

# Algorithmische Mathematik II

Dozent

PROFESSOR DR. PATRIK FERRARI

Mitschrift

MAXIMILIAN KESSLER

Version

12. Mai 2021 11:00

## Zusammenfassung

Bei folgenden Vorlesungsnotizen handelt es sich um (inoffizielle) Mitschriften zur Vorlesung 'Algorithmische Mathematik II', die im Sommersemester 2021 an der Universität Bonn gehalten wird. Ich garantiere weder für Korrektheit noch Vollständigkeit dieser Notizen, und bin dankbar für jegliche Art von Korrektur, sowohl inhaltlich, als auch Tippfehler.

Bemerkungen, die nicht zum eigentlichen Vorlesungsinhalte gehören, wurden mit einem \* gekennzeichnet. Sie werden nach eigenem Ermessen hinzugefügt, um weitere Details oder evtl. mündliche Anmerkungen beizufügen.

Manche Umgebungen sind mit einem <sup>†</sup> versehen. Das ist dann der Fall, wenn ihr Inhalt so, oder zumindest in sehr ähnlicher Form, in der Vorlesung vorkam (unter Umständen auch mündlich), ich aber die Umgebung der Aussage geändert habe. Das ist z.B. dann der Fall, wenn ich aus Aussagen, die einfach erwähnt werden, ein **Lemma**<sup>†</sup> mache, um sie hervorzuheben.

Weitere Informationen finden sich bei [GitHub](#) oder auf der [Vorlesungs-homepage](#)

# Inhaltsverzeichnis

<b>Übersicht der Vorlesungen</b>	<b>3</b>
<b>1 Diskrete Stochastik</b>	<b>4</b>
1.1 Einleitung	4
1.2 Ereignisse und Wahrscheinlichkeiten	6
1.3 Diskrete Verteilungen	10
1.4 Die Gleichverteilung	14
1.5 Die empirische Verteilung	15
1.6 Zufallsvariablen	16
1.6.1 Die Bernoulli-Verteilung	18
1.6.2 Die Binomial-Verteilung	19
1.6.3 Die Poisson-Verteilung	21
1.6.4 Die geometrische Verteilung	23
1.7 Simulation von Gleichverteilung	23
1.7.1 Lineare Kongruenzgeneratoren (LCG)	23
1.7.2 Zufallsvariablen aus $[0, 1)$	24
1.7.3 Zufallspermutationen	24
1.7.4 Geometrische Verteilung	25
1.8 Erwartungswert und Varianz	25
<b>2 Bedingte Wahrscheinlichkeit und Unabhängigkeit</b>	<b>32</b>
2.1 Bedingte Wahrscheinlichkeit	32
2.2 Baye'sche Regel	36
2.3 Mehrstufige Modelle	38
2.3.1 Produktmodelle	39
2.3.2 Markovketten (MK)	41
2.4 Unabhängige Zufallsvariablen	46
2.4.1 Unabhängige Ereignisse	46
2.4.2 Unabhängige Zufallsvariablen	47
<b>Stichwortverzeichnis</b>	<b>48</b>

# Übersicht der Vorlesungen

<b>Vorlesung 1 (Mo 12 Apr 2021 10:16)</b>	<b>4</b>
Motivationsfragen. Brown'sche Bewegung. Ereignisse, Wahrscheinlichkeiten, Modell von Zufallsexperimenten.	
<b>Vorlesung 2 (Mi 14 Apr 2021 10:17)</b>	<b>7</b>
$\sigma$ -Algebren, Messräume. Wahrscheinlichkeitsverteilungen, Wahrscheinlichkeitsräume. Einschluss-Ausschluss-Prinzip. Endliche (diskrete) Wahrscheinlichkeitsräume.	
<b>Vorlesung 3 (Mo 19 Apr 2021 10:23)</b>	<b>11</b>
Diskrete Verteilungen. Gleichverteilung. Fixpunkte von Permutationen. Empirische Verteilung.	
<b>Vorlesung 4 (Mi 21 Apr 2021 10:15)</b>	<b>16</b>
Diskrete Zufallsvariablen. Massfunktion. Bernoulli - Binomialverteilung. Unabhängigkeit von Ereignissen. Poissonverteilung.	
<b>Vorlesung 5 (Mo 26 Apr 2021 10:17)</b>	<b>22</b>
Geometrische Verteilung. Lineare Kongruenzgeneratoren (LCGs). Knuth. Simulation von Zufallsvariablen, Zufallspermutationen und der geometrischen Verteilung. Empirischer Mittelwert. Erwartungswert.	
<b>Vorlesung 6 (Sa 01 Mai 2021 09:18)</b>	<b>26</b>
Erwartungswerte von Binomialt und Poissonverteilung. Funktionen von Zufallsvariablen. Linearität und Monotonie des Erwartungswerts. Varianz. Varianz von Bernoulli-, Binomial-, Poisson- und geometrischer Verteilung. Beweis der Einschluss-Ausschluss-Formel mit Erwartungswerten.	
<b>Vorlesung 7 (Mo 03 Mai 2021 10:17)</b>	<b>32</b>
Bedingte Wahrscheinlichkeit und Wahrscheinlichkeitsverteilung. A-priori- und A-posteriori-Verteilung, Baye'sche Regel. Anwendung: Aktuelle Corona-Zahlen und -Tests.	
<b>Vorlesung 8 (Mi 05 Mai 2021 10:11)</b>	<b>38</b>
Mehrstufige Modelle. Produktmodelle.	
<b>Vorlesung 9 (Mo 10 Mai 2021 10:15)</b>	<b>41</b>
Markov-Ketten. Übergangsmatrizen.	
<b>Vorlesung 10 (Mi 12 Mai 2021 10:16)</b>	<b>45</b>

- Es gibt ein Helpdesk, auch explizit für Studentinnen
- die Vorlesung wird aufgenommen, und zwar ohne Videos der Teilnehmenden sowie des Dozenten, die Aufzeichnung werden anschließend in Sciebo hochgeladen.
- Es gibt ein Diskussionsforum für Fragen (auf eCampus).
- Ab heute Abend, 18 Uhr (Mo 12 Apr 2021 18:00), kann man sich auf eCampus für die Übungsgruppen registrieren und endet am Dienstag Abend um 24 Uhr (Di 12 Apr 2021 24:00), es wird versucht, die Studenten gleichmäßig zu verteilen.
- Falls ihr in der Warteliste landet und gewünscht ist, in der Gruppe abzugeben, schreibt eine Mail mit den gewünschten Abgabepartnern, dann kann eine gemeinsame Einteilung erfolgen.
- Es gibt auch das Modul **AlmaITb**. Registriert euch noch nicht, dies ist für den 2. Teil der Vorlesung notwendig.
- Die Abgabe der Übungsblätter erfolgt einheitlich jeden Freitag um 12 Uhr.
- Gruppenabgaben sind erlaubt, bis zu einer Größe von maximal 4 StudentInnen.
- Das 1. Blatt ist freiwillig und gibt Bonuspunkte.
- Für die Klausurzulassung werden 50% der Punkte benötigt. Von den Programmieraufgaben müssen mindestens 4 von 6 zufriedenstellend bearbeitet werden.
- Programmieraufgaben gibt es ab dem 2. Übungsblatt auf jedem 2. Blatt. Die Bearbeitungszeit beträgt dann 2 Wochen.

## Einleitung

In der Vorlesung werden wir sehen:

**Teil 1: Diskrete Stochastik** • Zufallsvariablen

- Bedingte Wahrscheinlichkeiten
- Unabhängigkeit von Variablen
- Monte-Carlo Methoden

**Teil 2: Numerische Analysis** • Iterative Verfahren

- Interpolation von Daten (durch Polynome, trigonometrische Funktionen, ...)
- Numerische Verfahren für die Integration

## 1 Diskrete Stochastik

### 1.1 Einleitung

**Ziel.** Beschreibung von Systemen, die einen Anteil an **Zufall** haben, d.h. nicht 100% deterministisch sind.

- Beispiel.** • Spiele: Kartenspiele, Glücksspiele, ...
- Statistik: Umfragen, Versicherung
  - Komplexe Systeme: Wettermodelle, Finanzmärkte

Was sind Quellen von Zufall?

- Zu komplexe Systeme. Dann sieht der Gesamteffekt zufällig aus.
- Fehlende Informationen (z.B. bei einem Kartenspiel)
- Chaotische Systeme (Wetter)

- Intrinsisch unvorhersagbare Systeme (z.B. radioaktiver Zerfall)

**Frage.** (1) Wie modelliert man ein System mit Zufall?  
 (2) Wie simuliert man ein System mit Zufall? (anwendungstechnischer)  
 (3) Welche Voraussagen kann man machen?

**Beispiel.** Die **Brown'sche Bewegung**. Das System ist implizit ein Pollen mit vielen Wassermolekülen ( $\sim 10^{23}$ ), die sich im Prinzip deterministisch bewegen.

$\Rightarrow$  Wir erhalten ein Gleichungssystem mit  $(N + 1) \cdot 6$  (3 Positionen, 3 Geschwindigkeit) Variablen. Dieses ist de facto unlösbar.

Was wollen wir hier eigentlich untersuchen?  $\rightarrow$  Die Bewegung des Pollens, jedoch nicht die der einzelnen Wassermoleküle.

In einer **Modellierung** ersetzt man die Stöße, die durch die Wassermoleküle entstehen durch **zufällige Stöße**.

Diskretes Modell: Die Zeit bewegt sich in  $n \in \{0, 1, 2, \dots\}$ . Sei

$$Z(n) := (\text{Position des Pollens zur Zeit } n) \in \mathbb{Z}^3.$$

OBdA setzen wir  $Z(0) = 0$ .

Dynamik:  $Z(n + 1) = Z(n) + \xi_n$ , wobei wir  $\xi_n$  aus dem Ergebnis eines Würfelwurfs bestimmen werden:

$$\xi_n = \begin{cases} (1, 0, 0) & \text{wenn Würfel} = 1 \\ (-1, 0, 0) & \text{wenn Würfel} = 2 \\ (0, 1, 0) & \text{wenn Würfel} = 3 \\ (0, -1, 0) & \text{wenn Würfel} = 4 \\ (0, 0, 1) & \text{wenn Würfel} = 5 \\ (0, 0, -1) & \text{wenn Würfel} = 6 \end{cases}.$$

**Frage.** Welche Fragen können wir mit solch einem System nun beantworten? Was passiert, wenn  $n \gg 1$ ?

- (a) Typischerweise erhalten wir  $|Z(n)| = O(\sqrt{n})$
- (b) Wenn wir die Frequenz von  $[Z(n)]_i$  betrachten, (d.h. bei welcher Koordinate in Richtung  $i$  befinden wir uns nach  $n$  Würfeln) sehen wir typischerweise: Für  $n \gg 1$  sieht diese Verteilung dann ungefähr wie die Gaussglocke aus.

Skalierung: Wir setzen nun

$$B(t) = \lim_{n \rightarrow \infty} \frac{Z(\lfloor nt \rfloor)}{\sqrt{n}}.$$

und dies ist dann die Brownsche Bewegung.

Nun möchten wir Vorhersagen treffen können:

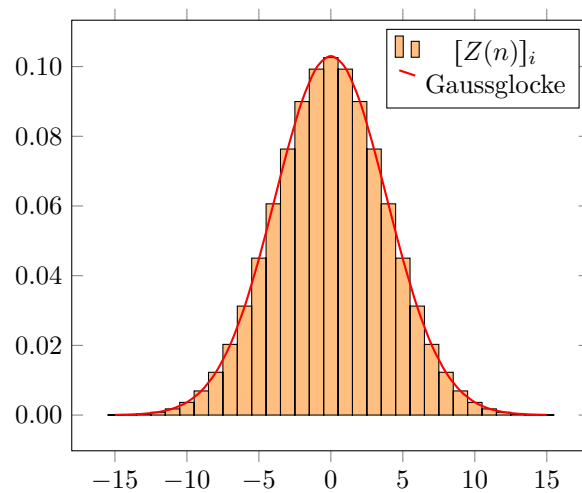


Abbildung 1: Binomialverteilung und Gaussglocke

**Frage.** Ist  $Z(n)$  in einer gegebenen Menge  $A$ ?

Das kann man (im Allgemeinen) nicht einfach mit 'Ja' oder 'Nein' beantworten. Stattdessen müssen wir fragen:

**Frage.** Wenn man  $Z(n)$  beobachtet, wie häufig wird  $Z(n)$  in  $A$  sein?

Diese Frage lässt sich mit einer Zahl  $\in [0, 1]$  beantworten.

## 1.2 Ereignisse und Wahrscheinlichkeiten

Wir benötigen 3 Grundelemente:

- (1) Die Menge  $\Omega$  von möglichen **Ergebnissen**. die Elemente von  $\Omega$  heißen auch **Elementarereignisse**.
- (2) Die Menge  $\mathcal{F}$  der **Ereignisse**. Ein Ereignis  $E$  ist eine Eigenschaft, die mit einer Teilmenge  $G \subseteq \Omega$  assoziiert ist:  $\omega \in G \Leftrightarrow$  Eigenschaft  $E$  ist erfüllt.
- (3) Eine **Wahrscheinlichkeitsverteilung (auch W-maß)**:

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

**Bemerkung\*.** Wir werden noch sehen, dass gewisse Dinge für unsere Begriffe erfüllt sein müssen, dazu aber später mehr.

**Beispiel.** Eine Urne hat 12 nummerierte Kugeln (von 1 bis 12).

- (1) Das **Zufallsexperiment** besteht daraus, dass wir eine Kugel aus der Urne ziehen und die Zahl notieren, die wir sehen. D.h.

$$\Omega = \{1, \dots, 12\}.$$

Ein Elementarereignis ist nun z.B. gegeben durch  $\omega = \{5\} \equiv 5$  (wir vereinfachen die Notation).

(2) Mögliche Ereignisse sind z.B:

$$\begin{aligned} A &= \text{'Die Zahl ist gerade'} \\ B &= \text{'Die Zahl ist } \leq 5 \text{' } \\ C &= \text{'Die Zahl ist 8'} \end{aligned} \quad (1)$$

Die assoziierten Mengen sind dann

$$\begin{aligned} A &= \{2, 4, 6, 8, 10, 12\} \\ B &= \{1, 2, 3, 4, 5\} \\ C &= \{8\} \end{aligned} \quad (2)$$

(3) Für die Wahrscheinlichkeiten nehmen wir an, dass jede Kugel die gleiche Chance hat, gezogen zu werden, d.h.

$$\forall G \in \mathcal{F} : \mathbb{P}(G) = \frac{|G|}{|\Omega|}.$$

Wir erhalten nun die Wahrscheinlichkeiten

$$\mathbb{P}(A) = \frac{6}{12} = \frac{1}{2} \quad \mathbb{P}(B) = \frac{5}{12} \quad \mathbb{P}(C) = \frac{1}{12}.$$

**Notation.**  $A \equiv \{\omega \in \Omega \mid \omega \in A\} \equiv \{\omega \in A\} \equiv \{A \text{ tritt ein}\}$

Wir kennen nun die Grundbegriffe  $\Omega, \mathcal{F}, \mathbb{P}$  zur Beschreibung von Zufallsexperimenten, die wir uns nun genauer ansehen wollen:

**Frage.** Welche Struktur muss  $\mathcal{F}$  besitzen.

Seien  $A, B \in \mathcal{F}$ , dann können wir das Ereignis  $A \cap B$  betrachten, d.h. beide der Eigenschaften treten ein. Genauso sollte

$$A^c := \Omega \setminus A.$$

, das **Komplement von A**, bzw. das **Gegenereignis** von A ebenfalls in  $\mathcal{F}$  sein. Aus den beiden vorherigen Eigenschaften folgt bereits, dass

$$A \cup B = (A^c \cap B^c)^c.$$

ebenfalls in  $\mathcal{F}$  sein wird.

Eine Menge  $\mathcal{F}$  mit solchen Eigenschaften heißt **Algebra**.

**Notation<sup>†</sup>.** Seien nun  $A, B$  und  $(A_i)_{i \in I}$  Ereignisse, wobei  $I$  endlich oder abzählbar sei. Dann notieren wir die folgenden Ereignisse:

- (a)  $A \cup B : \omega \in A \cup B \Leftrightarrow \omega \in A \vee \omega \in B$ , d.h.  $A \cup B$  tritt ein, genau dann, wenn A eintritt oder B eintritt.
- (b)  $\bigcup_{i \in I} A_i : \omega \in \bigcup_{i \in I} A_i$ , wenn es ein  $i \in I$  gibt, sodass  $\omega \in A_i$
- (c)  $A \cap B : \omega \in A \cap B \Leftrightarrow A \text{ und } B \text{ treten ein.}$
- (d)  $\bigcap_{i \in I} A_i : \omega \in \bigcap_{i \in I} A_i \Leftrightarrow \forall i \in I : A_i \text{ tritt ein.}$
- (e)  $A = \emptyset$  ist das Ereignis, das nie eintritt.

Vorlesung 2  
Mi 14 Apr 2021 10:17

$A = \Omega$  ist das Ereignis, dass immer eintritt.

**Definition 1.1** ( $\sigma$ -Algebra). Eine  $\sigma$ -Algebra ist eine nicht leere Menge  $\mathcal{F}$  von Teilmengen von  $\Omega$  mit den Eigenschaften:

- (a)  $\Omega \in \mathcal{F}$
  - (b)  $\forall A \in \mathcal{F}: A^c \in \mathcal{F}$ .
  - (c) Falls  $(A_i)_{i \in I} \in \mathcal{F}$ , dann auch  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
- Wir nennen  $(\Omega, \mathcal{F})$  dann einen **Messraum**.

**Lemma 1.2.** Sei  $\mathcal{F}$  eine  $\sigma$ -Algebra, dann ist:

- (a)  $\emptyset \in \mathcal{F}$
- (b)  $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$  und  $A \cap B \in \mathcal{F}$ .
- (c)  $(A_i)_{i \in I} \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ .

*Beweis.* (a)  $\emptyset = \Omega^c \in \mathcal{F}$  nach Eigenschaften (a) und (b) aus der Definition.

(b)  $A \cup B = A \cup B \cup \emptyset \cup \emptyset \dots \in \mathcal{F}$  nach Eigenschaften (b) und (c).  
 $A \cap B = (A^c \cup B^c)^c \in \mathcal{F}$

(c)  $\bigcap_{i=1}^{\infty} A_i = \left( \bigcup_{i=1}^{\infty} A_i^c \right)^c \in \mathcal{F}$  nach (b) und (c). □

Wir haben nun  $(\Omega, \mathcal{F})$  näher untersucht, es fehlt nun noch  $\mathbb{P}$ .

**Frage.** Welche Eigenschaften soll  $\mathbb{P}$  (das Wahrscheinlichkeitsmaß bzw. die Wahrscheinlichkeitsverteilung) besitzen?

Seien  $A, B \in \mathcal{F}$  mit  $A \cap B = \emptyset$ , d.h.  $A$  und  $B$  können nicht gleichzeitig eintreten. Dann fordern wir

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) \quad (\text{endliche Additivität}).$$

Dazu wollen wir, dass  $\Omega \in \mathcal{F}$  immer eintritt, d.h.  $\mathbb{P}(\Omega) = 1 \equiv 100\%$  (Normierung).

**Definition 1.3** (Wahrscheinlichkeitsverteilung). Sei  $(\Omega, \mathcal{F})$  ein Messraum. Eine Abbildung  $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}_+$  ist eine **Wahrscheinlichkeitsverteilung** auf  $(\Omega, \mathcal{F})$ , falls

- (1)  $\mathbb{P}(\Omega) = 1$
- (2) Sind  $(A_i)_{i \in I} \in \mathcal{F}$  paarweise disjunkt, so ist:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \quad (\sigma\text{-Additivität}).$$

**Bemerkung\*.** Die Definition macht implizit Gebrauch davon, dass die linke Seite überhaupt definiert ist. Dies folgt jedoch daraus, dass  $\mathcal{F}$  eine  $\sigma$ -Algebra ist.



**Definition 1.4** (Wahrscheinlichkeitsraum). Ein **Wahrscheinlichkeitsraum**  $(\Omega, \mathcal{F}, \mathbb{P})$  besteht aus einer Menge  $\Omega$ , einer  $\sigma$ -Algebra  $\mathcal{F} \subseteq \mathbb{P}(\otimes)$  und einem Wahrscheinlichkeitsmass  $\mathbb{P}$  auf  $(\Omega, \mathcal{F})$

**Lemma 1.5.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum. Dann ist

(a)  $\mathbb{P}(\emptyset) = 0$

(b)  $\forall A, B \in \mathcal{F}$  mit  $A \cap B = \emptyset$  ist

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

(c)  $\forall A, B \in \mathcal{F}$  mit  $A \subseteq B$  ist

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

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

$$\mathbb{P}(A) \leq \mathbb{P}(B) \leq 1$$

(d)  $\forall A, B \in \mathcal{F}$  ist

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &\leq \mathbb{P}(A) + \mathbb{P}(B) \end{aligned}$$

(e) Wenn  $A_n \nearrow A$ , d.h.  $A_1 \subseteq A_2 \subseteq \dots$  mit  $\bigcup_{i=1}^{\infty} A_i = A$  (monotone Konvergenz von Mengen), oder  $A_n \searrow A$  (d.h.  $A_1 \supseteq A_2 \supseteq \dots$  mit  $\bigcap_{i=1}^{\infty} A_i = A$ ), so ist

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n \rightarrow \infty} A_n\right) = \mathbb{P}(A).$$

*Beweis.* (a) Wir wissen:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset \cup \emptyset \cup \emptyset \dots) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \dots$$

subtrahieren von  $\mathbb{P}(\Omega) = 1$  liefert dann  $\mathbb{P}(\emptyset) = 0$ .

(b) Sei  $A \cap B = \emptyset$ , dann ist:

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \cup B \cup \emptyset \cup \emptyset \cup \dots) \\ &\stackrel{\sigma\text{-Additivitat}}{=} \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \dots \\ &= \mathbb{P}(A) + \mathbb{P}(B) \end{aligned}$$

(c) Sei  $A \subseteq B$ . Dann ist  $B = A \cup (B \setminus A)$  eine disjunkte Vereinigung, also erhalten wir

$$\mathbb{P}(B) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B \setminus A)}_{\geq 0} \geq \mathbb{P}(A).$$

Mit  $B = \Omega$  ergibt sich  $1 = \mathbb{P}(A) + \mathbb{P}(A^c)$

(d) Es ist

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}((A \cup B) \setminus A) \\ &= \mathbb{P}(A) + \mathbb{P}(B \setminus (A \cap B)) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \underbrace{\mathbb{P}(A \cap B)}_{\geq 0} \\ &\geq \mathbb{P}(A) + \mathbb{P}(B) \end{aligned} \tag{3}$$

ⓔ Übung

□

**Korollar 1.6** (Einschluss-Ausschluss-Prinzip). Seien  $A_1, \dots, A_n \in \mathcal{F}$ . Dann gilt

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}).$$

*Beweis.* Per Induktion, der Induktionsanfang lautet  $\mathbb{P}(A_1) = \mathbb{P}(A_1)$  und ist offensichtlich wahr.

Die Aussage gelte nun für ein  $n \in \mathbb{N}$ , dann erhalten wir

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mathbb{P}\left(\left(\bigcup_{i=1}^n A_i\right) \cup A_{n+1}\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\left(\bigcup_{i=1}^n A_i\right) \cap A_{n+1}\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\bigcup_{i=1}^n \underbrace{(A_i \cap A_{n+1})}_{=: \tilde{A}_i}\right) \\ &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) + \mathbb{P}(A_{n+1}) \\ &\quad - \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(\underbrace{\tilde{A}_{i_1} \cap \dots \cap \tilde{A}_{i_k}}_{A_{i_1} \cap \dots \cap A_{i_k} \cap A_{n+1}}) \end{aligned} \quad (4)$$

Andererseits ist aber auch:

$$\begin{aligned} &\sum_{k=1}^{n+1} (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n+1} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) \quad \left. \vphantom{\sum_{k=1}^n} \right\} \text{Terme mit } i_k \leq n \\ &+ \underbrace{\sum_{k=2}^{n+1} (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_{k-1} \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_{k-1}} \cap A_{n+1})}_{\substack{l:=k-1 \\ \sum_{l=1}^n (-1)^{l-1} \sum_{1 \leq i_1 < \dots < i_l \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_l} \cap A_{n+1})}} \quad \left. \vphantom{\sum_{k=2}^{n+1}} \right\} \text{Terme mit } i_k = n+1 \text{ und } k \geq 2 \\ &+ \mathbb{P}(A_{n+1}) \quad \left. \vphantom{\mathbb{P}(A_{n+1})} \right\} \text{Terme mit } i_k = n+1 \text{ und } k=1 \end{aligned}$$

und damit sehen wir, dass die beiden Ausdrücke übereinstimmen, also ist der Induktionsschritt erbracht. □

### 1.3 Diskrete Verteilungen

- Sei nun  $\Omega$  endlich oder abzählbar.

- Falls wir  $\mathcal{F}$  nicht explizit angeben, dann wird  $\mathcal{F} = \mathcal{P}(\Omega)$  gewählt, d.h.

$$\text{Card}(\mathcal{P}(\Omega)) \equiv |\mathcal{P}(\Omega)| = 2^{|\Omega|}.$$

**Beispiel** (Münzwurf). Es sei  $\Omega = \{K, Z\}$ , wobei  $K$  für Kopf stehe und  $Z$  für Zahl. Dann ist

$$\mathcal{F} = \{\{K\}, \{Z\}, \{Z, K\}, \emptyset\}.$$

Sei  $p \in [0, 1]$  die Wahrscheinlichkeit, dass man Kopf erhält. Da  $\mathbb{P}$  für alle Element aus  $\mathcal{F}$  definiert sein muss, erhalten wir

$$\mathbb{P}(\emptyset) = 0 \quad \mathbb{P}(K) = p, \quad \mathbb{P}(Z) = \mathbb{P}(K^c) = 1-p \quad \mathbb{P}(\{Z, K\}) = \mathbb{P}(\Omega) = 1.$$

**Frage** (Charakterisierung von diskreter Wahrscheinlichkeit). Was müssen wir fordern, sodass  $\mathbb{P}$  auf  $\mathcal{P}(\Omega)$  gibt?

**Beispiel.**  $\Omega = \{1, 2, \dots, 10\}$  würde genügen, da dann  $|\mathcal{P}(\Omega)| = 2^{|\Omega|} = 2^{10} = 1024$  endlich (diskret) ist.

Wir stellen fest, dass es im letzten Beispiel auch genügt hätte,  $\mathbb{P}(\{k\})$  für  $k = 1, \dots, 10$  anzugeben, das motiviert Folgendes:

Vorlesung 3  
Mo 19 Apr 2021 10:23

**Satz 1.7.** Sei  $\Omega$  abzählbar.

- (a) Sei  $p(\omega) \in [0, 1], \omega \in \Omega$ , sodass

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

Dann ist  $\mathbb{P}$  definiert durch:

$$\mathbb{P} : \begin{cases} \mathcal{P}(\Omega) & \longrightarrow & [0, 1] \\ A & \longmapsto & \sum_{\omega \in A} p(\omega) \end{cases}$$

eine Wahrscheinlichkeitsverteilung auf  $(\Omega, \mathcal{P}(\Omega))$ .

- (b) Jede Wahrscheinlichkeitsverteilung  $\mathbb{P}$  auf  $(\Omega, \mathcal{P}(\Omega))$  hat obige Form, wobei  $p(\omega) = \mathbb{P}(\{\omega\})$ .

**Bemerkung.**  $p : \Omega \rightarrow [0, 1]$  heißt Massenfunktion der Wahrscheinlichkeitsverteilung  $\mathbb{P}$ .

**Warnung.** Der Satz gilt nicht für  $\Omega$  überabzählbar.

**Bemerkung.** Sei  $A$  abzählbar und  $p(\omega) \geq 0$  für  $\omega \in A$ . Dann definiert

$$\sum_{\omega \in A} p(\omega) := \sum_{k \geq 1} p(\omega_k).$$

mit einer beliebigen Abzählung  $\omega_1, \omega_2, \dots, \omega_k$  von  $A$  eine wohldefinierte Summe der  $p(\omega)$ . Es ist wichtig, dass hier  $p(\omega) \geq 0$ , sonst ist obiges nicht wohldefiniert.

**Lemma 1.8.** Sei  $A$  abzählbar.

(a) Sei  $p(\omega) \in [0, 1]$  für alle  $\omega$ . Dann ist

$$\sum_{\omega \in A} p(\omega) \in [0, \infty].$$

wohldefiniert. Setzen wir

$$\mathbb{P}(A) := \sum_{\omega \in A} p(\omega).$$

so gilt

$$P(A) = \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F).$$

und  $P(A) \leq P(B)$  für  $A \subseteq B$ .

(b) Ist  $A = \bigsqcup_{k=1}^{\infty} A_k$  eine disjunkte Vereinigung, so ist

$$P(A) = \sum_{k=1}^{\infty} P(A_k).$$

*Beweis.* (a) Sei  $\omega_1, \omega_2, \dots$  eine beliebige Abzählung von  $A$ . Dann ist die Funktion

$$n \mapsto \sum_{k=1}^n p(\omega_k).$$

monoton wachsend. Also ist

$$\sum_{k=1}^{\infty} p(\omega_k) := \lim_{n \rightarrow \infty} \sum_{k=1}^n p(\omega_k) = \sup_{n \in \mathbb{N}} \sum_{k=1}^n p(\omega_k) \in [0, \infty].$$

wohldefiniert.

Wir wollen nun noch zeigen, dass

$$P(A) := \sum_{k=1}^{\infty} p(\omega_k) \stackrel{!}{=} \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F).$$

Die Ungleichung ' $\leq$ ' folgt sofort, da wir mit  $F_n := \{\omega_1, \dots, \omega_n\}$  feststellen, dass

$$\sum_{k=1}^n p(\omega_k) = \sum_{\omega \in F_n} p(\omega) = P(F_n) \leq \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F).$$

Also ergibt sich im Limes genau wie gewünscht

$$P(A) = \lim_{n \rightarrow \infty} \sum_{k=1}^n p(\omega_k) \leq \lim_{n \rightarrow \infty} \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F) = \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F).$$

Für ' $\geq$ ' stellen wir fest, dass es für jedes  $F \subseteq A$  endlich ein  $n \in \mathbb{N}$  gibt, sodass  $F \subseteq \{\omega_1, \dots, \omega_n\}$ , und somit ist

$$P(F) = \sum_{\omega \in F} p(\omega) \leq \sum_{k=1}^n p(\omega_k) \leq \sum_{k=1}^{\infty} p(\omega_k) = P(A).$$

und somit ist das Supremum der  $P(F)$  für  $F \subseteq A, |F| < \infty$  durch  $P(A)$  beschränkt.

Für die letzte Behauptung sehen wir mit  $A \subseteq B$  leicht, dass

$$P(A) = \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F) \leq \sup_{\substack{F \subseteq B \\ |F| < \infty}} P(F) = P(B).$$

(b) ( $\sigma$ -Additivität) Wir unterscheiden zwei Fälle:

1) Falls  $|A| < \infty$ , so ist  $A = \bigsqcup_{k=1}^n A_k$  für ein  $n$ , und somit ist

$$\begin{aligned} P(A) &= \sum_{l=1}^{|A|} p(\omega_l) = \sum_{l=1}^{|A|} \sum_{k=1}^n p(\omega_l) \mathbb{1}_{A_k}(\omega_l) \\ &= \sum_{k=1}^n \sum_{l=1}^{|A|} p(\omega_l) \mathbb{1}_{A_k}(\omega_l) = \sum_{k=1}^n P(A_k) \end{aligned} \quad (5)$$

2) Sei nun  $|A| = \infty$ . Wir zeigen zunächst ' $\leq$ '. Für ein endliches  $F \subseteq A$  ist

$$F = \bigcup_{k=1}^{\infty} (F \cap A_k).$$

eine disjunkte Vereinigung mit endlich vielen Termen, also ist

$$P(F) = \sum_{k=1}^{\infty} P(F \cap A_k) \leq \sum_{k=1}^{\infty} P(A_k).$$

und somit liefert das Supremum über beide Seiten, dass

$$P(A) = \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F) \leq \sum_{k=1}^{\infty} P(A_k).$$

Wir zeigen nun ' $\geq$ '.

**Idee.** Wir können  $P(A_k) = \sup_{\substack{F_k \subseteq A_k \\ |F_k| < \infty}} P(F_k)$  schreiben und

'optimieren' nun jedes einzelne  $F_k$ .

Seien also  $F_k \subseteq A_k$  jeweils endlich. Dann ist  $F_k \cap F_l \subseteq A_k \cap A_l = \emptyset$ , also sind auch die  $F_k$  paarweise disjunkt, und wir lernen

$$\sum_{k=1}^n P(A_k) = \sum_{k=1}^n \sup_{\substack{F_k \subseteq A_k \\ |F_k| < \infty}} P(F_k) = \sup_{\substack{F_1 \subseteq A_1 \\ \text{abs } F_1 < \infty}} \dots \sup_{\substack{F_k \subseteq A_k \\ |F_k| < \infty}} \sum_{k=1}^n P(F_k) \quad (6)$$

Also ist

$$\sum_{k=1}^n P(F_k) = P\left(\bigcup_{k=1}^n F_k\right) \leq P\left(\bigcup_{k=1}^{\infty} F_k\right) \leq P\left(\bigcup_{k=1}^{\infty} A_k\right) \stackrel{\text{def}}{=} P(A).$$

setzen wir dies nun in die rechte Seite von (1) ein, so ergibt sich

$$\sum_{k=1}^n P(A_k) \leq P(A) \quad \Rightarrow \quad \sum_{k=1}^{\infty} P(A_k) \leq P(A).$$

□

Beweis von **Satz 1.7.** (a) Es gilt

$$\sum_{\omega \in \Omega} p(\omega) = P(\Omega) = 1.$$

nach Voraussetzung. Die  $\sigma$ -Additivität folgt nun aus **Lemma 1.8.**  
Deswegen ist  $P(A)$  eine Wahrscheinlichkeitsverteilung.

(b) Da  $P$   $\sigma$ -additiv ist, ist  $\forall A \subseteq \Omega$ :

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

und dies hat genau die angegebene Form mit  $p(\omega) := P(\{\omega\})$

□

## 1.4 Die Gleichverteilung

Sei  $\Omega$  endlich ( $\neq \emptyset$ ) und betrachte  $\sigma$ -Algebra  $\mathcal{F} = \mathcal{P}(\Omega)$ .

Die **Gleichverteilung** ist die Wahrscheinlichkeitsverteilung, die ein uniformes "Gewicht" (Massenfunktion) auf die Elementarereignisse verteilt:

$$\forall \omega \in \Omega : p(\omega) = \mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}.$$

Aus **Satz 1.7** folgt dann bereits, dass  $\forall A \subseteq \Omega$ :

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega) = \frac{|A|}{|\Omega|}.$$

**Beispiel.** (a) Betrachte  $n$  Würfe eines fairen Würfels. In diesem Fall ist  $\Omega = \{1, 2, 3, 4, 5, 6\}^n = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_k \in \{1, \dots, 6\}\}$  und somit  $|\Omega| = 6^n$  und die Gleichverteilung ist gegeben durch

$$\mathbb{P}(\omega) = \frac{1}{6^n}.$$

(b) (Zufällige Permutationen).

- Eine Permutation  $\sigma \in \mathfrak{S}_n$  von  $\{1, \dots, n\}$  ist eine Abbildung von  $\{1, \dots, n\}$  nach  $\{1, \dots, n\}$ , die bijektiv ist. Oft schreiben wir

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}.$$

und meinen damit  $\sigma(1) = 4, \sigma(2) = 3, \sigma(3) = 1, \sigma(4) = 2$ .  
Manschal schreiben wir dann auch

$$\sigma = (4, 3, 1, 2) = (\sigma_1, \sigma_2, \sigma_3, \sigma_4).$$

- Sei  $\Omega = \mathfrak{S}_n$  die Menge aller Permutation von  $\{1, \dots, n\}$ . Dann ergibt sich

$$|\mathfrak{S}_n| = n!.$$

Also ergibt sich für die Gleichverteilung eine Wahrscheinlichkeit von

$$\mathbb{P}(\sigma) = \frac{1}{n!} \quad \forall \sigma \in \mathfrak{S}_n.$$

**Beispiel.** Sei  $N$  die Anzahl von Karten eines Kartenspiels, die gut gemischt sind, d.h. jede Reihenfolge ist gleich wahrscheinlich.

- (1) Was ist die Wahrscheinlichkeit, dass die  $k$ -te Karte auf der  $l$ . Stelle ist? D.h, was ist:

$$\mathbb{P}(\{\omega \in \mathfrak{S}_n \mid \omega(k) = l\}).$$

*Lösung.* Es ergibt sich

$$\mathbb{P}(\{\omega \in \mathfrak{S}_n \mid \omega(k) = l\}) = \frac{|\{\omega \in \mathfrak{S}_n \mid \omega(k) = l\}|}{|\Omega|} = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

□

- (2) Was ist die Wahrscheinlichkeit, dass eine Karte 'auf ihrer Stelle' ist, dh.

$$\mathbb{P}(\{\omega \mid \exists k: \omega(k) = k\}).$$

*Lösung.* Definiere die Ereignisse  $A_k := \{\omega(k) = k\}$ . Diese sind nicht disjunkt für verschiedene  $k$ . Es ergibt sich:

$$\begin{aligned} \mathbb{P}(\exists k: \omega(k) = k) &= \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) \\ &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \underbrace{\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})}_{= \frac{(n-k)!}{n!}} \\ &= \sum_{k=1}^n (-1)^{k-1} \frac{(n-k)!}{n!} \underbrace{\sum_{1 \leq i_1 < \dots < i_k \leq n} 1}_{= \binom{n}{k}} \\ &= \sum_{k=1}^n (-1)^{k-1} \frac{(n-k)!}{n!} \cdot \frac{n!}{(n-k)!k!} \\ &= - \sum_{k=1}^n \frac{(-1)^k}{k!} \\ &= 1 - \frac{1}{e} + \sum_{k=n+1}^{\infty} \frac{(-1)^k}{k!} \end{aligned} \tag{7}$$

Für  $n \rightarrow \infty$  geht das gegen  $1 - \frac{1}{e} \in (0, 1)$ . □

**Bemerkung\*.** Wir verwenden hier die Exponentialreihe, d.h.

$$e^x := \exp(x) := \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

konvergiert absolut und auf ganz  $\mathbb{R}$ , insbesondere für  $x = -1$  gegen  $\frac{1}{e}$

## 1.5 Die empirische Verteilung

Diese wird aus den Beobachtungen definiert.

**Definition** (Empirische Verteilung). Seien  $x_1, x_2, \dots, x_n \in \Omega$   $n$  Beobachtungen. Setze

$$N(A) := |\{k \in \{1, \dots, n\} \mid x_k \in A\}|.$$

Denn definiert

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

die **empirische Häufigkeit** von  $A$ .  $\mathbb{P}$  nennen wir die **empirische Verteilung**. Zudem ist

$$p(\omega) = \frac{N(\{\omega\})}{n}.$$

die **relative Häufigkeit** von  $\omega \in \Omega$ .

**Beispiel.** Die empirische Verteilung von  $n$  Zufallswürfeln eines Würfels wird gegeben durch  $x_1, \dots, x_n \in \{1, \dots, 6\}$ . Die Plots für  $p_k := \frac{N(k)}{n}$  für verschieden  $n$  sehen wie folgt aus:

Plots einfügen

## 1.6 Zufallsvariablen

Wir werden Funktionen der Ergebnisse betrachten:

**Definition 1.9** (Diskret Zufallsvariable). Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum. Eine **diskrete Zufallsvariable** ist eine messbare Abbildung

$$X : \Omega \longrightarrow \mathcal{S}.$$

mit  $\mathcal{S}$  abzählbar (denke: 'diskret').  
Messbar bedeutet hierbei, dass

$$\forall s \in \mathcal{S}: X^{-1}(s) = \{\omega \in \Omega \mid X(\omega) = s\} \in \mathcal{F}.$$

**Notation.** Wir schreiben auch kurz:

$$X^{-1}(s) = \{X(\omega) = s\} = \{X = s\}.$$

**Definition 1.10.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum. und  $X : \Omega \rightarrow \mathcal{S}$  eine diskrete Zufallsvariable.

Die **Verteilung von  $X$**  ist die Wahrscheinlichkeitsverteilung  $\mu_X$  auf  $\mathcal{S}$ ,  $\mathcal{P}(\mathcal{S}) = \tilde{\mathcal{F}}$ , s.d.  $\forall B \in \tilde{\mathcal{F}}: \mu_X(B) := \mathbb{P}(X^{-1}(B))$ .  
 $\mu_X$  hat eine **Massenfunktion**

$$p_X(s) := \mathbb{P}.$$

**Beispiel** (Werfen von  $n$  Münzen). Betrachte folgende Situation:



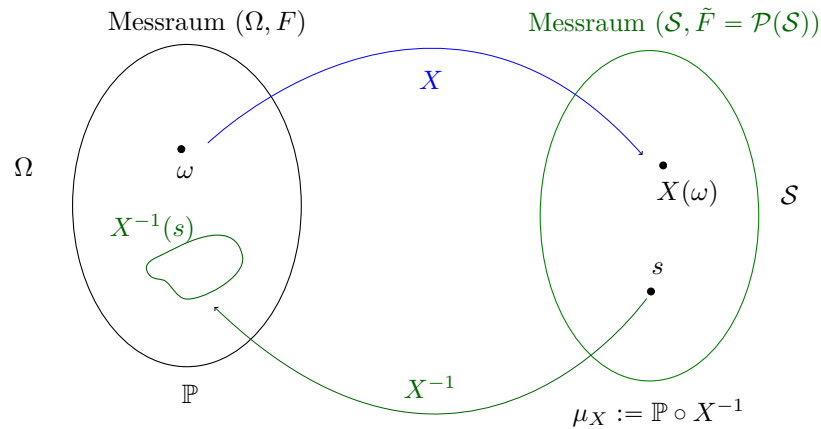


Abbildung 2: Diskrete Zufallsvariable

- Sei  $\Omega = \{\omega = (\omega_1, \dots, \omega_n \mid \omega_i \in \{0, 1\} \text{ für } 1 \leq i \leq n)\}$  wobei

$$\omega_k = \begin{cases} 0 & \text{falls } k\text{-ter Wurf ist Zahl} \\ 1 & \text{falls } k\text{-ter Wurf ist Kopf} \end{cases}.$$

- $\mathcal{F} = \mathcal{P}(\Omega)$  und  $\mathbb{P}$  die Gleichverteilung

(1) Setze

$$X_k : \begin{cases} \Omega & \longrightarrow \mathcal{S} = \{0, 1\} \\ \omega & \longmapsto \omega_k \end{cases}$$

für  $k = 1, \dots, n$ . Dies ist eine diskrete Zufallsvariable mit Verteilung  $\mu_{X_k}$  mit

$$p_{X_k}(s) = \mathbb{P}(X_k = s) = \frac{2^{n-1}}{2^n} = \frac{1}{2}.$$

Wir sehen also, dass  $X_k$  gleichverteilt ist.

(2) Definiere

$$Y : \begin{cases} \Omega & \longrightarrow \mathcal{S} := \{0, 1, \dots, n\} \\ \omega & \longmapsto \omega_1 + \dots + \omega_n \end{cases}$$

d.h.

$$Y(\omega) = \# \{\text{geworfene Köpfe}\}.$$

Es hat nun  $\mu_Y$  die Massenfunktion:

$$p_Y(k) = \frac{1}{2^n} |\{\omega \mid \omega_1 + \dots + \omega_n = k\}| = \frac{\binom{n}{k}}{2^n}.$$

Diese Verteilung sieht wie folgt aus:

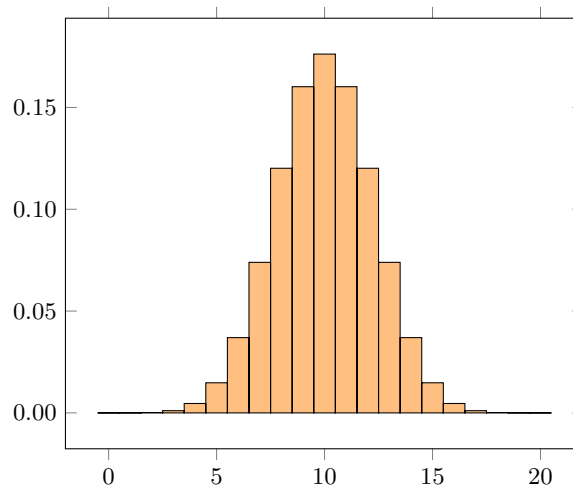


Abbildung 3: Massenfunktion  $p_Y(k)$   
Diese sind Sonderfälle der Bernoulliverteilung und der Binomialverteilung

### 1.6.1 Die Bernoulli-Verteilung

**Definition 1.11** (Bernoulli-Verteilung). Sei  $p \in [0, 1]$  gegeben. Die Wahrscheinlichkeitsverteilung auf  $\{0, 1\}$  mit Massenfunktion

$$p(k) = \begin{cases} p & k = 1 \\ 1 - p & k = 0 \end{cases}.$$

heißt **Bernoulli-Verteilung mit Parameter  $p$** .

**Notation.** Wir notieren auch  $\text{Ber}(p)$  für die Bernoulli-Verteilung mit Parameter  $p$ .

**Beispiel.** (a) Eine Münze, die mit Wahrscheinlichkeit  $p$  Kopf zeigt. Hier ist

$$\Omega = \{\text{Zahl}, \text{Kopf}\} \quad \mathbb{P}(\text{Kopf}) = p = 1 - \mathbb{P}(\text{Zahl}).$$

Sei

$$X(\omega) = \begin{cases} 1 & \omega = \text{Kopf} \\ 0 & \omega = \text{Zahl} \end{cases}.$$

Dann ist  $\mathbb{P}(X = 1) = p$  und  $X \sim \text{Ber}(p)$ .

**Notation.** Wir schreiben  $X \sim \text{Ber}(p)$ , wenn  $X$  die Verteilung  $\text{Ber}(p)$  hat.

(b) In einer Urne befinden sich  $n$  blaue Kugeln und  $m$  rote Kugeln. Wir ziehen eine Kugel aus der Urne (Annahme: Gleichverteilung).

lung). Dann ist

$$\Omega = \{\omega = (\omega_1, \dots, \omega_{n+m}) \mid \omega_i \in \{\text{blau}, \text{rot}\} \text{ mit } n \text{ mal blau}\}.$$

Setze  $\mathcal{F} = \mathcal{P}(\Omega)$  und wähle  $\mathbb{P}$  als die Gleichverteilung. Betrachte

$$X(\omega) = \begin{cases} 1 & \text{falls } \omega_i \text{ ist blau} \\ 0 & \text{sonst} \end{cases}.$$

Diese hat also die Verteilung

$$\mathbb{P}(X = 1) = \frac{\binom{m+n-1}{n-1}}{\binom{m+n}{n}} = \frac{(m+n-1)!}{(n-1)!m!} \frac{n!m!}{(m+n)!} = \frac{n}{m+n}.$$

Also ist  $X \sim \text{Ber}\left(\frac{n}{n+m}\right)$

### 1.6.2 Die Binomial-Verteilung

**Definition 1.12** (Binomial-Verteilung). Seien  $n \in \mathbb{N}$  und  $p \in [0, 1]$  gegeben. Die Wahrscheinlichkeitsverteilung auf  $\{0, 1, \dots, n\}$  mit Massenfunktion

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

für  $k = 0, \dots, n$  heißt **Binomialverteilung mit Parametern  $n$  und  $p$** .

**Notation.** Wir notieren  $\text{Bin}(n, p)$  für die Binomialverteilung mit Parametern  $n$  und  $p$ .

**Beispiel** (Ziehen mit Zurücklegen). • Seien  $m$  Kugeln in einer Urne, davon  $p \cdot m \in \mathbb{N}$  weiße Kugeln und  $(1-p)m$  schwarze Kugeln.

- Wir ziehen eine Kugel, notieren uns die Farbe und legen sie wieder zurück.
- Wir mischen die übrigen Kugeln wieder gut
- Wir wiederholen die vorherigen Schritte, bis wir  $n$  Ziehungen durchgeführt haben.
- Dies modellieren wir durch

$$\Omega = \{0, 1\}^n.$$

wobei  $\omega = (\omega_1, \dots, \omega_n) \in \Omega$  gegeben ist durch

$$\omega_i = \begin{cases} 1 & \text{falls Farbe der } i\text{-ten Kugel weiß ist} \\ 0 & \text{sonst} \end{cases}.$$

- Sei nun  $X(\omega) = \sum_{k=1}^n \omega_k = \# \{\text{weiße Kugeln}\}$

Dann behaupten wir, dass  $X \sim \text{Bin}(n, p)$ . In der Tat:

$$\begin{aligned} \frac{|\{\omega \in \Omega \mid X(\omega) = l\}|}{|\Omega|} &= \frac{\binom{n}{l} \cdot (pm)^l \cdot ((1-p)m)^{n-l}}{m^n} \\ &= \frac{\binom{n}{l} p^l (1-p)^{n-l} \cdot m^n}{m^n} = \binom{n}{l} p^l (1-p)^{n-l} \end{aligned} \quad (8)$$

**Bemerkung.** Wir haben hier den Begriff der **Unabhängigkeit** genutzt, den wir nun genauer kennenlernen wollen.

**Definition 1.13** (Unabhängige Ereignisse). Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum. Die Ereignisse  $E_1, E_2, \dots, E_n$  heißen **unabhängig**, falls

$$\mathbb{P}(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_k}) = \prod_{l=1}^k \mathbb{P}(E_{i_l}).$$

für alle  $2 \leq k \leq n$  und  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ .

**Beispiel.** • Betrachte zwei Würfelwürfe, d.h.  $\Omega = \{1, \dots, 6\}^2$  und notiere  $\omega = (\omega_1, \omega_2)$ . Dann können wir

$$E_1 = \{\omega_1 = 3\} \quad E_2 = \{\omega_2 \geq 4\}.$$

betrachten. Wir rechnen nach, dass

$$\mathbb{P}(\omega_1 = 3 \cap \omega_2 \geq 4) = \frac{3}{36} = \frac{1}{6} \cdot \frac{3}{6} = \mathbb{P}(\omega_1 = 3) \cdot \mathbb{P}(\omega_2 \geq 4).$$

also sind die beiden Ereignisse unabhängig voneinander. Das macht auch semantisch Sinn, weil wir durch das Ergebnis des einen Würfelwurfs keine Informationen über das Ergebnis des zweiten Würfelwurfs erhalten.

- Falls  $E_1, E_2, \dots, E_n$  unabhängige Ereignisse sind, mit  $\mathbb{P}(E_i) = p$  für  $1 \leq i \leq n$ , dann ist

$$\mathbb{P}(\text{genau } k \text{ der Ereignisse treten ein}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Dies rechnen wir nach. Setze hierzu

$$A_{(i_1, \dots, i_k)} = \{\omega \in \Omega \mid E_{i_1}, \dots, E_{i_k} \text{ treten ein, die anderen nicht}\}$$

Dann ist

$$\tilde{A} = \bigsqcup_{1 \leq i_1 < \dots < i_k \leq n} A_{i_1, \dots, i_k}.$$

eine disjunkte Vereinigung, also erhalten wir

$$\begin{aligned}
 \mathbb{P}(\tilde{A}) &= \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{(i_1, \dots, i_k)}) \\
 &\stackrel{\text{unabhängig}}{=} \sum_{1 \leq i_1 < \dots < i_k \leq n} \prod_{j \in \{i_1, \dots, i_k\}} \mathbb{P}(E_j) \cdot \prod_{l \notin \{i_1, \dots, i_k\}} \mathbb{P}(E_l^c) \\
 &= \sum_{1 \leq i_1 < \dots < i_k \leq n} p^k (1-p)^{n-k} \\
 &= \binom{n}{k} p^k (1-p)^{n-k}
 \end{aligned} \tag{9}$$

**Bemerkung.** Strenggenommen haben wir in der letzten Rechnung verwendet, dass mit  $E_1, \dots, E_n$  unabhängig auch  $F_1, \dots, F_n$  für  $F_i = E_i$  oder  $F_i = E_i^c$  unabhängig voneinander sind. Dies müssten wir noch einmal nachrechnen, dazu für den Fall  $n = 2$  ist z.B:

$$\mathbb{P}(E_1 \cap E_2^c) + \mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1 \cap (E_2 \cup E_2^c)) = \mathbb{P}(E_1).$$

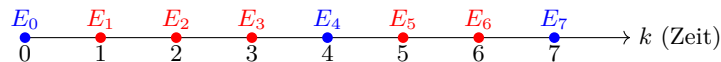
Also ergibt sich

$$\mathbb{P}(E_1 \cap E_2^c) = \mathbb{P}(E_1) - \mathbb{P}(E_1) \mathbb{P}(E_2) = \mathbb{P}(E_1)(1 - \mathbb{P}(E_2)) = \mathbb{P}(E_1) \mathbb{P}(E_2^c).$$

wie zu zeigen war.

### 1.6.3 Die Poisson-Verteilung

Betrachte Ereignisse  $E_1, \dots, E_n$ , die unabhängig voneinander sind und jeweils Wahrscheinlichkeit  $p$  haben, eintreten. Wir skizzieren dies mit  $\bullet$  für ein eingetretenes Ereignis und  $\bullet$  für das Nicht-Eintreten des entsprechenden Ereignisses

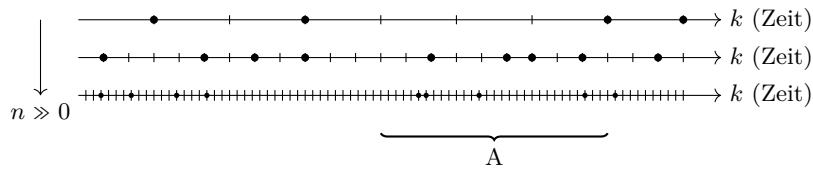


**Frage.** Was passiert wenn  $n \gg 1$ , d.h. wie viele Erfolge werden wir (ca.) unter diesen  $n$  Ereignissen haben?

Typischerweise haben wir dann  $\mathcal{O}(pn)$  Erfolge in  $E_1, \dots, E_n$ . Die Erfolgswahrscheinlichkeit  $p$  hänge nun von  $n$  ab, d.h.  $p = p(n)$ . Wir wollen den Erwartungswert  $p \cdot n$  festhalten und  $n$  groß werden lassen, d.h. sei  $\lambda \in (0, \infty)$  so, dass

$$\lim_{n \rightarrow \infty} p(n) \cdot n = \lambda.$$

Wir können uns das ganze so vorstellen, dass wir in einem Zeitintervall von 1 erwarten, dass  $\lambda$  Ereignisse eintreten, und wir nun mit einem kleinen Zeitintervall  $\delta = \frac{1}{n}$  für große  $n$  das kontinuierliche Zeitintervall durch  $n$  unabhängige Ereignisse ersetzen, die jeweils mit Wahrscheinlichkeit  $\frac{\lambda}{n}$  eintreten, und uns nun fragen, wie wahrscheinlich es also ist, dass wir eine gewisse Anzahl an Ereignissen im Zeitintervall beobachten.



Es stellt sich also

**Frage.** Was ist für ein Zeitintervall  $A$  die Wahrscheinlichkeit, dass wir im Intervall genau  $k$  Ereignisse beobachten, d.h. was ist

$$\lim_{n \rightarrow \infty} \mathbb{P}(\exists k \text{ Erfolge in } A).$$

?

- Sei  $p = p(n)$  sodass  $\lim_{n \rightarrow \infty} pn = \lambda \in (0, \infty)$
- Wähle Zeiteinheit  $\delta = \frac{1}{n}$

Die Antwort gibt folgender

**Satz 1.14.** Sei  $\lambda \in (0, \infty)$ . Dann ist

$$\lim_{n \rightarrow \infty} \text{Bin}\left(n, \frac{\lambda}{n}\right)(k) = \frac{e^{-\lambda} \lambda^k}{k!}.$$

für  $k = 0, 1, 2, \dots$

*Beweis.* Sei  $k$  fest. Dann ist

$$\begin{aligned} \text{Bin}\left(n, \frac{\lambda}{n}\right)(k) &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \underbrace{\frac{n(n-1) \cdots (n-k+1)}{n^k}}_{\rightarrow 1} \frac{\lambda^k}{k!} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow (1-0)^{-k} = 1} \\ &= \frac{\lambda^k}{k!} \cdot e^{-\lambda} \end{aligned} \tag{10}$$

□

**Definition 1.15** (Poisson-Verteilung). Sei  $\lambda \in (0, \infty)$  fest. Die Wahrscheinlichkeitsverteilung auf  $\{0, 1, 2, \dots\}$  mit Massenfunktion

$$p(k) = \frac{e^{-\lambda} \lambda^k}{k!}.$$

heißt **Poisson-Verteilung mit Parameter  $\lambda$** .

**Notation.** Wir schreiben auch  $\text{Poi}(\lambda)$  für die Poisson-Verteilung zum Parameter  $\lambda$ .

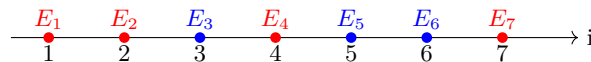
### 1.6.4 Die geometrische Verteilung

- Seien  $E_1, E_2, \dots$  unabhängige Ereignisse mit

$$\mathbb{P}(E_i \text{ tritt ein}) = 1 - q$$

$$\mathbb{P}(E_i \text{ tritt nicht ein}) = q$$

Wir verwenden wieder  $\bullet$  um das Eintreten eines Ereignisses zu notieren, und  $\bullet$  für das Gegenteil.



Setze nun  $T_0 = 0$  sowie  $T_{l+1} = \min \{i > T_l : E_i \text{ tritt ein}\}$  als die Eintrittszeitpunkte des  $k$ -ten Erfolgs. Dann definieren

$$N_l := T_l - T_{l-1} - 1 = \# \{ \bullet \text{ zwischen } T_{l-1} \text{ und } T_l \}.$$

die Abstände zwischen den jeweiligen Erfolgen. Es ergibt sich:

$$\begin{aligned} \mathbb{P}(N_l = k) &= \mathbb{P}(N_1 = k) = \mathbb{P}(T_1 = k + 1) \\ &= \mathbb{P} \left( \bigcap_{i=1}^k \{E_i \text{ tritt nicht ein}\} \cap \{E_{k+1} \text{ tritt ein}\} \right) \\ &= \left( \prod_{i=1}^k \mathbb{P}(E_i \text{ tritt nicht ein}) \right) \cdot \mathbb{P}(E_{k+1} \text{ tritt ein}) \\ &= (1 - q)q^k \end{aligned} \tag{11}$$

für  $k \geq 0$ . Das motiviert nun

**Definition 1.16** (Geometrische Verteilung). Sei  $q \in [0, 1)$ . Die Wahrscheinlichkeitsverteilung auf  $\{0, 1, 2, \dots\}$  mit Massenfunktion

$$p(k) = (1 - q)q^k.$$

heißt **geometrische Verteilung** mit Parameter  $q$ .

**Notation.** Wir notieren hierfür  $\text{Geo}(q)$

**Warnung.** Eine andere Version setzt  $p(k) = (1 - q)q^{k-1}$ , wobei  $k = 1, 2, \dots$ . Hier ist nur der Index verschoben. Wenn nicht anders genannt, ist für uns aber immer obige Definition gemeint.

## 1.7 Simulation von Gleichverteilung

**Frage.** Wie werden Zufallsvariablen simuliert?

Typischerweise benutzen wir folgende Situation:

**Input** Zahl(en), z.B. Redinerzeit

**Output** 'Zufällige Zahl' in  $\{0, \dots, m - 1\}$

### 1.7.1 Lineare Kongruenzgeneratoren (LCG)

**Startwert**  $x_0 \in \mathbb{N}$  gegeben.

**Parameter**  $a, c, m \in \mathbb{N}$

**Schritt** Setze  $x_{n+1} := (a \cdot x_n + c) \bmod m$ .  
Dieses Vorgehen produziert eine scheinbar zufällige Folge.

**Beispiel.** Es gibt folgende LCGs:

	m	a	c
ZX81	$2^{16} + 1$	75	0
RANDN	$2^{31}$	65539	0
Marsaglia	$2^{32}$	69069	1

**Beispiel** (Eine schlechte Wahl). Wenn wir  $a = 4, c = 1, m = 31$  wählen sowie  $x_0 = 3$ , so erreichen wir Periode 9, und somit werden nicht alle Zahlen erreichen / generieren.

Plot einfügen

**Lemma 1.17** (Knuth). Die Periode eines LCG ist gleich  $m$ , genau dann, wenn

- (a)  $c$  und  $m$  haben keine gemeinsamen Primfaktoren
- (b) Jeder Primfaktor von  $m$  ist ein Teiler von  $a - 1$
- (c) Falls  $4 \mid m$ , dann  $4 \mid a - 1$ .

*Beweis.* Kein Beweis. □

Beispiele von LCG's einfügen

### 1.7.2 Zufallsvariablen aus $[0, 1)$

- Sei  $(x_n)_{n \geq 1}$  eine Folge von (Pseudo)zufallszahlen aus  $\{0, 1, \dots, m-1\}$ .  
Dann ist

$$u_n := \left( \frac{x_n}{m} \right)_{n \geq 1}.$$

eine Folge von Pseudozahlen in  $[0, 1)$ . Gut ist aber nur der Fall, wenn  $m \approx 10^N$ , wobei  $N = \text{Rechnergenauigkeit, d.h. \#Ziffern}$ .

### 1.7.3 Zufallspermutationen

**Frage.** Wie erzeugt man eine gleichverteilte Permutation von  $\{1, \dots, N\}$ ?

---

#### Algorithmus 1.18 : Zufallspermutationen

---

**Eingabe :** Möglichkeit, aus endlicher Menge gleichverteilt zufällige Zahlen zu ziehen

**Ausgabe :** Eine zufällige Permutation von  $\{1, \dots, N\}$

```

Setze  $\sigma_0 := \{1, \dots, N\}$ 
for  $i = 1$  to  $n - 1$  do
    wähle  $k \in \{i, \dots, N\}$  gleichverteilt
    Setze  $\sigma_k := \sigma_{k-1} \circ \tau_{i,k}$ 
    
```

---

**Lemma 1.19.** Der Algorithmus erzeugt eine zufällige gleichverteilte Permutation.



*Beweis.* Der Algorithmus benutzt eine Gleichverteilung auf

$$\Omega_n := \{1, \dots, N\} \times \{2, \dots, n\} \times \{n-1, n\}.$$

Für  $\omega = (w_1, \dots, w_{N-1}) \in \Omega_N$  ist

$$\sigma(\omega) = \tau_{N-1, \omega_{N-1}} \circ \dots \circ \tau_{1, w} \circ \underbrace{(1, \dots, N)}_{\sigma_0}.$$

Es genügt also zu zeigen, dass  $\sigma : \Omega_N \rightarrow \mathcal{S}_N$  eine Bijektion ist. Wir sehen:

- (a)  $|\Omega_N| = |\mathcal{S}_N| = N!$
- (b) Sei  $w \neq \tilde{w}$  und setze  $k = \min \{j \mid \omega_j \neq \tilde{\omega}_j\}$ . Dann ist  $\sigma(\omega)_k \neq \sigma(\tilde{\omega})_k$  und somit ist die Funktion injektiv

Damit ist die Abbildung sogar bijektiv und wir sind fertig.  $\square$

### 1.7.4 Geometrische Verteilung

Sei  $X \sim \text{Geo}(q)$ , d.h.

$$\mathbb{P}(X = k) = (1 - q)q^k.$$

**Frage.** Wie simuliert man nun  $X$ ?

Erzeuge zunächst  $n \sim U[0, 1)$  als gleichverteilte Zufallsvariable auf  $[0, 1)$ .

Sei  $T_k := \mathbb{P}(X < k)$ . Falls  $n \in [T_k, T_{k+1})$ , dann setze  $X = k$ . Wir berechnen also

$$\begin{aligned} T_k &= \mathbb{P}(X < k) = 1 - \mathbb{P}(X \geq k) \\ &= 1 - \sum_{x \geq k} (1 - q)q^x \\ &= 1 - q^k \end{aligned} \tag{12}$$

Wegen

$$u = 1 - q^x \Leftrightarrow x = \frac{\ln(1 - u)}{\ln(q)}.$$

erhalten wir also für  $k$  den Wert

$$k := \left\lfloor \frac{\ln(1 - u)}{\ln(q)} \right\rfloor = \left\lfloor \frac{\ln(\tilde{u})}{\ln(q)} \right\rfloor.$$

wobei wir ebenfalls  $\tilde{u} \sim U[0, 1)$  gleichverteilen.

## 1.8 Erwartungswert und Varianz

- Sei  $X$  eine reellwertige diskrete Zufallsverteilung. Sei

$$X : \Omega \rightarrow \mathcal{S} \subseteq \mathbb{R}.$$

eine diskrete Zufallsvariable, d.h.  $\mathcal{S}$  abzählbar.

**Definition 1.20** (Empirischer Mittelwert). Seien  $x_1, \dots, x_n \in \mathcal{S}$   $n$  Beobachtungen einer Zufallsvariable  $X$ . Der **empirische Mittel-**

wert ist durch

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

definiert.

**Ziel.** Wir wollen eine Sorte von Mittelwert definieren, der nur von  $X$  abhängig ist, und nicht von den Beobachtungen.

Folgende Forderungen ergeben sich an solch einen Mittelwert:

- Falls  $X(\omega) = x$  für jedes  $\omega$ , dann muss der Mittelwert von  $X$  gleich  $x$  sein.
- Jeder Wert  $x \in \mathcal{S}$  muss bezüglich der Massenfunktion  $p_X(x)$  gewichtet sein.

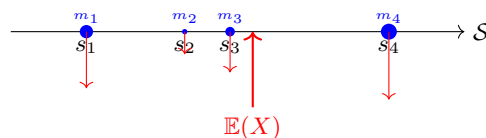
Das motiviert folgende

**Definition 1.21** (Erwartungswert). Der **Erwartungswert** von  $X$  bzgl.  $\mathbb{P}$  ist durch

$$\mathbb{E}(X) = \sum_{s \in \mathcal{S}} s \cdot \mathbb{P}(X = s) = \sum_{s \in \mathcal{S}} s \cdot p_X(s).$$

definiert. Dies ist wohldefiniert, falls die Reihe absolut gegen einen Wert  $< \infty$  konvergiert.

Stellen wir uns  $\mathcal{S}$  als Balken vor und tragen an den Stellen  $s_i \in \mathcal{S}$  jeweils das Gewicht  $m_i := p_X(s_i)$  an, so ist  $\mathbb{E}(X)$  derjenige Punkt, mit dem der Balken ausbalanciert ist:



**Bemerkung.** Nicht alle Wahrscheinlichkeitsverteilungen besitzen einen endlichen Mittelwert, das zeigt folgendes

**Beispiel.** Sei  $X$  auf  $\{1, 2, \dots\}$  verteilt mit

$$\mathbb{P}_X(s) = \frac{6}{\pi^2 s^2}.$$

dann ergibt sich für den Erwartungswert:

$$\mathbb{E}(X) = \sum_{s \geq 1} s \cdot \frac{6}{\pi^2 s^2} = \frac{6}{\pi^2} \cdot \sum_{s \geq 1} \frac{1}{s} \rightarrow \infty.$$

Vorlesung 6  
Sa 01 Mai 2021 09:18

**Beispiel** (Zufallsvariablen mit Werten in  $\{0, 1\}$ ). Sei  $A \in \mathcal{F}$  ein Ereignis, und definiere  $X$  durch

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

Dann ist

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

*Beweis.* Nach Definition ist

$$\begin{aligned}\mathbb{E}(X) &= 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) \\ &= \mathbb{P}(A)\end{aligned}$$

□

**Beispiel** (Binomialverteilung). Sei  $X \sim \text{Bin}(n, p)$ . Wir wollen zeigen, dass  $\mathbb{E}(X) = p \cdot n$ .

*Beweis.*

$$\mathbb{E}(X) = \sum_{k=0}^n \underbrace{\binom{n}{k} p^k (1-p)^{n-k}}_{=\mathbb{P}(X=k)} \cdot k.$$

Wir wollen nun allgemein, in Anlehnung an die Binomialformel, den Wert von

$$\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \cdot k.$$

berechnen. Dazu stellen wir fest, dass

$$p \cdot \frac{d}{dp} \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = p \cdot \sum_{k=0}^n \binom{n}{k} p^{k-1} q^{n-k} \cdot k.$$

unser Ausdruck ist, also suchen wir

$$p \cdot \frac{d}{dp} (p+q)^n = p \cdot (p+q)^{n-1} \cdot n.$$

Nun können wir  $q = 1-p$  auf beiden Seiten setzen, und somit erhalten wir

$$\mathbb{E}(X) = p \cdot (p + (1-p))^{n-1} \cdot n = p \cdot n.$$

wie gewünscht. □

**Beispiel** (Poisson-Verteilung). Sei  $X \sim \text{Poi}(\lambda)$ , dann behaupten wir, dass  $\mathbb{E}(X) = \lambda$ .

*Beweis.*

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k \geq 0} \frac{e^{-\lambda} \lambda^k}{k!} \cdot k \\ &= e^{-\lambda} \cdot \lambda \cdot \underbrace{\sum_{k \geq 1} \frac{\lambda^{k-1}}{(k-1)!}}_{\rightarrow e^{\lambda}} \\ &= \lambda\end{aligned}$$

□

**Bemerkung\*.** Diese Feststellung passt auch zur Konstruktion der Poisson-Verteilung.

**Bemerkung.** Oft kann man definieren

$$\psi(z) := \sum_{k \in \mathcal{S}} p(k) z^k.$$

Dann werden wir uns

$$\frac{d}{dz} \psi(z) = \sum_{k \in \mathcal{S}} p(k) k z^{k-1}$$

ansehen und bei  $z = 1$  evaluieren, um  $\mathbb{E}(X)$  zu berechnen.

Das ganze funktioniert, wenn  $X$  durch  $\mathbb{P}(X = k) = p(k)$  verteilt ist und natürlich nur, wenn alle Objekte wohldefiniert sind.

Wir wollen auch Funktionen von  $X$  betrachten.

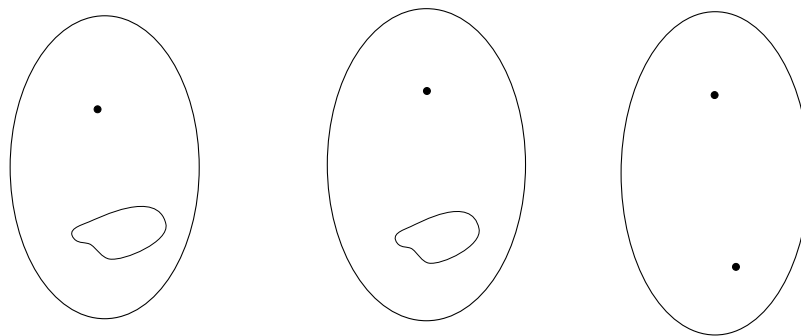


Abbildung 4: Funktionen von Zufallsvariablen

**Satz 1.22** (Transformationssatz). Sei  $X : \Omega \rightarrow \mathcal{S}$  eine diskrete Zufallsvariable und  $f : \mathcal{S} \rightarrow \mathbb{R}$  eine Funktion. Dann ist  $f(X) := f \circ X : \Omega \rightarrow \mathbb{R}$  auch eine Zufallsvariable und

$$\mathbb{E}(f(X)) = \sum_{s \in \mathcal{S}} f(s) \mathbb{P}(X = s).$$

falls die Summe wohldefiniert ist.

*Beweis.* Messbarkeit: Es ist

$$\{f(X) = a\} = \bigcup_{s \in f^{-1}(a)} \{X = s\} \in \mathcal{F}.$$

weil

$$\{\omega \mid X(\omega) = s\} \in \mathcal{F}.$$

da  $X$  eine Zufallsvariable ist. Nach Definition ist nun

$$\begin{aligned}
 \mathbb{E}(f(X)) &= \sum_{a \in f(\mathcal{S})} a \cdot \mathbb{P}(f(X) = a) \\
 &= \sum_{a \in f(\mathcal{S})} a \cdot \mathbb{P}\left(\bigcup_{s \in f^{-1}(a)} \{X = s\}\right) \\
 &= \sum_{a \in f(\mathcal{S})} a \cdot \sum_{s \in f^{-1}(a)} \mathbb{P}(X = s) \\
 &= \sum_{a \in f(\mathcal{S})} \sum_{s \in f^{-1}(a)} f(s) \mathbb{P}(X = s) \\
 &= \sum_{s \in \mathcal{S}} f(s) \mathbb{P}(X = s)
 \end{aligned}$$

□

Der Erwartungswert ist LINEAR und MONOTON:

**Satz 1.23** (Linearität des Erwartungswerts). Seien  $X_1 : \Omega \rightarrow \mathcal{S}_1$  und  $X_2 : \Omega \rightarrow \mathcal{S}_2$  zwei diskrete Zufallsvariablen auf  $(\Omega, \mathcal{F}, \mathbb{P})$ . Falls  $\mathbb{E}(|X_1|) < \infty$  und  $\mathbb{E}(|X_2|) < \infty$ , dann ist  $\forall \lambda_1, \lambda_2 \in \mathbb{R}$ :

$$\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 \mathbb{E}(X_1) + \lambda_2 \mathbb{E}(X_2).$$

*Beweis.* Es ist

$$|\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2)| \leq |\lambda_1| \mathbb{E}(|X_1|) + |\lambda_2| \mathbb{E}(|X_2|) < \infty.$$

(nach Dreiecksungleichung). Nun rechnen wir aus:

$$\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2) = \mathbb{E}(f(X_1, X_2)).$$

wobei  $f(x_1, x_2) = \lambda_1 x_1 + \lambda_2 x_2$ , also

$$\begin{aligned}
 &= \sum_{\substack{x_1 \in \mathcal{S}_1 \\ x_2 \in \mathcal{S}_2}} f(x_1, x_2) \mathbb{P}(X_1 = x_1 \cap X_2 = x_2) \\
 &= \lambda_1 \sum_{x \in \mathcal{S}_1} x_i \sum_{x_2 \in \mathcal{S}_2} \mathbb{P}(X_1 = x_1 \cap X_2 = x_2) + \text{sym.} \\
 &= \lambda_1 \sum_{x \in \mathcal{S}_1} x_i \mathbb{P}(X_1 = x_1) + \text{sym.} \\
 &= \lambda_1 \mathbb{E}(X_1) + \lambda_2 \mathbb{E}(X_2)
 \end{aligned}$$

□

**Korollar 1.24** (Monotonie des Erwartungswerts). Seien  $X_1, X_2$  reellwertige Zufallsvariablen mit  $X_1(\omega) \leq X_2(\omega)$  für alle  $\omega \in \Omega$ . Dann ist

$$\mathbb{E}(X_1) \leq \mathbb{E}(X_2).$$

*Beweis.* Da  $X_2(\omega) - X_1(\omega) \geq 0$ , also ist trivialerweise  $\mathbb{E}(X_2 - X_1) \geq 0$ . Wegen der Linearität ist nun  $\mathbb{E}(X_2 - X_1) = \mathbb{E}(X_2) - \mathbb{E}(X_1)$  und somit sind wir fertig. □

**Beispiel.** Seien  $A_1, A_2, \dots, A_n \in \mathcal{F}$  mit  $\mathbb{P}(A_i) = p$  für alle  $i$ . Sei  $X_i := 1$ . Dann ist  $X_i \sim \text{Ber}(p)$  und  $\mathbb{E}(X_i) = p$ . Sei

$$\mathcal{S}_n := \sum_{i=1}^n X_i.$$

Dann ist

$$\mathbb{E}(\mathcal{S}_n) = \sum_{i=1}^n \mathbb{E}(X_i) = \sum_{i=1}^n p = np.$$

Das ist eine Verallgemeinerung des Falles, in dem  $A_1, \dots, A_n$  unabhängig sind, weil dann  $\mathcal{S}_n \sim \text{Bin}(n, p)$ .

Oft interessieren wir uns auch dafür, wie weit eine Zufallsvariable von ihrem Ursprungswert entfernt ist. Ist z.B.  $\mathbb{P}(X = k) = \mathbb{P}(X = -k)$ , so ist  $\mathbb{E}(X) = 0$ , so sind immer noch folgende Fälle denkbar:

Plot einfügen

**Frage.** Wie weit sind die Werte vom  $X$  Mittelwert ( $\mathbb{E}(X)$ ) entfernt?

Die Antwort liefert die sogenannte Varianz  $\text{Var}(X)$ :

**Definition 1.25** (Varianz). Sei  $X$  eine reellwertige Zufallsvariable auf  $(\Omega, \mathcal{F}, \mathbb{P})$  mit  $\mathbb{E}(X^2) < \infty$ . Die **Varianz** von  $X$  ist durch

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

definiert.

**Lemma\*** (Eigenschaften der Varianz). Es gilt folgendes:

- i)  $\text{Var}(X) \geq 0$ , und  $\text{Var}(X) = 0 \Leftrightarrow \mathbb{P}(X = \mathbb{E}(X)) = 1$ , d.h.  $X(\omega)$  ist in diesem Fall eine Konstante.
- ii) Es ist

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

- iii) Die Varianz hängt nicht vom Erwartungswert ab, d.h.

$$\text{Var}(X) = \text{Var}(X + a) \quad \forall a \in \mathbb{R}.$$

*Beweis\*.* i)  $\text{Var}(X) \geq 0$  ist klar, weil Quadrate nichtnegativ sind, Gleichheit gilt, wenn  $X = \mathbb{E}(X)$  für jedes  $X$ , also wenn  $\mathbb{P}(X = \mathbb{E}(X)) = 1$ , und dann ist  $X(\omega)$  konstant  $\mathbb{E}(X)$ .

- ii) Wir rechnen nach:

$$\begin{aligned} \mathbb{E}((X - \mathbb{E}(X))^2) &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + \mathbb{E}(\mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X) \cdot \mathbb{E}(X) + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \end{aligned}$$

Für die 2. Behauptung ist

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

iii) Es ist

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

□

**Beispiel.** (a) Ist  $X \sim \text{Ber}(p)$ , so ist  $\text{Var}(X) = p(1 - p)$ .

(b) Ist  $X \sim \text{Bin}(n, p)$ , so ist  $\text{Var}(X) = n \cdot p(1 - p)$ .

(c) Ist  $X \sim \text{Poi}(\lambda)$ , so ist  $\text{Var}(X) = \lambda$

(d) Ist  $X \sim \text{Geo}(q)$ , so ist  $\text{Var}(X) = \frac{q}{(1-q)^2}$

*Beweis.* (a) Wir benutzen  $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ . Nun ist

$$\mathbb{E}(X^2) = 0^2 \mathbb{P}(X = 0) + 1^2 \mathbb{P}(X = 1) = p.$$

und somit  $\text{Var}(X) = p - p^2 = p(1 - p)$

(b) Sei  $X \sim \text{Bin}(n, p)$ , wir wissen bereits  $\mathbb{E}(X) = p \cdot n$ . Nun ist

$$\mathbb{E}(X^2) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} k^2.$$

Wir benutzen den gleichen Trick wie vorher nochmal, indem wir feststellen:

$$\begin{aligned} \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k^2 &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k(k-1) + \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k \\ &= p^2 \frac{d^2}{dp^2} (p+q)^n + p \frac{d}{dp} (p+q)^n \\ &= p^2 (p+q)^{n-1} n(n-1) + p(p+q)^{n-1} \cdot n \end{aligned}$$

Einsetzen von  $q = 1 - p$  liefert nun:

$$\mathbb{E}(X^2) = p^2 n(n-1) + p \cdot n = p^2 n^2 - p^2 n + pn.$$

Damit erhalten wir schlussendlich

$$\text{Var}(X) = p^2 n^2 - p^2 n + pn - (pn)^2 = np(1 - p).$$

(c) Ist  $X \sim \text{Poi}(\lambda)$ , so wissen wir bereits  $\mathbb{E}(X) = \lambda$ . Nun ist

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{k \geq 0} k^2 \cdot \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k \geq 0} k(k-1) \frac{e^{-\lambda} \lambda^k}{k!} + \sum_{k \geq 0} k \frac{e^{-\lambda} \lambda^k}{k!} \\ &= \lambda^2 e^{-\lambda} \sum_{k \geq 2} \frac{\lambda^{k-2}}{(k-2)!} + e^{-\lambda} \lambda \sum_{k \geq 1} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda^2 + \lambda \end{aligned}$$

Und damit ergibt sich für die Varianz:

$$\text{Var}(X) = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

□

**Bemerkung.** Sei  $X : \Omega \rightarrow \mathcal{S}$  eine Zufallsvariable. Wir beobachten  $X$   $n$  mal. Der Erwartungswert des Empirischen Masses ist genau der empirische Mittelwert, also für Beobachtungen  $x_1, \dots, x_n$  genau

$$m_n := \frac{1}{n} \sum_{i=1}^n x_i.$$

**Beispiel** (Anwendung). Ein alternativer Beweis des **Einschluss-Ausschluss-Prinzips** lässt sich mit der Linearität des Erwartungswerts führen:

Es ist

$$\begin{aligned} \mathbb{P}((A_1 \cup \dots \cup A_n)^c) &= \mathbb{P}(A_1^c \cap \dots \cap A_n^c) \\ &= E(1_{A_1^c \cap \dots \cap A_n^c}) = (1_{A_1^c} \cdot \dots \cdot 1_{A_n^c}) \\ &= E((1 - 1_{A_1}) \cdot \dots \cdot (1 - 1_{A_n})) \\ &\stackrel{\text{Linearität}}{=} \sum_{k=0}^n (-1)^k \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{E}(1_{A_{i_1}} \cdot \dots \cdot 1_{A_{i_k}}) \\ &= \sum_{k=0}^n (-1)^k \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) \end{aligned}$$

Komplementbildung liefert nun das gewünschte Ergebnis:

$$\begin{aligned} \mathbb{P}(A_1 \cup \dots \cup A_n) &= 1 - \mathbb{P}((A_1 \cap \dots \cap A_n)^c) \\ &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}). \end{aligned}$$

Vorlesung 7  
Mo 03 Mai 2021 10:17

## 2 Bedingte Wahrscheinlichkeit und Unabhängigkeit

### 2.1 Bedingte Wahrscheinlichkeit

**Beispiel.** Es werden statistische Daten über Kleinkinder gemessen: Wann sie krabbeln und wann sie laufen. Betrachten wir die Ereignisse

$$A = \{\text{Kind läuft vor dem 10. Monat}\} \quad B = \{\text{Kind krabbelt vor dem 6. Monat}\}.$$

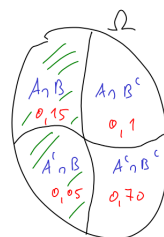
Aus den Daten geht hervor, dass  $\mathbb{P}(A) = 25\%$  und  $\mathbb{P}(B) = 20\%$ .

**Frage.** Sei ein Kind, das mit 6 Monaten krabbelt, gegeben. Wie hoch ist die Wahrscheinlichkeit, dass es mit 10 Monaten schon läuft?

Wir brauchen mehr Information als die obige, um die Frage beantworten zu können! Wir gehen also davon aus, dass wir sogar folgende Daten zur Verfügung haben:



$\cap$	$A$	$A^c$
$B$	0,15	0,05
$B^c$	0,10	0,70



Wir wissen, dass  $B$  eintritt, also befinden wir uns bereits im Zustandsraum  $\Omega_B := \{\omega \in \Omega \mid \omega \in B\}$ . Ziel ist es also, eine neue Massenfunktion  $\mathbb{P}(\cdot|B)$  (auf  $\Omega$ ) zu definieren, die die Information ' $\omega \in B$ ' berücksichtigt. Insbesondere muss also gelten:

$$\mathbb{P}(\omega|B) = 0 \quad \forall \omega \in \Omega_B^c.$$

Zudem wollen wir, dass die Information ' $\omega \in B$ ' dieselbe ist für alle  $\omega \in \Omega_B$ , d.h.

$$p(\omega|B) = c \cdot p(\omega) \quad \forall \omega \in \Omega_B.$$

wobei  $p(\omega)$  die Massenfunktion von  $\mathbb{P}$  ist. Wegen Normierung ergibt sich also bereits

$$1 = \sum_{\omega \in \Omega} \mathbb{P}(\omega|B) = c \cdot \sum_{\omega \in \Omega_B} p(\omega) \Leftrightarrow c = \frac{1}{\mathbb{P}(B)}.$$

Also ergibt sich, dass

$$p(\omega | B) = \begin{cases} \frac{p(\omega)}{\mathbb{P}(B)} & \text{falls } \omega \in B \\ 0 & \text{sonst} \end{cases}.$$

Wir können das ganze so darstellen:

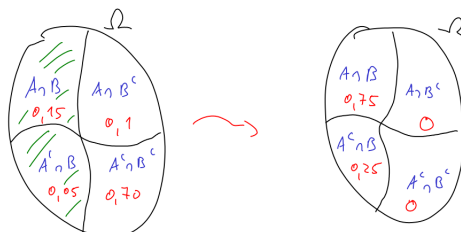


Abbildung 5: Änderung des Zustandsraums bei bedingten Wahrscheinlichkeiten

Wir erhalten also:

**Antwort.** Ein Kind, das mit 6 Monaten krabbelt, wird mit einer Wahrscheinlichkeit von 75% mit 10 Monaten laufen können.

**Definition 2.1** (Bedingte Wahrscheinlichkeit). Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum. Seien  $A, B \in \mathcal{F}$  Ereigniss mit  $\mathbb{P}(B) \neq 0$ . Dann definieren wir

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

Besser  
Skizzen  
machen

und nennen diese die **bedingte Wahrscheinlichkeit von  $A$  gegeben  $B$** .

**Bemerkung.** Die Abbildung

$$\mathbb{P}(\cdot | B) : \begin{cases} \mathcal{F} & \longrightarrow \mathbb{R}_+ \\ A & \longmapsto \mathbb{P}(A | B) \end{cases}$$

ist eine Wahrscheinlichkeitsverteilung auf  $(\Omega, \mathcal{F})$ , die wir auch die **bedingte Wahrscheinlichkeitsverteilung gegeben  $B$**  nennen.

**Definition 2.2** (Bedingter Erwartungswert). Sei  $X : \Omega \rightarrow \mathcal{S} \subseteq \mathbb{R}$  eine (diskrete) Zufallsvariable mit Verteilung  $\mathbb{P}(\cdot | B)$ . Dann hat  $X$  den Erwartungswert

$$\sum_{s \in \mathcal{S}} s \cdot \mathbb{P}(X = s | B) =: \mathbb{E}(X | B).$$

Dieser heißt **bedingter Erwartungswert von  $X$  gegeben  $B$** .

**Beispiel.** Wir werfen eine faire Münze  $N$  mal, dabei beobachten wir  $n$  mal das Ergebnis 'Zahl'.

**Frage.** Was ist die Wahrscheinlichkeit, dass bei den ersten  $m$  Würfeln immer 'Zahl' gefallen ist?

Ohne Weitere Informationen (dass insgesamt  $n$  mal Zahl gefallen ist) würden wir hier  $\mathbb{P} \equiv \frac{1}{2^m}$  erhalten. Betrachte nun den Zustandsraum

$$\Omega = \{\omega = (x_1, \dots, x_N) \mid x_i \in \{0, 1\}, 1 \leq i \leq N\}.$$

wobei

$$x_k := \begin{cases} 1 & \text{falls } k\text{-ter Wurf ist 'Zahl'} \\ 0 & \text{falls } k\text{-ter Wurf ist 'Kopf'} \end{cases}.$$

und versee ihn mit  $\mathcal{F} = \mathcal{P}(\Omega)$  sowie  $\mathbb{P}$  als Gleichverteilung auf  $\Omega$ . Mit  $X_k(\omega) := x_k$  interessieren wir uns also für

$$\mathbb{P}\left(X_1 = X_2 = \dots = X_m = 1 \mid \sum_{k=1}^N X_k = n\right).$$

Nach Definition ist dies

$$\begin{aligned}
 & \mathbb{P} \left( (X_1 = \dots = X_m = 1) \cap \left( \sum_{k=m+1}^N X_k = n - m \right) \right) \\
 &= \frac{\mathbb{P} \left( \sum_{k=1}^N X_k = n \right)}{\mathbb{P} \left( \sum_{k=1}^m X_k = m \right)} \\
 &= \frac{\frac{1}{2^N} \binom{N-m}{n-m}}{\frac{1}{2^N} \binom{N}{n}} \\
 &= \frac{(N-m)!}{(n-m)!(N-n)!} \cdot \frac{(N-n)!n!}{N!} \\
 &= \frac{(N-m)n!}{N!(n-m)!}
 \end{aligned}$$

**Notation.** Zur Vereinfachung der Notation schreiben wir oft

$$\mathbb{P}(X_1 = X_2 = a) = \mathbb{P}(\{\omega \in \Omega \mid X_1(\omega) = a\} \cap \{\omega \in \Omega \mid X_2(\omega) = a\}).$$

sowie

$$\mathbb{P}(X_1 = a, X_2 = b) = \mathbb{P}(\{\omega \in \Omega \mid X_1(\omega) = a\} \cap \{\omega \in \Omega \mid X_2(\omega) = b\}).$$

Wir haben gerade aus einer Wahrscheinlichkeitsverteilung  $\mathbb{P}$  die Verteilung  $\mathbb{P}(\cdot \mid B)$  gewonnen. Das ganze geht auch umgekehrt:

**Satz 2.3.** Sei  $\Omega = \bigcup_{k \in I} H_k$  eine **disjunkte** Zerlegung von  $\Omega$  in (abzählbar viele) Ereignisse  $H_k, k \in I$ , wobei  $\mathbb{P}(H_k) \neq 0$ . Dann ist  $\forall A \in \mathcal{F}$ :

$$\mathbb{P}(A) = \sum_{k \in I} \mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k).$$

*Beweis.*  $\forall A \in \mathcal{F}$  ist

$$A = A \cap \bigcup_{k \in I} H_k = \bigsqcup_{k \in I} (A \cap H_k).$$

eine disjunkte Vereinigung. Also folgt aus  $\sigma$ -Additivität, dass

$$\begin{aligned}
 \mathbb{P}(A) &= \sum_{k \in I} \mathbb{P}(A \cap H_k) \\
 &= \sum_{\substack{k \in I \\ \mathbb{P}(H_k) \neq 0}} \mathbb{P}(A \cap H_k) \\
 &= \sum_{\substack{k \in I \\ \mathbb{P}(H_k) \neq 0}} \mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k)
 \end{aligned}$$

□

**Beispiel.** Eine Urne  $A$  enthält **2 rote** und **3 blaue** Kugeln. In Urne  $B$  liegen umgekehrt **3 rote** und nur **2 blaue** Kugeln. Wir gehen davon aus, dass die Urnen immer gut gemischt sind. Nun machen wir Folgendes:

- ① Wir ziehen eine Kugel  $K_1$  aus  $A$  und legen sie in  $B$

② Wir ziehen eine Kugel  $K_2$  aus  $B$  und lg

**Frage.** Was ist die Wahrscheinlichkeit, dass  $K_2$  rot ist?

Wir erhalten nun

$$\begin{aligned}\mathbb{P}(K_2 \text{ ist rot}) &= \mathbb{P}(K_2 \text{ rot} \mid K_1 \text{ rot}) \cdot \mathbb{P}(K_1 \text{ rot}) \\ &\quad + \mathbb{P}(K_2 \text{ rot} \mid K_1 \text{ blau}) \cdot \mathbb{P}(K_1 \text{ blau}) \\ &= \frac{4}{6} \cdot \frac{2}{5} + \frac{3}{6} \cdot \frac{3}{5} = \frac{17}{30}\end{aligned}$$

Graphik  
einfügen

## 2.2 Baye'sche Regel

In der Baye'schen Statistik ist  $\mathbb{P}(H_k)$  auch die **a-priori-Einschätzung** der Wahrscheinlichkeit einer Hypothese  $H_k$ , das könnte z.B. sein

$H_k = \{\text{Die Unfallkosten pro Jahr liegen im Bereich } [100k, 100(k+1))\}$ .

Aus statistischen Daten weiß man, dass ein Ereignis  $A \in \mathcal{F}$  mit einer Wahrscheinlichkeit  $\mathbb{P}(A) \neq 0$  eintritt, also z.B.

$A = \text{'Es handelt sich um einen Auffahrunfall'}$ .

Dazu kennt man  $\mathbb{P}(A \mid H_k)$ . Falls  $A$  eintritt, werden die Versicherungskosten neu berechnet, auf der Basis von

$$\mathbb{P}(H_k \mid A).$$

Dies nennt man dann auch **a-posteriori-Verteilung** von  $H_k$ .

**Korollar 2.4** (Baye'sche Regel). Für  $A \in \mathcal{F}$  mit  $\mathbb{P}(A) \neq 0$  gilt

$$\mathbb{P}(H_k \mid A) = \frac{\mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k)}{\sum_{\substack{l \in I \\ \mathbb{P}(H_l) \neq 0}} \mathbb{P}(A \mid H_l) \cdot \mathbb{P}(H_l)}.$$

*Beweis.* Es ist

$$\mathbb{P}(H_k \mid A) = \frac{\mathbb{P}(H_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k)}{\mathbb{P}(A)}.$$

Aus **Satz 2.3** erhalten wir nun aber genau

$$\mathbb{P}(A) = \sum_{\substack{l \in I \\ \mathbb{P}(H_l) \neq 0}} \mathbb{P}(A \mid H_l) \cdot \mathbb{P}(H_l).$$

und wir sind fertig.  $\square$

**Beispiel.** Eine Krankheit  $K$  tritt selten auf, mit einer Häufigkeit von  $10^{-4}$ , also bei 10 von 100.000 Menschen. Ein Test zur Erkennung der Krankheit ist positiv (+) bei 96% der Kranken und 0,1% der Gesunden.

Der Test liefert also 0,1% falsch positive und 4% falsch negative Ergebnisse.

**Frage.** Was ist die Wahrscheinlichkeit, dass jemand krank ist, sofern er positiv getestet wurde?

- Die **a-priori-Wahrscheinlichkeit** beträgt  $\mathbb{P}(K) = 10^{-4}$  sowie  $\mathbb{P}(K^c) = 1 - 10^{-4}$ .
- Wir kennen die **bedingten Wahrscheinlichkeiten**  $\mathbb{P}(+ | K) = 0,96$  sowie  $\mathbb{P}(+ | K^c) = 0,001$ .
- Als **A-posteriori Wahrscheinlichkeit** erhalten wir nun:

$$\begin{aligned}\mathbb{P}(K | +) &= \frac{\mathbb{P}(+ | K) \cdot \mathbb{P}(K)}{\mathbb{P}(+ | K) \cdot \mathbb{P}(K) + \mathbb{P}(+ | K^c) \cdot \mathbb{P}(K^c)} \\ &= \frac{0,96 \cdot 10^{-4}}{0,96 \cdot 10^{-4} + 0,001 \cdot (1 - 10^{-4})} \\ &\approx 9,6\%\end{aligned}$$

**Antwort.** Die Wahrscheinlichkeit, dass man krank ist, wenn man positiv getestet ist, beträgt also (nur) 9,6%.

**Bemerkung\*.** Ein Test hat üblicherweise eine Sensitivität und ein Spezifität. Die Sensitivität gibt an, welcher Anteil der tatsächlich infizierten positiv getestet werden. Die Spezifität gibt an, welcher Anteil der gesunden Menschen auch negativ getestet wird.

**Beispiel** (Aktuelle Corona-Zahlen). Bei den aktuellen Schnelltests gibt es (in etwa) eine falsch-positiven Rate von 2%, also  $\mathbb{P}(B | K^c) = 2\%$ , und eine falsch-negativen Rate von 20%, also  $\mathbb{P}(- | K) = 20\%$ . Bei einer Inzidenz von 150-200 pro 100.000 Einwohner pro Woche, einer Dunkelziffer nah bei 2 ergibt sich eine Schätzung der aktuell infizierten von

$$\mathbb{P}(K) \in [0.005, 0.01].$$

(Zum Vergleich: Die aktuell gemeldeten positiven Fälle liegen bei 0,0035).

Nun können wir wieder berechnen:

$$\begin{aligned}\mathbb{P}(K | +) &= \frac{\mathbb{P}(+ | K) \cdot \mathbb{P}(K)}{\mathbb{P}(+ | K) \cdot \mathbb{P}(K) + \mathbb{P}(+ | K^c) \cdot \mathbb{P}(K^c)} \\ &= \frac{0,8 \cdot \mathbb{P}(K)}{0,8\mathbb{P}(K) + 0,02(1 - \mathbb{P}(K))}\end{aligned}$$

Wir erhalten

- (a) Mit  $\mathbb{P}(K) = 0,005$  eine Wahrscheinlichkeit von  $\mathbb{P}(K | +) \approx 17\%$
- (b) Mit  $\mathbb{P}(K) = 0,01$  eine Wahrscheinlichkeit von  $\mathbb{P}(K | +) \approx 29\%$
- (c) Mit  $\mathbb{P}(K) = 0,001$  eine Wahrscheinlichkeit von  $\mathbb{P}(K | +) \approx 3,8\%$

**Frage.** Lohnt es sich die Schnelltests in den Schulen zu machen?  
Also: Was ist  $\mathbb{P}(K | -)$

Auch das lässt sich mit der gleichen Formel beantworten, mit  $\mathbb{P}(- | K) = 0,2$  und  $\mathbb{P}(- | K^c) = 0,98$  erhalten wir

- (a) Für  $\mathbb{P}(K) = 0,005$  eine Wahrscheinlichkeit von  $\mathbb{P}(K | -) \approx 0,1\%$

- (b) Für  $\mathbb{P}(K) = 0,01$  eine Wahrscheinlichkeit von  $\mathbb{P}(K \mid -) \approx 0,2\%$
- (c) Für  $\mathbb{P}(K) = 0,001$  eine Wahrscheinlichkeit von  $\mathbb{P}(K \mid -) \approx 0,02\%$

Vorlesung 8  
Mi 05 Mai 2021 10:11

## 2.3 Mehrstufige Modelle

Sei eine Folge von  $n$  Zufallsexperimenten in den Wahrscheinlichkeitsräumen  $\Omega_1, \Omega_2, \dots, \Omega_n$  gegeben. Wir definieren ein  **$n$ -stufiges Zufallsexperiment** durch

- $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_k \in \Omega_k, 1 \leq k \leq n\}$
- $\mathcal{F} = \mathcal{P}(\Omega)$ .
- Definiere die Zufallsvariablen

$$X_k(\omega) = \omega_k \quad 1 \leq k \leq n.$$

Den Index  $k$  interpretieren wir hierbei als Zeit.  $k \mapsto X_k$  ist eine Trajektion von  $X = (X_1, \dots, X_n)$  ?

- $\mathbb{P} = ?$ . Wir konstruieren  $\mathbb{P}$  auf  $(\Omega, \mathcal{F})$  mit
  - (a) Der Anfangsverteilung  $\mathbb{P}(X_1 = x_1) := p_1(x_1)$  für alle  $x_1 \in \Omega_1$ .
  - (b) Den bedingten Verteilungen

$$\mathbb{P}(X_k = x_k \mid X_1 = x_1, X_2 = x_2, \dots, X_{k-1} = x_{k-1}) =: p_k(x_k \mid x_1, \dots, x_{k-1}).$$

für alle  $x_l \in \Omega_l, 1 \leq l \leq k-1$ , sodass  $\mathbb{P}(X_1 = x_1, \dots, X_{k-1} = x_{k-1}) \neq 0$ .

i

**Bemerkung\*.** Man kann das Allgemeiner machen, indem wir  $\mathcal{F}$  als die Produktsigmaalgebra der  $\mathcal{F}_i$  wählen.

**Satz 2.5.** Sei  $p_i(\cdot)$  die Massenfunktion einer Wahrscheinlichkeitsverteilung auf  $\Omega_i$  und  $p_k(\cdot \mid x_1, \dots, x_{k-1})$  für alle  $1 \leq k \leq n$  mit  $x_1 \in \Omega_1, \dots, x_{k-1} \in \Omega_{k-1}$  eine Massenfunktion auf  $\Omega_k$ .

Dann existiert eine eindeutige Wahrscheinlichkeitsverteilung  $\mathbb{P}$  auf  $(\Omega, \mathcal{F})$ , sodass

- (a)  $\mathbb{P}(X_1 = x_1) = p_1(x_1) \quad \forall x_1 \in \Omega_1$
- (b)  $\mathbb{P}(X_k = x_k \mid X_1 = x_1, \dots, X_{k-1} = x_{k-1}) = p_k(x_k \mid x_1, \dots, x_{k-1})$

Die Wahrscheinlichkeitsverteilung  $\mathbb{P}$  hat die Massenfunktion

$$p(x_1, \dots, x_n) = p_1(x_1)p_2(x_2 \mid x_1) \cdot \dots \cdot p_n(x_n \mid x_1, \dots, x_{n-1}).$$

*Beweis.* 1) Nimm zunächst an, dass solch ein Maß existiert, wir zeigen die letzte Aussage. Sei  $\mathbb{P}$  sodass (a) und (b) erfüllt sind. Dann ist

$$\forall 1 \leq k \leq n: \mathbb{P}(X_1 = x_1, \dots, X_k = x_k) = p(x_1, \dots, x_k).$$

- Für  $k = 1$  gilt das (aus (a)).
- Falls es für  $k-1$  gilt, so haben wir die Fälle
- $p(x_1, \dots, x_{k-1}) = 0$ , dann ist  $0 = 0$  wahr.
- Falls  $p(x_1, \dots, x_{k-1}) \neq 0$ , so ist

$$\begin{aligned} & \mathbb{P}(X_1 = x_1, \dots, X_k = x_k) \\ &= \mathbb{P}(X_k = x_k \mid X_1 = x_1, \dots, X_{k-1} = x_{k-1}) \cdot \mathbb{P}(X_1 = x_1, \dots, X_{k-1} = x_{k-1}) \quad . \\ &= p_1(x_1) \cdot \dots \cdot p_{k-1}(x_{k-1} \mid x_1, \dots, x_{k-2}) \cdot p_k(x_k \mid x_1, \dots, x_{k-1}) = p(x_1, \dots, x_k) \end{aligned}$$

Normierung:  $\forall x \in \Omega, x = (x_1, \dots, x_n)$  mit  $x_k \in \Omega_k$  ist

$$\begin{aligned} \sum_{x \in \Omega} p(x) &= \sum_{x_1 \in \Omega_1} \dots \sum_{x_n \in \Omega_n} p(x_1, \dots, x_n) \\ &= \sum_{x_1 \in \Omega_1} p(x_1) \sum_{x_2 \in \Omega_2} p(x_2 | x_1) \dots \sum_{x_n \in \Omega_n} p(x_n | x_1, \dots, x_{n-1}) = 1. \end{aligned}$$

Für Eigenschaft (b) ist

$$\begin{aligned} \mathbb{P}(X_1 = x_1, \dots, X_k = x_k) &= \sum_{x_{k+1} \in \Omega_{k+1}} \dots \sum_{x_n \in \Omega_n} p(x_1, \dots, x_n) \\ &= p_1(x_1) \dots p_{k-1}(x_{k-1} | x_1, \dots, x_{k-2}) p_k(x_k | x_1, \dots, x_{k-1}) \\ &= \mathbb{P}(X_1 = x_1, \dots, X_{k-1} = x_{k-1}) p_k(x_k | x_1, \dots, x_{k-1}) \end{aligned}$$

Also erhalten wir

$$\mathbb{P}(X_k = x_k | X_1 = x_1, \dots, X_{k-1} = x_{k-1}) = p_k(x_k | x_1, \dots, x_{k-1}).$$

□

**Anmerkung.** Mir ist noch nicht klar, wo wir im Beweis des Satzes jetzt gezeigt haben wollen, dass solch ein Wahrscheinlichkeitsmaß existiert, das muss ich noch ausarbeiten.

Beweis  
nochmal  
sortieren

**Bemerkung.** Falls  $p_k(x_k | x_1, \dots, x_{k-1})$  nur eine Funktion von  $x_{k-1}, \dots, x_{k-m-1}$ , dann sagen wir, dass unser Modell ein Gedächtnis von  $m$  Schritten hat.

### 2.3.1 Produktmodelle

Falls  $p_k(x_k | x_1 = \dots = x_{k-1}) = p_k(x_k)$ , d.h.  $x_k$  hängt nicht von den Werten  $x_1, \dots, x_{k-1}$  ab. Dann erhalten wir aus **Satz 2.5**, dass

$$p(x_1, \dots, x_n) = \prod_{k=1}^n p_k(x_k).$$

**Definition 2.6** (Produktmodell). Die Wahrscheinlichkeitsverteilung  $\mathbb{P}$  auf  $\Omega = \Omega_1 \times \dots \times \Omega_n$  mit Massenfunktion

$$p(x_1, \dots, x_n) = \prod_{k=1}^n p_k(x_k).$$

heißt **Produkt von  $\mathbb{P}_1, \dots, \mathbb{P}_n$** . ( $\mathbb{P}_k$  hat Massenfunktion  $p_k$ ).

**Notation.** Wir schreiben  $\mathbb{P} = \mathbb{P} \otimes \mathbb{P}_2 \otimes \dots \otimes \mathbb{P}_n$ , wenn  $\mathbb{P}$  das Produkt von  $\mathbb{P}_1, \dots, \mathbb{P}_n$  ist.

**Beispiel.** Seien  $n$  unabhängige 0-1-Experimente mit Erfolgswahrscheinlichkeit  $p$  gegeben. Also

$$\Omega_1 = \dots = \Omega_n = \{0, 1\}.$$

und  $p_k(1) = p = 1 - p_k(0)$  für  $k = 1, \dots, n$ . dann ist  $p_k(x) =$

$(1-p) \left(\frac{p}{1-p}\right)^x$  für  $x \in \{0, 1\}$ . Die entstehende Verteilung

$$p(x_1, \dots, x_n) = (1-p)^n \prod_{k=1}^n \left(\frac{p}{1-p}\right)^{x_k}.$$

ist die  **$n$ -dimensionale Bernoulli-Verteilung mit Parameter  $p$** .

**Satz 2.7.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Produktmodell. Dann ist für beliebige Ereignisse  $A_k \subseteq \Omega_k$ ,  $k = 1, \dots, n$ :

$$\mathbb{P}(A_1 \times \dots \times A_n) = \prod_{k=1}^n \mathbb{P}_k(A_k).$$

und  $\mathbb{P}(\tilde{A}_k) = \mathbb{P}_k(A_k)$ , wobei

$$\tilde{A}_k := \Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n.$$

Deswegen sind  $\tilde{A}_1, \dots, \tilde{A}_n$  unabhängige Ereignisse.

*Beweis.* Es ist

$$\begin{aligned} \mathbb{P}(A_1 \times \dots \times A_n) &= \mathbb{P}((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) \\ &= \sum_{(x_1, \dots, x_n) \in A_1 \times \dots \times A_n} p(x_1, \dots, x_n) \\ &= \sum_{x_1 \in A_1} \dots \sum_{x_n \in A_n} p_1(x_1) \cdot \dots \cdot p_n(x_n) \\ &= \prod_{k=1}^n \sum_{x_k \in A_k} p_k(a_k) \\ &= \prod_{k=1}^n \mathbb{P}(A_k) \end{aligned}$$

Es ergibt sich leicht

$$\begin{aligned} \mathbb{P}(\tilde{A}_k) &= \mathbb{P}(X_k \in A_k, X_l \in \Omega_l \ \forall l \neq k) \\ &= \left( \prod_{l \neq k} \underbrace{\mathbb{P}_l(X_l \in \Omega_l)}_{=1} \right) \cdot \mathbb{P}_k(X_k \in A_k) \\ &= \mathbb{P}_k(A_k) \end{aligned}$$

Damit ergibt sich schlussendlich für beliebiges  $I = \{i_1, \dots, i_l\}$ :

$$\begin{aligned} \mathbb{P}(\tilde{A}_{i_1} \cap \dots \cap \tilde{A}_{i_l}) &= \mathbb{P} \left( \bigcap_{i \in I} \{X_i \in A_i\} \cap \bigcap_{j \neq I} \{x_j \in \Omega_j\} \right) \\ &= \prod_{i \in I} \mathbb{P}_i(A_i) \cdot \prod_{j \neq I} \underbrace{\mathbb{P}_j(\Omega_j)}_{=1} \\ &= \prod_{i \in I} \mathbb{P}_i(A_i) \\ &\stackrel{\mathbb{P}(A_i) = \mathbb{P}(\tilde{A}_i)}{=} \prod_{k=1}^l \mathbb{P}(\tilde{A}_{i_k}) \end{aligned}$$

□



**Bemerkung.** • Es ist  $\Omega = \Omega_1 \times \dots \times \Omega_n$

- Eigentlich müssen wir  $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$  als entsprechende Wahrscheinlichkeitsräume betrachten, wir unterdrücken aber oft die Notation  $\mathcal{F}_i, \mathbb{P}_i$

Im Allgemeinen setzen wir

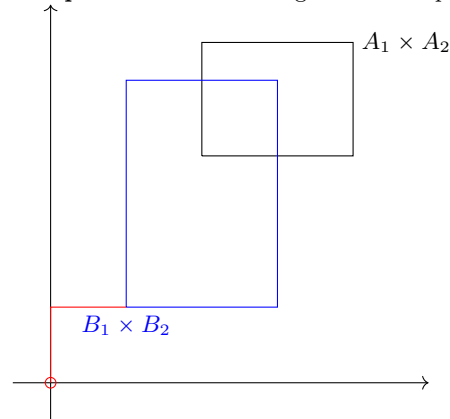
$$\mathcal{F} = \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n.$$

wobei  $\mathcal{F}$  dann die  $\sigma$ -Algebra ist, die von  $A_1 \times \dots \times A_n$  mit  $A_i \in \mathcal{F}_i$  erzeugt ist. Im Spezialfall  $\mathcal{F}_i = \mathcal{P}(\Omega_i)$  ergibt sich insbesondere wieder der uns bekannte Fall  $\mathcal{F} = \mathcal{P}(\Omega)$ .

Für Produktmodelle erhalten wir also  $\Omega = \Omega_1 \times \dots \times \Omega_n$ ,  $\mathcal{F} = \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n$  sowie  $\mathbb{P} = \mathbb{P}_1 \otimes \dots \otimes \mathbb{P}_n$ .

Beachte, dass  $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n \neq \mathcal{F}_1 \times \dots \times \mathcal{F}_n$  im Allgemeinen, wie das folgende Beispiel zeigt.

**Beispiel.** Im Fall  $n = 2$  ergibt sich beispielsweise folgende Situation:



Vorlesung 9  
Mo 10 Mai 2021 10:15

### 2.3.2 Markovketten (MK)

- Setze  $X = (X_0, X_1, X_2, \dots, X_n)$ . Die Zeit beginnt hier bei  $k = 0$ .
- Betrachte  $\Omega_k = \mathcal{S}$  für festes  $\mathcal{S}$ , also

$$\Omega = \mathcal{S}^{n+1} = \{(x_0, \dots, x_n) \mid x_i \in \mathcal{S}, 0 \leq i \leq n\}.$$

**Definition 2.8** (Markovkette). Eine **Markovkette** (abgekürzt: MK) ist ein mehrstufiges Modell mit der Eigenschaft

$$p_k(x_k \mid x_0, \dots, x_{k-1}) = p_k(x_k \mid x_{k-1}).$$

**Frage.** Sei  $\mathcal{S}$  abzählbar. Wie beschreibt man die Übergänge von  $X_k$  nach  $X_{k+1}$ ?

**Definition 2.9** (Stochastische Matrix). Eine Matrix  $P = [\mathbb{P}(x, y)]_{x, y \in \mathcal{S}}$  mit den Eigenschaften

- (a)  $\forall x \forall y: \mathbb{P}(x, y) \geq 0$
  - (b)  $\forall x \in \mathcal{S}: \sum_{y \in \mathcal{S}} \mathbb{P}(x, y) = 1$
- heißt **stochastische Matrix**.

**Bemerkung\***. Beachte, dass die Matrix in obiger Definition nicht zwingend endlich sein muss, Definitionen verallgemeinern sich kanonisch. Wir fordern aber, dass  $\mathcal{S}$  abzählbar ist.

**Lemma 2.10**. Die Matrix  $P_k$  mit den Einträgen

$$P_k(x, y) = p_k(Y \mid X) \quad \forall x, y \in \mathcal{S}.$$

ist eine stochastische Matrix.

*Beweis.* Offenbar ist  $p_k(Y \mid X) \geq 0$ . Zudem

$$\begin{aligned} \sum_{y \in \mathcal{S}} P_k(x, y) &= \sum_{y \in \mathcal{S}} \mathbb{P}(X_k = Y \mid X_{k-1} = x) \\ &= \mathbb{P}\left(\bigcup_{y \in \mathcal{S}} \{X_k = y\} \mid X_{k-1} = x\right) \\ &= \mathbb{P}(\Omega_k \mid X_{k-1} = x) \\ &= 1 \end{aligned}$$

weil es sich bei  $\mathbb{P}(\cdot \mid X_{k-1} = x)$  um eine Wahrscheinlichkeitsverteilung handelt, und  $\mathcal{S} = \bigsqcup_{y \in \mathcal{S}} \{y\}$  eine disjunkte Vereinigung ist.  $\square$

**Bemerkung.**  $P_k$  ist eine sogenannte **Übergangsmatrix**. Sie beschreibt den Übergang der Markovkette von  $\Omega_k$  nach  $\Omega_{k+1}$

Die Massenfunktion einer Markovkette ist

$$p(x_0, x_1, \dots, x_n) = p_0(x_0) \cdot P_1(x_0, x_1) \cdot \dots \cdot P_n(x_{n-1}, x_n).$$

wobei  $p_0$  die sogenannte **Anfangsverteilung** ist.

**Bemerkung.** Falls  $P_k = \mathbb{P}$ , d.h. die Übergangsmatrix hängt nicht von  $k$  ab, dann heißt die Markovkette (zeitlich) **homogen**.

**Bemerkung.** Seien  $P, Q$  zwei stochastische Matrizen. Dann ist auch  $P \cdot Q$  eine stochastische Matrix, wobei

$$(P \cdot Q)(x, y) = \sum_{z \in \mathcal{S}} P(x, z) \cdot Q(z, y).$$

**Frage.** Was ist

- 1)  $\mathbb{P}(X_n = x)$
- 2)  $\lim_{n \rightarrow \infty} \mathbb{P}(X_n = x)$  (Existiert dieser überhaupt?)
- 3) Ist  $\lim_{n \rightarrow \infty} \mathbb{P}(X_n = x)$  von  $x_0$  abhängig?

**Satz 2.11** (Massenfunktion in Markovketten). Sei  $\mu_0$  der Zeilenvektor mit Elementen  $p_0(x), x \in \mathcal{S}$ . Seien dazu  $P_1, P_2, \dots, P_n$  die Übergangsmatrizen einer Markovkette  $X = (X_0, X_1, \dots, X_n)$  auf  $\mathcal{S}$ . Dann hat die Wahrscheinlichkeit von  $X_n$  die Massenfunktion

$$\mu_n(x) := \mathbb{P}(X_n = x) = (\mu_0 \cdot P_1 \cdot \dots \cdot P_n)(x) \quad \forall x \in \mathcal{S}.$$

*Beweis.*

$$\begin{aligned} \mathbb{P}(X_n = x) &= \sum_{x_0, \dots, x_{n-1} \in \mathcal{S}} \mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) \\ &= \mu_0(x_0) P_1(x_0, x_1) \dots P_n(x_{n-1}, x) \end{aligned}$$

□

**Beispiel.** (a) Produktmodelle sind Markovketten.

- (b) Irrfahrten (eng: 'Random Walk') auf  $\mathbb{Z}^d$  sind Markovketten:  
Sei  $\mathcal{S} = \mathbb{Z}^d$  mit  $d \in \mathbb{N}$  fest. Die (symmetrische) Irrfahrt ist eine homogene Markovkette mit

$$P_k(x, y) = P(x, y) = \begin{cases} \frac{1}{2d} & \text{falls } \|x - y\| = 1 \\ 0 & \text{sonst} \end{cases}.$$

In jedem Schritt bewegen wir uns also auf einen benachbarten Gitterpunkt.

- (c) **Urnenmodell von Ehrenfest.** Wir haben ein System mit  $N$  Teilchen, die auf zwei Urnen  $A, B$  verteilt sind. Zu jedem Zeitpunkt  $t \in \mathbb{N}$  wechselt eine zufällig ausgewählte Kugel die Urne.

In der Makroskopischen Modellierung sei  $n_A := \# \text{Teilchen in } A$ . Dann ist

$$\rho_A := \frac{n_A}{N} \in \left\{ 0, \frac{1}{N}, \dots, \frac{N-1}{N}, 1 \right\} =: \mathcal{S}.$$

und die Markovkette, die die Zeitentwicklung von  $\rho_A$  beschreibt, hat die Übergangsmatrix

$$P(x, y) = \begin{cases} x & \text{falls } y = x - \frac{1}{N} \\ 1 - x & \text{falls } y = x + \frac{1}{N} \\ 0 & \text{sonst} \end{cases}.$$

Die Fälle spiegeln wieder, dass wir ein Teilchen aus  $A$  bzw.  $B$  gezogen haben, wobei der dritte Fall diejenigen  $y$  abdeckt, die wir nicht erreichen können, weil sich  $n_A$  immer nur um  $\pm 1$  ändert.

In der Mikroskopischen Modellierung nummerieren wir die Teilchen, d.h.

$$\mathcal{S} = \{0, 1\}^n = \{\sigma = (\sigma_1, \dots, \sigma_N) \mid \sigma_k \in \{0, 1\}, 1 \leq k \leq N\}.$$

wobei

$$\sigma_k = \begin{cases} 1 & \text{falls Teilchen } k \text{ ist in } A \\ 0 & \text{falls Teilchen } k \text{ ist in } B \end{cases}.$$

Die Dynamik von  $\sigma \in \mathcal{S}$  wird durch die Markovkette beschrieben:

$$P(\sigma, \tilde{\sigma}) = \begin{cases} \frac{1}{N} & \text{falls } \sum_{k=1}^N |\sigma_k - \tilde{\sigma}_k| = 1 \\ 0 & \text{sonst} \end{cases}.$$

**Bemerkung.** Die Notation  $\sum_{k=1}^N |\sigma_k - \tilde{\sigma}_k| = 1$  ist nur eine kurze (elegante) Möglichkeit auszudrücken, dass sich  $\sigma_k$  nach  $\tilde{\sigma}_k$  in nur einem Eintrag ändert.

Es ergibt sich also eine Irrfahrt auf dem Hyperwürfel  $\{0, 1\}^N$ .

**Frage.** Was ist  $P(X_l = y \mid X_k = x)$  für  $k < l$  beliebig?

Intuitiv sollten wir hier  $P_{k+1} \cdot \dots \cdot P_l(x_k, x_l)$  erhalten, indem wir die Übergangsmatrizen multiplizieren.

**Satz 2.12** (Markoveigenschaft). Sei  $X$  eine Markovkette auf  $\mathcal{S}$ . Für alle  $0 \leq k \leq l \leq n$  und  $x_0, \dots, x_l \in \mathcal{S}$  mit  $\mathbb{P}(X_0 = x_0, \dots, X_k = x_k) \neq 0$  ist

$$\begin{aligned} \mathbb{P}(X_l = x_l \mid X_0 = x_0, \dots, X_k = x_k) &= \mathbb{P}(X_l = x_l \mid X_k = x_k) \\ &= (P_{k+1} \cdot \dots \cdot P_l)(x_k, x_l) \end{aligned} \quad (13)$$

Diese Eigenschaft (erstes Gleichheitszeichen) heißt auch **Markov-Eigenschaft**.

*Beweis.* Es ergibt sich

$$\begin{aligned} & P(X_l = x_l \mid X_0 = x_0, \dots, X_k = x_k) \\ \stackrel{\text{Def.}}{=} & \frac{P(X_0 = x_0, \dots, X_k = x_k, X_l = x_l)}{P(X_0 = x_0, \dots, X_k = x_k)} \\ = & \frac{\sum_{x_{k+1}, \dots, x_{l-1}} p_0(x_0) P_1(x_0, x_1) \cdot \dots \cdot P_k(x_{k-1}, x_k) \cdot P_{k+1}(x_k, x_{k+1}) \cdot \dots \cdot P_l(x_{l-1}, x_l) \cdot P_{l+1}(x_l, x_{l+1}) \cdot \dots \cdot P_n(x_n)}{\sum_{x_{l+1}, \dots, x_n \in \mathcal{S}}} \\ = & \frac{\sum_{x_{k+1}, \dots, x_{l-1} \in \mathcal{S}} p_0(x_0) \cdot \dots \cdot P_k(x_{k-1}, x_k) \cdot P_{k+1}(x_k, x_{k+1}) \cdot P_l(x_{l-1}, x_l)}{\sum_{x_{k+1}, \dots, x_{l-1} \in \mathcal{S}} p_0(x_0) \cdot \dots \cdot P_k(x_{k-1}, x_k)} \\ = & \sum_{x_{k+1}, \dots, x_{l-1} \in \mathcal{S}} P_{k+1}(x_k, x_{k+1}) \cdot P_l(x_{l-1}, x_l) \\ = & (P_{k+1} \cdot \dots \cdot P_l)(x_k, x_l) \end{aligned}$$

Das zeigt die erste Gleichheit. Dazu ist

$$\begin{aligned}
 & \mathbb{P}(X_l = x_l \mid X_k = x_k) \\
 &= \frac{\mathbb{P}(X_l = x_l, X_k = x_k)}{\mathbb{P}(X_k = x_k)} \\
 &= \frac{\sum_{x_{k+1}, \dots, x_{l-1} \in \mathcal{S}} (\mu_0 P_1 \dots P_k)(x_k) \cdot P_{k+1}(x_k, x_{k+1}) \cdot P_l(x_{l-1}, x_l)}{(\mu_0 P_1 \dots P_k)(x_k)} \\
 &= (P_{k+1} \dots P_l)(x_k, x_l)
 \end{aligned}$$

□

**Beispiel.** Ist  $\mathcal{S} = \{1, \dots, N\}$ , so ergibt sich mit  $\mu_0 = (\mu_0(1), \dots, \mu_0(N))$

...

$$= \left( \sum_k \mu_0(k) P(k, 1), \sum_k \mu_0(k) P(k, 2), \dots \right) = ((\mu P)(1), (\mu P)(2), \dots).$$

Beispiel  
fertig

Vorlesung 10  
Mi 12 Mai 2021 10:16

**Beispiel.** Sei  $\mathcal{S} = \{0, 1\}$  und  $\alpha, \beta \in (0, 1)$ . Setze

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

Oft machen wir eine graphische Darstellung: Wir wollen also  $P^n$  ausrechnen, um das Verhalten der Markovkette zu studieren.

- (a) Man könnte nun  $P$  diagonalisieren, also  $P = U\Lambda U^{-1}$ , wobei  $\Lambda$  Diagonalform hat, so ist  $P^n = U\Lambda^n U^{-1}$ , und die Potenzen von  $\Lambda$  können wir leicht berechnen.
- (b)  $P^n$  ist stochastisch, also ist auch  $P^n(0, 0) + P^n(0, 1) = 1$ ,  $P^n(1, 1) + P^n(1, 0) = 1$ , weil  $P^n$  stochastisch ist. Wir wollen uns nun, welchen Wert  $P^n(0, 0)$  hat, dazu machen wir den Rekursion Ansatz:
  - Sind wir nach  $n - 1$  Schritten wieder bei der 0, so müssen wir von 0 nach 0 gehen
  - Sind wir nach  $n - 1$  Schritten bei der 1, so müssen wir im  $n$ -ten von 1 nach 0 gehen

Also ergibt sich

$$\begin{aligned}
 P^n(0, 0) &= P^{n-1}(0, 0) \cdot P(0, 0) + P^{n-1}(0, 1) \cdot P(1, 0) \\
 &= P^{n-1}(0, 0)(1 - \alpha) + \underbrace{P^{n-1}(0, 1)}_{=1 - P^{n-1}(0, 0)} \cdot \beta \\
 &= P^{n-1}(0, 0)(1 - \alpha - \beta) + \beta
 \end{aligned}$$

Als Rekursion suchen wir also eine Funktion  $n \mapsto f(n)$ , sodass

$$\begin{cases} f(n) &= \beta + (1 - \alpha - \beta)f(n-1) \\ f(1) &= 1 - \alpha \end{cases}.$$

Als Lösung ergibt sich (Theorie der linearen Rekursionen!):

$$f(n) = \frac{\beta}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta}(1 - \alpha - \beta)^n.$$

Damit ergibt sich

$$P^n = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix} + (1 - \alpha - \beta)^n \begin{pmatrix} \frac{\alpha}{\alpha+\beta} & \frac{-\alpha}{\alpha+\beta} \\ \frac{-\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}.$$

Wir beschränken uns auf  $\alpha, \beta \in (0, 1)$ . Da  $(1 - \alpha - \beta)^n \rightarrow 0$  exponentiell schnell, ergibt sich

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}.$$

Wir stellen fest, dass diese Matrix Rang 1 hat. Ebenso ergibt sich

$$\lim_{n \rightarrow \infty} \mu_0 P^n = \left( \frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right).$$

und dieser ist von  $\mu_0$  unabhängig.

Grafik

## 2.4 Unabhängige Zufallsvariablen

### 2.4.1 Unabhängige Ereignisse

**Definition 2.13** (Unabhängigkeit). Eine Familie von Ereignissen  $\{A_k\}_{k \in I}$  heißt unabhängig, falls

$$P(A_{i_1} \cap \dots \cap A_{i_n}) = \prod_{k=1}^n P(A_{i_k}).$$

für alle verschiedenen  $i_1, \dots, i_n \in I$  und für alle  $n \leq |I|$ .

Seien  $A, B \in \mathcal{F}$ , sodass  $\mathbb{P}(A) \neq 0, \mathbb{P}(B) \neq 0$ . Dann sind  $A$  und  $B$  unabhängig gwd

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

**Warnung.** Falls  $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$  für  $i \neq j \in I$ , so folgt noch nicht, dass  $\{A_i\}_{i \in I}$  unabhängig ist!

**Beispiel.** Betrachte 3 faire Münzen, also

$$\Omega = \{0, 1\}^3 \text{ f } \{(\omega_1, \omega_2, \omega_3) \mid \omega_i \in \{0, 1\} \text{ für } i = 1, 2, 3\}.$$

wobei

$$\omega_i = \begin{cases} 0 & \text{Münze } i \text{ ist Kopf} \\ 1 & \text{Münze } i \text{ ist Zahl} \end{cases}.$$

Wir setzen  $\mathcal{F} = \mathcal{P}(\Omega)$  und  $\mathbb{P}$  als die Gleichverteilung und betrachten die Ereignisse

$$A_1 = \{\omega_1 = \omega_2\} \quad A_2 = \{\omega_2 = \omega_3\} \quad A_3 = \{\omega_3 = \omega_1\}.$$

und erhalten unmittelbar

$$\mathbb{P}(A_i) = \frac{4}{8} = \frac{1}{2} \quad \mathbb{P}(A_i \cap A_j) = \frac{2}{8} = \frac{1}{4} = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j) \quad \forall i \neq j.$$

allerdings ist auch

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{2}{8} \neq \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

also sind die Ereignisse nicht unabhängig, allerdings paarweise unabhängig. Dies liegt daran, dass  $A_1 \wedge A_2 \Rightarrow A_3$  etc.

### 2.4.2 Unabhängige Zufallsvariablen

**Definition 2.14** (Gemeinsame Verteilung). Seien  $X_k : \Omega \rightarrow \mathcal{S}_k$  für  $1 \neq k \leq n$  diskret Zufallsvariablen auf  $(\Omega, \mathcal{F}, \mathbb{P})$ . Die Verteilung  $\mu_{X_1, \dots, X_n}$  der Zufallsvektoren  $(X_1, \dots, X_n)$  mit Massenfunktion

$$p_{X_1, \dots, X_n}(a_1, \dots, a_n) := \mathbb{P}(X_1 = a_1, \dots, X_n = a_n).$$

heißt die **gemeinsame Verteilung** von  $X_1, \dots, X_n$ .

**Definition 2.15** (Unabhängigkeit). Die diskreten Zufallsvariablen  $X_1, \dots, X_n$  heißen **unabhängig**, falls

$$\mathbb{P}(X_1 = a_1, \dots, X_n = a_n) = \prod_{k=1}^n \mathbb{P}(X_k = a_k).$$

für alle  $a_1 \in \mathcal{S}_1, \dots, a_n \in \mathcal{S}_n$ .

## Stichwortverzeichnis

- $\sigma$ -Algebra, 8
- $n$ -dimensionale
  - Bernoulli-Verteilung mit Parameter  $p$ , 40
- $n$ -stufiges Zufallsexperiment, 38
- a-posteriori-Verteilung, 36
- a-priori-Einschätzung, 36
- Algebra, 7
- Anfangsverteilung, 42
- Bernoulli-Verteilung mit Parameter  $p$ , 18
- Binomialverteilung mit Parametern  $n$  und  $p$ , 19
- Brown'sche Bewegung, 5
- diskrete Zufallsvariable, 16
- Elementarereignisse, 6
- empirische Häufigkeit, 16
- empirische Mittelwert, 26
- empirische Verteilung, 16
- Ereignisse, 6
- Ergebnissen, 6
- Erwartungswert, 26
  - bedingter, 34
- Gegenereignis, 7
- gemeinsame Verteilung, 47
- Gleichverteilung, 14
- homogen, 42
- Komplement von  $A$ , 7
- Markov-Eigenschaft, 44
- Markovkette, 41
- Massenfunktion, 16
- Messraum, 8
- Modellierung, 5
- Poisson-Verteilung mit Parameter  $\lambda$ , 22
- Produkt von  $\mathbb{P}_1, \dots, \mathbb{P}_n$ , 39
- relative Häufigkeit, 16
- stochastische Matrix, 42
- unabhängig, 20, 47
- Unabhängigkeit, 20
- Urnenmodell von Ehrenfest, 43
- Varianz, 30
- Verteilung
  - geometrische, 23
- Verteilung von  $X$ , 16
- Wahrscheinlichkeit
  - bedingte, 34
- Wahrscheinlichkeitsraum
  - $(\Omega, \mathcal{F}, \mathbb{P})$ , 9
- Wahrscheinlichkeitsverteilung, 8
  - bedingte, 34
- Wahrscheinlichkeitsverteilung (auch W-maß), 6
- Zufall, 4
- Zufallsexperiment, 6
- zufällige Stöße, 5
- Übergangsmatrix, 42