$  berechnen als

$$\mathbb{P}(K \cap P) = \mathbb{P}(K) \cdot \mathbb{P}(P|K) = 0.02 \cdot 0.99 = 0.0198.$$

Mit dieser Wahrscheinlichkeit ist also eine zufällig ausgewählte Person krank, und der Test zeigt bei dieser Person die Krankheit an.

Wie wir aus Satz 1.7 wissen, berechnet sich die Wahrscheinlichkeit eines Ereignisses durch Summation über die Ergebnisse die in dem Ereignis enthalten sind. Die Definition der bedingten Wahrscheinlichkeit, oder alternativ die obige Multiplikationsregel, führt deshalb auf die sogenannte Formel von der *totalen Wahrscheinlichkeit* (oder Formel von der *Gesamtwahrscheinlichkeit*).

**Satz 2.9** (Formel von der Gesamtwahrscheinlichkeit). *Seien  $A, B \subset \Omega$  Ereignisse, und gelte  $0 < \mathbb{P}(B) < 1$ . Dann gilt*

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

*Allgemeiner seien  $A$  ein Ereignis, und  $B_1, \dots, B_n$  eine disjunkte Zerlegung von  $\Omega$ . Dann gilt*

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Disjunkte Zerlegung bedeutet dabei  $\mathbb{P}(B_i) > 0$  für alle  $i = 1, \dots, n$ ,  $B_i \cap B_j = \emptyset$  für  $i \neq j$ , und  $\bigcup_{i=1}^n B_i = \Omega$ .

*Beweis.* Da  $B^c$  das Komplement von  $B$  bezeichnet, gilt  $A = (A \cap B) \cup (A \cap B^c)$ , und  $(A \cap B) \cap (A \cap B^c) = \emptyset$ . (siehe Skizze) Bild. Mit der Eigenschaft (P2) aus 1.5 und der Definition der bedingten Wahrscheinlichkeit folgt deshalb

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

□

Diese Formel kann man auch in eine *Additionsregel für mehrstufige Experimente* übersetzen, dies wird im Bild klar:

Bild siehe Vorlesung

**Beispiel 2.10** (Test auf Krankheit: Fortsetzung von Beispiel 2.6). Im Beispiel 2.6 gilt also für die Wahrscheinlichkeit eines positiven Tests

$$\mathbb{P}(P) = \mathbb{P}(P|K)\mathbb{P}(K) + \mathbb{P}(P|K^c)\mathbb{P}(K^c) = 0.99 \cdot 0.02 + 0.03 \cdot 0.98 = 0.0492.$$

Dies ist die Wahrscheinlichkeit, dass der Test bei einer zufällig ausgewählten Person positiv ist (ohne zu wissen, ob die Person krank oder gesund ist).



### 2.3. Unabhängigkeit von Ereignissen.

**Beispiel 2.11** (Urnenmodell). Urnenmodelle sind ein wichtiges Hilfsmittel in der Wahrscheinlichkeitstheorie und der stochastischen Modellierung. Ihr Nutzen liegt darin, dass sie als (Teil-)Modell in komplexeren Situationen zum Einsatz kommen. Hier betrachten wir zwei einfache Situationen. Nehmen wir eine Urne, in der sich  $n$  rote und  $m$  blaue Kugeln befinden. Dabei stellen wir uns vor, dass die Kugeln in der Urne gut gemischt werden können, und dass wir ohne hinzusehen eine Kugel aus der Urne ziehen. Das heißt, das das Experiment “ziehe eine Kugel” ein Zufallsexperiment ist, und dass jede der insgesamt  $n + m$  Kugeln dieselbe Wahrscheinlichkeit hat, gezogen zu werden. Damit gilt

$$\mathbb{P}(\text{“gezogene Kugel ist rot”}) = \frac{n}{n + m}$$

und

$$\mathbb{P}(\text{“gezogene Kugel ist blau”}) = \frac{m}{n + m}.$$

Was passiert jetzt, wenn wir mehrmals hintereinander Kugeln aus der Urne ziehen? Hierfür müssen wir erst spezifizieren, ob wir die gezogenen Kugeln jeweils wieder in die Urne zurücklegen (und die Kugeln in der Urne wieder gut durchmischen), oder ob wir die gezogenen Kugeln nicht zurücklegen. Wir bezeichnen mit  $R_1, R_2, \dots$  die Ereignisse, dass die erste, bzw. zweite, ... gezogene Kugel rot ist, und mit  $B_1, B_2, \dots$  die Ereignisse dass die erste, zweite, ... gezogene Kugel blau ist. Wenn wir nun zweimal ziehen, und die Wahrscheinlichkeit berechnen wollen, dass die erste Kugel rot und die zweite blau ist, bestimmen wir also die Wahrscheinlichkeit  $\mathbb{P}(R_1 \cap B_2)$ . In der Baumdarstellung:

**Bild**

Wenn wir die Kugeln nun nach dem Ziehen jeweils wieder zurücklegen, so gibt es auch beim zweiten Ziehen  $n$  rote und  $m$  blaue Kugeln. Die Wahrscheinlichkeit, im zweiten Zug blau zu ziehen, wird also nicht durch die Wahrscheinlichkeit des Ergebnisses im ersten Zug beeinflusst. Deshalb ist nach unseren Rechenregeln im Fall *mit Zurücklegen*

$$\mathbb{P}(R_1 \cap B_2) = \mathbb{P}(R_1)\mathbb{P}(B_2 | R_1) = \frac{n}{n + m} \cdot \frac{m}{n + m}.$$

Legen wir die Kugel jedoch nicht zurück, so befinden sich nach dem Ziehen einer roten Kugel beim ersten Versuch nur noch  $n - 1$  rote Kugeln in der Urne, aber immer noch  $m$  blaue, also insgesamt  $n - 1 + m$  Kugeln. Im Fall *ohne Zurücklegen* haben wir also

$$\mathbb{P}(R_1 \cap B_2) = \mathbb{P}(R_1)\mathbb{P}(B_2 | R_1) = \frac{n}{n + m} \cdot \frac{m}{n - 1 + m}.$$

Beim Ziehen mit Zurücklegen bleibt die Wahrscheinlichkeit für das Ergebnis im zweiten Zug unbeeinflusst durch das Ergebnis im ersten Zug, beim Ziehen ohne Zurücklegen ändert sich die Wahrscheinlichkeit. Mit Zurücklegen sind die verschiedenen Züge voneinander *unabhängig*, ohne Zurücklegen sind sie *abhängig*.

**Beispiel 2.12** (Vererbung von Farbenblindheit). **siehe Vorlesung**

**Definition 2.13** (Unabhängigkeit von Ereignissen). Sei  $(\Omega, \mathbb{P})$  ein endlicher Wahrscheinlichkeitsraum, und seien  $A, B \subseteq \Omega$  Ereignisse.  $A$  und  $B$  heißen *unabhängig*, falls

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

gilt.

**Satz 2.14.** Zwei Ereignisse  $A$  und  $B$  in einem endlichen Wahrscheinlichkeitsraum  $(\Omega, \mathbb{P})$  mit  $\mathbb{P}(B) > 0$  sind genau dann unabhängig, wenn

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

gilt.

*Beweis.* Seien  $A$  und  $B$  unabhängig. Nach Definition bedeutet das  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Setzen wir das in die Definition der bedingten Wahrscheinlichkeit ein, so folgt

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

Die umgekehrte Beweisrichtung folgt ebenfalls aus dieser Rechnung.  $\square$

Natürlich gilt genauso, dass  $A$  und  $B$  genau dann unabhängig sind, wenn  $\mathbb{P}(B|A) = \mathbb{P}(B)$  erfüllt ist.

Der Begriff der Unabhängigkeit bedeutet, dass die Wahrscheinlichkeit für das Eintreten von  $A$  nicht durch das Wissen um das Eintreten von  $B$  beeinflusst wird (und umgekehrt). D.h. sagt uns jemand, dass das Ereignis  $B$  eingetreten ist, so ist dieses Zusatzwissen irrelevant bei der Berechnung der Wahrscheinlichkeit für das Eintreten von  $B$ . Im Urnenmodell in Beispiel 2.11 sieht man das gut: Wenn wir mit Zurücklegen ziehen, dann haben wir jedes Mal dieselbe Wahrscheinlichkeit, eine rote Kugel zu ziehen – die Ereignisse “die erste Kugel ist rot” und “die zweite Kugel ist blau” sind unabhängig. Ziehen wir aber *ohne* Zurücklegen, dann verändert die Information, dass die erste gezogene Kugel rot ist, die Wahrscheinlichkeit dafür, dass die zweite gezogene Kugel blau ist, denn mit der Formel von der totalen Wahrscheinlichkeit gilt

$$\begin{aligned} \mathbb{P}(B_2) &= \mathbb{P}(B_2 | R_1)\mathbb{P}(R_1) + \mathbb{P}(B_2 | B_1)\mathbb{P}(B_1) \\ &= \frac{m}{n-1+m} \cdot \frac{n}{n+m} + \frac{m-1}{n+m-1} \cdot \frac{m}{n+m} \\ &= \frac{m}{n+m}, \end{aligned}$$

aber

$$\mathbb{P}(B_2 | R_1) = \frac{m}{n-1+m} \neq \frac{m}{n+m} = \mathbb{P}(B_2).$$

*Bemerkung.* Vorsicht: Der Begriff der Unabhängigkeit wird manchmal mit der Eigenschaft dass  $A$  und  $B$  disjunkt sind verwechselt. Diese Begriffe sind aber *nicht* gleichbedeutend – eher im Gegenteil: Wenn  $A$  und  $B$  disjunkt sind, d.h.  $A \cap B = \emptyset$ , dann können sie nicht gleichzeitig eintreten. Das bedeutet, dass  $\mathbb{P}(A \cap B) = 0$  ist. In anderen Worten, wenn wir wissen dass  $B$  eingetreten ist, dann kann  $A$  nicht mehr eintreten. Somit gilt  $\mathbb{P}(A|B) = 0$ , auch wenn  $\mathbb{P}(A) \neq 0$  ist.

**Definition 2.15.** Seien  $A_1, \dots, A_n \subseteq \Omega$  Ereignisse. Die  $n$  Ereignisse heißen *unabhängig*, falls für alle  $2 \leq k \leq n$  und Indizes  $i_1, \dots, i_k \in \{1, \dots, n\}, i_l \neq i_j$  für  $l \neq j$  gilt

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_{i_2}) \cdot \dots \cdot \mathbb{P}(A_{i_k}).$$

Beispiel: Unabhängigkeit von drei Ereignissen  $A, B, C$ . Diese sind genau dann unabhängig, wenn *alle* folgenden Gleichungen erfüllt sind:

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

und

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

Man kann Beispiele finden (siehe Tutorium/Hausaufgaben), in denen die Gleichungen

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

alle erfüllt sind, jedoch  $A, B$  und  $C$  trotzdem nicht unabhängig sind. Ebenso gibt es Beispiele, in denen  $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$  gilt, aber die drei Ereignisse nicht unabhängig sind.

**2.4. Bayes'sche Umkehrformel.** Nicht jede bedingte Wahrscheinlichkeit ist einfach zu berechnen. In praktischen Anwendungen steht man oft vor der Situation, dass man z.B. die bedingte Wahrscheinlichkeit  $\mathbb{P}(A|B)$  (leicht) berechnen kann (z.B. aus einer geeigneten Baumdarstellung), sich jedoch eigentlich für die umgekehrte Wahrscheinlichkeit  $\mathbb{P}(B|A)$  interessiert. In solchen Fällen ist die Bayes'sche Umkehrformel nützlich.

**Satz 2.16** (Bayes-Formel). *Seien  $A, B$  Ereignisse mit  $0 < \mathbb{P}(B) < 1$ . Dann gilt*

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

*Beweis.* Wir können die Definition der bedingten Wahrscheinlichkeit verwenden, und erhalten so

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

Dabei haben wir im letzten Schritt die Definition der bedingten Wahrscheinlichkeit noch ein zweites Mal verwendet. Mit der Formel von der Gesamtwahrscheinlichkeit folgt

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

Dies setzen wir in der obigen Gleichung auf der rechten Seite ein, und erhalten so das gewünschte Ergebnis. □

**Beispiel 2.17** (Test auf Krankheit, Fortsetzung von Beispiel 2.6). In Beispiel 2.6 können wir uns nun folgende Frage stellen: Wenn eine Person positiv auf besagte Krankheit getestet wird, wie wahrscheinlich ist es dann, dass sie diese Krankheit tatsächlich hat?

Wir interessieren uns also für die bedingte Wahrscheinlichkeit  $\mathbb{P}(K|P)$ . Diese können wir nun aus den bekannten Werten  $\mathbb{P}(K) = 0.98$ ,  $\mathbb{P}(P|K) = 0.99$  und  $\mathbb{P}(P|K^c) = 0.03$  mit Hilfe der Bayes-Formel berechnen. Es gilt nämlich

$$\mathbb{P}(K|P) = \frac{\mathbb{P}(P|K)\mathbb{P}(K)}{\mathbb{P}(K)\mathbb{P}(P|K) + \mathbb{P}(K^c)\mathbb{P}(P|K^c)} = \frac{0.99 \cdot 0.98}{0.99 \cdot 0.98 + 0.03 \cdot 0.02} \approx 0.999$$

Mit anderen Worten, die Wahrscheinlichkeit dass jemand bei positivem Test wirklich krank ist, beträgt ungefähr 0.4. Das bedeutet aber auch, dass die Wahrscheinlichkeit, tatsächlich krank zu sein, auch bei einem positiven Test gerade mal 40% beträgt, also keineswegs nahe an 1 liegt! Das Ergebnis wirkt vielleicht zuerst einmal nicht intuitiv, aber kann dadurch erklärt werden, dass eben nur 1% der Bevölkerung tatsächlich krank ist, wenn man also nach dem Zufallsprinzip testet, man meistens gesunde Personen testet, bei denen wiederum der Test mit einer gewissen (kleinen, aber nicht vernachlässigbaren) Wahrscheinlichkeit falsch positiv ist.

Dieses Beispiel sollten Sie sich immer vor Augen halten wenn Sie mit bedingten Wahrscheinlichkeiten arbeiten – gerade wenn man sich für relativ wenig wahrscheinliche Ereignisse interessiert, können bedingte Wahrscheinlichkeiten leicht um Größenordnungen anders aussehen, als unsere Intuition zuerst denken würde. Es ist deshalb wichtig, die Rechenregeln für bedingte Wahrscheinlichkeiten und insbesondere die Bayes-Formel korrekt anzuwenden!

Versprechen wie “der Test hat eine Genauigkeit von 99.999%” sollten ebenso mit Vorsicht betrachtet werden. Dabei ist zu hinterfragen, was für eine Genauigkeit genau gemeint ist. Beispielsweise kann auch  $\mathbb{P}(P|K)$  sehr klein sein, und  $\mathbb{P}((P \cap K) \cup (P^c \cap K^c))$  trotzdem mindestens 0.99999 sein. Das ist dann jedoch nicht die Wahrscheinlichkeit, die normalerweise relevant ist.

*Bemerkung* (Allgemeine Bayes-Formel). Mit Hilfe der allgemeinen Formel von der totalen Wahrscheinlichkeit erhält man eine allgemeinere Variante der Bayes-Formel: Sei  $A$  ein Ereignis und  $B_1, \dots, B_n$  eine disjunkte Zerlegung von  $\Omega$ . Dann gilt

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

**Beispiel 2.18** (Signalübertragung). Siehe Vorlesung

**Beispiel 2.19** (Markov-Ketten). siehe Vorlesung

### 3. DISKRETE ZUFALLSVARIABLEN UND VERTEILUNGEN

In Anwendungen kommen zufällige Ereignisse oder Abläufe bzw. Wahrscheinlichkeiten oft in komplexen Situationen vor. Einige mögliche Beispiele:

- Ausfallwahrscheinlichkeit eines Bauteils
- Wahrscheinlichkeit dass eine Website von  $x$  Personen besucht wird
- Position, Intensität und Kapazität von sich zufällig bewegenden Sendern
- Wahrscheinlichkeit für die Wirkung eines Medikaments, Nebenwirkungen
- Quantifizierung von Risiken
- Genetische Typen und ihre Mutationswahrscheinlichkeit

In solchen Situationen ist es oft nicht möglich oder praktikabel, einen Wahrscheinlichkeitsraum explizit anzugeben. Man behilft sich stattdessen in vielen Fällen mit der Untersuchung von gewissen *Zufallsvariablen*. Oft kann man unter geeigneten Modellannahmen Aussagen über deren *Verteilung* treffen, welche es erlauben, wichtige Aspekte des Problems zu beschreiben. Diese beiden zentralen Begriffe werden wir in diesem Kapitel einführen und diskutieren. Die **Lernziele** dieses Kapitels sind:

- Mit den Begriffen Zufallsvariable, Verteilung und Verteilungsfunktion umgehen können
- Wichtige diskrete Verteilungen und ihr Auftreten kennen und damit rechnen können

Als **Vorkenntnisse** benötigen Sie in diesem Abschnitt vor allem Kenntnisse über Funktionen und Abbildungen, wie sie in der Analysis I unterrichtet werden. Insbesondere mit den Begriffen Definitionsbereich, Wertebereich, Bild und Urbild sollten sie souverän umgehen können.

**Hinweis:** Dieses Kapitel enthält etliche neue Begriffe und Notation, die in der ganzen Vorlesung immer wieder verwendet werden, insbesondere die Begriffe Zufallsvariable, Verteilung und Verteilungsfunktion. Prägen Sie sich bitte die Definitionen gut ein, und machen Sie sich klar, welche Notation welche Art von mathematischen Objekten bezeichnet, und wie sie zusammenhängen. Ohne die sichere Beherrschung der Begriffe aus diesem Abschnitt wird es schwer, der Vorlesung zu folgen.

**3.1. Zufallsvariablen, Verteilung, Verteilungsfunktion.** Bevor wir die zentralen Begriffe dieses Kapitels einführen, verallgemeinern wir den Begriff des Wahrscheinlichkeitsraumes (vgl. Definition 1.5).

**Definition 3.1** (Diskreter Wahrscheinlichkeitsraum). Ein *diskreter Wahrscheinlichkeitsraum* ist ein Paar bestehend aus einer endlichen oder abzählbar unendlichen *Ergebnismenge*  $\Omega$ , und einem *Wahrscheinlichkeitsmaß*  $\mathbb{P}$ , welches jedem Ereignis  $A \subseteq \Omega$  eine Zahl  $\mathbb{P}(A) \in [0, 1]$  zuordnet. Diese Zahl ist die *Wahrscheinlichkeit* des Ereignisses  $A$ . Formal gesprochen ist ein Wahrscheinlichkeitsmaß eine Abbildung aus der Menge aller Teilmengen von  $\Omega$  nach  $[0, 1]$  mit den Eigenschaften

$$(P0) \quad 0 \leq \mathbb{P}(A) \leq 1 \text{ für alle } A \subseteq \Omega$$

- (P1)  $\mathbb{P}(\Omega) = 1$   
 (P2') Für  $(A_n)_{n \in \mathbb{N}}$  mit  $A_n \subset \Omega$  für alle  $n \in \mathbb{N}$ , so dass  $A_n \cap A_m = \emptyset$  für alle Paare  $n \neq m$  ist, gilt  $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ .

Diese Definition unterscheidet sich von Definition 1.5 nur dadurch, dass  $\Omega$  nun nicht mehr endlich sein muss, sondern abzählbar unendlich sein darf, sowie durch die Eigenschaft (P2'), welche eine Verallgemeinerung von (P2) ist. Wenn (P2') erfüllt ist, gilt automatisch auch die Eigenschaft (P2).

*Bemerkung.* Ein *allgemeiner Wahrscheinlichkeitsraum* wird definiert wie ein diskreter Wahrscheinlichkeitsraum, wobei die Ergebnismenge  $\Omega$  nicht endlich oder abzählbar unendlich sein muss, sondern eine beliebige Menge sein darf, z.B.  $\Omega = \mathbb{R}$ . Der einzige Unterschied besteht darin, dass streng mathematisch gesehen im Allgemeinen nicht jede Teilmenge von  $\Omega$  ein Ereignis sein kann, sondern nur bestimmte, "geeignete" Teilmengen von  $\Omega$ . Wir gehen jedoch in dieser Vorlesung nicht auf diese Problematik ein. Für unsere Zwecke, wie überhaupt für die meisten praktischen Anwendungen, spielt sie keine Rolle. Der Einfachheit halber nehmen wir auch im allgemeinen Fall an, alle Teilmengen von  $\Omega$  seine geeignete Ereignisse, und verwenden die Notation  $(\Omega, \mathbb{P})$  auch für beliebige Wahrscheinlichkeitsräume. Interessierte Leser\*innen verweisen wir auf die mathematische Literatur zur maßtheoretischen Stochastik.

Die Resultate aus Kapitel 2 gelten auch für allgemeine Wahrscheinlichkeitsräume, wir werden sie auch so benutzen.

**Definition 3.2** (Zufallsvariable). Sei  $(\Omega, \mathbb{P})$  ein Wahrscheinlichkeitsraum. Eine (eindimensionale) *Zufallsvariable* auf  $(\Omega, \mathbb{P})$  ist eine Abbildung

$$X : \Omega \rightarrow \mathbb{R}.$$

Eine Zufallsvariable ist also erst einmal einfach eine Abbildung, oder Funktion, deren Definitionsbereich die Ergebnismenge eines Wahrscheinlichkeitsraumes ist, und deren Wertebereich in den reellen Zahlen enthalten ist. Dies ist erst einmal eine recht allgemeine und abstrakte mathematische Definition, deren Nutzen sich vermutlich hier noch nicht erschließt. Dieser tritt erst hervor, wenn wir über Verteilungen gesprochen haben. Zuerst aber noch einige Bemerkungen.

*Notation:* Zufallsvariablen werden normalerweise mit Großbuchstaben vom Ende des Alphabets bezeichnet, also  $X, Y, Z, V, \dots$  oder auch mit  $X_1, X_2, \dots$ .

Der *Wertebereich* einer Zufallsvariable  $X$  wird mit  $X(\Omega)$  bezeichnet und ist eine Teilmenge von  $\mathbb{R}$ , anders ausgedrückt,

$$X(\Omega) = \{x : \exists \omega \in \Omega : X(\omega) = x\} \subseteq \mathbb{R}.$$

Wir werden noch einige weitere Begriffe in diesem Zusammenhang verwenden.

**Definition 3.3** (Diskrete Zufallsvariablen, Zufallsvektor).

- (a) Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable. Wenn  $X(\Omega)$  endlich oder abzählbar ist, so heißt  $X$  *diskrete* Zufallsvariable.

- (b) Eine Abbildung  $X : \Omega \rightarrow \mathbb{R}^d$  heißt *d-dimensionale Zufallsvariable* oder *Zufallsvektor*. Wir können dann auch  $X = (X_1, \dots, X_d)$  schreiben, wobei  $X_1, \dots, X_d$  (eindimensionale) Zufallsvariablen sind.

Falls  $(\Omega, \mathbb{P})$  ein *diskreter* Wahrscheinlichkeitsraum ist, so ist jede Zufallsvariable auf  $\Omega$  diskret. Beispielsweise ist  $\mathbb{N}_0$  ein möglicher Wertebereich einer diskreten Zufallsvariablen.

**Beispiel 3.4** (Würfel). Wir nehmen einen fairen Würfel, und werfen ihn zwei Mal hintereinander. Wir wissen bereits aus Beispiel 1.12, dass ein geeigneter Laplace-Wahrscheinlichkeitsraum gegeben ist durch

$$\Omega = \{\omega = (\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1, \dots, 6\}\}.$$

Auf diesem Wahrscheinlichkeitsraum können wir verschiedene Zufallsvariablen definieren, zum Beispiel:

- $X_1 : \Omega \rightarrow \mathbb{R} : X_1(\omega) = \omega_1$ . In Worten ausgedrückt beschreibt  $X_1$  also das Ergebnis des ersten Wurfs.
- Analog können wir natürlich als  $X_2 : \Omega \rightarrow \mathbb{R} : X_2(\omega) = \omega_2$  das Ergebnis des zweiten Wurfs definieren.
- Sei  $Z : \Omega \rightarrow \mathbb{R}$  definiert als  $Z(\omega) = \omega_1 + \omega_2$ . Das bedeutet,  $Z$  beschreibt die Summe der beiden Ergebnisse.
- Mit  $Y$  wollen wir das Minimum der beiden Ergebnisse bezeichnen, also  $Y(\omega) = \min\{\omega_1, \omega_2\}$ .

Wie sehen die Wertebereiche dieser Zufallsvariablen aus? Jede dieser Zufallsvariablen kann nur bestimmte Werte annehmen. Für  $X_1$  sind das die Werte von  $\omega_1$ , da ja  $X(\omega) = \omega_1$  ist. Nun ist aber  $\omega_1 \in \{1, \dots, 6\}$  nach Definition. Somit ist also

$$X_1(\Omega) = \{1, \dots, 6\}.$$

Natürlich kann  $X_2$  die gleichen Werte annehmen wie  $X_1$ . Betrachten wir also die Zufallsvariable  $Z$ . Hier haben wir  $Z(\omega) = \omega_1 + \omega_2$ , wobei  $\omega_1$  und  $\omega_2$  jeweils aus der Menge  $\{1, \dots, 6\}$  stammen. Die Summe kann also die Werte zwischen  $1 + 1 = 2$  und  $6 + 6 = 12$  annehmen, also

$$Z(\Omega) = \{2, 3, \dots, 12\}.$$

Als kleine Übung können Sie sich außerdem davon überzeugen, dass  $Y(\Omega) = \{1, \dots, 6\}$  ist.

**Beispiel 3.5** (Wartezeit). siehe Vorlesung

**Beispiel 3.6** (Indikatorfunktion). Betrachten wir ein Ereignis, das heißt, eine Teilmenge  $A \subset \Omega$ . Die *Indikatorfunktion* von  $A$  ist definiert als die Abbildung  $1_A : \Omega \rightarrow \{0, 1\}$ ,

$$1_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A, \\ 0 & \text{sonst.} \end{cases}$$

$1_A$  bezeichnet also die Funktion auf  $\Omega$ , welche genau dann den Wert 1 annimmt, wenn  $\omega$  in  $A$  liegt, und andernfalls gleich 0 ist. Das bedeutet, dass  $1_A$  anzeigt, ob ein Ergebnis zum Ereignis  $A$  gehört oder nicht. Daher der Name Indikatorfunktion. Indikatorfunktionen sind sehr nützlich zur Formalisierung vieler Sachverhalte, und werden uns immer wieder

begegnen. Als Beispiel betrachten wir noch einmal den zweifachen Münzwurf. Ein Ereignis ist beispielsweise

$$A = \{\omega \in \Omega : \omega_1 + \omega_2 \leq 3\},$$

also das Ereignis dass die Summe der beiden Würfe höchstens 3 ergibt. Für die Zufallsvariable  $1_A$  gilt also  $1_A((1, 2)) = 1$ , da für  $\omega = (1, 2)$  gilt  $\omega_1 + \omega_2 = 1 + 2 = 3$ , aber  $1_A((4, 1)) = 0$ , da hier die Summe strikt größer als 3 ist.

**Beispiel 3.7** (Funktionen von Zufallsvariablen, Fortsetzung von Beispiel 3.4). Aus Zufallsvariablen kann man sich neue basteln: Betrachte die Zufallsvariablen  $X_1$  und  $X_2$  aus dem Beispiel mit dem Münzwurf. Die Abbildung  $X_1 + X_2$  ist wiederum eine Abbildung von  $\Omega \rightarrow \mathbb{R}$ , also wieder eine Zufallsvariable. Es gilt  $(X_1 + X_2)(\omega) = X_1(\omega) + X_2(\omega) = \omega_1 + \omega_2$ . Damit ist also  $X_1 + X_2$  dieselbe Zufallsvariable wie  $Z$ , was nicht weiter überrascht. Ähnliche Konstruktionen sind auch mittels Verknüpfung möglich. Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable, und  $f : X(\Omega) \rightarrow \mathbb{R}$  eine Funktion. Dann ist  $f(X) : \Omega \rightarrow \mathbb{R}$ ,  $f(X)(\omega) = f(X(\omega))$  wieder eine Zufallsvariable. Mit  $f(x) = x^2$  ist beispielsweise  $f(X) = X^2$  wieder eine Zufallsvariable.

*Bemerkung.* Sei  $X$  eine Zufallsvariable auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathbb{P})$ . Für jede Menge  $E \subseteq X(\Omega)$  ist das *Urbild*

$$X^{-1}(E) = \{\omega \in \Omega : X(\omega) \in E\}$$

ein *Ereignis*.

*Notation.* Für  $E \subseteq X(\Omega)$  bzw. für  $x \in X(\Omega)$  schreiben wir kurz

- $\{X \in E\} := \{\omega \in \Omega : X(\omega) \in E\} = X^{-1}(E)$
- $\{X = x\} := \{\omega \in \Omega : X(\omega) = x\} = X^{-1}(\{x\})$
- $\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}([-\infty, x])$

etc. Außerdem schreiben wir für die entsprechenden Wahrscheinlichkeiten kurz

- $\mathbb{P}(X \in E) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in E\})$ ,
- $\mathbb{P}(X \leq x) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$

etc.

Damit kommen wir nun zum zweiten zentralen Begriff dieses Kapitels.

**Definition 3.8** (Verteilungsfunktion). Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable. Die (*kumulative*) *Verteilungsfunktion*  $F_X$  von  $X$  ist definiert als

$$F_X : \mathbb{R} \rightarrow [0, 1], \quad F_X(x) := \mathbb{P}(X \leq x).$$

Wenn wir die Verteilungsfunktion einer Zufallsvariablen kennen, so können wir damit also (im Prinzip) die Wahrscheinlichkeiten von Ereignissen der Form  $\{X \leq x\}$ ,  $x \in \mathbb{R}$ , bestimmen. Mit unseren Grundprinzipien für Wahrscheinlichkeitsmaße gelten außerdem

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x) = 1 - F_X(x),$$



sowie

$$\mathbb{P}(X \in (a, b]) = \mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F_X(b) - F_X(a).$$

Die Verteilungsfunktion darf nicht verwechselt werden mit einem ähnlich klingenden Begriff:

**Definition 3.9** (Diskrete Verteilung). Sei  $X : \Omega \rightarrow \mathbb{R}$  eine *diskrete* Zufallsvariable. Die *Verteilung* (oder Wahrscheinlichkeitsfunktion)  $p_X$  von  $X$  ist definiert als

$$p_X : \mathbb{R} \rightarrow [0, 1], p_X(x) := \mathbb{P}(X = x).$$

*Bemerkung.* Falls  $x \notin X(\Omega)$  ist, so gilt  $p_X(x) = 0$ . Die Verteilung von  $X$  wird also durch Angabe von  $X(\Omega)$  sowie von  $p_X(x)$  für alle  $x \in X(\Omega)$  festgelegt. Für eine diskrete Zufallsvariable  $X$  gilt folgender Zusammenhang zwischen Verteilungsfunktion und Verteilung:

$$F_X(x) = \sum_{y \leq x} p_X(y).$$

Beachten Sie, dass der Begriff der Verteilungsfunktion somit für beliebige Zufallsvariablen definiert ist, der Begriff der Verteilung hier nur für diskrete.

**Beispiel 3.10** (Summe beim zweifachen Würfeln, Fortsetzung von Beispiel 3.7). Wir werfen zwei faire Würfel, und bezeichnen mit  $Z$  die Summe der beiden gewürfelten Augenzahlen, siehe auch Beispiel 3.4. Wie dort wählen wir als Wahrscheinlichkeitsraum  $\Omega = \{(\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}, i = 1, 2\}$ . Das ist besonders nützlich, da es sich um einen Laplace-Raum handelt, und wir wissen, dass  $\mathbb{P}(\omega) = \frac{1}{36}$  für jedes  $\omega$  gilt.  $X$  ist also eine Zufallsvariable auf  $\Omega$ , und es gilt  $Z(\Omega) = \{2, 3, \dots, 12\}$ . Um für  $x \in Z(\Omega)$  die jeweiligen Wahrscheinlichkeitsgewichte  $p_Z(x)$  zu bestimmen, überlegen wir uns, wie jeweils die Menge  $\{\omega \in \Omega : Z(\omega) = x\}$  aussieht. Betrachten wir den Fall  $x = 2$ . Die Summe zweier Würfe kann genau dann 2 sein, wenn beide Würfe 1 gezeigt haben. Somit ist  $\{\omega \in \Omega : Z(\omega) = 2\} = \{(1, 1)\}$ , und deshalb  $p_Z(2) = \mathbb{P}(\{(1, 1)\}) = \frac{1}{36}$ . Das Ergebnis  $x = 3$  kann jedoch auf zwei Arten erreicht werden: Erster Wurf 1 und zweiter Wurf 2, oder erster Wurf 2 und zweiter Wurf 1. Somit ist  $p_Z(3) = \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{2}{36} = \frac{1}{18}$ . In der Tabelle sind alle Ergebnisse und die zugehörigen Wahrscheinlichkeiten aufgelistet.

$x$	2	3	4	5	6	7	8	9	10	11	12
$\{\omega : Z(\omega) = x\}$	(1,1)	(1,2) (2,1)	(1,3) (3,1) (2,2)	(1,4) (4,1) (2,3) (1,4)	(1,5) (5,1) (2,4) (4,2) (3,3)	(1,6) (6,1) (2,5) (5,2) (3,4) (4,3)	(2,6) (6,2) (3,5) (5,3) (4,4)	(3,6) (6,3) (4,5) (5,4)	(4,6) (6,4) (5,5)	(5,6) (6,5)	(6,6)
$p_Z(x)$	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36

Wir sehen also, dass  $Z = 7$  die höchste Wahrscheinlichkeit hat – eine Tatsache, die aus Gesellschaftsspielen bekannt sein dürfte. Der Grund ist, dass  $Z = 7$  aus der höchsten Zahl von möglichen Ergebnissen zusammengesetzt werden kann.

**Beispiel 3.11** (Diskrete Gleichverteilung). Ein wichtiges Beispiel für eine diskrete Verteilung ist gegeben, wenn  $X$  mit gleicher Wahrscheinlichkeit jeden der Werte in einer Menge mit  $n \in \mathbb{N}$  Elementen annehmen kann. Ohne Beschränkung der Allgemeinheit können wir dann  $X(\Omega) = \{1, \dots, n\}$  wählen, und es gilt

$$\mathbb{P}(X = k) = \frac{1}{n}, \quad k \in \{1, \dots, n\}.$$

Da jedes  $k$  dieselbe Wahrscheinlichkeit hat, entspricht das einem Laplace-Raum. Wenn man umgangssprachlich davon spricht, dass etwas “zufällig” geschieht, dann meint man oft implizit, dass eine (diskrete) Gleichverteilung vorliegt.

**3.2. Häufigkeiten, Histogramme und Verteilungen.** Genau wie Wahrscheinlichkeiten sind auch Zufallsvariablen (und Verteilungen) ein mathematisches Konzept, und können in der Wirklichkeit nicht direkt beobachtet werden. Beobachten können wir dagegen einzelne konkrete Ergebnisse von Zufallsexperimenten, bzw. deren jeweiligen Häufigkeiten. Beobachtet man ein Ergebnis eines Zufallsexperiments und bestimmt daraus den Wert einer dadurch bestimmten Zufallsvariablen, so nennt man das Ergebnis daraus eine *Realisierung* einer Zufallsvariablen. Führt man das Experiment mehrfach aus, kann man so mehrere Realisierungen bestimmen, und somit wiederum *Häufigkeiten* für die auftretenden Werte bestimmen. Häufigkeiten werden gern in *Histogrammen dargestellt*:

Bild siehe Vorlesung

Werden eine hinreichend große Anzahl an (unabhängigen) Realisierungen derselben diskreten Zufallsvariablen beobachtet, so sollte die Form des Histogramms der Verteilung der Zufallsvariablen sehr ähnlich sein – denn wir hatten in Kapitel 1.2 gesehen, dass die Häufigkeiten jeweils den Wahrscheinlichkeiten entsprechen.

**Beispiel 3.12** (Summe beim Würfeln). Wir werfen zwei faire Würfel jeweils zwei Mal, und notieren uns die Summe der Augenzahlen. Das machen wir insgesamt 100 Mal (bzw. lassen es von einem Rechner simulieren). Wir berechnen die resultierenden Häufigkeiten, und zeichnen ein Histogramm:

Bild siehe Vorlesung

Zum Vergleich die Verteilung, die wir in Beispiel 3.10 berechnet haben:

Bild siehe Vorlesung

Betrachtet man eine nicht-diskrete Zufallsvariable, z.B. eine Wartezeit, so wird man üblicherweise lauter verschiedene Werte beobachten. Um ein Histogramm zu zeichnen, werden normalerweise ähnliche Werte gruppiert.

Beispiel siehe Vorlesung

### 3.3. Bernoulli-Verteilung, Binomialverteilung, geometrische Verteilung.

**Definition 3.13** (Bernoulli-Experiment). Ein *Bernoulli-Experiment* ist ein Zufallsexperiment, welches genau zwei mögliche Ausgänge hat.

Ein Bernoulli-Experiment ist also sozusagen ein besonders einfaches Zufallsexperiment. Die beiden möglichen Ausgänge bezeichnet man oft mit “Erfolg” und “Misserfolg”, oder auch mit 0 und 1. Da es nur zwei Möglichkeiten gibt, muss im Prinzip nur die Wahrscheinlichkeit für eins der beiden möglichen Ergebnisse bekannt sein. Normalerweise wählt man die Bezeichnung  $p = \mathbb{P}(1) = \mathbb{P}(\text{“Erfolg”})$  als die sogenannte “Erfolgswahrscheinlichkeit”. Damit gilt dann automatisch  $\mathbb{P}(0) = 1 - \mathbb{P}(1) = 1 - p$  als Wahrscheinlichkeit für das andere mögliche Ergebnis, also den Misserfolg.

**Definition 3.14** (Bernoulli-Verteilung). Eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  heißt *Bernoulli-verteilt zum Parameter  $p \in [0, 1]$*  (bzw. hat die Bernoulli-Verteilung zum Parameter  $p$ ), falls

$$X(\Omega) = \{0, 1\} \quad \text{und} \quad \mathbb{P}(X = 1) = p$$

gelten.

Für die Bernoulli-Verteilung gilt also  $\mathbb{P}(X = 0) = 1 - p$ , oder in der Notation des letzten Abschnitts,  $p_X(1) = p$  und  $p_X(0) = 1 - p$ .

**Beispiel 3.15.** Siehe Vorlesung.

Die Bernoulli-Verteilung ist also besonders einfach, und daher geeignet, als “Baustein” für kompliziertere Situationen zu dienen.

**Definition 3.16** (Binomialverteilung). Seien  $n \in \mathbb{N}$  und  $p \in [0, 1]$ . Eine Zufallsvariable  $X$  heißt *binomialverteilt mit Parametern  $n$  und  $p$*  falls gilt:  $X(\Omega) = \{0, \dots, n\}$ , und

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

Hierbei ist  $\binom{n}{k}$  der Binomialkoeffizient, der aus der Schule oder der Analysis I bekannt sein sollte, Stichwort: Binomischer Lehrsatz. Zur Erinnerung: Es gilt

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{1 \cdot 2 \cdot \dots \cdot n}{1 \cdot \dots \cdot k \cdot 1 \cdot \dots \cdot (n-k)}.$$

Diese Verteilung sieht schon etwas komplizierter aus als die Bernoulli-Verteilung. Bevor wir sehen, welche Bedeutung die Binomialverteilung hat, rechnen wir erst nach, dass es sich tatsächlich um eine Verteilung handelt, d.h. dass gilt  $\sum_{k=0}^n p_X(k) = 1$ . (Siehe Vorlesung).

**Satz 3.17** (Binomialverteilung im wiederholten Bernoulli-Experiment). *Ein Bernoulli-Experiment mit Parameter  $p$  wird  $n$  mal unter gleichbleibenden Bedingungen wiederholt. Sei  $X$  die Anzahl Erfolge. Dann ist  $X$  binomialverteilt mit Parameter  $n$  und  $p$ .*

*Beweis.* Siehe Vorlesung. □

**Beispiel 3.18** (Urnenmodell, Binomialverteilung). Wir betrachten eine Urne, welche 20 rote und 30 weiße Kugeln enthält. Daraus ziehen wir 4 Kugeln mit Zurücklegen, d.h. wir ziehen eine Kugel, notieren ob sie rot oder weiß ist, legen sie zurück, mischen gut, ziehen

wieder eine Kugel, usw. bis wir 4 Kugeln gezogen haben. Wie groß ist die Wahrscheinlichkeit, dabei mindestens drei Mal eine weiße Kugel zu ziehen?

Wir bezeichnen die Anzahl weißer Kugeln mit  $X$ . Bei diesem Experiment handelt es sich um ein wiederholtes Bernoulli-Experiment, denn wir haben in jedem Versuch genau zwei Möglichkeiten, nämlich eine rote oder eine weiße Kugel zu ziehen. Nennen wir das Ergebnis “weiß” den Erfolg. Was ist dann die Erfolgswahrscheinlichkeit  $p$ ? Dies ist nach Definition die Wahrscheinlichkeit, mit der in *einem* Versuch, also einmal Ziehen, Erfolg, also weiß kommt. Somit ist

$$p = \frac{\text{Anzahl weißer Kugeln in der Urne}}{\text{Gesamtzahl der Kugeln in der Urne}} = \frac{30}{50} = \frac{3}{5}.$$

Da wir viermal Ziehen, ist die Anzahl der Wiederholungen  $n = 4$ . Somit ist also nach Satz 3.17  $X$  binomialverteilt mit Parametern  $p = 3/5$  und  $n = 4$ , also gilt,

$$\mathbb{P}(X = k) = \binom{4}{k} \left(\frac{3}{5}\right)^k \left(1 - \frac{3}{5}\right)^{4-k}, \quad k = 0, \dots, 4.$$

Da wir die Wahrscheinlichkeit für  $X \geq 3$  berechnen wollen, erhalten wir

$$\mathbb{P}(X \geq 3) = \mathbb{P}(X = 3) + \mathbb{P}(X = 4) = \binom{4}{3} \left(\frac{3}{5}\right)^3 \left(\frac{2}{5}\right)^1 + \binom{4}{4} \left(\frac{3}{5}\right)^4 \left(\frac{2}{5}\right)^0 = \frac{297}{625}$$

**Beispiel 3.19** (Terminierung eines Algorithmus). Von einem zufälligen Algorithmus ist bekannt, dass er mit Wahrscheinlichkeit 0.8 terminiert. Wird der Algorithmus 40 Mal aufgerufen, so ist die Anzahl Läufe, bei denen der Algorithmus terminiert, binomialverteilt mit  $n = 40$  und  $p = 0.8$ .

**Beispiel 3.20.** In einem Experiment mit *Drosophila* befinden sich 100 Fruchtfliegen in einem Behälter. Davon tragen  $m$  die Mutation “rote Augen”. Wir untersuchen der Reihe nach 15 Fruchtfliegen, und geben die untersuchte Fliege jedesmal gleich wieder zurück in den Behälter. Mit welcher Wahrscheinlichkeit finden wir weniger als 5 Mal rote Augen? Dieses Vorgehen kann man genau mit einem Urnen-Modell beschreiben, indem wir uns den Behälter als Urne und die *Drosophila* als Kugeln vorstellen. Die Erfolgswahrscheinlichkeit  $p$  beim einmaligen Durchführen ist dann

$$p = \frac{m}{100}$$

da wir insgesamt 100 Fliegen haben, und davon  $m$  rote Augen haben – “rote Augen” ist also das Erfolgsereignis. Somit ist, wenn  $Y$  die Anzahl untersuchter Fliegen mit roten Augen bezeichnet,

$$\mathbb{P}(Y < 5) = \mathbb{P}(Y = 0) + \mathbb{P}(Y = 1) + \mathbb{P}(Y = 2) + \mathbb{P}(Y = 4),$$

wobei die Anzahl Wiederholungen  $n = 15$  ist, also

$$\mathbb{P}(Y = k) = \binom{15}{k} p^k (1 - p)^{15-k}.$$

Die Binomialverteilung tritt also als Verteilung der Anzahl Erfolge im wiederholten Bernoulli-Experiment auf, wenn die Anzahl Wiederholungen vorgegeben ist. Wir können uns aber auch eine andere Frage stellen, nämlich “wie viele Versuche muss ich machen, bis ich zum ersten Mal einen Erfolg sehe?”, oder im Urnenmodell von Beispiel 3.18, “wie viele Kugeln muss ich (mit Zurücklegen, also ohne gegenseitige Beeinflussung der Teilexperimente) aus der Urne ziehen, bis ich zum ersten Mal eine weiße Kugel ziehe?” Wir werden gleich sehen, dass diese Größe von der *geometrischen Verteilung* beschrieben wird.

**Definition 3.21** (Geometrische Verteilung). Sei  $p \in [0, 1]$ . Eine Zufallsvariable  $X$  heißt *geometrisch verteilt mit Parameter  $p$*  falls gilt:  $X(\Omega) = \mathbb{N}$ , und

$$p_X(k) = \mathbb{P}(X = k) = p(1 - p)^{k-1}, \quad k \in \mathbb{N}.$$

Eine geometrische Zufallsvariable hat also einen abzählbaren, keinen endlichen Wertebereich. Wie immer müssen wir erst überprüfen, dass es sich hier tatsächlich um eine Verteilung handelt, d.h. dass  $\sum_{k=1}^{\infty} p_X(k) = 1$  gilt. Dies folgt mit einer einfachen Rechnung:

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

Dabei haben wir beim zweiten Gleichheitszeichen den Faktor  $p$  vor die Summe gezogen und den Summationsindex um eins verschoben, und beim dritten Gleichheitszeichen die aus der Analysis bekannte Formel für eine *geometrische Reihe* benutzt (!). Die Tatsache, dass hier eine geometrische Reihe auftritt, ist auch der Grund dafür, dass die Verteilung *geometrisch* heißt.

**Satz 3.22.** *Wir betrachten ein Bernoulli-Experiment mit Parameter  $p$ . Sei  $X$  die Anzahl Versuche, die gemacht werden müssen, bis zum ersten Mal “Erfolg” eintritt. Dann ist  $X$  geometrisch verteilt mit Parameter  $p$ .*

*Beweis.* Da wir es mit einem wiederholten Bernoulli-Experiment zu tun haben, ist bei jedem Versuch die Wahrscheinlichkeit Erfolg zu haben gleich  $p$ , und die Wahrscheinlichkeit für Misserfolg gleich  $1 - p$ . Um genau beim  $k$ -ten Versuch zum ersten Mal Erfolg zu haben, müssen also zuerst  $k - 1$  Misserfolge und dann ein Erfolg eintreten. Da die Ereignisse “1. Zug Misserfolg”, “2. Zug Misserfolg” ... “ $k$ -ter Zug Erfolg” unabhängig sind, erhalten wir

$$\mathbb{P}(X = k) = \underbrace{(1 - p) \cdot \dots \cdot (1 - p)}_{k-1 \text{ Mal}} p = (1 - p)^{k-1} p.$$

□

**Beispiel 3.23.** Bei einem stochastischen Algorithmus, der mit Wahrscheinlichkeit 0.8 terminiert, möchten wir wissen, wie viele Versuche wir machen müssen, bis der Algorithmus zum ersten Mal terminiert. Gemäß Satz 3.22 ist dies eine geometrisch verteilte Zufallsvariable  $X$  mit Parameter  $p = 0.8$ , also ist

$$\mathbb{P}(X = 1) = 0.8, \quad \mathbb{P}(X = 2) = 0.2 \cdot 0.8 = 0.16, \quad \mathbb{P}(X = 3) = (0.2)^2 \cdot 0.8 \approx 0.032.$$

**Beispiel 3.24.** Im Drosophila-Beispiel müssen wir also eine geometrische Anzahl Versuche mit Parameter  $m/100$  machen, bis wir eine Fliege mit roten Augen erwischen. Die Wahrscheinlichkeit, dass dies nach genau 3 Versuchen zum ersten Mal passiert, ist also

$$\frac{m}{100} \left(1 - \frac{m}{100}\right)^2.$$

**Beispiel 3.25** (Simulation von Zufallsvariablen mit R). Mit R können Zufallsvariablen mit einer bestimmten Verteilung simuliert werden. Beispielsweise erhält man mit dem Befehl `rbinom(k,n,p)` als Ausgabe  $k$  Realisierungen einer binomialverteilten Zufallsvariablen mit Parametern  $n$  und  $p$ . Mit `rgeom(k, p)` erhält man  $k$  Realisierungen einer geometrischen Zufallsvariablen mit Parameter  $p$ . Auch andere Verteilungen sind programmiert.

### 3.4. Poisson-Verteilung.

**Definition 3.26** (Poisson-Verteilung). Sei  $\lambda > 0$ . Eine Zufallsvariable  $X$  heißt *Poisson-verteilt* mit Parameter  $\lambda$ , falls  $X(\Omega) = \mathbb{N}_0$ , und

$$p_X(k) = \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Die Poisson-Verteilung tritt oft in Situationen auf, in denen die Anzahl (seltener) Ereignisse in einem bestimmten Zeitraum untersucht werden. Der Parameter  $\lambda$  bezeichnet die *Rate* mit der die Ereignisse eintreten – je größer  $\lambda$ , desto kürzer die Abstände, d.h. desto mehr Ereignisse pro Zeiteinheit.

Beispiele für Situationen die mit Hilfe der Poisson-Verteilung modelliert werden:

- Anzahl Anrufe, die in einer Telefonzentrale pro Minute eingehen
- Anzahl Autos, die pro Stunde eine bestimmte Brücke befahren
- Anzahl Datenbankabfragen, die pro Zeiteinheit eingehen
- Anzahl Mutationen auf einem Stück DNS
- Anzahl Schäden pro Tag in einem Versicherungsunternehmen

Mit `rpois(n, λ)` kann man sich in R  $n$  Realisierungen einer Poisson-Verteilung zum Parameter  $\lambda$  ausgeben lassen.

Ein Grund für die Bedeutung der Poisson-Verteilung liegt in folgendem Satz:

**Satz 3.27** (Poisson-Grenzwertsatz). Sei  $n \in \mathbb{N}$ , und sei  $(p_n)_{n \in \mathbb{N}}$  eine Folge von Zahlen aus  $[0, 1]$  mit  $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda \in ]0, \infty[$ . Sei  $X_n$  binomial verteilt mit Parametern  $n$  und  $p_n$ , und sei  $X$  Poisson-verteilt mit Parameter  $\lambda$ . Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k).$$

Für einen Beweis siehe z.B. Klenke, Wahrscheinlichkeitstheorie (Springer, 2. Auflage 2008).

Den Inhalt des Poisson-Grenzwertsatzes kann man folgendermaßen zusammenfassen: Wenn die Anzahl Versuche groß und die Erfolgswahrscheinlichkeit klein ist, so ist die Anzahl Erfolge ungefähr Poisson-verteilt. Deshalb beschreibt die Poisson-Verteilung auch die Anzahl seltener Ereignisse.

**Beispiel 3.28.** Siehe Vorlesung

**3.5. Zusammenfassung.** Wir fassen nun die Ergebnisse über die bisher bekannten diskreten Verteilungen zusammen.

Name	Parameter	$X(\Omega)$	$\mathbb{P}(X = k)$	Notation
Gleichverteilung	$n \in \mathbb{N}$	$\{1, \dots, n\}$	$1/n$	$X \sim \text{Unif}(n)$
Bernoulli	$p \in [0, 1]$	$\{0, 1\}$	$\mathbb{P}(X = 1) = p$	$X \sim \text{Ber}(p)$
Binomial	$p \in [0, 1], n \in \mathbb{N}$	$\{0, \dots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$X \sim \text{Bin}(n, p)$
Geometrisch	$p \in [0, 1]$	$\mathbb{N}$	$(1-p)^{k-1} p$	$X \sim \text{Geo}(p)$
Poisson	$\lambda > 0$	$\mathbb{N}_0$	$\frac{e^{-\lambda} \lambda^k}{k!}$	$X \sim \text{Poi}(\lambda)$

Zum Auftreten der Verteilungen:

Bernoulli	Bernoulli-Experiment
Binomial	Anzahl Erfolge im wiederholten Bernoulli-Experiment
Geometrisch	Anzahl Versuche bis zum ersten Erfolg im Bernoulli-Experiment
Poisson	Anzahl (seltener) Ereignisse in einem gegebenen Zeitraum

**3.6. Weitere diskrete Verteilungen und ihre Anwendungen.** Zipf-Verteilung, Hypergeometrische Verteilung siehe Vorlesung.

## 4. GEMEINSAME VERTEILUNG, UNABHÄNGIGKEIT VON ZUFALLSVARIABLEN

In komplexeren Situationen wird es oft nicht möglich sein, einen ganzen Sachverhalt mit nur einer Zufallsvariablen zu beschreiben, sondern man benötigt möglicherweise zwei oder noch mehr Zufallsvariablen, die sich unter Umständen gegenseitig beeinflussen. In diesem Kapitel lernen wir, wie man zwei oder mehr Zufallsvariablen mit Hilfe ihrer gemeinsamen Verteilung beschreibt, und wie man “gegenseitigen stochastischen Einfluss” von zwei Zufallsvariablen quantifiziert. Die **Lernziele** dieses Kapitels sind:

- Gemeinsame Verteilungen kennen und berechnen können
- Bedingte Verteilungen und Randverteilungen berechnen können
- Zufallsvariablen auf Unabhängigkeit überprüfen können.

## 4.1. Gemeinsame Verteilung, bedingte Verteilung.

**Definition 4.1** (Gemeinsame Verteilung). Seien  $X$  und  $Y$  zwei diskrete Zufallsvariablen, welche auf demselben Wahrscheinlichkeitsraum  $(\Omega, \mathbb{P})$  definiert sind. Die *gemeinsame Verteilung* von  $X$  und  $Y$  ist gegeben durch  $p_{X,Y} : X(\Omega) \times Y(\Omega) \rightarrow [0, 1]$ ,

$$p_{X,Y}(x, y) := \mathbb{P}(X = x, Y = y), \quad x \in X(\Omega), y \in Y(\Omega).$$

Eine gemeinsame Verteilung von zwei Zufallsvariablen ist also eine *Funktion von zwei Variablen*, welche für jedes Paar von Werten  $x \in X(\Omega), y \in Y(\Omega)$  angibt, wie wahrscheinlich es ist, dass die Zufallsvariable  $X$  den Wert  $x$  und gleichzeitig die Zufallsvariable  $Y$  den Wert  $y$  annimmt.

Beispiele: Größe und Gewicht eines Menschen; Anzahl Bestellungen und Bearbeitungszeit in einem Logistikzentrum; Anzahl und Höhe von Schäden in einem Versicherungsunternehmen, usw.

Die gemeinsame Verteilung von mehr als zwei Zufallsvariablen wird analog definiert: Seien  $X_1, \dots, X_n$  diskrete Zufallsvariablen. Ihre gemeinsame Verteilung ist gegeben durch

$$p_{X_1, \dots, X_n} : X_1(\Omega) \times \dots \times X_n(\Omega) \rightarrow [0, 1] : p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

**Beispiel 4.2** (Urnenmodell). In einer Urne befinden sich 4 Kugeln, numeriert von 1 bis 4. Es werden zwei Kugeln mit Zurücklegen gezogen. Sei  $X$  die Summe der beiden Zahlen, und  $Y$  das Minimum der beiden gezogenen Zahlen. Was ist die gemeinsame Verteilung von  $X$  und  $Y$ ?

Zuerst müssen wir die Wertebereiche bestimmen.  $X(\Omega)$  ist die Menge aller Zahlen, die als Summe zweier Zahlen zwischen 1 und 4 dargestellt werden können, also  $X(\Omega) = \{2, \dots, 8\}$ . Für  $Y(\Omega)$  finden wir  $Y(\Omega) = \{1, \dots, 4\}$ . Dabei ist es sinnvoll, den Wahrscheinlichkeitsraum

$$\Omega = \{1, 2, 3, 4\}^2$$

zugrunde zu legen, d.h.  $\omega \in \Omega$  hat die Form  $(\omega_1, \omega_2)$ , mit  $\omega_1, \omega_2 \in \{1, \dots, 4\}$ . Dies ist ein Laplace-Raum mit  $\mathbb{P}(\omega) = \frac{1}{16}$  für jedes  $\omega \in \Omega$ .



Wir können formal schreiben

$$X(\omega) = \omega_1 + \omega_2, \text{ und } Y(\omega) = \min\{\omega_1, \omega_2\}.$$

Nun interessieren wir uns für die Wahrscheinlichkeiten

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y).$$

Diese müssen wir nun für alle möglichen Werte  $x \in X(\Omega)$  und  $y \in Y(\Omega)$  bestimmen, z.B. also für  $x = 2$  und  $y = 1$ . Wie kann dieses Ereignis realisiert werden? Die Summe ist genau dann 2, wenn zweimal die Kugel 1 gezogen wurde, d.h. wenn  $\omega = (1, 1)$  ist. In diesem Fall ist automatisch  $Y = 1$ , also ist  $\{X = 2, Y = 1\} = \{(1, 1)\}$ . Somit erhalten wir

$$p_{X,Y}(2, 1) = \mathbb{P}(X = 2, Y = 1) = \mathbb{P}((1, 1)) = \frac{1}{16}.$$

Für  $x = 3, y = 1$  gibt es zwei Möglichkeiten: Erste Kugel 1 und zweite 2, oder umgekehrt, also  $\{X = 3, Y = 1\} = \{(1, 2), (2, 1)\}$ .

$$p_{X,Y}(3, 1) = \mathbb{P}(X = 3, Y = 1) = \mathbb{P}((1, 2), (2, 1)) = \frac{2}{16} = \frac{1}{8}.$$

Andererseits ist  $x = 3, y = 3$  unmöglich zu erreichen, denn wenn die Summe 2 ist, kann keine Zahl größer als 2 sein, also insbesondere kann das Minimum nicht gleich 3 sein. Somit ist

$$\mathbb{P}(X = 3, Y = 3) = p_{X,Y}(3, 3) = 0.$$

Analog kann man die anderen Wahrscheinlichkeiten ausrechnen. Als kompakte Darstellung der gemeinsamen Verteilung bietet sich eine Tabelle an:

Gemeinsame Verteilung  $p_{X,Y}$  von  $X$  und  $Y$

$Y \backslash X$	2	3	4	5	6	7	8
1	1/16	1/8	1/8	1/8	0	0	0
2	0	0	1/16	1/8	1/8	0	0
3	0	0	0	0	1/16	1/8	0
4	0	0	0	0	0	0	1/16

**Beispiel 4.3** (Zufälliger Graph). Siehe Vorlesung

**Definition 4.4** (Randverteilung). Sei die gemeinsame Verteilung von  $X$  und  $Y$  gegeben. Die *Randverteilungen* von  $X$  bzw. von  $Y$  sind die Verteilungen der einzelnen Zufallsvariablen  $X$  bzw.  $Y$ ,

$$p_X(x) := \mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} p_{X,Y}(x, y),$$

$$p_Y(y) := \mathbb{P}(Y = y) = \sum_{x \in X(\Omega)} p_{X,Y}(x, y).$$

**Beispiel 4.5** (Urnenmodell, Fortsetzung von Beispiel 4.2). Die Tabelle der gemeinsamen Verteilung wird um die Randverteilungen ergänzt, indem man unten eine Zeile und rechts eine Spalte einfügt, und dort die jeweiligen Randverteilungen einträgt. Die Zahlen ergeben sich als Summe der Werte oberhalb bzw. rechts davon.

Gemeinsame Verteilung und Randverteilung								
$\begin{array}{c} X \\ Y \end{array}$	2	3	4	5	6	7	8	$P(Y = \cdot)$
1	1/16	1/8	1/8	1/8	0	0	0	7/16
2	0	0	1/16	1/8	1/8	0	0	5/16
3	0	0	0	0	1/16	1/8	0	3/16
4	0	0	0	0	0	0	1/16	1/16
$P(X = \cdot)$	1/16	1/8	3/16	1/4	3/16	1/8	1/16	1

Die 1 in der unteren linken Ecke kann als Kontrollwert verstanden werden, sie ergibt sich durch Summation sowohl der Spalte ganz rechts als auch der Zeile ganz unten.

**Definition 4.6** (Bedingte Verteilung). Seien  $X, Y$  zwei Zufallsvariablen welche auf demselben Wahrscheinlichkeitsraum  $(\Omega, \mathbb{P})$  definiert sind. Sei  $y \in Y(\Omega)$ . Die *bedingte Verteilung* von  $X$  gegeben  $Y$  ist definiert als

$$p_{X|Y}(x|y) := \mathbb{P}(X = x | Y = y), \quad x \in X(\Omega), y \in Y(\Omega).$$

Wird auf einen *festen* Wert von  $Y$  bedingt, so berechnet man die *bedingte Verteilung* von  $X$  gegeben  $Y = y$  als

$$p_{X|Y=y}(x) := \mathbb{P}(X = x | Y = y), \quad x \in X(\Omega).$$

Nach Definition der bedingten Wahrscheinlichkeit und der gemeinsamen Verteilung kann man die bedingte Verteilung folgendermaßen berechnen:

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

Beachten Sie, dass die Rollen von  $X$  und  $Y$  hier unterschiedlich sind, der Ausdruck ist nicht symmetrisch, d.h. Vertauschen von  $X$  und  $Y$  führt auf unterschiedliche Ergebnisse!

**Beispiel 4.7** (Test auf Krankheit, Fortsetzung von Beispiel 2.6). **siehe Vorlesung**

**Beispiel 4.8** (Urnenmodell, Fortsetzung von Beispiel 4.2). Wie oben betrachten wir eine Urne mit 4 Kugeln, numeriert von 1 bis 4. Wir ziehen zwei Kugeln mit Zurücklegen. Sei wieder  $X$  die Summe der beiden Zahlen, und  $Y$  das Minimum der beiden gezogenen Zahlen. Aufgabe: Bestimme die bedingte Verteilung von  $Y$  gegeben  $X$ .

Wir betrachten erst die bedingte Verteilung von  $Y$  gegeben  $X = 2$ . Aus der Tabelle in 4.5 sehen wir:  $p_X(2) = \frac{1}{16}$ , und  $p_{X,Y}(2,1) = \frac{1}{16}$ ,  $p_{X,Y}(2,2) = p_{X,Y}(2,3) = p_{X,Y}(2,4) = 0$ . Somit erhalten wir

$$p_{Y|X=2}(1) = \frac{1/16}{1/16} = 1, \quad p_{Y|X=2}(2) = p_{Y|X=2}(3) = p_{Y|X=2}(4) = \frac{0}{1/16} = 0.$$

Analog berechnen wir  $p_{Y|X=3}, p_{Y|X=4}, \dots, p_{Y|X=8}$ , und fassen die Werte in einer Tabelle zusammen:

Bedingte Verteilung  $\mathbb{P}(Y|X)$  von  $X$  gegeben  $Y$

$Y \backslash X$	2	3	4	5	6	7	8
1	1	1	$2/3$	$1/2$	0	0	0
2	0	0	$1/3$	$1/2$	$2/3$	0	0
3	0	0	0	0	$1/3$	1	0
4	0	0	0	0	0	0	1

#### 4.2. Unabhängigkeit von Zufallsvariablen.

Zur Erinnerung (siehe Kapitel 2): Zwei *Ereignisse*  $A$  und  $B$  heißen unabhängig, falls  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$  gilt, oder äquivalent  $\mathbb{P}(A|B) = \mathbb{P}(A)$  bzw.  $\mathbb{P}(B|A) = \mathbb{P}(B)$  gelten, d.h. die Wahrscheinlichkeit für das eine Ereignis wird nicht durch das Eintreten (oder Nicht-Eintreten) des anderen beeinflusst.

Nun definieren wir den analogen Begriff für *Zufallsvariablen*. Obwohl die beiden Begriffe viel miteinander zu tun haben, ist es wichtig, sich die Unterschiede vor Augen zu führen!

**Definition 4.9** (Unabhängigkeit von Zufallsvariablen). Zwei Zufallsvariablen  $X$  und  $Y$  definiert auf demselben Wahrscheinlichkeitsraum heißen *unabhängig*, falls für alle  $x \in X(\Omega), y \in Y(\Omega)$  die *Ereignisse*  $\{X = x\}$  und  $\{Y = y\}$  unabhängig sind.

Daraus folgt unmittelbar mit der Definition der Unabhängigkeit von Ereignissen:

**Satz 4.10.** Zwei Zufallsvariablen  $X$  und  $Y$  sind genau dann unabhängig, wenn für alle  $x \in X(\Omega), y \in Y(\Omega)$  gilt

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

d.h. genau dann, wenn für alle  $x \in X(\Omega), y \in Y(\Omega)$  gilt

$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y).$$

**Beispiel 4.11** (Urnenmodell, Fortsetzung von Beispiel 4.5). Sind  $X$  und  $Y$  aus dem Urnen-Beispiel 4.2 unabhängig?

Zur Erinnerung die Tabelle der gemeinsamen Verteilung von  $X$  und  $Y$ :

$Y \backslash X$	2	3	4	5	6	7	8	$P(Y = \cdot)$
1	$1/16$	$1/8$	$1/8$	$1/8$	0	0	0	$7/16$
2	0	0	$1/16$	$1/8$	$1/8$	0	0	$5/16$
3	0	0	0	0	$1/16$	$1/8$	0	$3/16$
4	0	0	0	0	0	0	$1/16$	$1/16$
$P(X = \cdot)$	$1/16$	$1/8$	$3/16$	$1/4$	$3/16$	$1/8$	$1/16$	1

Beispielsweise sehen wir gleich in der ersten Zeile, dass  $\mathbb{P}(X = 2, Y = 1) \neq \mathbb{P}(X = 2)\mathbb{P}(Y = 1)$ , also können die Zufallsvariablen *nicht* unabhängig sein. Bei unabhängigen Zufallsvariablen müsste *jeder Eintrag im Inneren der Tabelle* gleich dem Produkt der beiden zugehörigen Randwahrscheinlichkeiten sein.

Man kann sich auch anschaulich klar machen, dass  $X$  und  $Y$  nicht unabhängig sein können, da beispielsweise die Information  $X = 8$  zwingend festlegt, dass  $Y = 4$  sein muss. Aus der Kenntnis von  $X$  lassen sich also Rückschlüsse auf  $Y$  ziehen, und umgekehrt.

Wir wissen aus Beispiel 3.7, dass Funktionen von Zufallsvariablen wieder Zufallsvariablen sind. Die Unabhängigkeit bleibt dabei erhalten.

**Satz 4.12** (Funktionen von unabhängigen Zufallsvariablen). *Seien  $X$  und  $Y$  unabhängige (diskrete) Zufallsvariablen, und seien  $f : X(\Omega) \rightarrow \mathbb{R}$  und  $g : Y(\Omega) \rightarrow \mathbb{R}$  zwei Funktionen. Dann sind die Zufallsvariablen  $f(X)$  und  $g(Y)$  ebenfalls unabhängig.*

(ohne Beweis)

### Beispiel 4.13.

Wie zu erwarten, kann man Unabhängigkeit auch von drei oder mehr Zufallsvariablen untersuchen. Dabei muss man jedoch etwas aufpassen, wie wir in einem Beispiel in den Hausaufgaben sehen werden.

**Definition 4.14** (Unabhängigkeit von mehr als zwei Zufallsvariablen). Seien  $X_1, X_2, \dots, X_n$  diskrete Zufallsvariablen auf  $(\Omega, \mathbb{P})$ . Diese heißen *unabhängig*, falls

$$\mathbb{P}(X_{i_1} = a_{i_1}, \dots, X_{i_k} = a_{i_k}) = \mathbb{P}(X_{i_1} = a_{i_1}) \cdot \dots \cdot \mathbb{P}(X_{i_k} = a_{i_k})$$

für alle  $k \leq n$ , für alle paarweise verschiedenen  $i_1, \dots, i_k \in \{1, \dots, n\}$ , und für alle  $a_{i_1} \in X_{i_1}(\Omega), \dots, a_{i_k} \in X_{i_k}(\Omega)$  gilt.

Eine analoge Definition gilt für Folgen  $(X_n)_{n \in \mathbb{N}}$  von Zufallsvariablen.

Wenn zwei Zufallsvariablen unabhängig sind, kann man eine Formel für die Verteilung ihrer Summe herleiten.

**Satz 4.15** (Faltungsformel). *Seien  $X$  und  $Y$  zwei unabhängige diskrete Zufallsvariablen auf  $(\Omega, \mathbb{P})$ . Dann hat die Zufallsvariable  $X + Y$  die Verteilung*

$$\mathbb{P}(X + Y = k) = \sum_{m \in X(\Omega)} \mathbb{P}(X = m) \cdot \mathbb{P}(Y = k - m).$$

Wichtig: Wenn  $X$  und  $Y$  nicht unabhängig sind, ist diese Formel falsch!

*Beweis.* Beachte dass  $(X + Y)(\Omega) = \{m + l : m \in X(\Omega), l \in Y(\Omega)\}$  gilt. Wir haben für  $k \in (X + Y)(\Omega)$ ,

$$\begin{aligned} \mathbb{P}(X + Y = k) &= \sum_{m \in X(\Omega)} \mathbb{P}(X = m, X + Y = k) \\ &= \sum_{m \in X(\Omega)} \mathbb{P}(X = m, Y = k - m) \\ &= \sum_{m \in X(\Omega)} \mathbb{P}(X = m) \mathbb{P}(Y = k - m). \end{aligned}$$

Dabei haben wir im letzten Schritt die Unabhängigkeit verwendet. □

## 5. KENNGRÖSSEN FÜR ZUFALLSVARIABLEN

Wir wissen aus dem letzten Kapitel, dass die *Verteilung* im Prinzip alle Informationen über die zugehörigen Zufallsvariablen enthält. Wir haben aber auch gemerkt, dass es nicht immer einfach und teilweise aufwändig ist, die Verteilung komplett zu bestimmen, und dass man auch nicht immer auf den ersten Blick die für ein konkretes Problem relevanten Angaben erkennt. Ziel dieses Kapitels ist es deshalb, einige *Kenngrößen* zu identifizieren, die bereits zentrale Informationen über eine Zufallsvariable enthalten, und die hoffentlich in vielen Fällen einfach zu berechnen sind. Konkret werden wir den Erwartungswert und die Varianz von Zufallsvariablen kennenlernen, sowie Kovarianz und Korrelation als Kenngrößen für den Zusammenhang zweier Zufallsvariablen.

Die **Lernziele** dieses Kapitel sind:

- Die Begriffe Erwartungswert, Varianz, Standardabweichung kennen
- Erwartungswert und Varianz für diskrete Verteilungen berechnen können
- Informationen über die Verteilung daraus herleiten können
- Kovarianz und Korrelation berechnen können, und wissen, welche Informationen daraus erhalten werden können.

Vorkenntnisse: Neben dem bisher gelernten Stoff sowie den dabei benötigten Grundlagen werden wir in diesem Kapitel insbesondere mehrfach Grenzwerte von unendlichen Reihen berechnen.

## 5.1. Erwartungswert.

**Beispiel 5.1.** Wir werfen einen fairen Würfel 100 Mal, und notieren die Ergebnisse. Sei  $y_i$  das Ergebnis des  $i$ -ten Wurfs. Wie groß ist dann der Mittelwert über alle Würfe, d.h.

$$\frac{1}{100} \sum_{i=1}^{100} y_i?$$

Da die Ergebnisse  $\{1, \dots, 6\}$  gleich wahrscheinlich sind, erwarten wir im Durchschnitt jedes der 6 möglichen Ergebnisse gleich oft. Das Mittel über viele (z.B. 100) Versuche sollte damit ungefähr gleich dem arithmetischen Mittel (Durchschnitt) der möglichen Werte sein, d.h.

$$\frac{1}{100} \sum_{i=1}^{100} y_i \approx \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3.5.$$

Sie können dieses Experiment entweder mit einem echten Würfel oder mit einem geeigneten Programm am Rechner durchführen, und werden feststellen, dass tatsächlich eine Zahl in der Nähe von 3.5 dabei herauskommt.

Diese Beobachtung ist ein Beispiel für das sogenannte “Gesetz der großen Zahlen”, welches wir später in der Vorlesung kennen lernen werden, und welches besagt, dass das Mittel über viele Ergebnisse eines wiederholten Zufallsexperiments ungefähr gleich dem *Erwartungswert* ist. Im obigen Beispiel ist der Erwartungswert genau das arithmetische Mittel der möglichen Ergebnisse. Allgemein ist er folgendermaßen definiert:

**Definition 5.2** (Erwartungswert einer diskreten Zufallsvariablen). Sei  $X$  eine diskrete Zufallsvariable. Der *Erwartungswert* von  $X$  ist definiert als

$$\mathbb{E}[X] := \sum_{k \in X(\Omega)} k \cdot \mathbb{P}(X = k),$$

sofern  $\sum_{k \in X(\Omega)} |k| \cdot \mathbb{P}(X = k) < \infty$  ist. Wenn letzters erfüllt ist, so sagen wir, dass der *Erwartungswert existiert*.

*Bemerkung.* Falls der Wertebereich  $X(\Omega)$  endlich ist, d.h. falls die Zufallsvariable  $X$  nur endlich viele Werte annehmen kann (z.B. beim Würfeln), ist die Bedingung  $\sum_{k \in X(\Omega)} |k| \cdot \mathbb{P}(X = k) < \infty$  stets erfüllt. Ist jedoch der Wertebereich unendlich, so kann es vorkommen, dass  $\sum_{k \in X(\Omega)} |k| \cdot \mathbb{P}(X = k) = \infty$  ist. In diesem Fall ist der Erwartungswert nicht definiert, man sagt *der Erwartungswert existiert nicht*. Diese Situation kann tatsächlich vorkommen, wie wir noch sehen werden.

**Beispiel 5.3** (Würfeln, Fortsetzung von Beispiel 5.1). Betrachten wir noch einmal die Situation aus Beispiel 5.1.  $X$  bezeichne die Augenzahl beim Werfen eines fairen Würfels. Es gilt also  $X(\Omega) = \{1, 2, \dots, 6\}$ , und  $\mathbb{P}(X = x) = \frac{1}{6}$  für  $x \in \{1, \dots, 6\}$ . Dann erhalten wir aus der Definition des Erwartungswerts

$$\mathbb{E}[X] = \sum_{k \in X(\Omega)} k \mathbb{P}(X = k) = \sum_{k=1}^6 k \cdot \mathbb{P}(X = k) = \sum_{k=1}^6 k \cdot \frac{1}{6} = \frac{1}{6}(1 + 2 + \dots + 6) = 3.5.$$

Das ist genau der Wert, den wir oben bereits erraten hatten.

**Beispiel 5.4** (Zufälliger Graph). siehe Vorlesung

**Beispiel 5.5** (Geometrische Verteilung). Sei  $X$  eine Zufallsvariable mit geometrischer Verteilung zum Parameter  $p$ . Wir wollen  $\mathbb{E}[X]$  berechnen. Einsetzen in die Definition liefert, da  $X(\Omega) = \mathbb{N}$  und  $\mathbb{P}(X = k) = p(1-p)^{k-1}$  ist, erhalten wir

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot p \cdot (1-p)^{k-1} = p \sum_{k=0}^{\infty} k(1-p)^{k-1}.$$

Um die unendliche Summe zu berechnen, verwenden wir nun einen Trick. Wir wissen aus der Analysis, dass

$$\sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}$$

ist (geometrische Reihe). Diese Gleichung können wir nun links und rechts nach  $p$  ableiten, und bekommen

$$(*) \quad \sum_{k=0}^{\infty} k(1-p)^{k-1}(-1) = -\frac{1}{p^2},$$

wobei wir einerseits die Kettenregel verwendet haben, und andererseits die Tatsache, dass bei einer Potenzreihe innerhalb des Konvergenzradius die Ableitung mit der Summation

vertauscht werden können. Die linke Seite dieser Gleichung (\*) ist fast (bis auf das Vorzeichen und den konstanten Vorfaktor  $p$ ) die rechte Seite der obigen Gleichung für  $\mathbb{E}[X]$ , wir erhalten also zusammengefasst

$$\mathbb{E}[X] = p \sum_{k=0}^{\infty} k(1-p)^{k-1} = -p \sum_{k=0}^{\infty} k(1-p)^{k-1}(-1) \stackrel{(*)}{=} (-p) \cdot \left(-\frac{1}{p^2}\right) = \frac{1}{p}.$$

Also ist der Erwartungswert der geometrischen Verteilung mit Parameter  $p$  genau  $\frac{1}{p}$ .

Ist also etwa  $p = 1/2$ , ist der entsprechende Erwartungswert 2. Bei  $p = 1/10$  ist er 10. Für kleine  $p$  muss man also im Mittel länger auf den ersten Erfolg warten, was nicht überrascht.

**Beispiel 5.6** (Zipf-Verteilung). *siehe Vorlesung*

Die Definition des Erwartungswertes kann leicht umgeschrieben werden, um das folgende Ergebnis zu erhalten:

**Satz 5.7.** *Falls  $\Omega$  höchstens abzählbar ist und der Erwartungswert von  $X$  existiert, so gilt*

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega).$$

In den meisten Situationen ist es aber einfacher, mit der Definition des Erwartungswertes zu arbeiten als mit diesem Satz, da man üblicherweise die Verteilung kennt.

Der Erwartungswert hat eine Reihe wichtiger Eigenschaften.

**Satz 5.8** (Eigenschaften des Erwartungswertes). *Seien  $X$  und  $Y$  diskrete Zufallsvariablen auf  $(\Omega, \mathbb{P})$ . Dann gelten*

- (a)  $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- (b)  $\mathbb{E}[a \cdot X] = a \cdot \mathbb{E}[X]$  für jedes  $a \in \mathbb{R}$
- (c) Falls  $X(\omega) \geq 0$  für alle  $\omega$ , so ist  $\mathbb{E}[X] \geq 0$ .
- (d) Falls  $X$  konstant ist, d.h.  $X(\omega) = c \in \mathbb{R}$  für alle  $\omega$ , dann ist  $\mathbb{E}[X] = c$ .
- (e) Sei  $f : \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion. Dann gilt

$$\mathbb{E}[f(X)] = \sum_{x \in X(\Omega)} f(x) \mathbb{P}(X = x).$$

- (f) Falls  $X$  und  $Y$  unabhängig sind, so gilt

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

*Beweis. siehe Vorlesung*

□

Diesen Satz kann man nun beispielsweise benutzen, um den Erwartungswert der Binomialverteilung schmerzlos zu berechnen.

**Beispiel 5.9** (Erwartungswert von Bernoulli- und Binomialverteilung). Wir beginnen mit dem Erwartungswert der Bernoulli-Verteilung, und verwenden dann die Beziehung zwischen Binomialverteilung und Bernoulli-Verteilung. Sei  $Y$  Bernoulli-verteilt zum Parameter  $p$ . Dann gilt

$$\mathbb{E}[Y] = 0 \cdot \mathbb{P}(Y = 0) + 1 \cdot \mathbb{P}(Y = 1) = 0 + 1 \cdot p = p.$$

Sei nun  $X$  binomialverteilt mit Parametern  $n$  und  $p$ . Dank Satz 3.17 können wir  $X$  schreiben als

$$X = \sum_{i=1}^n Y_i,$$

wenn  $Y_i, i = 1, \dots, n$  unabhängige, Bernoulli-verteilte Zufallsvariable mit Parameter  $p$  sind. Denken Sie einen Moment über diese Darstellung nach, wir werden sie noch mehrfach verwenden!

Nach Satz 5.8 (a) gilt dann

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n \mathbb{E}[Y_i] = \sum_{i=1}^n p = np.$$

Der Erwartungswert einer binomialverteilten Zufallsvariablen mit Parametern  $n$  und  $p$  ist also genau  $np$ , d.h. beim  $n$ -fach wiederholten Bernoulli-Experiment mit Erfolgswahrscheinlichkeit  $p$  sehen wir *im Mittel*  $np$  Erfolge.

**5.2. Anwendung: Laufzeit von Quicksort.** siehe Vorlesung (zweiter Teil zum Thema stochastische Algorithmen)

**5.3. Varianz.**

**Beispiel 5.10** (Glücksspiel). Wir betrachten ein simples Glücksspiel, wobei einfach eine Münze geworfen wird. Fällt “Kopf”, so bekommt Spieler A einen Euro, fällt “Zahl”, so muss er einen bezahlen. In jeder Spielrunde ist also die Wahrscheinlichkeit für einen Gewinn gleich  $\frac{1}{2}$ . Sei  $X$  die Gewinnsumme eines Spielers nach 5 Spielen. Offensichtlich gilt für den Erwartungswert  $\mathbb{E}[X] = 0$ .

Nun wird der Einsatz auf 1000 Euro pro Spielrunde erhöht. Natürlich bleibt der Erwartungswert dabei gleich. Trotzdem ändert sich etwas – viele von uns würden möglicherweise beim ersten Spiel mitspielen, bei der zweiten Variante aber dankend verzichten. Was ist der Grund? Im zweiten Fall kann ein Spieler bis zu 5000 Euro verlieren (allerdings auch gewinnen). In anderen Worten, es gibt eine höhere *Streuung*.

Dieses Beispiel zeigt, dass der Erwartungswert durchaus nicht alle wichtigen Informationen enthalten muss. Um eine Zufallsvariable genauer zu beschreiben, benötigen wir ein Maß für die Streuung.

**Definition 5.11** (Varianz). Sei  $X$  eine Zufallsvariable. Die *Varianz* von  $X$  ist definiert als

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

falls dieser Erwartungswert existiert, in diesem Fall sagen wir dass die *Varianz existiert*.



**Beispiel 5.12** (Forsetzung von Beispiel 5.10). Wir berechnen nun die Varianz in den beiden Varianten des obigen Glücksspiels. *siehe Vorlesung*

**Satz 5.13.** *Sei  $X$  eine Zufallsvariable, deren Varianz existiert. Dann gilt*

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

*Beweis.* Wir formen die Definition der Varianz durch ausmultiplizieren um, und verwenden die Linearität des Erwartungswerts. Damit erhalten wir

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

□

Je nach Situation kann diese Formel einfacher anzuwenden sein als die Definition. Hier ist es wichtig, die Reihenfolge von Quadrieren und Erwartungswertbildung zu beachten! Man kann außerdem zeigen, dass die Varianz genau dann existiert, wenn  $\mathbb{E}[X^2]$  existiert.

**Definition 5.14** (Standardabweichung). Sei  $X$  eine Zufallsvariable, deren Varianz existiert. Dann heißt

$$\sigma(X) := \sqrt{\mathbb{V}(X)}$$

*Standardabweichung von  $X$ .*

Hier sind einige wichtige Eigenschaften der Varianz.

**Satz 5.15** (Eigenschaften der Varianz). *Seien  $X$  und  $Y$  zwei Zufallsvariablen deren Varianzen existieren.*

(a) *Für  $a, b \in \mathbb{R}$  gilt*

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

(b) *Falls  $X$  und  $Y$  unabhängig sind, so gilt*

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

(c) *Falls  $X$  konstant ist, so gilt  $\mathbb{V}(X) = 0$ .*

*Beweis. siehe Vorlesung*

□

Beachten Sie: Falls  $X$  und  $Y$  nicht unabhängig sind, so gilt im allgemeinen *nicht*  $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$ . Wir werden im Abschnitt über Kovarianzen sehen, was stattdessen gilt.

In der Vorlesung berechnen wir die Varianzen einiger wichtiger Verteilungen. Hier sind sie zusammengefasst in einer Tabelle mit den Erwartungswerten dargestellt.

Verteilung	Parameter	Erwartungswert	Varianz
Bernoulli	$p$	$p$	$p(1 - p)$
Binomial	$n, p$	$np$	$np(1 - p)$
Geometrisch	$p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	$\lambda$	$\lambda$	$\lambda$

**5.4. Markov- und Chebyshev-Ungleichung.** Wir haben die Varianz eingeführt, um ein Maß für die *Streuung* zu bekommen, und auch in Beispiel 5.10 gesehen, dass tatsächlich eine breite Streuung durch einen großen Wert der Varianz angezeigt wird. Ist die Varianz also groß, so kann die Zufallsvariable stark von ihrem Erwartungswert abweichen. Für diesen bisher etwas ungenau formulierten Zusammenhang formulieren geben wir nun eine konkrete Abschätzung an.

**Satz 5.16** (Chebyshev-Ungleichung). *Sei  $X$  eine Zufallsvariable deren Varianz existiert. Dann gilt für jedes  $\varepsilon > 0$*

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

In Worten ausgedrückt ist also die Wahrscheinlichkeit, dass eine Zufallsvariable um mehr als den Wert  $\varepsilon$  vom Erwartungswert abweicht (nach oben oder unten), höchstens so groß wie die Varianz dividiert durch  $\varepsilon^2$ .

*Beweis.* Wir beweisen diesen Satz für den Fall dass  $X$  eine diskrete Zufallsvariable ist, und auf einem diskreten Wahrscheinlichkeitsraum definiert ist. Er gilt aber allgemeiner, wobei der Beweis ganz ähnlich läuft, nur haben wir die Methoden dafür noch nicht zur Verfügung. In unserem Fall gilt mit Definition 5.11 und Satz 5.7

$$\begin{aligned} \mathbb{V}(X) &= \sum_{\omega \in \Omega} (X(\omega) - \mathbb{E}[X])^2 \mathbb{P}(\omega) \geq \sum_{\omega \in \Omega: |X(\omega) - \mathbb{E}[X]| > \varepsilon} (X(\omega) - \mathbb{E}[X])^2 \mathbb{P}(\omega) \\ &\geq \sum_{\omega \in \Omega: |X(\omega) - \mathbb{E}[X]| > \varepsilon} \varepsilon^2 \mathbb{P}(\omega) \geq \varepsilon^2 \mathbb{P}(|X(\omega) - \mathbb{E}[X]| > \varepsilon). \end{aligned}$$

Auflösen der obigen Ungleichung nach  $\mathbb{P}(|X(\omega) - \mathbb{E}[X]| > \varepsilon)$  liefert das Ergebnis.  $\square$

Die Markov-Ungleichung ist eine Verallgemeinerung der Chebyshev-Ungleichung. Wir beweisen sie nicht, der Beweis ist sehr ähnlich wie der obige, und kann als Übungsaufgabe gemacht werden. Ebenfalls übungshalber können Sie die Chebyshev-Ungleichung aus der Markov-Ungleichung herleiten.

**Satz 5.17** (Markov-Ungleichung). *Sei  $X$  eine Zufallsvariable, und sei  $f : \mathbb{R} \rightarrow \mathbb{R}$  eine monoton wachsende Funktion. Falls  $\mathbb{E}[f(X)]$  existiert, so gilt*

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[f(X)]}{f(a)}.$$

**Beispiel 5.18.** siehe Vorlesung

**5.5. Kovarianz und Korrelation.** Wir haben soeben zwei wichtige Kenngrößen für Zufallsvariablen kennengelernt: Der Erwartungswert gibt an, welchen Wert eine Zufallsvariable im Mittel annimmt, und die Varianz ist ein Maß dafür, wie stark die Zufallsvariable um dieses Mittel herum streut. Nun ist es aber in vielen Situationen wichtig, dass man nicht nur eine Zufallsvariable betrachtet, sondern mehrere. Wir haben in diesem Zusammenhang bereits den Begriff der Unabhängigkeit kennen gelernt. Wenn nun zwei Zufallsvariablen nicht unabhängig sind, so möchte man eine Aussage über den Grad der Abhängigkeit treffen. Dafür führen wir nun einen weiteren Begriff ein.

**Beispiel 5.19** (Geysir-Ausbruch). In der Statistik-Software R ist ein Datensatz namens “Faithful” verfügbar, welcher für einen bestimmten Geysir im Yellowstone-Nationalpark Dauer eines Ausbruchs sowie die Wartezeit bis zum jeweiligen Ausbruch angibt. Im Plot entspricht jeder Punkt einem Ausbruch, der  $x$ -Achsenwert gibt die Dauer an, der  $y$ -Achsenwert die Wartezeit seit dem letzten Ausbruch (in Minuten). Dabei springt ins Auge dass hier ein Zusammenhang besteht: Ist der  $x$ -Wert eher klein, so ist der  $y$ -Wert auch eher klein, und umgekehrt. Bild (siehe Vorlesung)

**Definition 5.20** (Kovarianz). Seien  $X$  und  $Y$  zwei Zufallsvariablen auf  $(\Omega, \mathbb{P})$ . Die Kovarianz von  $X$  und  $Y$  ist definiert als

$$\text{cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

falls der Erwartungswert auf der rechten Seite existiert.

Die Kovarianz zweier Zufallsvariablen kann Werte zwischen  $-\infty$  und  $\infty$  annehmen, also positiv oder negativ sein. Ist sie positiv, d.h.  $\text{cov}(X, Y) > 0$ , so bedeutet das nach der Definition  $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] > 0$ . Da unter dem Erwartungswert ein Produkt  $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$  steht, und Produkte genau dann positiv sind, wenn beide Faktoren das gleiche Vorzeichen haben, ist die Kovarianz also genau dann positiv, wenn *im Mittel* entweder  $(X - \mathbb{E}[X]) > 0$  und gleichzeitig  $(Y - \mathbb{E}[Y]) > 0$  gilt, oder  $(X - \mathbb{E}[X]) < 0$  und gleichzeitig  $(Y - \mathbb{E}[Y]) < 0$ , d.h. wenn im Mittel  $X$  und  $Y$  beide entweder kleiner oder größer als ihr jeweiliger Erwartungswert ist. Ist umgekehrt  $\text{cov}(X, Y) < 0$ , so bedeutet das, dass im Mittel einer der beiden Werte  $X$  bzw.  $Y$  über und der andere unter dem Erwartungswert liegt. Dies ist bereits eine nützliche Information, die in der Kovarianz steckt, und anzeigt, dass diese in der Tat in einem gewissen Sinn die Abhängigkeit von  $X$  und  $Y$  misst.

Wir unterscheiden drei Fälle:

**Definition 5.21.** Seien  $X$  und  $Y$  Zufallsvariablen, deren Kovarianz existiert.

- Ist  $\text{cov}(X, Y) > 0$ , so heißen  $X$  und  $Y$  *positiv korreliert*.
- Ist  $\text{cov}(X, Y) < 0$ , so heißen  $X$  und  $Y$  *negativ korreliert*.
- Ist  $\text{cov}(X, Y) = 0$ , so heißen  $X$  und  $Y$  *unkorreliert*.

Wie bei der Varianz gibt es eine alternative Möglichkeit, die Kovarianz zu berechnen, die man durch einfache Umformung der Definition und Ausnutzung der Eigenschaften des Erwartungswerts erhält.

**Satz 5.22.** Seien  $X$  und  $Y$  Zufallsvariablen deren Kovarianz existiert. Dann gilt

$$\text{cov}(X, Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X]\mathbb{E}[Y].$$

**Beispiel 5.23.** Wir betrachten zwei Zufallsvariablen  $X$  und  $Y$ , deren gemeinsame Verteilung von  $X$  und  $Y$  als Tabelle gegeben ist:

Gemeinsame Verteilung von  $X$  und  $Y$ 

$Y \backslash X$	0	1	2	$\Sigma$
0	1/16	1/4	1/8	7/16
1	3/16	1/4	1/8	9/16
$\Sigma$	1/4	1/2	1/4	1

Nun möchten wir die Kovarianz dieser Zufallsvariablen bestimmen. Wir verwenden Satz 5.22 um  $\text{cov}(X, Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X]\mathbb{E}[Y]$  zu schreiben, und berechnen die involvierten Erwartungswerte. Nach Definition des Erwartungswerts gilt (zur Erinnerung: Die Verteilung von  $X$  steht in der letzten Zeile)

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x) = \sum_{x=0}^2 x \mathbb{P}(X = x) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

und

$$\mathbb{E}[Y] = 0 \cdot \frac{7}{16} + 1 \cdot \frac{9}{16} = \frac{9}{16}.$$

Zur Berechnung von  $\mathbb{E}[XY]$  müssen wir erst den Wertebereich von  $XY$  bestimmen: Da  $X$  die Werte 0,1,2 und  $Y$  die Werte 0,1 annehmen kann, kann also  $X \cdot Y$  die Werte 0,1,2 annehmen. Es gilt also

$$\mathbb{E}[XY] = 0 \cdot \mathbb{P}(XY = 0) + 1 \cdot \mathbb{P}(XY = 1) + 2 \cdot \mathbb{P}(XY = 2) = \mathbb{P}(XY = 1) + 2 \cdot \mathbb{P}(XY = 2).$$

Aus der Tabelle sehen wir

$$\mathbb{P}(XY = 1) = \mathbb{P}(X = 1, Y = 1) = \frac{1}{4},$$

und

$$\mathbb{P}(XY = 2) = \mathbb{P}(X = 2, Y = 1) = \frac{1}{8}.$$

(Beachten Sie:  $\mathbb{P}(XY = 0)$  brauchen wir gar nicht zu berechnen, da diese Zahl bei der Berechnung des Erwartungswerts ohnehin mit 0 multipliziert wird, ihr Wert ist uns also egal. Als kleine Übungsaufgabe sollten Sie sie jedoch trotzdem kurz ausrechnen!)

Zusammengenommen erhalten wir also

$$\mathbb{E}[XY] = \mathbb{P}(XY = 1) + 2 \cdot \mathbb{P}(XY = 2) = \frac{1}{4} + 2 \cdot \frac{1}{8} = \frac{1}{2}.$$

Damit folgt aus Satz 5.22

$$\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{1}{2} - 1 \cdot \frac{9}{16} = \frac{8}{16} - \frac{9}{16} = -\frac{1}{16}.$$

Diese beiden Zufallsvariablen sind also negativ korreliert.

Wir geben nun einige Eigenschaften der Kovarianz an, welche relativ direkt aus der Definition hergeleitet werden können.

**Satz 5.24** (Eigenschaften der Kovarianz). *Seien  $X, Y, Z$  Zufallsvariablen auf  $(\Omega, \mathbb{P})$ , deren Kovarianz existiert, und seien  $a, b \in \mathbb{R}$ . Dann gelten*

- (a)  $\text{cov}(X, Y) = \text{cov}(Y, X)$ ,
- (b)  $\text{cov}(X, X) = \text{var}(X)$ ,
- (c)  $\text{cov}(aX + bY, Z) = a \cdot \text{cov}(X, Z) + b \cdot \text{cov}(Y, Z)$ ,
- (d) Falls  $X$  und  $Y$  unabhängig sind, so gilt  $\text{cov}(X, Y) = 0$ .

*Beweis.* siehe Vorlesung □

In Worten kann man diese Eigenschaften folgendermaßen formulieren:

- (a) Die Kovarianz ist *symmetrisch*
- (b) Die Kovarianz einer Zufallsvariablen mit sich selbst ist die Varianz
- (c) Die Kovarianz ist *bilinear*, d.h. linear in beiden Argumenten.
- (d) Unabhängigkeit impliziert Unkorreliertheit. Wichtig: Die Umkehrung dieser Aussage gilt im Allgemeinen nicht!

In der linearen Algebra haben Sie gelernt, dass eine symmetrische Bilinearform auf einem Vektorraum ein *Skalarprodukt* ist. Da die Menge aller Zufallsvariablen auf einem Wahrscheinlichkeitsraum ein Vektorraum bildet, haben wir also gemeinsam mit der Kovarianz einen Euklidischen Vektorraum. Auch wenn das vielleicht von eher theoretischem Interesse ist, hilft es uns bei der Interpretation der Bedeutung der Kovarianz, wie wir weiter unten sehen werden.

Zuerst jedoch der folgende wichtige Satz über die Varianz zweier (nicht notwendigerweise unabhängiger) Zufallsvariablen.

**Satz 5.25** (Varianz der Summe zweier Zufallsvariablen, “binomische Formel” für die Varianz). *Seien  $X$  und  $Y$  Zufallsvariablen auf  $(\Omega, \mathbb{P})$ . Dann gilt*

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + 2\text{cov}(X, Y) + \mathbb{V}(Y).$$

*Beweis.* Wir wissen aus Satz 5.13, wegen der Linearität des Erwartungswert, und unter Verwendung der ersten binomischen Formel

$$\begin{aligned} \mathbb{V}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\ &= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2) \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\ &= \mathbb{V}(X) + 2\text{cov}(X, Y) + \mathbb{V}(Y). \end{aligned}$$

□

Aus diesem Satz und aus Satz 5.22 (d) folgt nun noch einmal Satz 5.13 (b).

**Beispiel 5.26.** siehe Vorlesung

Wir hatten oben gesehen, dass das Vorzeichen der Kovarianz angibt, ob zwei Zufallsvariablen positiv korreliert, negativ korreliert oder unkorreliert sind. Der absolute Wert einer Kovarianz gibt uns jedoch kaum Information. Wir *normieren* deshalb die Kovarianz in geeigneter Art und Weise.

**Definition 5.27.** Seien  $X, Y$  zwei Zufallsvariablen mit positiver Varianz auf  $(\Omega, \mathbb{P})$ . Der *Korrelationskoeffizient* oder die *Korrelation* von  $X$  und  $Y$  ist definiert als

$$\text{corr}(X, Y) := \frac{\text{cov}(X, Y)}{\sqrt{\mathbb{V}(X) \cdot \mathbb{V}(Y)}}$$

Notation: Wir schreiben manchmal auch  $\varrho(X, Y) := \text{corr}(X, Y)$ .

Warum ist diese Normierung mit der Wurzel aus dem Produkt der Varianzen die richtige? Man kann folgenden Satz beweisen.

**Satz 5.28** (Eigenschaften der Korrelation). *Seien  $X, Y$  zwei Zufallsvariablen mit positiver Varianz auf  $(\Omega, \mathbb{P})$ . Dann gelten*

- (a) *Das Vorzeichen der Korrelation stimmt mit dem Vorzeichen der Kovarianz überein.*
- (b) *Die Korrelation nimmt nur Werte in  $[-1, 1]$  an, d.h.  $-1 \leq \text{corr}(X, Y) \leq 1$ .*

*Beweis.* Teil (a) ist klar. Um Teil (b) zu beweisen, ist die obige Bemerkung hilfreich, dass die Kovarianz die Eigenschaften eines Skalarproduktes hat. Dann folgt aus der Cauchy-Schwarz-Ungleichung, welche aus der linearen Algebra bekannt sein sollte, dass  $\text{cov}(X, Y)^2 \leq \mathbb{V}(X)\mathbb{V}(Y)$  gilt. Daraus folgt durch Umformen (b).  $\square$

*Bemerkung* (Linearer Zusammenhang). **siehe Vorlesung**

Bis jetzt haben wir uns nur mit zwei Zufallsvariablen beschäftigt. Wenn man mehr als zwei Zufallsvariablen untersuchen möchte, schaut man sich die sogenannte *Kovarianzmatrix* an, welche folgendermaßen definiert ist:

**Definition 5.29.** Seien  $X_1, \dots, X_n$  Zufallsvariablen auf  $(\Omega, \mathbb{P})$ . Die Kovarianzmatrix von  $X_1, \dots, X_n$  ist

$$\begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \dots & \text{cov}(X_2, X_n) \\ \vdots & \ddots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \dots & \text{cov}(X_n, X_n) \end{bmatrix}$$

Beachten Sie dass nach Satz 5.22 (b) die Einträge auf der Diagonalen gleich den Varianzen sind, und dass die Matrix symmetrisch ist nach Satz 5.22 (a). Man kann im Übrigen zeigen, dass sie *positiv semidefinit* ist (siehe lineare Algebra).

Analog zur Kovarianzmatrix kann die Korrelationsmatrix aufgestellt werden.

## 6. ZUFALLSVARIABLEN MIT DICHTEN

Bisher haben wir uns (fast) nur mit diskreten Zufallsvariablen beschäftigt, d.h. mit Zufallsvariablen, deren Bild eine endliche oder abzählbar unendliche Menge ist. Wir haben gesehen, dass für solche Zufallsvariablen  $X$  die *Verteilung*  $p_X(x), x \in X(\Omega)$  im Prinzip alle relevanten Informationen enthält. Wir können außerdem Erwartungswert und Varianz von diskreten Zufallsvariablen berechnen.

Nicht alle in der Wirklichkeit auftretenden Zufallsvariablen sind jedoch diskret. Beispielsweise kann die Wartezeit auf das Ergebnis einer Datenbankabfrage im Prinzip jeden Wert zwischen 0 und  $\infty$  annehmen, also Werte im (überabzählbaren) Intervall  $]0, \infty[$ . Auch die Dauer eines Laserimpulses kann nicht nur bestimmte diskrete Werte annehmen. In diesem Kapitel führen wir deshalb zuerst einige Begriffe für allgemeine Zufallsvariablen ein, und lernen dann einige besonders wichtige Verteilungen kennen, die mit Hilfe einer *Dichte* beschrieben werden können.

Die **Lernziele** dieses Kapitel sind:

- Die Zusammenhänge zwischen Zufallsvariablen und Verteilungsfunktion kennen
- Dichten und ihre Eigenschaften kennen, und mit Dichten rechnen können
- Normalverteilung und Exponentialverteilung, ihr Auftreten, ihre Dichten und ihre Kenngrößen kennen.

Vorkenntnisse: Wir werden bald sehen, dass das Rechnen mit Dichten auf der Integralrechnung beruht. Um dieses Kapitel zu verstehen, sind deshalb solide Kenntnisse der Integral- und Differentialrechnung notwendig, einschließlich der Integrationstechniken durch Substitution und partielle Integration.

**6.1. Allgemeine Zufallsvariablen, Verteilungsfunktion, Dichten.** Zur Erinnerung: In Definition 3.8 wurde die (*kumulative*) *Verteilungsfunktion einer (diskreten) Zufallsvariablen*  $X$  ist definiert durch

$$F_X(x) = \mathbb{P}(X \leq x).$$

Aus dieser Definition kann man sich relativ leicht überzeugen, dass die Verteilungsfunktion einer Zufallsvariablen folgende Eigenschaften hat:

- (i) Die Verteilungsfunktion  $F : \mathbb{R} \rightarrow [0, 1]$  ist monoton nichtfallend, d.h.  $x \geq y$  impliziert  $F_X(x) \geq F_X(y)$
- (ii) Es gilt  $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- (iii) Ebenso gilt  $\lim_{x \rightarrow \infty} F_X(x) = 1$ .

In Definition 3.8 sind wir mit einer Zufallsvariablen gestartet, und haben ausgehend davon die zur Zufallsvariable gehörige Verteilungsfunktion definiert. Man kann nun die Situation auch umgekehrt angehen, und – erst einmal ohne über Zufallsvariablen zu sprechen – sagen, was eine Verteilungsfunktion sein soll. Dabei soll eine Verteilungsfunktion genau die obigen drei Eigenschaften haben. Also:

**Definition 6.1** (Verteilungsfunktion). Eine *Verteilungsfunktion* ist eine Funktion  $F : \mathbb{R} \rightarrow [0, 1]$  mit diesen drei Eigenschaften:

- (i)  $F : \mathbb{R} \rightarrow [0, 1]$  ist monoton nichtfallend
- (ii)  $\lim_{x \rightarrow -\infty} F(x) = 0$
- (iii)  $\lim_{x \rightarrow \infty} F(x) = 1$ .

Es kann bewiesen werden (was jedoch den Rahmen dieser Vorlesung sprengt), dass zu jeder solchen Funktion auch eine passende Zufallsvariable existiert.

**Satz 6.2.** *Zu jeder Verteilungsfunktion  $F$  existiert eine Zufallsvariable  $X$ , so dass  $F = F_X$  ist, d.h. jede Verteilungsfunktion ist die Verteilungsfunktion einer Zufallsvariablen.*

**Beispiel 6.3.** Um zu entscheiden, ob eine vorgegebene Funktion eine Verteilungsfunktion ist, müssen wir also diese drei Eigenschaften überprüfen. In der Vorlesung sehen wir dazu Beispiele.

Obiger Satz ist vielleicht eher von theoretischer Bedeutung. Der folgende Begriff ist jedoch insbesondere für konkrete Berechnungen wichtig.

**Definition 6.4** (Dichte). Sei  $X$  eine Zufallsvariable und  $F_X$  ihre Verteilungsfunktion.  $X$  bzw.  $F_X$  besitzt eine *Dichte*, falls eine Funktion  $f_X$  existiert, so dass für alle  $x \in \mathbb{R}$  gilt

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

Die Funktion  $f_X$  heißt dann *Dichte* von  $F_X$ , beziehungsweise von  $X$ .

**Satz 6.5** (Eigenschaften einer Dichte). *Eine Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  ist genau dann eine Dichte einer Zufallsvariablen, wenn folgende Eigenschaften erfüllt sind:*

- (i)  $f(t) \geq 0$  für alle  $t \in \mathbb{R}$ ,
- (ii)  $\int_{-\infty}^{\infty} f(t) dt = 1$ .

*Beweis.* Wenn  $f$  eine Dichte ist, so muss nach Definition die Funktion  $x \mapsto \int_{-\infty}^x f(t) dt$  die drei Eigenschaften aus Definition 6.1 erfüllen. Wäre  $f(t) \geq 0$  nicht erfüllt, so wäre  $x \mapsto \int_{-\infty}^x f(t) dt$  nicht monoton nichtfallend. Wäre  $\int_{-\infty}^{\infty} f(t) dt \neq 1$ , so wäre (iii) aus Definition 6.1 nicht erfüllt. Somit muss eine Dichte diese beiden Eigenschaften haben.

Hat umgekehrt  $f$  diese beiden Eigenschaften, so erfüllt  $x \mapsto \int_{-\infty}^x f(t) dt$  die drei Eigenschaften aus Definition 6.1, und ist somit eine Verteilungsfunktion.  $\square$

**Beispiel 6.6.** Seien  $a < b$ ,  $a, b \in \mathbb{R}$ . Die Funktion  $F : \mathbb{R} \rightarrow \mathbb{R}$ ,

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

erfüllt die Eigenschaften einer Verteilungsfunktion, wie man an einer Skizze leicht erkennen kann (siehe Vorlesung). Eine Zufallsvariable  $X$  mit dieser Verteilungsfunktion heißt *uniform (oder stetig gleichverteilt)* auf  $[a, b]$ .

Hat eine Zufallsvariable mit dieser Verteilung eine Dichte? In anderen Worten, existiert eine Funktion  $f$  mit  $F(x) = \int_{-\infty}^x f(t) dt$ ? Der Hauptsatz der Differential- und Integralrechnung sagt uns, dass, falls eine Dichte existiert, sie die *Ableitung* von  $F$  sein muss.



Wir müssen also überprüfen, ob  $F$  differenzierbar ist. Nach Definition ist  $F$  auf jedem Fall auf den Intervallen  $] - \infty, a[$  und  $]b, \infty[$  differenzierbar, die Ableitung ist dort 0 weil die Funktion konstant ist. Auf dem (offenen) Intervall  $]a, b[$  ist die Funktion als lineare Funktion ebenfalls differenzierbar, die Ableitung ist die konstante Funktion  $\frac{1}{b-a}$  (Übungsaufgabe!). Diese Funktion hat rechtsseitige Grenzwerte für  $x \searrow a$  und  $x \searrow b$ , so dass die Funktion

$$f(x) := \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x < b \\ 0 & x \geq b \end{cases}$$

wohldefiniert ist, und  $f(t) = F'(t)$  für  $x \in \mathbb{R} \setminus \{a\} \cup \{b\}$  gilt. Somit ist also  $f$  die *Dichte* einer auf  $[a, b]$  uniform verteilten Zufallsvariablen.

**Beispiel 6.7.** (Siehe Vorlesung)

**Definition 6.8.** Sei  $X$  eine Zufallsvariable mit Dichte  $f_X$ . Der *Erwartungswert* von  $X$  ist definiert als

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} t \cdot f_X(t) dt,$$

falls das Integral existiert. Ist  $g : \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion, so ist  $g(X)$  wieder eine Zufallsvariable, und wir definieren

$$\mathbb{E}[g(X)] := \int_{-\infty}^{\infty} g(t) \cdot f_X(t) dt.$$

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

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

falls dieser Erwartungswert existiert.

*Bemerkung.* Der Ausdruck für  $\mathbb{E}[g(X)]$  ist streng genommen keine Definition, sondern kann aus den Eigenschaften der Dichte und des Erwartungswerts hergeleitet werden. Wir verzichten jedoch darauf, und fassen es der Einfachheit halber einfach als Definition auf. Beachten Sie, dass die Definition der Varianz genau gleich lautet wie im Fall von diskreten Zufallsvariablen, vgl. Definition 5.11.

Die Eigenschaften von Erwartungswert und Varianz aus Satz 5.8 und Satz 5.15 gelten auch im Fall von Zufallsvariablen mit Dichten, ebenso wie die Chebyshev und die Markov-Ungleichung. Wir verzichten darauf, die Aussagen und Beweise zu wiederholen.

**Satz 6.9.** Sei  $X$  eine Zufallsvariable mit Dichte  $f_X$ . Für die Varianz gilt

$$\mathbb{V}(X) = \int_{-\infty}^{\infty} (t - \mathbb{E}[X])^2 f_X(t) dt = \int_{-\infty}^{\infty} t^2 \cdot f_X(t) dt - \left( \int_{-\infty}^{\infty} t \cdot f_X(t) dt \right)^2.$$

*Beweis.* Die erste Gleichung folgt aus Definition 6.8 mit  $g(x) = (x - \mathbb{E}[X])^2$ . Die zweite Gleichung folgt durch Ausmultiplizieren, vgl. Satz 5.13.  $\square$

**Beispiel 6.10** (Erwartungswert der uniformen Verteilung). *siehe Vorlesung*

**Satz 6.11** (Rechnen mit Dichten). *Sei  $X$  eine Zufallsvariable mit Dichte  $f_X$ . Seien  $a, b \in \mathbb{R}$  mit  $a < b$ .*

- (i)  $\mathbb{P}(X \leq b) = \mathbb{P}(X < b) = \int_{-\infty}^b f_X(t) dt$
- (ii)  $\mathbb{P}(X \geq a) = \mathbb{P}(X > a) = \int_a^{\infty} f_X(t) dt$
- (iii)  $\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in ]a, b[) = \mathbb{P}(X \in ]a, b]) = \mathbb{P}(X \in [a, b[) = \int_a^b f_X(t) dt$

*Beweis.* (i) Aus der Definition der Verteilungsfunktion, Def. 3.8, und der Definition der Dichte, Def. 6.4 folgt

$$\mathbb{P}(X \leq b) = F_X(b) = \int_{-\infty}^b f_X(t) dt.$$

Da  $x \rightarrow \int_{-\infty}^x f_X(t) dt$  nach dem Hauptsatz der Differential- und Integralrechnung stetig ist, folgt  $\mathbb{P}(X < b) = \mathbb{P}(X \leq b)$ .

(ii) folgt aus (i), da  $\mathbb{P}(X \geq a) = 1 - \mathbb{P}(X < a)$  und  $\mathbb{P}(X > a) = 1 - \mathbb{P}(X \geq a)$  gilt.

(iii) erhalten wir aus (i) und (ii) mit

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a) = \int_{-\infty}^b f_X(t) dt - \int_{-\infty}^a f_X(t) dt.$$

Der letzte Ausdruck auf der rechten Seite kann nun nach den Regeln der Integralrechnung zu  $\int_a^b f_X(t) dt$  zusammengefasst werden. Analog folgen die anderen Gleichungen.  $\square$

*Bemerkung.* Aus dem obigen Satz kann man folgendes schließen: Für eine Zufallsvariable mit Dichte ist die Wahrscheinlichkeit, genau einen bestimmten Wert  $x$  anzunehmen, gleich 0. Es gilt nämlich

$$\mathbb{P}(X = x) = \mathbb{P}(X \in [x, x]) = \int_x^x f(t) dt = 0.$$

Das mag auf den ersten Blick verwirrend sein, man sollte sich aber noch einmal vor Augen führen, wie das Integral konstruiert wird, dann leuchtet es eher ein. Man kann es sich auch so vorstellen: Angenommen, die Zufallsvariable  $X$  bezeichnet die Dauer eines Laser-Impulses. Die Wahrscheinlichkeit, dass der Impuls eine ganz exakt bemessene Zeit dauert, also z.B. 2/3 Sekunden, ist verschwindend klein. Natürlich dauert der Impuls aber dennoch jedes Mal eine gewisse Zeit, nur kann man diese nicht so exakt bemessen. Die Wahrscheinlichkeit, dass der Impuls eine Dauer zwischen z.B. 0.66 und 0.67 Sekunden hat, ist nun vermutlich nicht mehr 0, weil wir es nun nicht mehr mit einem ganz exakten Wert, also einem infinitesimal kleinen Intervall, zu tun haben, sondern mit einer *Spanne* von Werten.

**Beispiel 6.12.** siehe Vorlesung

## 6.2. Exponentialverteilung.

**Definition 6.13** (Exponentialverteilung). Eine Zufallsvariable  $X$  heißt *exponentialverteilt* mit Parameter  $\lambda > 0$ , falls  $X$  die Dichte

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

hat.

Damit diese Definition überhaupt sinnvoll ist, müssen wir erst einmal überprüfen, dass  $f_X$  überhaupt eine Dichte ist. Da  $\lambda > 0$  und die  $e$ -Funktion nur positive Werte annimmt, gilt auf jeden Fall schon  $f_X(t) \geq 0$  für alle  $t \in \mathbb{R}$ . Außerdem haben wir

$$\int_{-\infty}^{\infty} f_X(t) dt = \int_0^{\infty} \lambda e^{-\lambda t} dt = \left[ -e^{-\lambda} \right]_0^{\infty} = -0 + 1 = 1.$$

Somit ist  $f_X$  eine Dichte.

**Satz 6.14** (Exponentialverteilung).  *$X$  exponentialverteilt mit Parameter  $\lambda > 0$ . Dann ist die Verteilungsfunktion  $F_X$  gegeben durch*

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

und es gelten

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \text{sowie} \quad \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

*Beweis.* Nach Definition gilt  $F_X(x) = \int_{-\infty}^x f_X(t) dt$ , somit ist klar dass  $F_X(x) = 0$  für  $x < 0$  gelten muss, und für  $x \geq 0$  erhalten wir wie oben

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \int_0^x \lambda e^{-\lambda t} dt = \left[ -e^{-\lambda} \right]_0^x = -e^{-\lambda x} + 1.$$

Weiter gilt nach Definition 6.8

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t f_X(t) dt = \int_0^{\infty} t \lambda e^{-\lambda t} dt.$$

Dieses Integral kann mittels partieller Integration berechnet werden, und man findet (mit  $u'(t) = \lambda e^{-\lambda t}$ ,  $v(t) = 1$ , also  $u(t) = -e^{-\lambda t}$ ,  $v'(t) = 1$ ),

$$\int_0^{\infty} t \lambda e^{-\lambda t} dt = \left[ -t e^{-\lambda t} \right]_0^{\infty} - \int_0^{\infty} -e^{-\lambda t} dt = \left[ \frac{-1}{\lambda} e^{-\lambda t} \right]_0^{\infty} = \frac{1}{\lambda}.$$

Für die Varianz wird Satz 6.9 benutzt und zweimal partiell integriert (Übung).  $\square$

Die Exponentialverteilung ist in vielerlei Hinsicht das stetige Analog zur geometrischen Verteilung. Beispielsweise gilt folgende Aussage: Ist  $(X_n)_{n \in \mathbb{N}}$  eine Folge von unabhängigen Zufallsvariablen, so dass  $X_n$  geometrisch verteilt mit Parameter  $p_n$  ist, wobei  $\lim_{n \rightarrow \infty} n p_n = \lambda$  gilt. Dann gilt für alle  $x > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_n}{n} \leq x\right) = 1 - e^{-\lambda x},$$

d.h. die Verteilung der “reskalierten” geometrischen Zufallsvariablen  $X_n/n$  konvergiert gegen eine Exponentialverteilung, sofern die Parameter geeignet konvergieren. Dies ist ein Beispiel für einen Grenzwertsatz, ähnlich wie der Poisson-Grenzwertsatz 3.27. Wir beweisen dieses Ergebnis hier nicht, obwohl es nicht allzu schwer ist, wenn Sie im Umgang mit Limiten und der Exponentialfunktion geübt sind, können Sie sich den Beweis selbst

erarbeiten.

Eine weitere Analogie zur geometrischen Verteilung ist die sogenannte *Gedächtnislosigkeit*, siehe dazu die Übungsaufgaben.

Wie die geometrische Verteilung tritt auch die Exponentialverteilung oft in Situationen auf, in denen *Wartezeiten* eine Rolle spielen. Typischerweise sind z.B. folgende Größen exponentialverteilt:

- Wartezeit auf einen radioaktiven Zerfall
- Wartezeiten zwischen zwei Telefonanrufen/E-Mail-Anfragen
- Zeit zwischen zwei Unfällen auf einer bestimmten Straße
- Restlebenszeit eines Bauteils mit konstanter Ausfallrate
- Zeit zwischen zwei Mutationen auf der DNS

### 6.3. Normalverteilung.

**Definition 6.15.** Eine Zufallsvariable  $X$  heißt *normalverteilt* bzw. *Gauß-verteilt* zu den Parametern  $\mu \in \mathbb{R}, \sigma^2 > 0$ , falls  $X$  die Dichte

$$f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

hat. Falls  $\mu = 0$  und  $\sigma^2 = 1$  gelten, so nennt man  $X$  *standardnormalverteilt* ist.

Auch hier muss wiederum überprüft werden, dass es sich bei der angegebenen Funktion tatsächlich um eine Dichte handelt. Es ist nicht schwer zu sehen, dass die Funktion nicht-negative Werte annimmt. Um zu überprüfen, dass  $\int_{-\infty}^{\infty} f_X(t)dt = 1$  gilt, ist ein kleiner Trick sowie Integration durch Substitution notwendig. Wir verzichten hier erst einmal auf die Details.

**Satz 6.16** (Normalverteilung). *Sei  $X$  eine normalverteilte Zufallsvariable mit Parametern  $\mu \in \mathbb{R}$  und  $\sigma^2 > 0$ . Dann gelten*

$$\mathbb{E}[X] = \mu \quad \text{und} \quad \mathbb{V}(X) = \sigma^2.$$

*Beweis.* Siehe Vorlesung. □

Im Falle der Normalverteilung kann man keine geschlossene Form, d.h. keine explizite Formel, für die Verteilungsfunktion angeben. Die Werte der Verteilungsfunktion der *Standardnormalverteilung* sind jedoch in Tabellen aufgelistet, deren Verwendung etwas Übung bedarf (siehe Vorlesung/Übungen). Man kann sich die Werte auch von gängigen Statistik-Programmen (z.B. R) ausgeben lassen.

*Notation:* Für die Dichte der Normalverteilung verwendet man oft die Notation  $\varphi_{\mu,\sigma}$  oder  $\varphi_{\mu,\sigma^2}$ , für die Verteilungsfunktion  $\Phi_{\mu,\sigma^2}$ .

Ist  $X$  eine normalverteilte Zufallsvariable, so schreibt man auch  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

**Satz 6.17** (Standardisierung). *Sei  $X \sim \mathcal{N}(\mu, \sigma^2)$ . Dann ist*

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

D.h. ist eine normalverteilte Zufallsvariable  $X$  mit beliebigem (aber bekanntem) Erwartungswert und Varianz kann durch Subtraktion von  $\mu$  und Division durch  $\sigma$  (nicht  $\sigma^2$ !) auf eine *standardnormalverteilte* Zufallsvariable transformiert werden. Damit kann man aus der Standardnormalverteilung auch Werte für beliebige Normalverteilungen berechnen.

*Beweis.* siehe Vorlesung □

Da das Rechnen mit der Normalverteilung aufgrund der fehlenden expliziten Formel für die Verteilungsfunktion etwas umständlich ist, machen wir hier zwei ausführliche Rechenbeispiele.

**Beispiel 6.18.** Sei  $X \sim \mathcal{N}(-1.3, 4)$ . Aufgabe: Berechne  $\mathbb{P}(X > 0)$ .

Lösung: Es gilt also  $\mu = -1.3, \sigma^2 = 4$ . Wir formen die gesuchte Wahrscheinlichkeit erst so um, dass die Verteilungsfunktion auftritt:

$$\mathbb{P}(X > 0) = 1 - \mathbb{P}(X \leq 0) = 1 - \Phi_{-1.3,4}(0).$$

Nun können wir aber diesen Wert nicht berechnen. Deshalb verwenden wir Satz 6.17. Dazu subtrahieren wir im Ausdruck  $\mathbb{P}(X \leq 0)$  auf beiden Seiten den Erwartungswert und dividieren durch die Standardabweichung. Somit erhalten wir

$$\begin{aligned} \mathbb{P}(X > 0) &= 1 - \mathbb{P}(X \leq 0) \\ &= 1 - \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq \frac{0 - \mu}{\sigma}\right) \\ &= 1 - \mathbb{P}\left(\frac{X - (-1.3)}{2} \leq \frac{0 - (-1.3)}{2}\right). \end{aligned}$$

Sei nun  $Y$  eine standardnormalverteilte Zufallsvariable. Nach Satz 6.17 gilt

$$1 - \mathbb{P}\left(\frac{X - (-1.3)}{2} \leq \frac{0 - (-1.3)}{2}\right) = 1 - \mathbb{P}\left(Y \leq \frac{1.3}{2}\right) = 1 - \Phi_{0,1}(0.65),$$

was laut der Tabelle ungefähr gleich  $1 - 0.7422 = 0.2578$  ist.

**Beispiel 6.19.** Sei  $X \sim \mathcal{N}(-1.3, 4)$ . Aufgabe: Berechne  $\mathbb{P}(X > -2)$ .

Wir gehen vor wie im vorigen Beispiel, und erhalten

$$\begin{aligned} \mathbb{P}(X > -2) &= 1 - \mathbb{P}(X \leq -2) \\ &= 1 - \mathbb{P}\left(\frac{X - (-1.3)}{2} \leq \frac{-2 - (-1.3)}{2}\right) \\ &= 1 - \Phi_{0,1}(-0.35) \end{aligned}$$

Unser Problem ist nun: -0.35 steht nicht in der Tabelle. Was tun?

Die Lösung lautet hier: Symmetrie! Schaut man sich die Skizze der Normalverteilungsdichte an, so sieht man schnell: Für  $Y \sim \mathcal{N}(0, 1)$  gilt

$$\mathbb{P}(Y \leq -a) = \mathbb{P}(Y > a).$$

Dies kann man auch leicht formal aus der Formel für die Dichte beweisen, denn diese ist eine achsensymmetrische (gerade) Funktion.

Also erhalten wir

$$1 - \Phi_{0,1}(-0.35) = \Phi_{0,1}(0.35) \approx 0.6368.$$

Vorsicht: Die Symmetrie  $\mathbb{P}(Y \leq -a) = \mathbb{P}(Y > a)$  gilt nur für  $\mu = 0$ . Andernfalls ist die Spiegelachse die Gerade mit  $x = \mu$ .

Die Normalverteilung spielt eine enorm wichtige Rolle. Sie tritt nämlich als *universelle* Verteilung immer dann auf, wenn viele unabhängige, gleichartige Zufallsexperimente ausgeführt werden (und keine oder nur wenige ganz extreme Schwankungen vorkommen). Deshalb sind in Natur und Technik viele Größen (annähernd) normalverteilt, beispielsweise das Gewicht von Lebewesen, oder die Anzahl Reisender auf einer bestimmten Strecke, die Anzahl Nutzer eines bestimmten Online-Dienstes während einer festen Zeitspanne, usw.

Diese Tatsache werden wir im Kapitel [], genauer im *zentralen Grenzwertsatz* genauer formulieren, und im Statistik-Teil dieser Vorlesung immer wieder ausnutzen.

**6.4. Weitere Beispiele.** Pareto-Verteilung, Gamma-Verteilung,  $\chi^2$ -Verteilung,  $t$ -Verteilung, siehe Vorlesung

## 7. MARKOV-KETTEN

Die **Lernziele** dieses Kapitel sind:

- Den Begriff der Markov-Kette und den Zusammenhang zu stochastischen Matrizen kennen
- Mit stochastischen Matrizen rechnen können
- Invariante Verteilungen berechnen können, Kriterien für Existenz und Eindeutigkeit kennen, ihre Bedeutung für Markov-Ketten kennen
- Anwendungen von Markov-Ketten kennen.

In diesem Kapitel spielen insbesondere bedingte Wahrscheinlichkeiten eine wichtige Rolle. Außerdem ist der souveräne Umgang mit Matrizen und linearen Gleichungssystemen wichtig, sowie mit Folgen und Grenzwerten.

### 7.1. Markov-Ketten und stochastische Matrizen.

**Beispiel 7.1** (Erkunden eines Netzwerks). Wir betrachten ein Netzwerk, welches aus Knoten und Kanten besteht. Eine Kanten verbindet jeweils zwei Knoten miteinander, jedoch hat nicht jeder Knoten gleich viele Verbindungen.

**Bild siehe Vorlesung**

Wir möchten die Struktur des Netzwerks untersuchen, haben aber nicht die Möglichkeit, das ganze Netzwerk zu betrachten (z.B. weil wir uns selbst innerhalb des Netzwerks befinden). Wir können uns jedoch innerhalb des Netzwerks bewegen, und zwar können wir von Knoten zu Knoten springen, falls dazwischen eine Kante liegt. Folgendes soll gelten:

- Von einem Knoten innerhalb des Netzwerks sieht man nur seine direkten Nachbarn
- Von einem Knoten aus springt man zu einem zufällig ausgewählten Nachbarknoten, und zwar wählt man jeden Nachbarknoten mit gleicher Wahrscheinlichkeit.

Auch wenn dieses Beispiel sehr simpel ist, so bildet es die Grundlage des Google-Algorithmus “PageRank”. Dabei stellen die Knoten einzelne Websites dar, und die Kanten die Verbindungen dazwischen, also die Links. Das beschriebene Vorgehen zum Erkunden des Netzwerks entspricht dem sogenannten *Zufallssurfermodell*. Wir können uns beispielsweise folgende Fragen stellen:

- Wie lange dauert es, bis man jeden Knoten einmal besucht hat?
- Wie lange dauert es, bis man wieder beim Ausgangspunkt ist?
- Wie oft besucht man (auf lange Sicht) einen bestimmten Knoten/einen durchschnittlichen Knoten? Besucht man alle gleich oft?
- Welche Rolle spielt die Wahl des Startpunktes in den obigen Fragen?
- Was kann ich aus solchen Informationen über die Struktur des Netzwerks sagen?

Um dieses Beispiel mathematisch zu behandeln, wollen wir es zuerst noch etwas formalisieren:

- Ein Netzwerk wird beschrieben als ein *Graph* bestehend aus Knoten und Kanten, die Knoten sind nummeriert als  $\{1, 2, \dots, k\}$ .

- Start in (zufällig ausgewähltem) Knoten, mit  $X_0$  bezeichnen wir die Zufallsvariable, welche den Startknoten angibt, d.h.  $X_0$  nimmt Werte in der Knotenmenge  $\{1, \dots, k\}$  an.
- Wähle zufällig (gleichverteilt) einen Nachbarknoten aus (d.h. ein Knoten der mit  $X_0$  durch eine Kante verbunden ist), und springe dorthin: Knoten  $X_1$ .
- Iterativ:  $X_n$  die Nummer des Knoten, die man im  $n$ -ten Schritt besucht.

Dieses Vorgehen beschreibt einen *stochastischen Algorithmus*, dessen Ausgabe eine Folge von Zufallsvariablen  $X_0, X_1, \dots$  ist, welche jeweils Werte in  $\{1, \dots, k\}$  annehmen. Diese Folge hat eine ganz bestimmte Eigenschaft, die wir herausheben wollen, da sie in diesem Kapitel entscheidend ist:

Der Wert von  $X_n$  hängt vom Wert von  $X_{n-1}$  ab, aber nicht von  $X_{n-2}, X_{n-3}, \dots$ , denn welchen Knoten ich im nächsten Schritt besuche, wird dadurch entschieden, wo ich gerade bin, jedoch nicht, welche Knoten ich davor schon besucht habe. Formal kann man diese Eigenschaft so aufschreiben: Für jede Wahl von Werten  $a_0, \dots, a_n \in \{1, \dots, k\}$  gilt

$$\mathbb{P}(X_n = a_n | X_0 = a_0, \dots, X_{n-1} = a_{n-1}) = \mathbb{P}(X_n = a_n | X_{n-1} = a_{n-1}).$$

Außerdem wird der nächste Knoten immer nach der gleichen zufälligen Regel ausgewählt, egal wie viele Sprünge man schon gemacht hat, es gilt also auch für alle Werte  $a, b \in \{1, \dots, k\}$  und für alle  $n \in \mathbb{N}$

$$\mathbb{P}(X_n = b | X_{n-1} = a) = \mathbb{P}(X_1 = b | X_0 = a).$$

**Definition 7.2** (Markov-Kette). Sei  $S$  eine (höchstens abzählbare) Menge. Eine (*homogene*) *Markov-Kette* auf  $S$  ist eine Folge von Zufallsvariablen  $X_0, X_1, X_2, \dots$  mit Werten in  $S$ , so dass für alle  $n \in \mathbb{N}_0$  und für alle  $a_0, a_1, \dots, a_n \in S$  gilt:

$$\begin{aligned} \mathbb{P}(X_n = a_n | X_0 = a_0, \dots, X_{n-1} = a_{n-1}) &= \mathbb{P}(X_n = a_n | X_{n-1} = a_{n-1}) \\ &= \mathbb{P}(X_1 = a_n | X_0 = a_{n-1}). \end{aligned}$$

Die Menge  $S$  heißt *Zustandsraum* der Markov-Kette, ein Element  $a \in S$  heißt *Zustand*.

**Beispiel 7.3** (Irrfahrten). Die Markovkette aus Beispiel 7.1 ist ein Beispiel für eine sogenannte *Irrfahrt*. Auf einem *Graphen* bestehend aus Knoten und Kanten definiert man eine Irrfahrt genau so wie oben, indem man über die Kanten von Knoten zu Knoten springt. Irrfahrten können auch auf Graphen mit unendlich großer Knotenmenge definiert werden: Beispielsweise kann man sich die ganzen Zahlen  $\mathbb{Z}$  als einen solchen Graphen vorstellen, wobei die Knoten gegeben sind durch die einzelnen Zahlen, und die Kanten sich jeweils zwischen benachbarten Zahlen befinden. Startet man also in  $k \in \mathbb{Z}$ , z.B.  $k = -217$ , so springt man mit jeweils gleicher Wahrscheinlichkeit zu einem Nachbarknoten, in diesem Fall sind das  $-216$  und  $-218$ . Formal ausgedrückt:

$$\mathbb{P}(X_n = k-1 | X_{n-1} = k) = \mathbb{P}(X_n = k+1 | X_{n-1} = k) = \frac{1}{2}, \quad \text{für alle } k \in \mathbb{Z}, \text{ für alle } n \in \mathbb{N}.$$



Da man mit gleicher Wahrscheinlichkeit nach links bzw. rechts springt, nennt man das auch die *symmetrische Irrfahrt* auf  $\mathbb{Z}$ . Die *asymmetrische Irrfahrt* springt mit einer Wahrscheinlichkeit  $p \in [0, 1]$  nach links, und mit der Gegenwahrscheinlichkeit nach rechts, also

$$\mathbb{P}(X_n = k - 1 \mid X_{n-1} = k) = p, \quad \mathbb{P}(X_n = k + 1 \mid X_{n-1} = k) = 1 - p, \quad k \in \mathbb{Z}, n \in \mathbb{N}.$$

Auf dem zweidimensionalen Gitter  $\mathbb{Z}^2$  hat man von jedem Knoten aus 4 Möglichkeiten, somit springt die symmetrische Irrfahrt in diesem Fall mit Wahrscheinlichkeit  $\frac{1}{4}$  zu jedem der 4 Nachbarknoten. Analog in höheren Dimensionen.

**Definition 7.4.** Sei  $X_0, X_1, \dots$  eine Markov-Kette auf einem Zustandsraum  $S$ . Seien  $a, b \in S$ . Dann heißt

$$p_{a,b} := \mathbb{P}(X_n = b \mid X_{n-1} = a)$$

*Übergangswahrscheinlichkeit* von  $a$  nach  $b$ .

*Notation:* Wir verwenden äquivalent die verschiedenen Schreibweisen

$$p_{a,b} = p_{ab} = p(a, b)$$

für die Übergangswahrscheinlichkeiten.

**Beispiel 7.5** (Irrfahrt). Wir betrachten eine Irrfahrt auf einem einfachen Graphen. Die Übergangswahrscheinlichkeiten können dann durch Zählen der Nachbarknoten abgelesen werden.

Bild siehe Vorlesung

**Beispiel 7.6** (Übergangsgraph). Markov-Ketten und ihre Übergangswahrscheinlichkeiten können gut an Hand ihrer *Übergangsgraphen* dargestellt werden, indem man mögliche Übergänge mit Pfeilen zwischen den Zuständen markiert, und die entsprechenden Wahrscheinlichkeiten über die Pfeile schreibt.

Bild siehe Vorlesung

**Beispiel 7.7** (Symmetrische Irrfahrt auf  $\mathbb{Z}^d$ ). Betrachte die symmetrische Irrfahrt auf  $\mathbb{Z}^d$ , siehe Beispiel 7.3. Dann gelten für die Übergangswahrscheinlichkeiten

$$p_{xy} = \begin{cases} \frac{1}{2d} & \text{falls } x \text{ und } y \text{ Nachbarn sind} \\ 0 & \text{sonst.} \end{cases}$$

**Beispiel 7.8** (Verzweigungsprozesse). Bei einem Verzweigungsprozess wird sukzessive ein zufälliger “Baum” erzeugt, bestehend aus einer “Wurzel”, von der aus eine Anzahl Äste wegführt, die sich wiederum verzweigen.

Bild siehe Vorlesung

Dabei gibt es verschiedene Ebenen von Knoten, die 0-te Ebene bildet die Wurzel, die erste Ebene sind diejenigen Knoten, welche eine Verbindung zur Wurzel haben, usw. Die Anzahl Knoten auf jeder Ebene ist zufällig (mit Ausnahme von Ebene 0). Dies können wir folgendermaßen durch eine Markov-Kette beschreiben: Wir setzen  $X_0 = 1$ , das bezeichnet

die einzige Wurzel. Mit  $X_n$  bezeichnen wir die Anzahl Knoten auf Ebene  $n \in \mathbb{N}_0$ . Diese Zufallsvariable wird folgendermaßen berechnet: Sei  $(q_i)_{i \in \mathbb{N}_0}$  eine diskrete Wahrscheinlichkeitsverteilung, d.h.  $0 \leq q_i \leq 1$  für alle  $i \in \mathbb{N}_0$ , und  $\sum_{i=0}^{\infty} q_i = 1$ . Wir nennen  $(q_i)_{i \in \mathbb{N}_0}$  die *Nachkommenverteilung*, da  $q_i$  die Wahrscheinlichkeit angibt, mit der ein Knoten genau  $i$  Nachkommen (d.h. Äste die zur nächsten Ebene von Knoten führen) hat. Wenn wir mit der Wurzel starten, so hat diese eine zufällige Anzahl von Nachkommen, die wiederum die Knoten von Ebene 1 bilden. Die Anzahl Knoten auf Ebene 1 bezeichnen wir mit  $X_1$ , und wir wollen also

$$\mathbb{P}(X_1 = k \mid X_0 = 1) = q_k.$$

Wenn wir nun auf Ebene 1 genau  $k$  Knoten haben, so hat jeder dieser  $k$  Knoten jeweils unabhängig von allen anderen Knoten wieder eine zufällige Anzahl an Nachkommen, und die Anzahl folgt wieder der Wahrscheinlichkeitsverteilung  $(q_i)_{i \in \mathbb{N}_0}$ . Nach diesem Prinzip wird der zufällige Baum rekursiv aufgebaut: Sind auf Ebene  $n$  genau  $k$  Knoten vorhanden, also  $X_n = k$ , so hat jeder dieser  $k$  Knoten unabhängig davon eine Anzahl Nachkommen, sagen wir,  $Z_n^{(i)}$ ,  $i = 1, \dots, k$  bezeichnet die Anzahl Nachkommen von Knoten  $i$  auf Ebene  $n$ . Dabei gelte  $\mathbb{P}(Z_n^{(i)} = j) = q_j$ ,  $j \in \mathbb{N}_0$ , und die einzelnen  $Z_n^{(i)}$  sind *unabhängige* Zufallsvariablen. Die Gesamtzahl an Knoten in Ebene  $n+1$  ist dann die Summe der Nachkommen der einzelnen Knoten:

$$X_{n+1} = \sum_{i=1}^k Z_n^{(i)}, \quad \text{falls } X_n = k \text{ ist.}$$

Wenn die Nachkommenverteilung  $(q_i)_{i \in \mathbb{N}_0}$  bekannt ist, kann die Übergangswahrscheinlichkeit

$$p_{km} = \mathbb{P}(X_{n+1} = m \mid X_n = k)$$

im Prinzip berechnet werden, da wir in Kapitel 4 gelernt haben, dass diese Wahrscheinlichkeit durch die *Faltungsformel* gegeben wird. Der Ausdruck wird aber im Allgemeinen etwas kompliziert.

**Beispiel 7.9** (Warteschlangenmodelle). *siehe Vorlesung*

**Definition 7.10.** Die *Übergangsmatrix* einer Markovkette auf einem endlichen Zustandsraum  $S = \{a_1, \dots, a_K\}$  ist gegeben durch

$$P := (p_{a_n, a_m})_{n, m=1, \dots, K} = \begin{bmatrix} p_{a_1, a_1} & p_{a_1, a_2} & \cdots & p_{a_1, a_K} \\ p_{a_2, a_1} & p_{a_2, a_2} & \cdots & p_{a_2, a_K} \\ \vdots & \ddots & & \vdots \\ p_{a_K, a_1} & p_{a_K, a_2} & \cdots & p_{a_K, a_K} \end{bmatrix}$$

**Satz 7.11** (Stochastische Matrix). *Sei  $P$  die Übergangsmatrix einer Markov-Kette. Dann gelten  $0 \leq p_{a,b} \leq 1$  für alle  $a, b \in S$ , sowie*

$$\sum_{b \in S} p_{ab} = 1 \quad \forall a \in S.$$

*Beweis.* Nach Definition gilt  $p_{ab} = \mathbb{P}(X_1 = b \mid X_0 = a)$ , also muss  $p_{ab}$  als Wahrscheinlichkeit zwischen 0 und 1 liegen. Außerdem haben wir

$$\sum_{b \in S} p_{ab} = \sum_{b \in S} \mathbb{P}(X_1 = b \mid X_0 = a) = \sum_{b \in S} \mathbb{P}(X_1 \in S \mid X_0 = a),$$

und die rechte Seite ist einfach die Wahrscheinlichkeit, bei Start in  $a$  nach einem Schritt in einem beliebigen Zustand zu sein. Da man aber auf jeden Fall in irgend einem Zustand sein muss, ist die rechte Seite gleich 1.  $\square$

Eine Matrix  $P$  mit den beiden Eigenschaften aus obigem Satz heißt auch *stochastische Matrix*. So wie man zu jeder Markov-Kette eine Übergangsmatrix findet, welche diese beiden Eigenschaften hat, existiert auch zu jeder stochastischen Matrix eine entsprechende Markov-Kette.

**Satz 7.12** (Übergangswahrscheinlichkeiten und eindimensionale Verteilungen). *Sei  $X_0, X_1, \dots$  eine homogene Markov-Kette auf einem Zustandsraum  $S$ . Dann gilt für alle  $b \in S$*

$$\mathbb{P}(X_n = b) = \sum_{a \in S} \mathbb{P}(X_{n-1} = a) \cdot p_{a,b},$$

und iterativ

$$\mathbb{P}(X_n = b) = \sum_{a_0 \in S} \dots \sum_{a_{n-1} \in S} \mathbb{P}(X_0 = a_0) p_{a_0, a_1} \cdot \dots \cdot p_{a_{n-1}, b}.$$

Dieser Satz besagt also dass die Kenntnis von  $\mathbb{P}(X_0 = a_0)$  und der Übergangsmatrix  $P$  im Prinzip ausreicht, um die Verteilung jedes  $X_n, n = 1, 2, \dots$  zu berechnen.

*Beweis.* Im Prinzip ist dieser Satz nichts anderes als die Formel von der Gesamtwahrscheinlichkeit – vgl. auch Beispiel 2.19 in Kapitel 2. Entsprechend ist der Beweis einsichtig, wenn man den geeigneten Baum aufzeichnet (siehe Vorlesung).  $\square$

**Definition 7.13.** Die Funktion  $\nu : S \rightarrow [0, 1]$  :

$$\nu(a) = \mathbb{P}(X_0 = a)$$

heißt *Startverteilung* der Markov-Kette. Äquivalente Schreibweise:  $\nu_a := \nu(a)$ .

Wir können die Startverteilung auch als (Spalten-)Vektor auffassen, indem wir  $\nu = (\nu_a)_{a \in S}$  schreiben. Wenn wir uns daran erinnern, dass die Übergangswahrscheinlichkeiten eine Matrix bilden, dann sehen wir, dass das obige Ergebnis

$$\mathbb{P}(X_n = b) = \sum_{a_0 \in S} \dots \sum_{a_{n-1} \in S} \nu_{a_0} \cdot p_{a_0, a_1} \cdot \dots \cdot p_{a_{n-1}, b}$$

nichts anderes ist als eine (wiederholte) Vektor-Matrix-Multiplikation, und wir können es viel kürzer schreiben als

$$\mathbb{P}(X_n = b) = (\nu^T P^n)(b),$$

wobei  $P^n$  die  $n$ -fache *Matrix*-Potenz von  $P$  bezeichnet (also das  $n$ -fache Matrix-Produkt von  $P$  mit sich selbst), und  $\nu^T$  der Zeilenvektor ist, den man durch Transposition von  $\nu$  erhält.

Durch Angabe von  $\nu$  und  $P$  kennt man also im Prinzip die ganze Markov-Kette, und kann Wahrscheinlichkeiten mit obiger Formel berechnen. Für praktische Zwecke ist für große  $n$  und großen Zustandsraum  $S$  aber der Rechenaufwand zu hoch. Deshalb werden wir uns nun mit strukturellen Eigenschaften von Markov-Ketten beschäftigen, welche es erlauben werden, das *Langzeitverhalten* einer Markov-Kette zu untersuchen.

**Beispiel 7.14.** Wir betrachten eine Markov-Kette mit folgender Übergangsmatrix:

$$P = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

Man überprüft leicht (wie?), dass das tatsächlich eine stochastische Matrix ist, die zu einer Markov-Kette mit Zustandsraum  $\{1, 2, 3\}$  gehört. Wählen wir die Startverteilung

$$\nu = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix}, \text{ d.h. } \mathbb{P}(X_0 = 1) = \mathbb{P}(X_0 = 3) = 1/2, \mathbb{P}(X_0 = 2) = 0.$$

Was ist dann die Verteilung von  $X_3$ ? Nach obiger Formel gilt

$$\mathbb{P}(X_3 = \cdot) = \nu^T P^3.$$

Wir berechnen also als erstes die dritte Potenz von  $P$  (z.B. mit Matab) und erhalten

$$P^3 = \begin{bmatrix} 1/2 & 3/16 & 5/16 \\ 1/2 & 13/64 & 19/64 \\ 1/2 & 7/32 & 9/32 \end{bmatrix}.$$

Damit erhalten wir

$$\mathbb{P}(X_3 = \cdot) = \nu^T P^3 = \begin{bmatrix} 1/2 & 0 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} 1/2 & 3/16 & 5/16 \\ 1/2 & 13/64 & 19/64 \\ 1/2 & 7/32 & 9/32 \end{bmatrix} = \begin{bmatrix} 1/2 & 13/64 & 19/64 \end{bmatrix}.$$

Wir haben also

$$\mathbb{P}(X_3 = 1) = 1/2, \quad \mathbb{P}(X_3 = 2) = 13/64, \quad \mathbb{P}(X_3 = 3) = 19/64.$$

## 7.2. Eigenschaften von Markov-Ketten.

**Beispiel 7.15** (Irrfahrt auf Graph). Wir haben bereits mehrfach Irrfahrten auf Graphen betrachtet. Graphen können verschiedene Eigenschaften haben, die Auswirkungen auf das Verhalten einer solchen Irrfahrt haben: Sie können endlich oder unendlich sein, aus einer oder mehreren Komponenten bestehen, regelmäßig wie ein Gitter geformt sein, usw. In der Vorlesung sehen wir dazu einige Beispiele.

**Definition 7.16.** Sei  $(X_n)_{n \in \mathbb{N}}$  eine Markov-Kette mit Zustandsraum  $S$ . Eine Folge von Zuständen  $(a_0, a_1, \dots, a_k)$  heißt *(guter) Pfad*, falls  $p_{a_i, a_{i+1}} > 0$  ist für alle  $i = 0, \dots, k-1$ . Das bedeutet, dass die Kette von  $a_0$  nach  $a_1$  und weiter nach  $a_2, a_3, \dots$  bis  $a_k$  tatsächlich springen kann. Eine Markov-Kette heißt *irreduzibel*, falls für jede Wahl von zwei Zuständen  $a, b \in S$  ein Pfad existiert, welcher  $a$  und  $b$  verbindet, d.h. ein Pfad mit  $a_0 = a$  und  $a_k = b$ . Die Länge  $k$  ist dabei egal.

**Definition 7.17.** Eine irreduzible Markov-Kette heißt *aperiodisch*, falls für alle  $a \in S$  gilt

$$\text{ggT}\{k \in \mathbb{N} : \exists \text{Pfad der Länge } k \text{ von } a \text{ nach } a\} = 1.$$

**Beispiel 7.18** (Irrfahrt auf Graph: Irreduzibilität, Aperiodizität). Für Irrfahrten auf Graphen kann man oft auf einen Blick erkennen, ob sie irreduzibel sind, und falls ja, ob sie aperiodisch oder periodisch sind. **siehe Vorlesung**

**Definition 7.19** (Rekurrenz, Transienz). Sei  $(X_n)_{n \in \mathbb{N}_0}$  eine irreduzible Markov-Kette mit Zustandsraum  $S$ . Die Kette heißt *rekurrent*, falls es ein  $a \in S$  gibt, so dass

$$\mathbb{P}(\exists n \in \mathbb{N} : X_n = a \mid X_0 = a) = 1$$

ist. Andernfalls heißt die Kette *transient*.

*Bemerkung.* Die Eigenschaft

$$\mathbb{P}(\exists n \in \mathbb{N} : X_n = a \mid X_0 = a) = 1$$

bedeutet, dass die Kette, wenn sie im Zustand  $a$  startet, mit Wahrscheinlichkeit 1, also sicher, irgendwann wieder in diesen Zustand zurückkehrt.

Man kann beweisen, dass für irreduzible Markov-Ketten die Eigenschaft dass es ein  $a \in S$  gibt mit  $\mathbb{P}(\exists n \in \mathbb{N} : X_n = a \mid X_0 = a) = 1$  bereits impliziert, dass *jeder* Zustand  $a \in S$  diese Eigenschaft hat, und dass es auch bedeutet, dass man, egal wo man startet, jeden Zustand mit Wahrscheinlichkeit 1 mindestens einmal besucht (sogar, dass man dann jeden Zustand unendlich oft besucht).

Ist die Kette *nicht irreduzibel*, so kann es Zustände geben, zu denen man mit Wahrscheinlichkeit 1 wieder zurückkehrt, und solche, zu denen man mit positiver Wahrscheinlichkeit nie wieder zurückkehrt.

**Beispiel 7.20** (Endlicher Zustandsraum). Man kann sich relativ leicht davon überzeugen, dass irreduzible Markov-Ketten auf einem endlichen Zustandsraum immer rekurrent sind. Wir betrachten in der **Vorlesung** ein Beispiel für eine nicht irreduzible Markovkette auf einem endlichen Zustandsraum, bei der beide möglichen Varianten auftreten.

**Beispiel 7.21** (Verzweigungsprozess: Rekurrenz und Transienz). **siehe Vorlesung**

**7.3. Invariante Verteilungen.** In diesem Unterkapitel befassen wir uns mit dem *Langzeitverhalten* von Markov-Ketten, das heißt, mit Fragen wie z.B. “Welche Zustände werden auf lange Sicht von einer Markov-Kette besonders oft besucht?” Dabei ist ein Begriff zentral.

**Definition 7.22** (Invariante Verteilung). Sei  $X_0, X_1, \dots$  eine homogene Markov-Kette mit Zustandsraum  $S$ . Ein (Spalten-)Vektor (bzw. eine Folge)  $(\pi_a)_{a \in S}$  heißt *invariante Verteilung* oder (für die Markov-Kette  $X = (X_n)_{n \geq 0}$ ), falls gelten

- (1)  $\pi_a = \sum_{b \in S} \pi_b p_{b,a}$ , für alle  $a \in S$ ,
- (2)  $\pi_a \geq 0$  für alle  $a \in S$ , und  $\sum_{a \in S} \pi_a = 1$ .

Eine invariante Verteilung ist also (wegen der Eigenschaft (2)) eine diskrete Wahrscheinlichkeitsverteilung auf dem Zustandsraum  $S$ . Statt dem Begriff invariante Verteilung benutzt man oft auch die folgenden äquivalenten Bezeichnungen: *Stationäre Verteilung* oder *Gleichgewichtsverteilung*.

Falls statt (2) nur  $\pi_a \geq 0$  für alle  $a \in S$  gilt, so heißt  $\pi$  auch *invarianter Vektor* bzw. invariante Folge (oder auch invariantes Maß).

**Satz 7.23.** Falls  $\pi$  eine stationäre Verteilung für  $(X_n)_{n \geq 0}$  ist, und  $\mathbb{P}(X_0 = a) = \pi_a$  für alle  $a \in S$  gilt, so gilt für alle  $n \in \mathbb{N}$ ,

$$\mathbb{P}_\pi(X_n = a) = \pi_a.$$

*Beweis.* Dies ist im Prinzip bloß eine Umformulierung der Eigenschaft (1) einer invarianten Verteilung, für Details siehe Vorlesung.  $\square$

Dieser Satz sagt aus, dass falls die *Startverteilung* der Kette gerade eine invariante Verteilung ist, die Verteilung der Kette zu jedem festen Zeitpunkt gegeben ist durch ebendiese invariante Verteilung. Mit anderen Worten, die Verteilung bleibt *invariant*, also unverändert, im Laufe der Zeit. Dies erklärt den Namen “invariant”.

**Theorem 7.24** (Existenz und Berechnung der invarianten Verteilung). Eine homogene Markov-Kette auf einem endlichen Zustandsraum  $S$  besitzt immer mindestens eine invariante Verteilung. Invariante Verteilungen sind Lösungen des Gleichungssystems

$$(P - I)^T \pi = 0, \quad \sum_{a \in S} \pi_a = 1,$$

wobei  $I$  die Einheitsmatrix (Identität) bezeichnet, d.h.  $I$  ist die quadratische Matrix der Größe  $|S|$ , welche auf der Diagonalen Einsen, und ansonsten nur Nullen enthält.  $(P - I)^T$  bezeichnet die transponierte Matrix zu  $P - I$ , d.h. die Matrix die man erhält indem man Zeilen zu Spalten macht.

*Beweis.* Gegeben die Übergangsmatrix  $P$  muss eine invariante Verteilung  $\pi = (\pi_a)_{a \in S}$  (als Zeilenvektor aufgefasst) nach Definition die folgenden Eigenschaften erfüllen:

- (1)  $\pi_a = \sum_{b \in S} \pi_b p_{b,a}$ , für alle  $a \in S$ ,
- (2)  $\sum_{a \in S} \pi_a = 1$ .

Die Bedingung (1) können wir in Vektor-Matrix-Schreibweise umformen zu

$$\pi^T = \pi^T P,$$

wobei  $\pi^T$  den *Zeilenvektor* bezeichnet, den man durch Transposition des *Spaltenvektors*  $\pi$  erhält. Äquivalenzumformungen führen auf

$$(P - I)^T \pi = 0.$$

Dies ist ein lineares Gleichungssystem wie aus der linearen Algebra bekannt. Lösen des Gleichungssystems unter der zusätzlichen Bedingung (2) liefert also eine invariante Verteilung. Mit den Methoden der linearen Algebra kann man zeigen, dass der Lösungsraum von (1) für stochastische Matrizen immer mindestens eindimensional ist.  $\square$

**Satz 7.25.** *Falls die Markov-Kette im obigen Fall irreduzibel ist, so ist die invariante Verteilung eindeutig.*

(ohne Beweis)

**Beispiel 7.26.** Betrachte wieder die Matrix

$$P = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

Man kann sich, z.B. durch Aufmalen des Übergangsgraphen, davon überzeugen, dass die zugehörige Markov-Kette irreduzibel und aperiodisch ist. Berechnen wir die invariante Verteilung. Dazu müssen wir die Gleichung

$$(P - I)^T \pi = 0$$

lösen, zuerst also  $(P - I)^T$  berechnen. Das ist

$$\left( \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)^T = \begin{bmatrix} -1/2 & 0 & 1/2 \\ 1/2 & -3/4 & 1/4 \\ 1/2 & 1/2 & -1 \end{bmatrix}^T = \begin{bmatrix} -1/2 & 1/2 & 1/2 \\ 0 & -3/4 & 1/2 \\ 1/2 & 1/4 & -1 \end{bmatrix}.$$

Das Gleichungssystem  $(P - I)^T \pi = 0$  wird also zu

$$\begin{bmatrix} -1/2 & 1/2 & 1/2 \\ 0 & -3/4 & 1/2 \\ 1/2 & 1/4 & -1 \end{bmatrix} \cdot \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Zusätzlich muss die Gleichung  $\pi_1 + \pi_2 + \pi_3 = 1$  erfüllt sein, wir haben also ein Gleichungssystem mit drei Unbekannten und vier Gleichungen. Lösen mittels den bekannten Methoden der linearen Algebra (z.B. Gauß-Elimination) liefert für die invariante Verteilung

$$\pi = \begin{bmatrix} 1/2 \\ 1/5 \\ 3/10 \end{bmatrix}.$$

**Beispiel 7.27** (Netzwerk). siehe Vorlesung

**Theorem 7.28.** Sei  $(X_n)_{n \in \mathbb{N}_0}$  eine irreduzible und aperiodische homogene Markov-Kette mit Zustandsraum  $S$ . Sei  $(\pi_a)_{a \in S}$  eine invariante Verteilung. Dann gilt für alle  $a \in S$ , unabhängig von der Startverteilung,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = a) = \pi_a.$$

*Beweis.* Siehe z.B. Georgii, Stochastik (de Gruyter, 4. Auflage 2009), Cor. 6.17. □

Der Satz besagt, dass, ganz egal wo die Kette startet, für große  $n$  gilt

$$\mathbb{P}(X_n = a) \approx \pi_a.$$

D.h. auf lange Frist ( $n \rightarrow \infty$ ) gibt  $\pi_a$  die Wahrscheinlichkeit an, mit der sich die Kette im Zustand  $a$  aufhält, oder, anders ausgedrückt, den Anteil an Zeitschritten, zu denen sich die Kette im Zustand  $a$  befindet.

Dies liefert wichtige Informationen über Struktur der Markov-Kette: Zustände mit großem Wert für  $\pi$  werden häufiger besucht als Zustände mit kleinem Wert. In unserem Einstiegsbeispiel mit dem Netzwerk (Bsp. 7.1) heißt das: Knoten mit vielen Kanten werden oft besucht, für solche Knoten  $a$  ist  $\pi_a$  größer als für solche die selten besucht werden.

Man kann beweisen, dass falls keine invariante Verteilung existiert, unter den Voraussetzungen des Satzes gilt  $\lim_{n \rightarrow \infty} \mathbb{P}(X_n = a) = 0$  für alle  $a \in S$ .

Die Voraussetzungen des Satzes implizieren dass die invariante Verteilung, falls sie existiert, eindeutig ist.

**7.4. Anwendungen.** Die Beispiele werden in der Vorlesung behandelt, je nach Zeitdruck auch erst im Teil über stochastische Algorithmen.

**Beispiel 7.29** (PageRank).

**Beispiel 7.30** (Qualitätstest von technischen Geräten).

**Beispiel 7.31** (Tiefe von Suchbäumen).

**Beispiel 7.32** (Stabilität von Warteschlangen).

**Fortsetzung des Skripts mit dem Statistik-Teil in Arbeit, folgt später!!!**