



universität  
wien

# DIPLOMARBEIT

Titel der Diplomarbeit

**Warum Betrunkene zurückfinden,  
Kinder im Klettergerüst hingegen nicht  
– Irrfahrten und Markov-Ketten**

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat)

Verfasser:	Dietmar Azesberger
Matrikelnummer:	0604208
Studienkennzahl lt. Studienblatt:	A 190 406 412
Studienrichtung lt. Studienblatt:	UF Mathematik und UF Physik
Betreuer:	Ao. Univ.-Prof. Dr. Peter Raith

Wien, im Mai 2010



# Vorwort und Einleitung

Warum Betrunkene zurückfinden, Kinder im Klettergerüst hingegen nicht – Dieses Thema passt auf den ersten Blick eher in den Bereich der Geistes- bzw. Sozialwissenschaften denn zur Mathematik.

Die vorliegende Arbeit beschäftigt sich allerdings nicht mit den psychischen, physischen und sozialen Ursachen für das Verhalten genannter Personen. Vielmehr widmet sie sich dem mathematischen Modell, das man den geschilderten Situationen zu Grunde legen könnte: Irrfahrten und Markov-Ketten, wie schon der zweite Teil des Titels verrät.

Die Arbeit gliedert sich in vier Kapitel, von denen die ersten beiden die theoretische Grundlage für die Kapitel drei und vier liefern.

Da Stochastik, so mein subjektiver Eindruck, nicht unbedingt das Lieblingsfach vieler Menschen zu sein scheint und daher nicht alle benötigten Inhalte so präsent sein dürften, enthält Kapitel eins die wahrscheinlichkeitstheoretischen Grundlagen, die später benötigt werden, aber auch teilweise spezielle Resultate die vor allem im zweiten Kapitel verwendet werden, dort aber den inhaltlichen Zusammenhang unterbrechen würden. Außerdem werden durch dieses Kapitel die hier verwendeten Symbole und Bezeichnungen geklärt, die sich teilweise von der oft uneinheitlichen Literatur unterscheiden.

Das zweite Kapitel widmet sich den offensichtlich weniger bekannten Markov-Ketten. Eigenschaften der Markov-Ketten als Prozesse mit extremem Kurzzeitgedächtnis und ohne Zeitgefühl werden angeführt, bewiesen und erläutert mit dem Hintergedanken, im nächsten Kapitel Aussagen über Betrunkene und Kinder treffen zu können.

Mit dem ersten Teil des Titels beschäftigt sich dann endlich das dritte Kapitel. Hier werden Irrfahrten auf  $\mathbb{Z}^d$  (diese sind spezielle Markov-Ketten) untersucht. Augenmerk wird dabei auf das Langzeitverhalten gelegt. Der teilweise stereotype Vergleich mit der uns umgebenden Welt darf durchaus mit einem Augenzwinkern verstanden werden. Das abschließende Kapitel bietet etwas für das Auge. Es wird versucht, die behandelten Irrfahrten mit dem CAS *Mathematica* zu simulieren und damit dem Ruf der Mathematik als Wissen-

schaft, die sich mit Dingen beschäftigt, die man sich ohnehin nicht vorstellen kann, etwas entgegen zu treten. Jenen, die jetzt noch immer Lust verspüren, die Arbeit zu lesen, wünsche ich viel Spaß beim Lesen, und dass die Erwartungen erfüllt werden.

Bleiben noch die obligaten und auch berechtigten Danksagungen:

Bedanken möchte ich mich bei meiner Familie, insbesondere bei meinen Eltern, die mir immer mit Rat und Tat zu Seite und voll hinter mir stehen und mich in jeder Hinsicht voll unterstützen. Ohne sie wäre es mir nicht möglich gewesen, meine Ausbildung so zu absolvieren und mein bisheriges Leben ganz nach meinen Vorstellungen zu gestalten.

Ein weiterer Dank gilt dem Betreuer meiner Diplomarbeit, Peter Raith, der mir bei Fragen immer sofort behilflich war, mir gleichzeitig aber alle Freiheiten ließ und mich fast nach Belieben schalten und walten ließ. Das ermöglichte mir ein selbstständiges und auch einigermaßen ökonomisches Erarbeiten des Themas.

# Inhaltsverzeichnis

<b>Vorwort und Einleitung</b>	<b>3</b>
<b>1 Grundlagen</b>	<b>7</b>
1.1 Wahrscheinlichkeitsräume . . . . .	7
1.2 Eigenschaften von Wahrscheinlichkeitsmaßen . . . . .	9
1.3 Zufallsvariablen . . . . .	13
1.4 Multinomialverteilung . . . . .	14
1.5 Bedingte Wahrscheinlichkeit . . . . .	15
1.6 Erwartungswert und Eigenschaften . . . . .	21
1.6.1 Erwartungswert . . . . .	21
1.6.2 Erzeugende Funktionen . . . . .	23
<b>2 Markov-Ketten</b>	<b>27</b>
2.1 Die Markov-Eigenschaft . . . . .	27
2.1.1 Was ist eine Markov-Kette? . . . . .	27
2.1.2 Ein einfaches Beispiel einer Markov-Kette . . . . .	30
2.1.3 Die Markov-Eigenschaft . . . . .	31
2.2 Stochastische Matrizen . . . . .	33
2.3 Absorptionswahrscheinlichkeiten . . . . .	36
2.4 Rückkehr zum Startpunkt . . . . .	40
2.4.1 Rekurrenz und Transienz . . . . .	40
2.4.2 Positive Rekurrenz und Nullrekurrenz . . . . .	45
<b>3 Irrfahrten</b>	<b>55</b>
3.1 Einfache Irrfahrten auf $\mathbb{Z}$ . . . . .	55
3.1.1 Verteilung zum Zeitpunkt $n$ . . . . .	58
3.1.2 Eintrittszeiten . . . . .	61
3.1.3 Eintritt in endlicher Zeit, Absorption am Rand . . . . .	65
3.1.4 Rekurrenz oder Transienz . . . . .	70
3.2 einfache Irrfahrten auf $\mathbb{Z}^d$ . . . . .	75

<b>4</b>	<b>Simulation von Irrfahrten auf <math>\mathbb{Z}^d</math></b>	<b>85</b>
4.1	Irrfahrten in einer Dimension . . . . .	85
4.1.1	symmetrische Irrfahrt . . . . .	85
4.1.2	asymmetrische Irrfahrt . . . . .	86
4.2	Irrfahrten in zwei Dimensionen . . . . .	86
4.2.1	symmetrische Irrfahrt . . . . .	86
4.2.2	asymmetrische Irrfahrt . . . . .	92
4.2.3	„Irrfahrt“ mit Betrag 1 . . . . .	92
4.3	Irrfahrten in drei Dimensionen . . . . .	95
4.3.1	symmetrische Irrfahrt . . . . .	95
4.3.2	asymmetrische Irrfahrt . . . . .	95
4.4	Irrfahrten in mehr als drei Dimensionen . . . . .	97
4.5	Irrfahrten mit absorbierendem Rand . . . . .	97
4.6	Irrfahrten mit abstoßendem Rand . . . . .	100
	<b>Literaturverzeichnis</b>	<b>101</b>
	<b>Anhang</b>	<b>103</b>

# Kapitel 1

## Wahrscheinlichkeitstheoretische Grundlagen

Dieses Kapitel beinhaltet die Grundlagen, die für die Behandlung eines wahrscheinlichkeitstheoretischen Themas notwendig sind und die später benötigt werden. Auch Sätze, die in späteren Kapiteln den inhaltlichen Zusammenhang unterbrechen würden, finden sich hier. Gleichzeitig soll dieses Kapitel dazu dienen, mit den in der Arbeit verwendeten Symbolen und Begriffen vertraut zu werden.

Die Sätze, Definitionen etc. wurden in [3], [5], [6], [7] und [12] gefunden. Da der größte Teil aus [5] stammt bzw. schlussendlich in einer ähnlichen Form wie in [5] verwendet wurde, werden nur bei den restlichen Teilen die Quellen explizit angegeben.

### 1.1 Wahrscheinlichkeitsräume

**Definition 1.**  $\sigma$ -Algebra.

Sei  $\Omega \neq \emptyset$ . Ein System  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  mit den Eigenschaften

- (a)  $\Omega \in \mathcal{A}$ ,
- (b)  $A \in \mathcal{A} \Rightarrow A^C := \Omega \setminus A \in \mathcal{A}$  („logische Verneinung“),
- (c)  $A_1, A_2, \dots \in \mathcal{A}$  (abzählbar viele)  $\Rightarrow \bigcup_{i \geq 1} A_i \in \mathcal{A}$  („logisches Oder“)

heißt eine  $\sigma$ -Algebra in  $\Omega$ . Das Paar  $(\Omega, \mathcal{A})$  heißt dann *Ereignisraum* oder *messbarer Raum*.

Im Folgenden wird die Abzählbarkeit meist nicht explizit erwähnt. Schreibweisen wie  $A_1, A_2, \dots$  oder  $(A_i)_{i \geq 0}$  implizieren aber immer (sofern nicht anders angegeben) eine abzählbare Indexmenge.

**Bemerkung 1.** Aus der Definition folgen sofort weitere Eigenschaften für  $A, B \in \mathcal{A}$ :

- $\emptyset \in \mathcal{A}$  weil  $\Omega^C = \emptyset$ ,
- $A \cup B = A \cup B \cup \emptyset \cup \dots \in \mathcal{A}$ ,
- $A \cap B = (A^C \cup B^C)^C \in \mathcal{A}$ ,
- $A \setminus B = A \cap B^C \in \mathcal{A}$ .

**Bemerkung 2.** Ist  $\Omega \neq \emptyset$  und  $\mathcal{G} \subseteq \mathcal{P}(\Omega)$  beliebig, so gibt es genau eine kleinste  $\sigma$ -Algebra  $\mathcal{A} = \sigma(\mathcal{G})$  in  $\Omega$  mit  $\mathcal{A} \supseteq \mathcal{G}$ .

*Beweis.* Wir bezeichnen mit  $\Sigma$  das System aller  $\sigma$ -Algebren  $\mathcal{F}$  in  $\Omega$ , die  $\mathcal{F} \supseteq \mathcal{G}$  erfüllen. Weil  $\mathcal{P}(\Omega) \in \Sigma$  gilt, ist  $\Sigma$  nichtleer. Wir setzen daher  $\mathcal{A} := \bigcap_{\mathcal{F} \in \Sigma} \mathcal{F}$ . Dieses  $\mathcal{A}$  erfüllt die in Definition 1 auf Seite 7 definierten Eigenschaften einer  $\sigma$ -Algebra, da die  $\mathcal{F}$   $\sigma$ -Algebren sind:

- (a)  $\Omega \in \mathcal{A}$ , denn  
 $\Omega \in \mathcal{F} \quad \forall \mathcal{F} \in \Sigma \Rightarrow \Omega \in \bigcap_{\mathcal{F} \in \Sigma} \mathcal{F} = \mathcal{A}$ .
- (b)  $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$ , denn  
 $A \in \mathcal{F} \quad \forall \mathcal{F} \in \Sigma \Rightarrow A^C \in \mathcal{F} \quad \forall \mathcal{F} \in \Sigma \Rightarrow A^C \in \mathcal{A}$ .
- (c)  $A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{i \geq 1} A_i \in \mathcal{A}$ , denn  
 $A_1, A_2, \dots \in \mathcal{F} \quad \forall \mathcal{F} \in \Sigma \Rightarrow \bigcup_{i \geq 1} A_i \in \mathcal{F} \quad \forall \mathcal{F} \in \Sigma \Rightarrow \bigcup_{i \geq 1} A_i \in \mathcal{A}$ .

Also ist  $\mathcal{A}$  in  $\Sigma$  und offenbar auch dessen kleinstes Element. □

**Definition 2.** *Erzeugung von  $\sigma$ -Algebren.*

Dieses  $\mathcal{A} = \sigma(\mathcal{G})$  aus Bemerkung 2 heißt *die von  $\mathcal{G}$  erzeugte  $\sigma$ -Algebra*, und  $\mathcal{G}$  heißt dann ein *Erzeuger von  $\mathcal{A}$* .

**Definition 3.** *Borel'sche  $\sigma$ -Algebra.*

Sei  $\Omega = \mathbb{R}^n$  und

$$\mathcal{G} = \left\{ \prod_{i=1}^n [a_i, b_i] : a_i < b_i \in \mathbb{Q} \right\}$$

das System aller achsenparallelen kompakten Quader in  $\mathbb{R}^n$  mit rationalen Eckpunkten. Dann heißt  $\mathcal{B}^n := \sigma(\mathcal{G})$  die *Borel'sche  $\sigma$ -Algebra* auf  $\mathbb{R}^n$  und jedes  $A \in \mathcal{B}^n$  eine *Borel-Menge*. Statt  $\mathcal{B}^1$  schreiben wir einfach  $\mathcal{B}$ .

**Definition 4.** *Produkt- $\sigma$ -Algebra.*

Sei  $\Omega$  ein kartesisches Produkt von Mengen  $E_i$ , d.h.  $\Omega = \prod_{i \in I} E_i$  für eine Indexmenge  $I \neq \emptyset$ . Sei  $\mathcal{E}_i$  eine  $\sigma$ -Algebra auf  $E_i$ ,  $X_i : \Omega \rightarrow E_i$  die Projektion auf die  $i$ -te Koordinate, und  $\mathcal{G} = \{X_i^{-1}A_i : i \in I, A_i \in \mathcal{E}_i\}$  das System aller Mengen in  $\Omega$ , die durch ein Ereignis in einer einzelnen Koordinate bestimmt sind. Dann heißt  $\bigotimes_{i \in I} \mathcal{E}_i := \sigma(\mathcal{G})$  die *Produkt- $\sigma$ -Algebra* der  $\mathcal{E}_i$  auf  $\Omega$ . Im Fall  $E_i = E$  und  $\mathcal{E}_i = \mathcal{E}$  für alle  $i$  schreibt man auch  $\mathcal{E}^{\otimes I}$  statt  $\bigotimes_{i \in I} \mathcal{E}_i$ .

**Definition 5.** *Wahrscheinlichkeitsmaß.*

Sei  $(\Omega, \mathcal{A})$  ein Ereignisraum. Eine Funktion  $P : \mathcal{A} \rightarrow [0, 1]$  heißt *Wahrscheinlichkeitsmaß* oder (*Wahrscheinlichkeits-*)*Verteilung* auf  $(\Omega, \mathcal{A})$ , falls folgende Eigenschaften (die sogenannten *Kolmogorov-Axiome*) gelten:

(N) Normierung:  $P(\Omega) = 1$

(A)  $\sigma$ -Additivität: Es gilt für abzählbar viele paarweise disjunkte Ereignisse  $A_1, A_2, \dots \in \mathcal{A}$ :

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

Das Tripel  $(\Omega, \mathcal{A}, P)$  heißt dann ein *Wahrscheinlichkeitsraum*.

## 1.2 Eigenschaften von Wahrscheinlichkeitsmaßen

**Proposition 1.** *Für die leere Menge  $\emptyset$  gilt  $P(\emptyset) = 0$ .*

*Beweis.* Wir verwenden die  $\sigma$ -Additivität (A) für abzählbar viele Mengen  $A_i = \emptyset$ . Dann ist  $\bigcup_{i \geq 1} \emptyset = \emptyset$  und damit  $P\left(\bigcup_{i \geq 1} \emptyset\right) = P(\emptyset) = \sum_{i \geq 1} P(\emptyset)$ . Das kann aber nur für  $P(\emptyset) = 0$  erfüllt sein. □

**Bemerkung 3.**

Aus Proposition 1 folgt unmittelbar die endliche  $\sigma$ -Additivität:

Hat man eine endliche Anzahl von Mengen  $A_1, \dots, A_n \in \mathcal{A}$ , so ergänzt man mit der leeren Menge und erhält

$$P\left(\bigcup_{i=1}^n A_i\right) = P\left(\bigcup_{i=1}^n A_i \cup \emptyset \cup \dots\right) = \sum_{i=1}^n P(A_i) + P(\emptyset) + \dots = \sum_{i=1}^n P(A_i).$$

**Satz 1.** Rechenregeln für Wahrscheinlichkeitsmaße. (vgl. [6])

Für jedes Wahrscheinlichkeitsmaß  $P$  auf  $(\Omega, \mathcal{A})$  gelten folgende Rechenregeln:

$$(a) P(A^C) = 1 - P(A) \quad \forall A \in \mathcal{A}.$$

$$(b) P(A \setminus B) = P(A) - P(A \cap B) \quad \forall A, B \in \mathcal{A}.$$

$$(c) P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \forall A, B \in \mathcal{A}.$$

$$(d) P(A) \leq P(B) \quad \forall A, B \in \mathcal{A} \text{ mit } A \subseteq B \text{ (Monotonie)}.$$

$$(e) P\left(\bigcup_{i \geq 1} A_i\right) \leq \sum_{i \geq 1} P(A_i) \quad \forall A_i \in \mathcal{A}$$

( $\sigma$ -Subadditivität bzw. Bonferroni-Ungleichung).

(f) Ist  $A_1 \subseteq A_2 \subseteq \dots$ , dann gilt  $\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{i=1}^{\infty} A_i\right)$  für eine Folge  $(A_n)_{n \in \mathbb{Z}_+}$  von Mengen aus  $\mathcal{A}$  ( $\sigma$ -Stetigkeit von unten).

(g) Ist  $A_1 \supseteq A_2 \supseteq \dots$ , dann gilt  $\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{i=1}^{\infty} A_i\right)$  für eine Folge  $(A_n)_{n \in \mathbb{Z}_+}$  von Mengen aus  $\mathcal{A}$  ( $\sigma$ -Stetigkeit von unten).

*Beweis.* Wir verwenden die Kolmogorov-Axiome (N) und (A):

$$(a) 1 = P(\Omega) = P(A \cup A^C) = P(A) + P(A^C).$$

$$(b) A = (A \setminus B) \cup (A \cap B) \Rightarrow P(A) = P(A \setminus B) + P(A \cap B).$$

$$(c) A \cup B = A \cup (B \setminus A) \Rightarrow P(A \cup B) = P(A) + P(B \setminus A) \\ = P(A) + P(B) - P(A \cap B).$$

$$(d) \text{ Für } A \subseteq B \text{ gilt } B = A \cup B \setminus A. \text{ Daraus folgt } P(B) = P(A \cup B \setminus A) \\ = P(A) + P(B \setminus A) \geq P(A).$$

(e) Wir können  $\bigcup_{i \geq 1} A_i$  darstellen als eine Vereinigung disjunkter Mengen:  
 $\bigcup_{i \geq 1} A_i = A_1 \cup A_2 \setminus A_1 \cup A_3 \setminus (A_1 \cup A_2) \cup \dots$ . Damit gilt:

$$P\left(\bigcup_{i \geq 1} A_i\right) = P\left(\bigcup_{i \geq 1} \left(A_i \setminus \bigcup_{j < i} A_j\right)\right) \\ = \sum_{i \geq 1} P\left(A_i \setminus \bigcup_{j < i} A_j\right) \leq \sum_{i \geq 1} P(A_i).$$

- (f) Wir setzen  $A_0 := \emptyset$  und definieren  $B_n := A_n \setminus A_{n-1}, n \in \mathbb{Z}_+$ . Diese Mengen  $B_n$  sind disjunkt und es gilt damit  $\bigcup_{i \geq 1} A_i = \bigcup_{i \geq 1} B_i$  und  $A_n = B_1 \cup \dots \cup B_n$ . Mit (A) ist dann

$$\begin{aligned} P\left(\bigcup_{i \geq 1} A_i\right) &= P\left(\bigcup_{i \geq 1} B_i\right) = \sum_{i \geq 1} P(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) \\ &= \lim_{n \rightarrow \infty} P(B_1 \cup \dots \cup B_n) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

- (g) Wegen (b) und  $A_1 \supseteq A_n$  ist  $P(A_1 \setminus A_n) = P(A_1) - P(A_n)$ . Außerdem ist  $A_1 \setminus A_n \subseteq A_1 \setminus A_{n+1}$ , weil  $A_n \supseteq A_{n+1}$ . Wir erhalten dann mit (f)

$$\begin{aligned} P(A_1) - P\left(\bigcap_{i \geq 1} A_i\right) &= P\left(A_1 \setminus \bigcap_{i \geq 1} A_i\right) = P\left(\bigcup_{i \geq 1} A_1 \setminus A_i\right) \\ &= \lim_{n \rightarrow \infty} P(A_1 \setminus A_n) = P(A_1) - \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

□

**Satz 2.** Eindeutigkeitsatz.

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum, und es gelte  $\mathcal{A} = \sigma(\mathcal{G})$  für ein Erzeugendensystem  $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ . Ist  $\mathcal{G}$   $\cap$ -stabil (d. h. mit  $A, B \in \mathcal{G}$  ist auch  $A \cap B \in \mathcal{G}$ ), so ist  $P$  bereits durch seine Einschränkung  $P|_{\mathcal{G}}$  eindeutig bestimmt.

*Beweis.*

Sei  $Q$  ein beliebiges Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  mit  $P|_{\mathcal{G}} = Q|_{\mathcal{G}}$ , und sei weiter  $\mathcal{D} = \{A \in \mathcal{A} : P(A) = Q(A)\}$  (insbesondere ist  $\mathcal{D} \subseteq \mathcal{A}$ ). Dann gilt:

- (a)  $\Omega \in \mathcal{D}$ ,
- (b) Sind  $A, B \in \mathcal{D}$  und  $A \subseteq B$ , so gilt  $B \setminus A \in \mathcal{D}$ .
- (c) Sind  $A_1, A_2, \dots \in \mathcal{D}$  paarweise disjunkt, so ist  $\bigcup_{i \geq 1} A_i \in \mathcal{D}$ .

Dabei folgt (a) aus dem Kolmogorov-Axiom (N) und (c) folgt aus (A). Eigenschaft (b) folgt daraus, dass  $P(B \setminus A) = P(B) - P(A)$  für  $A \subseteq B$  gilt. So ein System  $\mathcal{D}$  heißt Dynkin-System. Wegen der Voraussetzung  $P|_{\mathcal{G}} = Q|_{\mathcal{G}}$  gilt  $\mathcal{D} \supseteq \mathcal{G}$ . Definiert man das von  $\mathcal{G}$  erzeugte Dynkin-System  $d(\mathcal{G})$  so wie in Bemerkung 2 auf Seite 8 als kleinstes Dynkin-System, welches  $\mathcal{G}$  umfasst, dann umfasst  $\mathcal{D}$  natürlich  $d(\mathcal{D})$ , also  $\mathcal{D} \supseteq d(\mathcal{G})$ .

Unter der Annahme, dass  $d(\mathcal{G}) = \sigma(\mathcal{G})$  gilt, erhalten wir  $\mathcal{D} \supseteq \mathcal{A}$ , denn  $\sigma(\mathcal{G}) = \mathcal{A}$  laut Voraussetzung. Insgesamt ist dann  $\mathcal{D} = \mathcal{A}$  und damit auch  $P = Q$ . □

Für die Vollständigkeit des Beweises fehlt noch der Beweis von  $d(\mathcal{G}) = \sigma(\mathcal{G})$ :

**Proposition 2.** Erzeugtes Dynkin-System.

Für ein  $\cap$ -stabiles Mengensystem  $\mathcal{G}$  gilt  $d(\mathcal{G}) = \sigma(\mathcal{G})$ .

*Beweis.* So wie jede  $\sigma$ -Algebra ist auch  $\sigma(\mathcal{G})$  ein Dynkin-System. Aufgrund der Minimalität von  $d(\mathcal{G})$  folgt daraus schon  $\sigma(\mathcal{G}) \supseteq d(\mathcal{G})$ . Für die Gleichheit ist also noch  $\sigma(\mathcal{G}) \subseteq d(\mathcal{G})$  zu zeigen. Dies ist erfüllt, falls  $d(\mathcal{G})$  eine  $\sigma$ -Algebra ist, weil  $\sigma(\mathcal{G})$  minimal ist.

Zuerst zeigen wir, dass  $d(\mathcal{G})$   $\cap$ -stabil ist:

Wir definieren  $\mathcal{D}_1 := \{A \subseteq \Omega : A \cap B \in d(\mathcal{G}) \text{ für alle } B \in \mathcal{G}\}$ . Wie sich leicht überprüfen lässt, ist  $\mathcal{D}_1$  ein Dynkin-System. Weil  $\mathcal{G}$   $\cap$ -stabil ist, ist  $\mathcal{D}_1 \supseteq \mathcal{G}$ . (Alle  $A \in \mathcal{G}$  erfüllen  $A \cap B \in d(\mathcal{G})$  weil  $B \in \mathcal{G}$  und  $\mathcal{G} \subseteq d(\mathcal{G})$  gilt.) Weil  $d(\mathcal{G})$  das kleinste  $\mathcal{G}$  umfassende Dynkin-System ist, folgt  $\mathcal{D}_1 \supseteq d(\mathcal{G})$ . Für alle  $A \in d(\mathcal{G})$  und  $B \in \mathcal{G}$  gilt also  $A \cap B \in d(\mathcal{G})$ .

Auch  $\mathcal{D}_2 := \{A \subseteq \Omega : A \cap B \in d(\mathcal{G}) \text{ für alle } B \in d(\mathcal{G})\}$  ist ein Dynkin-System und es gilt  $\mathcal{D}_2 \supseteq \mathcal{D}_1 \supseteq d(\mathcal{G})$ , weil  $\mathcal{G} \subseteq d(\mathcal{G})$ . Das heißt aber, dass für alle  $A$  und  $B \in d(\mathcal{G})$  gilt:  $A \cap B \in d(\mathcal{G})$ . Also ist  $d(\mathcal{G})$   $\cap$ -stabil.

Wir müssen nun noch zeigen, dass mit  $A_1, A_2, \dots \in d(\mathcal{G})$  auch  $\bigcup_{i \geq 1} A_i \in d(\mathcal{G})$  gilt. Das heißt wir zeigen, dass die für Dynkin-Systeme nötige Einschränkung auf paarweise disjunkte  $A_1, A_2, \dots$  in diesem Fall nicht notwendig ist.

Dazu definieren wir

$$B_i := A_i \setminus \bigcup_{j < i} A_j = A_i \cap \bigcap_{j < i} \Omega \setminus A_j.$$

Die zweite Gleichheit folgt im Wesentlichen aus den „de Morganschen Regeln“. Weil  $d(\mathcal{G})$  ein Dynkin-System ist, ist mit  $\Omega$  und  $A_j \in d(\mathcal{G})$  auch  $\Omega \setminus A_j$  in  $d(\mathcal{G})$ . Da  $d(\mathcal{G})$   $\cap$ -stabil ist, ist daher auch  $B_i$  in  $d(\mathcal{G})$ . Weil die  $B_i$  paarweise disjunkt sind, ist die Vereinigung  $\bigcup_{i \geq 1} B_i$  in  $d(\mathcal{G})$ . Es ist aber auch  $\bigcup_{i \geq 1} A_i = \bigcup_{i \geq 1} B_i$  und daher  $\bigcup_{i \geq 1} A_i \in d(\mathcal{G})$ , und  $d(\mathcal{G})$  ist eine  $\sigma$ -Algebra.

Insgesamt haben wir  $\sigma(\mathcal{G}) \supseteq d(\mathcal{G})$  und  $\sigma(\mathcal{G}) \subseteq d(\mathcal{G})$  gezeigt. Daraus folgt  $\sigma(\mathcal{G}) = d(\mathcal{G})$ .  $\square$

**Bemerkung 4.** Die Abbildung  $\lambda^n : \mathcal{B}^n \rightarrow [0, \infty]$ , die jedem  $A \in \mathcal{B}^n$  sein  $n$ -dimensionales Volumen

$$\lambda^n(A) := \int 1_A(x) dx$$

zuordnet, erfüllt die  $\sigma$ -Additivitätseigenschaft (A), und es gilt  $\lambda^n(\emptyset) = 0$ . Folglich ist  $\lambda^n$  ein „Maß“ auf  $(\mathbb{R}^n, \mathcal{B}^n)$ .

**Definition 6.** *Lebesgue-Maß.*

Das Maß  $\lambda^n$  aus Bemerkung 4 heißt (n-dimensionales) *Lebesgue-Maß auf  $\mathbb{R}^n$* . Für  $\Omega \in \mathcal{B}^n$  heißt die Einschränkung  $\lambda_\Omega^n$  von  $\lambda^n$  auf  $\Omega$  das *Lebesgue-Maß auf  $\Omega$* .

**Definition 7.** *Zähldichte.*

Jede Folge  $(\varrho(\omega))_{\omega \in \Omega}$  in  $[0, 1]$  heißt eine *Zähldichte*, wenn gilt

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

## 1.3 Zufallsvariablen

**Definition 8.** *Zufallsvariablen.*

Seien  $(\Omega, \mathcal{A})$  und  $(\Omega', \mathcal{A}')$  zwei Ereignisräume. Jede Abbildung  $X : \Omega \rightarrow \Omega'$  mit der Eigenschaft

$$A' \in \mathcal{A}' \Rightarrow X^{-1}A' \in \mathcal{A} \quad (1.1)$$

heißt eine *Zufallsvariable von  $(\Omega, \mathcal{A})$  nach  $(\Omega', \mathcal{A}')$* .

**Definition 9.** *Diskrete Zufallsvariable.*

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $X : \Omega \rightarrow \mathbb{R}$  eine reelle Zufallsvariable. Wenn die Wertemenge  $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$  höchstens abzählbar ist, dann heißt  $X$  *diskret*.

**Proposition 3.** *Messbarkeitskriterium.*

*In der Situation von Definition 8 werde  $\mathcal{A}'$  erzeugt von einem Mengensystem  $\mathcal{G}'$ , d.h. es sei  $\mathcal{A}' = \sigma(\mathcal{G}')$ . Dann ist  $X : \Omega \rightarrow \Omega'$  bereits dann eine Zufallsvariable, wenn die Bedingung  $X^{-1}A' \in \mathcal{A}$  nur für alle  $A' \in \mathcal{G}'$  gilt.*

*Beweis.* Das System  $\mathcal{F}' := \{A' \subseteq \Omega' : X^{-1}A' \in \mathcal{A}\}$  („Menge aller Teilmengen von  $\Omega'$  mit Urbild in  $\mathcal{A}$ “) ist eine  $\sigma$ -Algebra, die nach Voraussetzung  $\mathcal{G}'$  umfasst: Wir haben vorausgesetzt, dass für alle  $A' \in \mathcal{G}'$  das Urbild  $X^{-1}A'$  in  $\mathcal{A}$  liegt. Da nach Voraussetzung  $\mathcal{A}'$  die kleinste  $\mathcal{G}'$  umfassende  $\sigma$ -Algebra ist, gilt auch  $\mathcal{F}' \supseteq \mathcal{A}'$ . Für alle  $A' \in \mathcal{A}'$  gilt also die Bedingung für die Elemente in  $\mathcal{A}'$  und (1.1) ist erfüllt.  $\square$

**Satz 3.** *Verteilung einer Zufallsvariablen.*

*Ist  $X$  eine Zufallsvariable von einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  in einen Ereignisraum  $(\Omega', \mathcal{A}')$ , so wird durch*

$$P'(A') := P(X^{-1}A') = P(\{X \in A'\}) \text{ für } A' \in \mathcal{A}'$$

*ein Wahrscheinlichkeitsmaß  $P'$  auf  $(\Omega', \mathcal{A}')$  definiert.*

Zur Vereinfachung der Schreibweise schreibt man statt  $P(\{X \in A'\})$  meist  $P(X \in A')$ .

*Beweis.* Wegen (1.1) ist die Definition von  $P'$  sinnvoll. Ferner erfüllt  $P'$  die Bedingungen (N) und (A):

Es ist  $P'(\Omega') = P(X \in \Omega') = P(\Omega) = 1$ .

Sind  $A'_1, A'_2, \dots \in \mathcal{A}'$  paarweise disjunkt, so sind auch die Urbilder  $X^{-1}A'_1, X^{-1}A'_2, \dots$  paarweise disjunkt, und deshalb gilt:

$$\begin{aligned} P'\left(\bigcup_{i \geq 1} A'_i\right) &= P\left(X^{-1}\bigcup_{i \geq 1} A'_i\right) = P\left(\bigcup_{i \geq 1} X^{-1}A'_i\right) \\ &= \sum_{i \geq 1} P(X^{-1}A'_i) = \sum_{i \geq 1} P'(A'_i). \end{aligned}$$

Also ist  $P'$  ein Wahrscheinlichkeitsmaß. □

### Definition 10.

- (a) Das Wahrscheinlichkeitsmaß  $P'$  in Satz 3 auf Seite 13 heißt die *Verteilung von  $X$  bei  $P$*  oder das *Bild von  $P$  unter  $X$*  und wird mit  $P \circ X^{-1}$  bezeichnet.
- (b) Zwei Zufallsvariablen heißen *identisch verteilt*, wenn sie dieselbe Verteilung haben.

## 1.4 Eine wichtige Verteilung - die Multinomialverteilung

### Definition 11. Multinomialkoeffizient.

Sei  $n \in \mathbb{Z}_+$  und  $\mathbf{k} = (k_a)_{a \in E} \in \mathbb{Z}_+^E$ . Wir schreiben

$$\binom{n}{\mathbf{k}} = \begin{cases} \frac{n!}{\prod_{a \in E} k_a!} & \text{falls } \sum_{a \in E} k_a = n, \\ 0 & \text{sonst} \end{cases}$$

für den *Multinomialkoeffizienten*.

Der Multinomialkoeffizient beschreibt die Anzahl der Möglichkeiten, bei der Auswahl von  $n$  Elementen (mit Zurücklegen)  $k_a$ -mal das Element  $a$  zu wählen für alle  $a \in E$ .

Wir definieren die Ergebnismenge  $\Omega = \{\mathbf{k} = (k_a)_{a \in E} \in \mathbb{Z}_+^E : \sum_{a \in E} k_a = n\}$ . (Fett gedruckte Buchstaben symbolisieren Vektoren.)

**Definition 12.** *Multinomialverteilung.*

Für jede Zähldichte  $\varrho$  auf  $E$  heißt das Wahrscheinlichkeitsmaß  $M_{n,\varrho}$  auf  $(\Omega, \mathcal{P}(\Omega))$  mit Zähldichte

$$M_{n,\varrho}(\{\mathbf{k}\}) = \binom{n}{\mathbf{k}} \cdot \prod_{a \in E} \varrho(a)^{k_a} \quad (1.2)$$

die *Multinomialverteilung* für  $n$  Stichproben mit Ergebniswahrscheinlichkeiten  $\varrho(a)$ ,  $a \in E$ .

Da  $M_{n,\varrho}(\{\mathbf{k}\})$  eine Zähldichte ist, gilt  $\sum_{\mathbf{k}} M_{n,\varrho}(\{\mathbf{k}\}) = 1$ .

## 1.5 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

**Proposition 4.** *Neubewertung von Ereignissen.*

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $A \in \mathcal{A}$  mit  $P(A) > 0$ . Dann gibt es genau ein Wahrscheinlichkeitsmaß  $P_A$  auf  $(\Omega, \mathcal{A})$  mit den Eigenschaften

- (a)  $P_A(A) = 1$ , d.h. das Ereignis  $A$  ist jetzt sicher.
- (b) Die neue Bewertung der Teilereignisse von  $A$  ist proportional zu ihrer ursprünglichen Bewertung, d.h. es existiert eine Konstante  $c_A > 0$  mit  $P_A(B) = c_A \cdot P(B)$  für alle  $B \in \mathcal{A}$  mit  $B \subseteq A$ .

nämlich

$$P_A(B) := \frac{P(A \cap B)}{P(A)} \text{ für } B \in \mathcal{A}$$

*Beweis.*  $P_A$  erfülle (a) und (b). Dann gilt für alle  $B \in \mathcal{A}$

$$P_A(B) = P_A(A \cap B) + P_A(B \setminus A) = c_A \cdot P(A \cap B),$$

denn wegen (a) ist  $P_A(B \setminus A) = 0$ . Für  $B = A$  folgt  $1 = P_A(A) = c_A \cdot P(A)$ , also  $c_A = 1/P(A)$ . Somit hat  $P_A$  die angegebene Gestalt.

Umgekehrt: Es ist klar, dass  $P_A$  (a) und (b) erfüllt, denn  $P(A \cap A) = P(A)$  und  $P(A \cap B) = P(B)$  wenn  $B \subseteq A$ .  $\square$

**Definition 13.** *Bedingte Wahrscheinlichkeit.*

In der Situation von Proposition 4 heißt für jedes  $B \in \mathcal{A}$

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

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

**Satz 4.** Fallunterscheidungs- und Bayes-Formel.

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $\Omega = \bigcap_{i \in I} B_i$  eine höchstens abzählbare Zerlegung von  $\Omega$  in paarweise disjunkte Ereignisse  $B_i \in \mathcal{A}$ . Dann gilt

(a) die Fallunterscheidungsformel: Für alle  $A \in \mathcal{A}$  gilt

$$P(A) = \sum_{i \in I} P(B_i) \cdot P(A|B_i).$$

(b) die Formel von Bayes: Für alle  $A \in \mathcal{A}$  mit  $P(A) > 0$  und alle  $k \in I$  gilt

$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{i \in I} P(B_i) \cdot P(A|B_i)}.$$

*Beweis.* (a) Aus der Definition 13 auf Seite 15 der bedingten Wahrscheinlichkeit und der  $\sigma$ -Additivität von  $P$  folgt

$$\sum_{i \in I} P(B_i) \cdot P(A|B_i) = \sum_{i \in I} P(A \cap B_i) = P(A).$$

(b) folgt aus (a) und der Definition:

$$\frac{P(B_k) \cdot P(A|B_k)}{\sum_{i \in I} P(B_i) \cdot P(A|B_i)} = \frac{P(B_k \cap A)}{P(A)} = P(B_k|A).$$

□

**Proposition 5.** Multiplikationsformel.

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $A_1, \dots, A_n \in \mathcal{A}$ . Dann gilt

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

*Beweis.* Verschwindet die linke Seite, ist auch der letzte Faktor rechts gleich null (wegen Definition 13 auf Seite 15). Andernfalls sind alle bedingten Wahrscheinlichkeiten rechts definiert und ungleich null. Sie bilden ein Teleskopprodukt; aufeinanderfolgende Zähler und Nenner heben sich beim Einsetzen der Definition 13 jeweils weg und nur der Ausdruck auf der linken Seite bleibt übrig:

$$\begin{aligned} & P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdots \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})} \\ &= P(A_1 \cap \dots \cap A_n). \end{aligned}$$

□

**Satz 5.** Konstruktion von Wahrscheinlichkeitsmaßen durch bedingte Wahrscheinlichkeiten.

Gegeben seien  $n$  abzählbare Ergebnisräume  $\Omega_1, \dots, \Omega_n \neq \emptyset, n \geq 2$ . Sei  $\varrho_1$  eine Zähldichte auf  $\Omega_1$ , und für  $k = 2, \dots, n$  und beliebige  $\omega_i \in \Omega_i$  mit  $i < k$  sei  $\varrho_{k|\omega_1, \dots, \omega_{k-1}}$  eine Zähldichte auf  $\Omega_k$ . Sei ferner  $\Omega = \prod_{i=1}^n \Omega_i$  der Produktraum und  $X_i : \Omega \rightarrow \Omega_i$  die  $i$ -te Projektion. Dann existiert genau ein Wahrscheinlichkeitsmaß  $P$  auf  $(\Omega, \mathcal{P}(\Omega))$  mit den Eigenschaften

(a) Für alle  $\omega_1 \in \Omega_1$  gilt  $P(X_1 = \omega_1) = \varrho_1(\omega_1)$ .

(b) Für alle  $k = 2, \dots, n$  und alle  $\omega_i \in \Omega_i$  gilt

$$P(X_k = \omega_k | X_1 = \omega_1, \dots, X_{k-1} = \omega_{k-1}) = \varrho_{k|\omega_1, \dots, \omega_{k-1}}(\omega_k),$$

sofern  $P(X_1 = \omega_1, \dots, X_{k-1} = \omega_{k-1}) \geq 0$ .

$P$  ist gegeben durch

$$P(\{\omega\}) = \varrho_1(\omega_1) \cdot \varrho_{2|\omega_1}(\omega_2) \cdot \varrho_{3|\omega_1, \omega_2}(\omega_3) \cdots \varrho_{n|\omega_1, \dots, \omega_{n-1}}(\omega_n) \quad (1.3)$$

für  $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ .

*Beweis.* Zuerst verifizieren wir, dass  $P$  nicht anders als durch (1.3) definiert werden kann. Es ist  $\{\omega\} = \bigcap_{i=1}^n \{X_i = \omega_i\}$ .

Die Multiplikationsformel für Ereignisse  $A_i$  besagt, dass

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

gilt. Die bedingten Wahrscheinlichkeiten auf der rechten Seite der Gleichung sind für die Ereignisse  $A_i = \{X_i = \omega_i\}$  aus (a) und (b) bekannt. Ersetzt man die  $A_i$  in der Multiplikationsformel durch  $\{X_i = \omega_i\}$ , erhält man die Gleichung in (1.3).

Somit ist die Eindeutigkeit von  $P$  bewiesen.

Nun sei  $P$  durch (1.3) definiert. Uns interessiert die Wahrscheinlichkeit des Ereignisses  $\{X_1 = \omega_1, \dots, X_k = \omega_k\}$ . Da nur die  $\omega_1, \dots, \omega_k$  fest sind, muss über alle  $\omega_{k+1}, \dots, \omega_n$  summiert werden, um die gesuchte Wahrscheinlichkeit zu erhalten. Für alle  $1 \leq k \leq n$  und  $\omega_1, \dots, \omega_k$  folgt also durch Summation

über  $\omega_{k+1}, \dots, \omega_n$

$$\begin{aligned}
P(X_1 = \omega_1, \dots, X_k = \omega_k) &= \sum_{\omega_{k+1} \in \Omega_{k+1}, \dots, \omega_n \in \Omega_n} P(\{(\omega_1, \dots, \omega_n)\}) \\
&= \varrho_1(\omega_1) \cdots \varrho_{k|\omega_1, \dots, \omega_{k-1}}(\omega_k) \cdot \\
&\quad \sum_{\omega_{k+1} \in \Omega_{k+1}} \varrho_{k+1|\omega_1, \dots, \omega_k}(\omega_{k+1}) \cdots \sum_{\omega_n \in \Omega_n} \varrho_{n|\omega_1, \dots, \omega_{n-1}}(\omega_n) \\
&= \varrho_1(\omega_1) \cdots \varrho_{k|\omega_1, \dots, \omega_{k-1}}(\omega_k).
\end{aligned}$$

Im zweiten Schritt wurde (1.3) eingesetzt, die von der Summe nicht betroffenen Faktoren herausgehoben und die Summation aufgespalten. Die Summen können nun von innen nach außen ausgewertet werden. Laut Voraussetzung von Satz 5 ist  $\varrho_{n|\omega_1, \dots, \omega_{n-1}}$  eine Zähldichte, die Summe hat also den Wert 1. Dasselbe ergibt sich für die anderen Summen. Die gesamte Mehrfachsumme ist daher gleich 1. Für  $k = 1$  folgt (a). Für  $k > 1$  ergibt sich (durch vollständige Induktion)

$$P(X_1 = \omega_1, \dots, X_k = \omega_k) = P(X_1 = \omega_1, \dots, X_{k-1} = \omega_{k-1}) \cdot \varrho_{k|\omega_1, \dots, \omega_{k-1}}(\omega_k). \quad (1.4)$$

Bekanntermaßen gilt für zwei Ereignisse  $A$  und  $B$ :

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

Das Ereignis  $A$  ist hier  $\{X_1 = \omega_1, \dots, X_{k-1} = \omega_{k-1}\}$ ;  $P(A)$  ist damit dann  $P(X_1 = \omega_1, \dots, X_{k-1} = \omega_{k-1})$ , der erste Faktor auf der rechten Seite in (1.4). Ereignis  $B$  ist  $\{X_k = \omega_k\}$ . Der Durchschnitt  $A \cap B$  ist dann also das Ereignis  $\{X_1 = \omega_1, \dots, X_k = \omega_k\}$ ;  $P(A \cap B)$  ist  $P(X_1 = \omega_1, \dots, X_k = \omega_k)$ , die linke Seite in (1.4). Der bedingten Wahrscheinlichkeit auf der rechten Seite  $P(X_k = \omega_k | X_1 = \omega_1, \dots, X_{k-1} = \omega_{k-1})$  bleibt also nichts anderes mehr übrig, als  $\varrho_{k|\omega_1, \dots, \omega_{k-1}}(\omega_k)$  zu sein. Das ist aber genau Eigenschaft (b).  $\square$

**Satz 6.** Konstruktion von Wahrscheinlichkeitsmaßen auf unendlichen Produkträumen.

Zu jedem  $i \in \mathbb{Z}_+$  sei  $\Omega_i \neq \emptyset$  eine abzählbare Menge. Sei  $\varrho_1$  eine Zähldichte auf  $\Omega_1$ , und für alle  $k \geq 2$  und  $\omega_i \in \Omega_i$  mit  $i < k$  sei  $\varrho_{k|\omega_1, \dots, \omega_{k-1}}$  eine Zähldichte auf  $\Omega_k$ . Sei  $\Omega = \prod_{i \geq 1} \Omega_i$ ,  $X_i : \Omega \rightarrow \Omega_i$  die Projektion auf die  $i$ -te Koordinate, und  $\mathcal{A} = \bigotimes_{i \geq 1} \mathcal{P}(\Omega_i)$  die Produkt- $\sigma$ -Algebra auf  $\Omega$ . Dann existiert genau ein Wahrscheinlichkeitsmaß  $P$  auf  $(\Omega, \mathcal{A})$  mit der Eigenschaft

$$P(X_1 = \omega_1, \dots, X_k = \omega_k) = \varrho_1(\omega_1) \cdot \varrho_{2|\omega_1}(\omega_2) \cdots \varrho_{k|\omega_1, \dots, \omega_{k-1}}(\omega_k) \quad (1.5)$$

für alle  $k \geq 1$  und  $\omega_i \in \Omega_i$ .

Gleichung (1.5) entspricht der Gleichung (1.3) in Satz 5 auf Seite 17 und ist äquivalent zu den dortigen Bedingungen (a) und (b).

*Beweis.* Die Eindeutigkeit folgt aus dem Eindeutigkeitssatz, welcher besagt, dass auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathcal{P})$  mit  $\cap$ -stabilem Erzeugendensystem  $\mathcal{G} \subseteq \mathcal{P}(\Omega)$   $P$  bereits durch die Einschränkung  $P|_{\mathcal{G}}$  eindeutig bestimmt ist.

Der  $\cap$ -stabile Erzeuger von  $\mathcal{A}$  ist in diesem Fall

$$\mathcal{G} = \{ \{X_1 = \omega_1, \dots, X_k = \omega_k\} : k \geq 1, \omega_i \in \Omega_i \} \cup \{ \emptyset \}.$$

Für die Existenz nutzen wir die Existenz des Lebesgue-Maßes  $\lambda = \mathcal{U}_{[0,1]}$  auf dem halboffenen Einheitsintervall  $[0, 1)$ . ( $\mathcal{U}_{\Omega}$  ist dabei die Gleichverteilung auf  $\Omega$ .) Dieses Intervall  $[0, 1)$  zerlegen wir in halboffene Intervalle  $(I_{\omega_1})_{\omega_1 \in \Omega_1}$  der Länge  $\varrho_1(\omega_1)$ ; wir nennen sie Intervalle der ersten Stufe. Jedes dieser  $I_{\omega_1}$  zerlegen wir wieder, und zwar in halboffene Intervalle  $(I_{\omega_1 \omega_2})_{\omega_2 \in \Omega_2}$  der Länge  $\varrho_1(\omega_1) \cdot \varrho_2|_{\omega_1}(\omega_2)$ , die Intervalle der zweiten Stufe. So machen wir weiter, das heißt, wenn ein Intervall  $I_{\omega_1 \dots \omega_{k-1}}$  der  $k-1$ -ten Stufe bereits definiert ist, so zerlegen wir es weiter in disjunkte Teilintervalle  $(I_{\omega_1 \dots \omega_k})_{\omega_k \in \Omega_k}$  der  $k$ -ten Stufe mit der Länge  $\lambda(I_{\omega_1 \dots \omega_{k-1}}) \varrho_k|_{\omega_1, \dots, \omega_{k-1}}(\omega_k)$ .

Für  $x \in [0, 1)$  gibt es zu jedem  $k$  genau ein Intervall der  $k$ -ten Stufe, welches  $x$  enthält. Das heißt es gibt genau eine Folge  $Z(x) = (Z_1(x), Z_2(x), \dots)$  in  $\Omega$  mit  $x \in I_{Z_1(x) \dots Z_k(x)}$  für alle  $k \geq 1$ . Die Abbildung  $Z : [0, 1) \rightarrow \Omega$  ist eine Zufallsvariable: Für  $A = \{X_1 = \omega_1, \dots, X_k = \omega_k\} \in \mathcal{G}$  ist nämlich

$$Z^{-1}A = \{x : Z_1(x) = \omega_1, \dots, Z_k(x) = \omega_k\} = I_{\omega_1 \dots \omega_k} \in \mathcal{B}_{[0,1)},$$

so dass die Behauptung aus Proposition 3 auf Seite 13 folgt. Nach Satz 3 auf Seite 13 ist daher  $P := \lambda \circ Z^{-1}$  ein wohldefiniertes Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$ , und dieses hat nach Konstruktion die verlangte Eigenschaft.  $\square$

**Definition 14.** *Unabhängigkeit.*

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Zwei Ereignisse  $A, B \in \mathcal{A}$  heißen (stochastisch) *unabhängig* bezüglich  $P$ , wenn  $P(A \cap B) = P(A) \cdot P(B)$ .

**Definition 15.**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $I \neq \emptyset$  eine beliebige Indexmenge. Eine Familie  $(A_i)_{i \in I}$  von Ereignissen in  $\mathcal{A}$  heißt *unabhängig* bezüglich  $P$ , wenn für jede endliche Teilmenge  $\emptyset \neq J \subseteq I$  gilt:

$$P \left( \bigcap_{i \in J} A_i \right) = \prod_{i \in J} P(A_i).$$

**Definition 16.**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $(\Omega', \mathcal{A}')$  ein Ereignisraum und  $I \neq \emptyset$  eine beliebige Indexmenge. Eine Familie  $(X_i)_{i \in I}$  von Zufallsvariablen heißt *unabhängig*, falls für jede endliche Teilmenge  $\emptyset \neq J \subseteq I$  und für alle  $(A'_i)_{i \in J} \in \mathcal{A}'$  gilt:

$$P\left(\bigcap_{i \in J} X_i^{-1} A'_i\right) = \prod_{i \in J} P(X_i^{-1} A'_i).$$

**Definition 17. Limes Superior und Inferior.** (vgl. [7])

Sei  $(A_k)_{k \in \mathbb{Z}_+}$  eine Folge von Ereignissen. Dann bezeichnet man

$$\begin{aligned} \limsup A_k &= \bigcap_{k \in \mathbb{Z}_+} \bigcup_{m \geq k} A_m \text{ als } \textit{Limes Superior} \text{ und} \\ \liminf A_k &= \bigcup_{k \in \mathbb{Z}_+} \bigcap_{m \geq k} A_m \text{ als } \textit{Limes Inferior}. \end{aligned}$$

Diese Definition impliziert

$$\begin{aligned} \liminf A_k &\subseteq \limsup A_k, \\ \omega \in \limsup A_k &\text{ genau dann, wenn } \omega \in A_k \text{ für unendlich viele } k, \\ \omega \in \liminf A_k &\text{ genau dann, wenn } \omega \in A_k \text{ für fast alle } k. \end{aligned}$$

**Satz 7. Lemma von Borel–Cantelli.**

Sei  $(A_k)_{k \geq 1}$  eine Folge von Ereignissen in einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  und

$$A := \{\omega \in \Omega : \omega \in A_k \text{ für unendlich viele } k\} = \limsup_{k \rightarrow \infty} A_k.$$

(a) Ist  $\sum_{k \geq 1} P(A_k) < \infty$ , so ist  $P(A) = 0$ .

(b) Ist  $\sum_{k \geq 1} P(A_k) = \infty$  und  $(A_k)_{k \geq 1}$  unabhängig, so ist  $P(A) = 1$ .

Zu beachten ist, dass (a) die Unabhängigkeit nicht benötigt.

*Beweis.* (a) Weil  $A \subseteq \bigcup_{m \geq k} A_m$  für alle  $k$  gilt, ist auch  $P(A) \leq \sum_{m \geq k} P(A_m)$  für alle  $k$ . Wenn  $\sum_{k \geq 1} P(A_k) < \infty$ , strebt  $\sum_{m \geq k} P(A_m)$  gegen 0 für  $k \rightarrow \infty$ . (b) Den Limes Superior von  $A_k$  kann man auch schreiben in der Form  $A = \bigcap_{k \geq 1} \bigcup_{m \geq k} A_m$ . Aufgrund der „de Morganschen Regeln“ ist dann das

Komplement  $A^C = \bigcup_{k \geq 1} \bigcap_{m \geq k} A_m^C$ . Dann ist

$$\begin{aligned}
 P(A^C) &\leq \sum_{k \geq 1} P\left(\bigcap_{m \geq k} A_m^C\right) = \sum_{k \geq 1} \lim_{n \rightarrow \infty} P\left(\bigcap_{m=k}^n A_m^C\right) \\
 &= \sum_{k \geq 1} \lim_{n \rightarrow \infty} \prod_{m=k}^n P(A_m^C) \\
 &= \sum_{k \geq 1} \lim_{n \rightarrow \infty} \prod_{m=k}^n [1 - P(A_m)] \\
 &\leq \sum_{k \geq 1} \lim_{n \rightarrow \infty} \exp\left[-\sum_{m=k}^n P(A_m)\right] = \sum_{k \geq 1} 0 = 0,
 \end{aligned}$$

wenn  $\sum_{m \geq 1} P(A_m) = \infty$ . Im Beweis wurde verwendet, dass auch die Ereignisse  $(A_m^C)_{m \geq 1}$  unabhängig sind, wenn  $(A_m)_{m \geq 1}$  unabhängig sind (ohne Beweis). Außerdem wurde  $1 - x \leq e^{-x}$  ausgenutzt.  $\square$

## 1.6 Erwartungswert und Eigenschaften

### 1.6.1 Erwartungswert

**Definition 18.** *Erwartungswert einer diskreten Zufallsvariablen.*

Sei  $X$  eine diskrete Zufallsvariable. Falls  $\sum_{x \in X(\Omega)} |x| \cdot P(X = x) < \infty$ , so sagt man  $X$  besitzt einen Erwartungswert.

Die Summe

$$\mathbb{E}(X) = \mathbb{E}_P(X) = \sum_{x \in X(\Omega)} x \cdot P(X = x)$$

ist dann wohldefiniert und heißt *Erwartungswert* von  $X$ . Wir schreiben dann  $X \in \mathcal{L}^1(P)$  bzw.  $X \in \mathcal{L}^1$ , wenn  $P$  nicht hervorgehoben zu werden braucht.

**Proposition 6.** Erwartungswert der Indikatorfunktion.

Für  $A \in \mathcal{A}$  ist  $1_A \in \mathcal{L}^1(P)$  und  $\mathbb{E}(1_A) = P(A)$ .

*Beweis.* Es gilt  $\mathbb{E}(1_A) = 0 \cdot P(1_A = 0) + 1 \cdot P(1_A = 1) = P(A)$ .  $\square$

**Satz 8.** Rechenregeln für Erwartungswerte.

Es seien  $X, Y, X_n, Y_n : \Omega \rightarrow \mathbb{R}$  diskrete Zufallsvariablen in  $\mathcal{L}^1$ . Dann gilt:

(a) *Monotonie:*

Ist  $X \leq Y$ , so ist  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .

(b) *Linearität:*

Für alle  $c \in \mathbb{R}$  ist  $cX + Y \in \mathcal{L}^1$  und es gilt  $\mathbb{E}(cX + Y) = c\mathbb{E}(X) + \mathbb{E}(Y)$ .

(c)  *$\sigma$ -Additivität und monotone Konvergenz:*

Gilt  $X_n > 0$  und ist  $X = \sum_{n \geq 1} X_n$  so gilt  $\mathbb{E}(X) = \sum_{n \geq 1} \mathbb{E}(X_n)$ . Wenn  $Y_n$  monoton gegen  $Y$  konvergiert, folgt  $\mathbb{E}(Y) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_n)$ .

(d) *Produktregel bei Unabhängigkeit:*

Sind  $X, Y$  unabhängig, so ist  $XY \in \mathcal{L}^1$  und  $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$ .

*Beweis.*

(a) Aus der  $\sigma$ -Additivität und aus Definition 18 auf Seite 21 folgt

$$\mathbb{E}(X) = \sum_{x \in X(\Omega), y \in Y(\Omega)} x \cdot P(X = x, Y = y).$$

Weil aber aufgrund der Voraussetzung nur für  $x \leq y$  die Wahrscheinlichkeit  $P(X = x, Y = y)$  ungleich 0 ist, ist obige Summe sicher kleiner gleich

$$\sum_{x \in X(\Omega), y \in Y(\Omega)} y \cdot P(X = x, Y = y) = \mathbb{E}(Y).$$

(b) Auch  $cX + Y$  ist eine diskrete Zufallsvariable und  $cX + Y \in \mathcal{L}^1$ :

$$\begin{aligned} \sum_z |z| \cdot P(cX + Y = z) &= \sum_{x,z} |z| \cdot P(cX = cx, Y = z - cx) \\ &\leq \sum_{x,z} (c|x| + |z - cx|) \cdot P(cX = cx, Y = z - cx) \\ &= c \sum_x |x| \cdot P(X = x) + \sum_y |y| \cdot P(Y = y) \\ &= c\mathbb{E}(|X|) + \mathbb{E}(|Y|). \end{aligned}$$

Schritt zwei verwendet die Dreiecksungleichung. Im dritten Schritt wurde  $y$  substituiert für  $z - cx$ , im nächsten die Summe aufgespalten. Nach Voraussetzung ist das Endergebnis endlich. Daher kann dieselbe Rechnung ohne Betragsstriche durchgeführt werden. Im zweiten Schritt gilt dann Gleichheit und man erhält das gesuchte Resultat.

(c) Es gilt  $X \geq S_N := \sum_{n=1}^N X_n$  und daher wegen (a) und (b) auch  $\mathbb{E}(X) \geq \sum_{n=1}^N \mathbb{E}(X_n)$ . Im Limes  $N \rightarrow \infty$  gilt dann  $\mathbb{E}(X) \geq \sum_{n \geq 1} \mathbb{E}(X_n)$ . Wie kommen wir nun zur umgekehrten Ungleichung? Wir wählen ein beliebiges  $c \in (0, 1)$  und betrachten die Zufallsvariable  $\tau = \inf\{N \geq 1 : S_N \geq cX\}$ .

Da  $S_N$  monoton gegen  $X < \infty$  konvergiert, ist auch  $\tau < \infty$ . Auch die zufällige Summe  $S_\tau$  ist eine diskrete Zufallsvariable. Ihr Wertebereich ist nämlich enthalten in  $S(\Omega) := \bigcup_{N \geq 1} S_N(\Omega)$ , der Vereinigung der Wertebereiche aller diskreten  $S_N$ . Mit (a), Definition 18 und der  $\sigma$ -Additivität erhalten wir

$$\begin{aligned} c\mathbb{E}(X) &\leq \mathbb{E}(S_\tau) = \sum_{x \in S(\Omega)} x \sum_{N \geq 1} P(\tau = N, S_N = x) \\ &= \sum_{N \geq 1} \sum_{x \in S(\Omega)} x \cdot P(\mathbf{1}_{\{\tau=N\}} \cdot S_N = x) = \sum_{N \geq 1} \mathbb{E}(\mathbf{1}_{\{\tau=N\}} S_N). \end{aligned}$$

Im letzten Schritt wurden die Summen vertauscht. Mit (b) folgt weiter

$$\begin{aligned} \sum_{N \geq 1} \sum_{n=1}^N \mathbb{E}(\mathbf{1}_{\{\tau=N\}} X_n) &= \sum_{n \geq 1} \sum_{N \geq n} \sum_{x \in X_n(\Omega)} x \cdot P(\tau = N, X_n = x) \\ &= \sum_{n \geq 1} \sum_{x \in X(\Omega)} x \cdot P(\tau \geq n, X_n = x) \leq \sum_{n \geq 1} \mathbb{E}(X_n). \end{aligned}$$

Im Limes  $c \rightarrow 1$  erhalten wir dann  $\mathbb{E}(X) \leq \sum_{n \geq 1} \mathbb{E}(X_n)$ .

Die zweite Aussage von (c) erhalten wir, indem wir die erste Aussage auf  $X_n = Y_{n+1} - Y_n$  und  $X = Y - Y_1$  anwenden.

(d) Auch  $XY$  ist eine diskrete Zufallsvariable. Wegen

$$\begin{aligned} \sum_z |z| \cdot P(XY = z) &= \sum_{z \neq 0} |z| \cdot \sum_{x \neq 0} P(X = x, Y = \frac{z}{x}) \\ &= \sum_{x \neq 0, y \neq 0} |x| \cdot |y| \cdot P(X = x) \cdot P(Y = y) \\ &= \mathbb{E}(|X|) \cdot \mathbb{E}(|Y|) < \infty. \end{aligned}$$

Im zweiten Schritt wurde die Unabhängigkeit von  $X$  und  $Y$  genutzt. Dann wurden die Summen getrennt. Da der Ausdruck endlich ist, kann die Rechnung ohne Betragsstriche wiederholt werden und man erhält das gesuchte Resultat.  $\square$

## 1.6.2 Erzeugende Funktionen

Wir betrachten Wahrscheinlichkeitsmaße  $P$  auf  $(\mathbb{Z}_+, \mathcal{P}(\mathbb{Z}))$ .

**Definition 19.** Ist  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{Z}_+, \mathcal{P}(\mathbb{Z}_+))$  und  $X$  eine  $\mathbb{Z}_+$ -wertige Zufallsvariable, so heißt die Funktion

$$g_X(r) := \mathbb{E}(r^X) = \sum_{n \geq 0} P(X = n) \cdot r^n, \quad 0 \leq r \leq 1$$

die *erzeugende Funktion* von  $X$ .

Die erzeugende Funktion  $g_X$  existiert für alle  $r \in [0, 1]$ , da

$$\sum_{n \geq 0} P(X = n) \cdot r^n \leq \sum_{n \geq 0} P(X = n) = 1 \quad \text{für } r \in [0, 1]$$

ist.

Mit der erzeugenden Funktion kann die gesamte Verteilung von  $X$  ermittelt werden. Die erzeugende Funktion ist nämlich unendlich oft differenzierbar und die  $k$ -te Ableitung ist

$$g_X^{(k)}(r) = \sum_{n=k} n \cdot (n-1) \cdots (n-k+1) \cdot r^{n-k} \cdot P(X = n).$$

Dann ist  $P(X = k) = \frac{g_X^{(k)}(0)}{k!}$ .

Die Verteilung von  $X$  ist mit  $g_X$  eindeutig bestimmt:

**Satz 9.** (vgl. [6])

Ist  $g_X(r) = g_Y(r)$  für alle  $r \in [0, \gamma]$  und alle  $\gamma \in (0, 1]$ , dann folgt daraus  $P(X = k) = P(Y = k)$  für alle  $k \in \mathbb{Z}_+$ .

*Beweis.*

Wir zeigen: Sind zwei Potenzreihen  $H(r) = \sum_{n=1}^{\infty} a_n \cdot r^n$  und  $K(r) = \sum_{n=1}^{\infty} b_n \cdot r^n$  gleich für alle  $r \in [0, \gamma]$  und  $\gamma \in (0, 1]$  und sind die Folgen  $(a_n)_{n \geq 0}$  und  $(b_n)_{n \geq 0}$  beschränkt, so folgt für alle Koeffizienten  $a_n = b_n$ .

Der Beweis erfolgt mittels vollständiger Induktion.

*Induktionsanfang:  $n = 0$*

Es ist  $a_0 = H(0)$  und  $b_0 = K(0)$ . Da aber  $H(0) = K(0)$  ist auch  $a_0 = b_0$ .

*Induktionsvoraussetzung:*

Angenommen  $a_0 = b_0, \dots, a_n = b_n$ . Wir zeigen, dass daraus schon  $a_{n+1} = b_{n+1}$  folgt:

*Induktionsschritt:*

Wir betrachten die Differenz  $H(r) - K(r)$  und dividieren durch  $r^{n+1}$ :

$$\begin{aligned} r^{-(n+1)} \cdot [H(r) - K(r)] &= r^{-(n+1)} \cdot \sum_{i=1}^{\infty} (a_i - b_i) \cdot r^i \\ &= \sum_{i=n+1}^{\infty} (a_i - b_i) \cdot r^{i-(n+1)} \\ &= (a_{n+1} - b_{n+1}) + \sum_{i=n+2}^{\infty} (a_i - b_i) \cdot r^{i-(n+1)}. \end{aligned}$$

Im zweiten Schritt wurde die Induktionsvoraussetzung verwendet. Mit der Dreiecksungleichung erhalten wir dann

$$|a_{n+1} - b_{n+1}| \leq r^{-(n+1)} \cdot |H(r) - K(r)| + \sum_{i=n+2}^{\infty} |a_i - b_i| \cdot r^{i-(n+1)}.$$

Laut Voraussetzung ist  $H(r) = K(r)$  und daher ist der erste Summand rechts gleich 0. Weil die Folgen  $(a_i)_{i \geq 0}$  und  $(b_i)_{i \geq 0}$  beschränkt sind, ist auch  $(|b_i - a_i|)_{i \geq 0}$  beschränkt und es gibt daher ein  $C < \infty$  mit  $|b_i - a_i| < C$  für alle  $i \geq 0$ . Damit ist aber für alle  $r \in [0, \gamma]$

$$|a_{n+1} - b_{n+1}| \leq C \cdot \sum_{i=n+2}^{\infty} r^{i-(n+1)} = C \cdot \sum_{i=1}^{\infty} r^i = K \cdot \frac{r}{1-r}.$$

Jetzt vollziehen wir den Grenzübergang  $r \searrow 0$ . Die rechte Seite geht dann gegen 0 und damit auch  $|a_{n+1} - b_{n+1}|$ . Die Koeffizienten  $a_{n+1}$  und  $b_{n+1}$  sind daher gleich.  $\square$

**Satz 10.**

Sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{Z}_+, \mathcal{P}(\mathbb{Z}_+))$  und  $X$  eine  $\mathbb{Z}_+$ -wertige Zufallsvariable. Dann gilt:

$\mathbb{E}(X)$  existiert genau dann, wenn  $g'_X(1)$  existiert. Dann gilt  $\mathbb{E}(X) = g'_X(1)$ .

*Beweis.* Wir zeigen, dass  $g'_X(1) = \sum_{n \geq 0} P(X = n) \cdot n$ . Wenn  $g'_X(1)$  existiert, dann ist auch  $\sum_{n \geq 0} P(X = n) \cdot n < \infty$  und damit existiert  $\mathbb{E}(X)$ .

$$\begin{aligned} g'_X(1) &= \lim_{r \nearrow 1} \frac{g_X(1) - g_X(r)}{1-r} \\ &= \lim_{r \nearrow 1} \sum_{n \geq 0} P(X = n) \cdot \frac{1-r^n}{1-r} \\ &= \lim_{r \nearrow 1} \sum_{n \geq 0} P(X = n) \cdot \sum_{k=0}^{n-1} r^k \\ &= \sum_{n \geq 0} P(X = n) \cdot n = \mathbb{E}(X). \end{aligned}$$

Der Beweis verwendet den Abel'schen Grenzwertsatz, welcher besagt, dass der Wert der Potenzreihe am (rechten) Rand des Konvergenzradius ( $= 1$  für die geometrische Reihe) gleich dem Grenzwert (von links) gegen den Rand ist, sofern sie dort konvergiert.  $\square$

Nun verfügen wir über das nötige stochastische Rüstzeug, um uns den Herausforderungen der nächsten Kapitel zu stellen.



# Kapitel 2

## Markov-Ketten

In diesem Kapitel werden Markov-Ketten, stochastische Prozesse ohne Erinnerungsvermögen, behandelt. Der Bogen spannt sich von einfachen Eigenschaften über das Langzeitverhalten bis zur möglichen Konvergenz gegen die stationäre Verteilung. Das Kapitel dient auch als theoretische Grundlage für die im nächsten Kapitel behandelten speziellen Markov-Ketten, die Irrfahrten. Hauptquelle für dieses Kapitel waren [5] und [6]. Abschnitte wurden in [9] gefunden. Die Werke [1] und [4] dienten vor allem zum Vergleich. Da große Teile in ähnlicher Form wie in [5] verwendet wurden, wird diese Quelle nicht extra angeführt.

### 2.1 Die Markov-Eigenschaft

#### 2.1.1 Was ist eine Markov-Kette?

**Definition 20.** *Stochastische Matrix.* (vgl. [8])

Sei  $E \neq \emptyset$  eine höchstens abzählbare Menge.

Eine Matrix  $\Pi = (\Pi(x, y))_{x, y \in E}$  heißt *stochastische Matrix*, falls jede Zeile  $\Pi(x, \cdot)$  der Matrix eine Zähldichte auf  $E$  ist. D. h.:  $\Pi(x, y) \in [0, 1]$  und es gilt:

$$\sum_{y \in E} \Pi(x, y) = 1 \quad \forall x \in E.$$

Alle Zeilensummen sind also 1.

Wir betrachten nun einen Zufallsprozess in  $E$ , der bei jedem Schritt mit Wahrscheinlichkeit  $\Pi(x, y)$  von  $x$  nach  $y$  springt.

**Definition 21.** *Markov-Ketten.*

Eine Folge  $X_0, X_1, \dots$  von Zufallsvariablen auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit Werten in  $E$  heißt (nach A. A. Markov, 1856-1922) eine

Markov-Kette mit Zustandsraum  $E$  und Übergangsmatrix  $\Pi$ , wenn für alle  $n \geq 0$  und alle  $x_0, \dots, x_{n+1} \in E$  gilt:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = \Pi(x_n, x_{n+1}) \quad (2.1)$$

sofern  $P(X_0 = x_0, \dots, X_n = x_n) > 0$ . Die Verteilung  $\alpha = P \circ X_0^{-1}$  von  $X_0$  heißt die *Startverteilung* der Markov-Kette. (vgl.[5])

Gleichung (2.1) besagt Folgendes: Die Einträge der Matrix lassen sich auch anders darstellen:

$$\Pi(x_n, x_{n+1}) = P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n).$$

Die Übergangswahrscheinlichkeit von  $x_n$  zu  $x_{n+1}$  ist gleich der bedingten Wahrscheinlichkeit,  $x_{n+1}$  zu erreichen, wenn direkt davor der Zustand  $x_n$  eingenommen wurde. Unter der Bedingung der bekannten „Vorgeschichte“  $x_1, \dots, x_n$  hängt die bedingte Verteilung von  $X_{n+1}$  also nur von der Gegenwart  $x_n$  ab. Diese Eigenschaft wird als *Markov-Eigenschaft* bezeichnet.

Eine Zusatzannahme vereinfacht die Behandlung. Man nimmt an, dass die Übergangswahrscheinlichkeiten nicht vom Zeitpunkt  $n$  abhängen. Aufgrund dieser Zeitinvarianz spricht man von *stationären Übergangswahrscheinlichkeiten*.

Zusammengefasst bedeutet dies, dass eine Markov-Kette  $(X_n)_{n \geq 0}$  ein stochastischer Prozess mit sehr kurzem Gedächtnis (erinnert sich nur an das soeben Vergangene) und ohne innere Uhr und Zeitgefühl ist.

**Bemerkung 5.** Setzt man in Satz 5 auf Seite 17  $\varrho_k | \omega_0, \dots, \omega_{k-1} = \Pi(\omega_{k-1}, \omega_k)$ , sieht man, dass die Markov-Eigenschaft (Gleichung (2.1)) ein Spezialfall der Bedingung (b) von Satz 5 ist. Auch in (a) steckt natürlich die Markov-Eigenschaft.

Gemäß Satz 6 auf Seite 18 gibt es zu jeder Verteilung  $\varrho_1$  auf  $\Omega_1$  genau ein Wahrscheinlichkeitsmaß  $P$  auf  $(\prod_{i \geq 1} \Omega_i, \otimes_{i \geq 1} \mathcal{P}(\Omega_i))$  mit der Eigenschaft

$$P(X_1 = \omega_1, \dots, X_k = \omega_k) = \varrho_1 \omega_1 \cdot \varrho_2 | \omega_1(\omega_2) \cdots \varrho_k | \omega_1, \dots, \omega_{k-1}(\omega_k) \quad (2.2)$$

für alle  $k \geq 1$  und  $\omega_i \in \Omega_i$ . (Für die genauen Voraussetzungen siehe Satz 6) Außerdem ist obige Gleichung äquivalent zu den Bedingungen (a) und (b) von Satz 5, die, wie oben erwähnt, die Markov-Eigenschaft als Spezialfall enthalten.

Wir setzen nun  $\varrho_1 = \alpha, \Omega_i = E \quad \forall i \geq 1$  und erneut  $\varrho_k | \omega_0, \dots, \omega_{k-1} = \Pi(\omega_{k-1}, \omega_k)$  für  $k \geq 1$ . Auf dem Produktereignisraum  $(\Omega, \mathcal{A}) := (\prod_{k \geq 0} E, \otimes_{k \geq 0} \mathcal{P}(E))$

gibt es wiederum genau ein Wahrscheinlichkeitsmaß  $P^\alpha$  mit der Eigenschaft (1.5) bzw. (2.2). Da Gleichung (2.2) äquivalent zu (a) und (b) in Satz 5 ist und mit der Ersetzung  $\varrho_k|\omega_0, \dots, \omega_{k-1} = \Pi(\omega_{k-1}, \omega_k)$ ,  $k \geq 1$  die Eigenschaft (b) in Satz 5 auf Seite 17 genau die Markov-Eigenschaft ist, gibt es also zu jeder Startverteilung  $\alpha$  genau ein Wahrscheinlichkeitsmaß  $P^\alpha$  auf  $(\Omega, \mathcal{A})$  derart, dass die Projektionen  $X_n : (\omega_k)_{k \geq 0} \mapsto \omega_n$  von  $\Omega$  nach  $E$  eine Markov-Kette zu  $\Pi$  und  $\alpha$  bilden.

Setzt man die Startverteilung  $\alpha = \delta_x$  ( $\delta_x = 1$  an der Stelle  $x$ , 0 sonst; das bedeutet man startet sicher in  $x$ ), schreibt man statt  $P^\alpha$  kurz  $P^x$ . Es gilt dann (vgl. Gleichung (1.3) auf Seite 17):

$$P^x(X_1 = x_1, \dots, X_n = x_n) = \Pi(x, x_1)\Pi(x_1, x_2) \cdots \Pi(x_{n-1}, x_n) \quad (2.3)$$

für beliebige  $n \geq 1$  und  $x_1, \dots, x_n \in E$ .

Interessiert nur die Wahrscheinlichkeit des Übergangs von  $x$  nach  $y$  in  $n$  Schritten, so erhält man

$$P^x(X_n = y) = \Pi^n(x, y) \quad \forall x, y \in E \quad (2.4)$$

durch Summation über  $x_1, \dots, x_{n-1} \in E$ .

In Gleichung (2.4) bezeichnet  $\Pi^n$  die  $n$ -te Matrixpotenz von  $\Pi$ . Das bedeutet, dass für jeden Schritt eine Multiplikation mit  $\Pi$  hinzukommt. Verständlich wird dies sofort, wenn man bedenkt, dass die Matrix in den einzelnen Einträgen  $\Pi(x, y)$  die Übergangswahrscheinlichkeiten von  $x$  nach  $y$  enthält, und zwar Einträge für alle Elemente  $x, y$  aus der Ergebnismenge  $E$ . Die Matrix  $\Pi$  enthält also die gesamte Information über alle Übergangswahrscheinlichkeiten nach einem Schritt. Multipliziert man nun diese Matrix mit sich selbst ( $\Pi^2$ ), macht man anschaulich gesprochen zwei Schritte. Die Einträge  $\Pi^2(x, y)$  der Matrix beschreiben dann die Übergangswahrscheinlichkeiten von  $x$  nach  $y$  in zwei Schritten. Wiederum existieren Einträge für alle Elemente in  $E$ . Für den dritten Schritt multipliziert man ein weiteres Mal mit der Matrix  $\Pi$ .

Für  $n$  Schritte nimmt man der Argumentation folgend die  $n$ -te Potenz der Matrix  $\Pi$ . Die Einträge  $\Pi^n(x, y)$  dieser Matrix sind nun die Übergangswahrscheinlichkeiten von  $x$  nach  $y$  in  $n$  Schritten. Mit anderen Worten: Das Matrixelement  $\Pi^n(x, y)$  steht für die Wahrscheinlichkeit, dass das  $n$ -te Element der Markov-Kette bei Start in  $x$  genau  $y$  ist. Und das wird in Gleichung (2.4) durch  $P^x(X_n = y)$  beschrieben. Natürlich existiert auch in dieser Matrix ein Eintrag für alle  $x, y \in E$ . Der Beweis wird später nachgeholt.

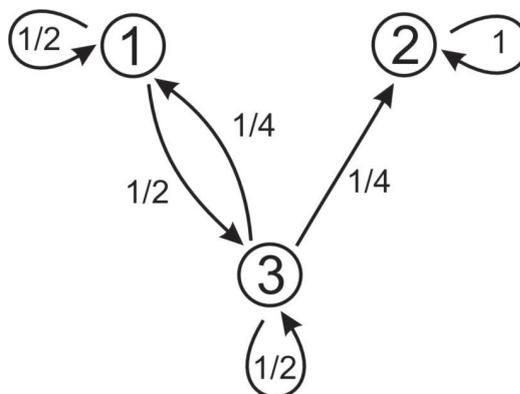


Abbildung 2.1: Übergangsgraph des Beispiels

### 2.1.2 Ein einfaches Beispiel einer Markov-Kette

Zur Veranschaulichung des Matrix-Formalismus soll das folgende Beispiel dienen.

Die Elemente des Ereignisraums  $E$  werden durch Zahlen dargestellt, um eine einfachere Handhabung zu erreichen:  $E = \{1, 2, 3\}$ . Das Matrixelement  $\Pi(1, 2)$  zum Beispiel bezeichnet dann die Übergangswahrscheinlichkeit von 1 nach 2, wie gehabt. Die Position dieses Elements in der Matrix in der ersten Zeile und der zweiten Spalte ist natürlich klar. Bei anderen Bezeichnungen im Ergebnisraum müsste man das explizit festlegen.

Betrachten wir also eine konkrete Matrix:

$$\Pi = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

Aus der Matrix  $\Pi$  kann man zum Beispiel herauslesen, dass die Wahrscheinlichkeit des Übergangs von 1 nach 2 gleich 0 ist. Der Eintrag  $\Pi(2, 2) = 1$  bedeutet, dass man von 2 aus mit Wahrscheinlichkeit 1 bei 2 bleibt.

Noch anschaulicher wird der Prozess, wenn man sogenannte Übergangsgraphen betrachtet. Abbildung 2.1 zeigt den zu diesem Beispiel passenden Übergangsgraphen. Die Elemente 1 bis 3 werden als Kreise dargestellt, die Übergänge durch Pfeile, die vom Ausgangselement zum angestrebten Element zeigen. Neben den Pfeilen ist jeweils die Übergangswahrscheinlichkeit angegeben.

Aus dem Übergangsgraphen lassen sich die Eigenschaften des Prozesses noch schöner ablesen. Von 2 zum Beispiel kommt man nie wieder weg, da der einzige Pfeil, der wegführt, wieder zu 2 zurückgeht. Bei 3 ist die Situation anders: Mit Wahrscheinlichkeit  $\frac{1}{2}$  bleibt man bei 3, mit Wahrscheinlichkeit  $\frac{1}{2}$

verlässt man 3, und zwar gleich wahrscheinlich in Richtung 1 oder 2.

Betrachten wir nun den Übergang in zwei Schritten. Die neue Übergangsmatrix ist nun das Quadrat der ursprünglichen Übergangsmatrix  $\Pi$  des Prozesses. Sie ergibt sich folgendermaßen:

$$\Pi^2 = \begin{pmatrix} \frac{3}{8} & \frac{1}{8} & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{4} & \frac{3}{8} & \frac{3}{8} \end{pmatrix}$$

Wie sich leicht überprüfen lässt, ist diese Matrix wiederum eine stochastische Matrix.

Wie man die Matrix  $\Pi^2$  berechnet ist klar. Allerdings kann man sich in diesem einfachen Fall auch überlegen, wie man zu den Einträgen kommt. Wie schon weiter oben angeführt, ist die Wahrscheinlichkeit eines Übergangs von  $x$  nach  $y$  in  $n$  Schritten die Summe über alle möglichen Übergangswahrscheinlichkeiten von  $x$  nach  $y$  in  $n$  Schritten. Da es sich im Beispiel um zwei Schritte handelt, kann man das leicht nachprüfen.

Für das Element  $\Pi^2(1, 1)$  werden die Überlegungen durchgeführt. Um nach zwei Schritten wieder bei 1 zu sein, gibt es zwei Möglichkeiten:

1. Man bleibt zweimal bei 1.
2. Man geht zu 3 und wieder zurück zu 1.

Die Wahrscheinlichkeit für den ersten Fall ist das „Quadrat des Pfeils von 1 nach 1“  $\Pi(1, 1) \cdot \Pi(1, 1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ . Im zweiten Fall erhält man die Wahrscheinlichkeit als „Produkt“ des Pfeiles von 1 nach 3 mit dem Pfeil von 3 nach 1,  $\Pi(1, 3) \cdot \Pi(3, 1) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$ .

Die Übergangswahrscheinlichkeit von 1 nach 2 in zwei Schritten ist dann die Summe:  $P^1(X_2 = 1) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8} = \Pi^2(1, 1)$ .

### 2.1.3 Die Markov-Eigenschaft

Die Markov-Eigenschaft (2.1) auf Seite 28 lässt sich zu der folgenden allgemeineren Aussage verschärfen, die sich nicht nur auf den jeweils folgenden Zeitpunkt  $n + 1$  bezieht, sondern auf die gesamte Zukunft nach der Zeit  $n$ : Bei bekanntem gegenwärtigen Zustand „vergisst“ die Markov-Kette die Vergangenheit und startet „neugeboren“.

**Satz 11.** Markov-Eigenschaft.

Ist  $(X_n)_{n \geq 0}$  eine Markov-Kette zu  $\Pi$  und  $\alpha$ , so gilt für alle  $n \geq 0$ ,

$A \in \mathcal{A} = \mathcal{P}(E)^{\otimes \mathbb{Z}_+}$ ,  $B \subset E^n$ , und  $x \in E$

$$P^\alpha((X_n, X_{n+1}, \dots) \in A | (X_0, \dots, X_{n-1}) \in B, X_n = x) = P^x(A),$$

sofern  $P^\alpha((X_0, \dots, X_{n-1}) \in B, X_n = x) > 0$ .

Satz 11 beschreibt die „Gedächtnislosigkeit“ der Markov-Kette. Die Wahrscheinlichkeit, bei Startverteilung  $\alpha$  eine Folge  $(X_{n+i})_{i \geq 0} \in A$  zu erhalten unter der Bedingung, dass die Folgeglieder  $(X_i)_{0 \leq i \leq n-1}$  in  $B$  liegen und  $X_n = x$  ist gleich der Wahrscheinlichkeit,  $A$  zu erhalten bei Start in  $x$ . Man könnte sagen, die Zeitrechnung beginnt von vorne.

*Beweis.* Wie in Bemerkung 5 auf Seite 28 vereinbart ist

$(\Omega, \mathcal{A}) = (E^{\mathbb{Z}_+}, \mathcal{P}(E)^{\otimes \mathbb{Z}_+})$  und  $X_n$  die  $n$ -te Projektion. Für beliebige  $k \geq 0$  und  $x_0, \dots, x_k \in E$  ergibt die Multiplikationsformel (2.3) auf Seite 29

$$\begin{aligned} & P^\alpha((X_0, \dots, X_{n-1}) \in B, X_n = x, X_{n+i} = x_i, \text{ für } 0 \leq i \leq k) \\ &= \sum_{(y_0, \dots, y_{n-1}) \in B} \alpha(y_0) \cdot \Pi(y_0, y_1) \cdots \Pi(y_{n-1}, x) \cdot \delta_{x, x_0} \cdot \Pi(x_0, x_1) \cdots \Pi(x_{k-1}, x_k) \\ &= P^\alpha((X_0, \dots, X_{n-1}) \in B, X_n = x) \cdot P^x(X_i = x_i \text{ für } 0 \leq i \leq k). \end{aligned} \quad (2.5)$$

Die zweite Zeile enthält die Wahrscheinlichkeit, bei Startverteilung  $\alpha$  die (endliche) Folge  $(X_i)_{0 \leq i \leq n+k}$  zu erhalten. Dazu wird die Wahrscheinlichkeit  $\alpha(y_0)$  des Startelements mit den jeweiligen Übergangswahrscheinlichkeiten multipliziert und über alle Möglichkeiten  $(y_0, \dots, y_{n-1}) \in B$  summiert. Das Kronecker-Delta  $\delta_{x, x_0}$  sorgt dafür, dass alle Folgen, die  $X_n = x = x_0$  nicht erfüllen, ausscheiden.

Zeile drei fasst die von  $\alpha$  abhängigen Faktoren zu  $P^\alpha(\dots)$  und den Rest zu  $P^x(\dots)$  zusammen.

Doch was wurde hier eigentlich gezeigt? Um das zu verstehen, definieren wir die beiden Ereignisse

$$\begin{aligned} R &:= \{X_{n+i} = x_i \text{ für } 0 \leq i \leq k\} \quad \text{und} \\ S &:= \{(X_0, \dots, X_{n-1}) \in B, X_n \in x\}. \end{aligned}$$

Die erste Zeile in (2.5) ist dann gleichbedeutend mit  $P^\alpha(R \cap S)$ , die letzte entspricht  $P^\alpha(S) \cdot P^x(X_i = x_i \text{ für } 0 \leq i \leq k)$ . Insgesamt ergibt das:

$$P^x(X_i = x_i \text{ für } 0 \leq i \leq k) = \frac{P^\alpha(R \cap S)}{P^\alpha(S)} = P^\alpha(R|S).$$

Die rechte Seite entspricht der linken Seite der Gleichung im Satz. Damit folgt für  $A = \{X_i = x_i \text{ für } 0 \leq i \leq k\}$  also die Behauptung. Der allgemeine Fall ergibt sich aus dem Eindeigkeitssatz 2 auf Seite 11, da die Mengen  $A$  dieser speziellen Form (zusammen mit  $\emptyset$ ) einen  $\cap$ -stabilen Erzeuger von  $\mathcal{A}$  bilden.  $\square$

### Kommunizieren – eine Äquivalenzrelation

Dieser Abschnitt entstammt [6].

Nun wollen wir noch eine Äquivalenzrelation auf den Zuständen von Markov-Ketten definieren. Es gibt Markov-Ketten, bei denen jeder Zustand  $y$  von jedem Zustand  $x$  in endlich vielen Schritten erreicht werden kann. Wir sagen ein Zustand  $y$  ist *erreichbar* von Zustand  $x$ , wenn für ein  $n \in \mathbb{N}_0$  gilt  $\Pi^n(x, y) > 0$ . Dann schreiben wir  $x \rightarrow y$ . Gilt sowohl  $x \rightarrow y$  als auch  $y \rightarrow x$ , sagt man, die Zustände  $x$  und  $y$  *kommunizieren* und schreibt dafür  $x \leftrightarrow y$ .

**Proposition 7.** *Die Relation „kommunizieren“ ist eine Äquivalenzrelation, denn es gilt*

- (a)  $x \leftrightarrow x$  (Reflexivität),
- (b)  $x \leftrightarrow y \implies y \leftrightarrow x$  (Symmetrie),
- (c)  $x \leftrightarrow y$  und  $y \leftrightarrow z \implies x \leftrightarrow z$  (Transitivität).

*Beweis.* Reflexivität gilt wegen  $\Pi^0(x, x) = 1$ .

Die Symmetrie folgt aus der Definition der Relation.

Für die Transitivität betrachten wir vorerst nur die „Erreichbarkeit“, also die Relation  $\rightarrow$ . Wegen  $x \rightarrow y$  und  $y \rightarrow z$  gibt es  $n$  und  $m$  mit  $\Pi^n(x, y) > 0$  und  $\Pi^m(y, z) > 0$ . Für  $x \rightarrow z$  muss  $\Pi^k(x, z)$  für ein  $k$  erfüllt sein. Wir betrachten  $\Pi^{(n+m)}(x, z)$ :

$$\begin{aligned} \Pi^{(n+m)}(x, z) &= \sum_{u \in E} \Pi^n(x, u) \cdot \Pi^m(u, z) \\ &= \Pi^n(x, y) \cdot \Pi^m(y, z) + \sum_{\substack{u \in E \\ u \neq y}} \Pi^n(x, u) \cdot \Pi^m(u, z) \\ &\geq \Pi^n(x, y) \cdot \Pi^m(y, z) > 0. \end{aligned}$$

Wir haben  $k = n + m$  gefunden, es gilt also  $x \rightarrow z$ . Ebenso kann man aus  $z \rightarrow y$  und  $y \rightarrow x$  zeigen, dass  $z \rightarrow x$ .  $\square$

## 2.2 Stochastische Matrizen

Dieser Abschnitt beruht auf eigenen Überlegungen.

Wesentlich für die Betrachtung von Markov-Ketten sind stochastischen Matrizen. Bisher wurden einige Eigenschaften von solchen Matrizen verwendet, ohne sie zu begründen. Deshalb soll in diesem Abschnitt kurz darauf eingegangen werden.

So wurden bisher stochastische Matrizen einfach miteinander multipliziert, ohne zu hinterfragen, ob das überhaupt zulässig ist. Deshalb formulieren wir folgende Proposition:

**Proposition 8.** Multiplikation stochastischer Matrizen.

*Das Produkt von stochastischen Matrizen ist wieder eine stochastische Matrix.*

*Beweis.* Definition 20 auf Seite 27 besagt, dass für eine stochastische Matrix  $A = (A(x, y))_{x, y \in E}$  folgende Eigenschaften gelten müssen:

1.  $A(x, y) \in [0, 1] \quad \forall x, y \in E,$
2.  $\sum_{y \in E} A(x, y) = 1 \quad \forall x \in E.$

Betrachten wir zwei stochastische  $n \times n$ -Matrizen  $A := (x_{ij})_{i, j=1, \dots, n}$  und  $B := (y_{ij})_{i, j=1, \dots, n}$  (Andere als  $n \times n$ -Matrizen kommen nicht in Frage, denn besteht der Ergebnisraum aus  $n$  Elementen, so können  $n$  verschiedene Elemente jeweils in ein anderes (inklusive sich selbst) übergehen. Dementsprechend besteht die Matrix aus  $n$  Zeilen. Ebenso gibt es aber auch  $n$  Elemente „in die übergegangen werden kann“. Daher gibt es auch  $n$  Spalten. Für jeden Übergang gibt es also in der Matrix einen Platz.)

Allgemein gilt für die Matrixmultiplikation:

$$(A \cdot B)_{ij} = \sum_{k=1}^n x_{ik} \cdot y_{kj} \quad (2.6)$$

Zuerst zeigen wir Eigenschaft 1:

Dass  $(A \cdot B)_{ij} \geq 0$  gilt, ist klar, da die Matrixelemente von  $A$  und  $B$  alle nichtnegativ sind. Und Produkte und Summen nichtnegativer Zahlen sind ebenfalls wieder nichtnegativ.

Es ist also nur zu zeigen, dass die Einträge der Produktmatrix kleiner oder gleich 1 sind.

Dazu betrachtet man ein Matrixelement der Produktmatrix:

$$(A \cdot B)_{ij} = \sum_{k=1}^n (x_{ik} \cdot \underbrace{y_{kj}}_{\leq 1 \text{ wg. 1}}) \leq \sum_{k=1}^n x_{ik} = 1 \quad \forall i, j = 1, \dots, n$$

$\underbrace{\hspace{10em}}_{=1 \text{ wg. 2}}$

Im ersten Schritt wurde verwendet, dass alle Einträge von  $B$  kleiner oder gleich 1 sind. Alle Summanden  $x_{ik} \cdot y_{kj}$  sind daher kleiner oder gleich  $x_{ik}$ .

Also ist auch die Summe kleiner als die Summe über die  $x_{ik}$ . Und die ist laut Eigenschaft 2 genau gleich 1. Insgesamt folgt, dass jedes Matrixelement  $(A.B)_{ij}$  kleiner oder gleich 1 ist. Die Einträge der Produktmatrix liegen also tatsächlich im Intervall  $[0, 1]$ .

Zu zeigen bleibt nun noch die Eigenschaft 2:

Man bildet die Summe über eine Zeile der Produktmatrix  $A.B$ :

$$\begin{aligned} \sum_{j=1}^n (A.B)_{ij} &= \sum_{j=1}^n \sum_{k=1}^n x_{ik} \cdot y_{kj} = \sum_{k=1}^n \sum_{j=1}^n x_{ik} \cdot y_{kj} \\ &= \sum_{k=1}^n x_{ik} \cdot \underbrace{\sum_{j=1}^n y_{kj}}_{=1 \text{ wegen 2}} = \sum_{k=1}^n x_{ik} = 1 \quad \forall i, j = 1, \dots, n \end{aligned}$$

Hier wurde für das Element der Produktmatrix (2.6) eingesetzt. Anschließend wurden die Summationszeichen vertauscht und die von  $j$  unabhängigen  $x_{ik}$  vor die zweite Summe geholt. Die zweite Summe beschreibt nun genau die Summe über eine Zeile von  $B$ , ist also gleich 1. Übrig bleibt die Summe über eine Zeile von  $A$ , welche wegen Eigenschaft 2 wiederum gleich 1 ist.

Das gilt für alle Zeilen der Produktmatrix. Diese Zeilen sind also alle Zählreihen. Die Eigenschaften einer stochastischen Matrix sind also erfüllt.  $\square$

Erweitert man Proposition 8 auf Seite 34 iterativ, so erhält man, dass *beliebige Produkte stochastischer Matrizen* wieder stochastische Matrizen sind. Insbesondere gilt dies für *Potenzen stochastischer Matrizen*.

Eine weitere Behauptung, die schon verwendet wurde, aber nicht bewiesen, steckt in Gleichung (2.4) auf Seite 29:

$$P^x(X_n = y) = \Pi^n(x, y) \quad \forall x, y \in E$$

durch Summation über  $x_1, \dots, x_{n-1} \in E$ .

Zu zeigen ist, dass die Summe

$$\sum_{x_1, \dots, x_{n-1} \in E} \Pi(x, x_1) \cdot \Pi(x_1, x_2) \cdots \Pi(x_{n-1}, x_n)$$

wirklich  $\Pi^n(x, y)$  ergibt.

*Beweis.* Der Beweis erfolgt mittels vollständiger Induktion nach  $n$ . Der Fall  $n = 1$  ist trivial. Im Fall  $n = 2$  ist  $\Pi^2(x, x_2) = \sum_{x_1 \in E} \Pi(x, x_1) \cdot \Pi(x_1, x_2)$

zu zeigen. Das ist aber die schon verwendete Formel für einen Eintrag einer Produktmatrix. Wir nehmen nun an, für  $n - 1$  ist die Behauptung schon gezeigt und führen  $n$  darauf zurück.

$$\begin{aligned}
 & \sum_{x_1, \dots, x_{n-1} \in E} \Pi(x, x_1) \cdot \Pi(x_1, x_2) \cdots \Pi(x_{n-1}, x_n) \\
 = & \sum_{x_{n-1} \in E} \Pi(x_{n-1}, x_n) \cdot \underbrace{\sum_{x_1, \dots, x_{n-2} \in E} \Pi(x, x_1) \cdots \Pi(x_{n-2}, x_{n-1})}_{= \Pi^{n-1}(x, x_{n-1}) \text{ lt. Induktionsvoraussetzung}} \\
 = & \sum_{x_{n-1} \in E} \Pi^{n-1}(x, x_{n-1}) \cdot \Pi(x_{n-1}, x_n) = (\Pi^{n-1} \cdot \Pi)(x, x_n) = \Pi^n(x, x_n).
 \end{aligned}$$

Im letzten Schritt wurde der Fall  $n = 2$  verwendet. □

## 2.3 Absorptionswahrscheinlichkeiten

**Definition 22.** *Absorbierende Zustände.*

Ein Zustand  $z \in E$  heißt *absorbierend* bezüglich einer Übergangsmatrix  $\Pi$ , wenn  $\Pi(z, z) = 1$ .

Ein absorbierender Zustand ist also eine „Falle“ aus der die Markov-Kette nicht mehr entkommen kann.

Meist ist die Situation nach  $n$  Schritten interessant. Daher formulieren wir folgende Proposition (die in der Literatur nicht gefunden wurde, vermutlich, weil sie ohnehin klar ist):

**Proposition 9.** *Ist  $z \in E$  absorbierend bezüglich  $\Pi$ , so ist  $z$  auch absorbierend bezüglich  $\Pi^n$ .*

Kommt man mit einem Schritt mit Sicherheit nicht weg, dann auch nicht mit  $n$  Schritten.

Der Beweis ist relativ einfach und kurz:

*Beweis.* Wir nehmen an, das  $z$  absorbierend ist, d.h.  $\Pi(z, z) = 1$ . Weil jede Zeile der Übergangsmatrix eine Zähldichte ist folgt aus  $\Pi(z, z) = 1$  schon  $\Pi(z, x) = 0 \quad \forall x \neq z$ . Zu zeigen ist, dass  $\Pi^n(z, z) = 1$  gilt.

Der Beweis erfolgt mittels vollständiger Induktion nach  $n$ .

Im Fall  $n = 1$  ist nichts zu zeigen.

Angenommen für  $n - 1$  gilt  $\Pi^{n-1}(z, z) = 1$ .

Wir zeigen nun, dass dann auch  $\Pi^n(z, z) = 1$  gilt:

$$\begin{aligned} \Pi^n(z, z) &= \sum_{x \in E} \Pi^{n-1}(z, x) \cdot \Pi(x, z) \\ &= \sum_{x \in E} \underbrace{\Pi^{n-1}(z, x)}_{\substack{= 1 \text{ falls } x=z \\ = 0 \text{ sonst} \text{ lt. IV}}} \cdot \Pi(x, z) = \Pi(z, z) = 1. \end{aligned}$$

Im ersten Schritt wurde die Matrixmultiplikation  $\Pi^{n-1} \cdot \Pi$  für das Element  $\Pi^n(z, z)$  angeschrieben. Dann wurde die Induktionsvoraussetzung eingesetzt. Da  $\Pi^{n-1}$  nur dann nicht 0 ist, wenn  $x = z$  gilt, bleibt von den zweiten Faktoren unter der Summe nur  $\Pi(z, z)$  übrig, was aber laut Voraussetzung der Proposition genau 1 ist.  $\square$

Weiters interessant ist natürlich die Wahrscheinlichkeit, mit der eine Markov-Kette mit absorbierendem Zustand  $z$  beim Start in  $x \in E$  schließlich in  $z$  landet. Dazu folgende Definition:

**Definition 23.** *Absorptionswahrscheinlichkeit.*

Ist  $z \in E$  ein absorbierender Zustand, so heißt

$$h_z(x) := P^x(X_n = z \text{ schließlich}) = P^x \left( \bigcup_{N \geq 0} \bigcap_{n \geq N} \{X_n = z\} \right)$$

die *Absorptionswahrscheinlichkeit* in  $z$  bei Start in  $x \in E$ .

Auch ohne Beweis ist intuitiv klar, dass die Absorptionswahrscheinlichkeit die Wahrscheinlichkeit ist, mit welcher der absorbierende Zustand  $z$  eingenommen wird. Um das zu bestätigen, definieren wir:

**Definition 24.** *Eintrittszeit.*

Für beliebiges  $z \in E$  heißt

$$\tau_z = \inf\{n \geq 1 : X_n = z\}$$

(mit der Vereinbarung  $\inf \emptyset := \infty$ ) die *Eintrittszeit* in den Zustand  $z$ . Bei Start in  $z$  spricht man von der *Rückkehrzeit* nach  $z$ . Ist der Startpunkt mit  $x$  fest, schreibt man auch

$$\tau_{xz} = \inf\{n \geq 1 : X_n = z | X_0 = x\}.$$

Die Rückkehrzeit nach  $z$  ist also  $\tau_{zz}$ .

**Bemerkung 6.** Für alle  $n \geq 1$  gilt offenbar

$$\{\tau_z = n\} = \{(X_0, \dots, X_{n-1}) \in B, X_n = z\} \quad \text{mit } B = E \times (E \setminus \{z\})^{n-1},$$

das heißt  $\{\tau_z = n\}$  hängt nur von  $(X_0, \dots, X_n)$  ab. So eine Abbildung  $\tau : \Omega \rightarrow \mathbb{Z}_+ \cup \{\infty\}$  mit dieser Eigenschaft des „nicht in die Zukunft Blickens“ heißt eine *Stoppzeit* oder *Optionszeit* bezüglich  $(X_n)_{n \geq 0}$ .

Nun formulieren wir den oben schon angedeuteten Satz:

**Satz 12.** Charakterisierung von Absorptionswahrscheinlichkeiten.

Für absorbierendes  $z \in E$  und alle  $x \in E$  mit der Übergangsmatrix  $\Pi$  gilt

$$h_z(x) = P(\tau_{xz} < \infty) = \lim_{n \rightarrow \infty} P^x(X_n = z),$$

und  $h_z$  ist die kleinste nichtnegative Funktion mit  $h_z(z) = 1$  und

$$\sum_{y \in E} \Pi(x, y) \cdot h_z(y) = h_z(x) \quad \text{für alle } x \in E.$$

Fasst man  $h_z$  als Spaltenvektor auf, kann man die Gleichung auch verkürzt in der Form  $\Pi \cdot h_z = h_z$  schreiben. Dann ist  $h_z$  ein rechter Eigenvektor von  $\Pi$  zum Eigenwert 1.

*Beweis.* Wir verwenden Proposition 9 auf Seite 36, allerdings etwas anders dargestellt:

$$P^z(X_i = z \text{ für alle } i \leq n) = \Pi^n(z, z) = 1$$

und für den Grenzübergang im Unendlichen

$$P^z(X_i = z \text{ für alle } i \geq 1) = \lim_{n \rightarrow \infty} \Pi^n(z, z) = 1.$$

Weiters gilt für jeden Startpunkt  $x \in E$ :

$$\begin{aligned} P^x(X_n = z) &= P^x(X_n = z) \cdot P^z(X_i = z \text{ für alle } i \geq 1) \\ &= P^x(X_{n+i} = z \text{ für alle } i \geq 0) = P^x \left( \bigcup_{n \geq N \geq 1} \bigcap_{i \geq N} \{X_i = z\} \right) \\ &\xrightarrow{n \rightarrow \infty} P^x \left( \bigcup_{N \geq 1} \bigcap_{i \geq N} \{X_i = z\} \right) = h_z(x). \end{aligned}$$

Im ersten Schritt wurde einfach mit der rechten Seite der vorhergehenden Gleichung mit Wert 1 multipliziert. Schritt zwei verwendet Satz 11 auf Seite 31, genauer gesagt die Umformungen im Beweis, nur diesmal von unten

nach oben. Hier steht  $P^x$  statt  $P^\alpha$  und  $P^z$  statt  $P^x$ . Die Aussage der zweiten Zeile ist die Wahrscheinlichkeit, dass bei Start in  $x$  ab dem  $n$ -ten Folgenglied alle gleich  $z$  sind.

Das Ereignis  $\{X_{N+i} = z \forall i \geq 0\}$  („ab  $N$  sind alle Folgenglieder gleich  $z$ “) lässt sich als Durchschnitt darstellen:  $\bigcap_{i \geq N} \{X_i = z\}$ . Da aber gefordert ist, dass ab  $n$  alle Folgenglieder gleich  $z$  sind, müssen alle Ereignisse, die schon früher ( $N \leq n$ ) in  $z$  landen, berücksichtigt werden. Das erreicht man durch die Vereinigung aller Ereignisse, die zu Zeitpunkten  $N \leq n$  in  $z$  landen. Der Grenzübergang bedeutet die Wahrscheinlichkeit, dass man bei Start in  $x$  „spätestens“ im Unendlichen in  $z$  landet. Der letzte Ausdruck ist dann die Definition der Absorptionswahrscheinlichkeit.

Die Gleichheit von  $h_z(x)$  und  $\lim_{n \rightarrow \infty} P^x(X_n = z)$  im Satz ist also gezeigt. Bleibt noch  $P^x(\tau_z < \infty)$ .

Wegen Satz 11 auf Seite 31 und Bemerkung 6 auf Seite 38 gilt:

$$\begin{aligned} P^x(X_n = z | \tau_z = k) &= P^x(X_n = z | (X_0, \dots, X_{k-1}) \in B, X_k = z) \\ &= P^z(X_{n-k} = z) = 1 \end{aligned}$$

für jedes  $k \leq n$ . ( $B := E \times (E \setminus \{z\})^{n-1}$ ;  $A$  in Satz 11 ist hier  $\{X_{n-k} = z\}$ , damit ist  $\{X_n = z\} \in A$ , weil  $z$  absorbierend ist.)

Mit der Fallunterscheidungsformel 4 (a) auf Seite 16 erhält man:

$$\begin{aligned} P^x(X_n = z) &= \sum_{k=1}^n P^x(\tau_z = k) \cdot P^x(X_n = z | \tau_z = k) \\ &= P^x(\tau_z \leq n) \xrightarrow{n \rightarrow \infty} P^x(\tau_z < \infty). \end{aligned}$$

Die erste Aussage des Satzes ist somit bewiesen.

Zum zweiten Teil:

Es gilt  $h_z \geq 0$  und  $h_z(x) = 1$ . Mit Satz 11 und der Fallunterscheidungsformel 4 (a) ergibt sich:

$$\begin{aligned} \sum_{y \in E} \Pi(x, y) \cdot h_z(y) &= \sum_{y \in E} P^x(X_1 = y) \cdot P^y(X_i = z \text{ schließlich}) \\ &= \sum_{y \in E} P^x(X_1 = y) \cdot P^x(X_{i+1} = z \text{ schließlich} | X_1 = y) = h_z(x). \end{aligned}$$

Es gilt also  $\Pi \cdot h_z = h_z$ . Für jede weitere Funktion  $h \geq 0$  mit  $\Pi \cdot h = h$  und  $h(z) = 1$  gilt

$$h = \Pi^n \cdot h \geq \Pi^n(x, z) = P^x(X_n = z).$$

Für  $n \rightarrow \infty$  folgt dann  $h(x) \geq h_z(x)$ .

Woher kommt das Ungleichungszeichen? Dazu überlegen wir, was  $\Pi^n \cdot h$  bedeutet:  $\Pi^n$  ist eine Matrix,  $h$  betrachten wir als Spaltenvektor. Die Multiplikation kann man dann schreiben als  $\sum_{y \in E} \Pi^n(x, y) \cdot h(y)$ . Alle Summanden sind  $\geq 0$  (die Matrixeinträge sind Wahrscheinlichkeiten,  $h \geq 0$  laut Voraussetzung). Da der Summand „ $y = z$ “ wegen  $h(z) = 1$  gleich  $\Pi^n(x, z)$  ist, ergibt sich die Ungleichung (denn „Summe  $\geq$  Summand“).  $\square$

Satz 12 auf Seite 38 ist nicht beschränkt auf absorbierende Zustände. Er lässt sich auch auf die Eintrittswahrscheinlichkeit in nicht absorbierende Zustände anwenden:

**Proposition 10.** Eintrittswahrscheinlichkeiten.

Sei  $z \in E$  ein beliebiger Zustand und  $h_z(x) = P^x(\tau_z < \infty)$  für  $x \neq z$  und  $h_z(z) = 1$ . Dann ist  $h_z$  die kleinste nichtnegative Funktion mit  $\Pi \cdot h(x) = h(x)$  für alle  $x \neq z$ .

*Beweis.* Wir definieren  $\tilde{X}_n := X_{\min(n, \tau_z)}$ ,  $n \geq 0$ . Die Folge  $(\tilde{X}_n)_{n \geq 0}$  ist dann die „zur Zeit  $\tau_z$  gestoppte Markov-Kette“, und zwar zur modifizierten Übergangsmatrix

$$\tilde{\Pi}(x, y) = \begin{cases} \Pi(x, y) & \text{falls } x \neq z, \\ 1 & \text{falls } x = y = z, \\ 0 & \text{sonst.} \end{cases}$$

wie sich leicht überprüfen lässt. Der Zustand  $z$  ist jetzt ein absorbierender. Für  $x \neq z$  ist  $h_z(x)$  gerade die Wahrscheinlichkeit, dass die Markov-Kette  $(\tilde{X}_n)$  bei Start in  $x$  in  $z$  absorbiert wird. Die Behauptung erfolgt somit aus Satz 12 auf Seite 38.  $\square$

## 2.4 Rückkehr zum Startpunkt

In diesem Kapitel betrachten wir eine Markov-Kette mit höchstens abzählbarem Zustandsraum und gegebener Übergangsmatrix  $\Pi$ . Wir beschäftigen uns mit dem Wiederkehrverhalten in einen Zustand  $x \in E$ , und zwar damit, ob die Markov-Kette mit Sicherheit wieder zurückkehrt, und wenn ja, wie lange es im Mittel bis zur ersten Rückkehr dauert.

### 2.4.1 Rekurrenz und Transienz

Für das Folgende benötigen wir eine besondere Funktion, deren Potential wir aber nicht ausschöpfen müssen. Ein spezieller Fall wird für uns genügen.

**Definition 25.** *Greenfunktion.* (vgl. [9])

Mit  $T_y^{(k)} := \sum_{n=0}^k 1_{\{X_n=y\}}$  bezeichnen wir die Anzahl der Besuche der Markov-Kette  $(X_n)_{n \geq 0}$  nach  $k$  Schritten in  $y$ . Die Funktion

$$G_k(x, y) = \mathbb{E}^x (T_y^{(k)}) = \sum_{n=0}^k \Pi^n(x, y)$$

heißt *Greenfunktion* der Markov-Kette. Sie beschreibt die erwartete Anzahl der Besuche in  $y$  beim Start in  $x$ . Wir schreiben  $G_\infty(x, y) = G(x, y)$  im Fall von  $k = \infty$ .

**Definition 26.** Für gegebenes  $x \in E$  sei

$$F_1(x, x) = P^x(\tau_x < \infty)$$

die Wahrscheinlichkeit einer Rückkehr nach  $x$  zu irgendeiner (zufälligen) endlichen Zeit und

$$F_\infty(x, x) = P^x(X_n = x \text{ für unendlich viele } n)$$

die Wahrscheinlichkeit, unendlich oft zurückzukehren.

Wir definieren damit nun

**Definition 27.** *Rekurrenz und Transienz.*

Gilt für ein  $x \in E$ , dass  $F_1(x, x) = P^x(\tau_x < \infty) = 1$  so heißt  $x$  *rekurrent* bezüglich  $\Pi$ .

Ist  $F_1(x, x) = P^x(\tau_x < \infty) < 1$ , so heißt  $x$  *transient* bezüglich  $\Pi$ .

Der folgende Satz beschreibt die beiden Fälle genauer:

**Satz 13.** *Wiederkehr von Markov-Ketten.*

*Für jedes  $x \in E$  gibt es zwei Möglichkeiten:*

- (a)  $F_1(x, x) = 1$ . Dann ist  $F_\infty(x, x) = 1$   
und  $G(x, x) = \infty$ .
- (b)  $F_1(x, x) < 1$ . Dann ist  $F_\infty(x, x) = 0$   
und  $G(x, x) = (1 - F_1(x, x))^{-1} < \infty$ .

*Beweis.* Sei  $\sigma = \sup\{n \geq 0 : X_n = x\}$  der Zeitpunkt des letzten Aufenthalts in  $x$ . Wegen  $X_0 = x$  ist  $\sigma$  wohldefiniert. Eventuell ist  $\sigma = \infty$ . Dann gilt

$$1 - F_\infty(x, x) = P^x(\sigma < \infty).$$

(Das Komplement zu „kehrt unendlich oft zurück“ ist ja „kehrt irgendwann nicht mehr zurück“)

Für jedes  $n \geq 0$  folgt außerdem aus Satz 11 auf Seite 31

$$\begin{aligned} P^x(\sigma = n) &= P^x(X_n = x) \cdot P^x(X_i \neq x \text{ für alle } i \geq 1) \\ &= \Pi^n(x, x) \cdot (1 - F_1(x, x)). \end{aligned}$$

Summiert man über  $n$ , erhält man

$$\begin{aligned} \sum_{n \geq 0} P^x(\sigma = n) &= P^x(\sigma < \infty) \\ &= \sum_{n \geq 0} \Pi^n(x, x) \cdot (1 - F_1(x, x)) = G(x, x) \cdot (1 - F_1(x, x)) \end{aligned}$$

und mit der ersten Gleichung

$$1 - F_\infty(x, x) = G(x, x) \cdot (1 - F_1(x, x)).$$

Im Fall (a) folgt aus der zweiten Gleichung  $P^x(\sigma = n) = 0$  für alle  $n$ , und daher aus der ersten Gleichung  $F_\infty(x, x) = 1$ . Das Borel–Cantelli-Lemma (Satz 7 auf Seite 20) liefert dann  $G(x, x) = \infty$ .

Im Fall (b) folgt aus der letzten Gleichung  $G(x, x) < \infty$ , und daher folgt aus dem Borel–Cantelli-Lemma  $F_\infty(x, x) = 0$ . Formt man nun die letzte Gleichung um, erhält man die Beziehung zwischen  $G$  und  $F_1$ .  $\square$

Für kommunizierende (vgl. Proposition 7 auf Seite 33) Markov-Ketten vereinfacht sich das Überprüfen auf Rekurrenz. Denn ist ein Zustand rekurrent, so sind es alle, da man von diesem rekurrenten Zustand in einer endlichen Anzahl von Schritten zu jedem anderen Zustand und wieder zurück gelangt. Dazu folgender Satz:

**Satz 14.** (vgl. [6])

*Miteinander kommunizierende Zustände einer Markov-Kette sind alle rekurrent oder alle transient.*

*Beweis.* Der Beweis beruht auf den erzeugenden Funktionen von  $\Pi^n(x, y)$  und  $P^x(\tau_y = n)$ . Wir definieren:

$$P_{xy}(r) := \sum_{n=0}^{\infty} \Pi^n(x, y) \cdot r^n, \quad (2.7)$$

$$R_{xy}(r) := \sum_{n=1}^{\infty} P^x(\tau_y = n) \cdot r^n \quad (2.8)$$

mit  $0 \leq r < 1$  (und  $\Pi^0(x, y) := \delta_{xy}$ ).

Es gilt außerdem der Zusammenhang

$$P_{xy}(r) = \delta_{xy} + R_{xy}(r) \cdot P_{yy}(r). \quad (2.9)$$

(Diese Gleichung wird im Anschluss an den Beweis hergeleitet.)

Damit erhalten wir die Beziehungen

$$P_{xx}(r) = \frac{1}{1 - R_{xx}(r)}, \quad (2.10)$$

$$P_{xy}(r) = R_{xy}(r) \cdot P_{xx}(r) \quad \forall i \neq j. \quad (2.11)$$

Es gilt weiters  $R_{xy}(1) \leq 1$  für alle  $x, y \in E$ , denn  $R_{xy}(1)$  ist ja gerade die Wahrscheinlichkeit, mit einer endlichen Anzahl an Schritten von  $x$  zu  $y$  zu gelangen. Und  $R_{xx}$  ist dann gleich  $F_1(x, x)$ . (Vergleiche dazu Definition 26 auf Seite 41.)

Somit ist ein Zustand  $x$  rekurrent, falls  $R_{xx} = 1$  und transient, falls  $R_{xx} < 1$ . Wegen Gleichung (2.10) gilt daher für einen rekurrenten Zustand  $P_{xx}(1) = \infty$  und für einen transienten Zustand  $P_{xx}(1) < \infty$ .

Nun seien  $x$  und  $y$  kommunizierende Zustände. Dann gibt es  $m, n \in \mathbb{Z}_+$  und  $\mu, \nu > 0$  mit

$$\Pi^m(x, y) = \mu, \quad \Pi^n(y, x) = \nu.$$

Damit erhalten wir für  $l \in \mathbb{Z}_+$ :

$$\Pi^{n+l+m}(x, x) \geq \Pi^n(x, y) \cdot \Pi^l(y, y) \cdot \Pi^m(y, x) = \nu\mu\Pi^l(y, y) \quad (2.12)$$

$$\Pi^{n+l+m}(y, y) \geq \Pi^n(y, x) \cdot \Pi^l(x, x) \cdot \Pi^m(x, y) = \nu\mu\Pi^l(x, x) \quad (2.13)$$

Nun bilden wir auf beiden Seiten der Ungleichung die Summe über  $l$ :

$$\sum_{l=0}^{\infty} \Pi^{n+l+m}(x, x) \geq \nu\mu \sum_{l=0}^{\infty} \Pi^l(y, y) \quad (2.14)$$

$$\sum_{l=0}^{\infty} \Pi^{n+l+m}(y, y) \geq \nu\mu \sum_{l=0}^{\infty} \Pi^l(x, x) \quad (2.15)$$

Jetzt müssen wir nur noch die Summen durch  $P_{xx}(1)$  bzw.  $P_{yy}(1)$  ersetzen:

$$P_{xx}(1) - \sum_{l=0}^{n+m-1} \Pi^l(x, x) \geq \nu\mu P_{yy}(1) \quad (2.16)$$

$$P_{yy}(1) - \sum_{l=0}^{n+m-1} \Pi^l(y, y) \geq \nu\mu P_{xx}(1) \quad (2.17)$$

Lässt man die abgezogene „Restsumme“ auf der rechten Seite der Ungleichungen weg, erhält man schließlich

$$P_{xx}(1) \geq \nu\mu P_{yy}(1), \quad (2.18)$$

$$P_{yy}(1) \geq \nu\mu P_{xx}(1). \quad (2.19)$$

Damit ist der Satz bewiesen, denn für einen rekurrenten Zustand  $x$  gilt  $P_{xx}(1) = \infty$  und somit aufgrund der zweiten Ungleichung und weil  $\nu > 0$  und  $\mu > 0$  auch  $P_{yy}(1) = \infty$  für jedes  $y$ . Das heißt also, dass auch  $y$  rekurrent ist. Ist  $x$  hingegen transient, so folgt aus der ersten Ungleichung und der Bedingung  $P_{xx} < \infty$  schon  $P_{yy} < \infty$  und damit die Transienz jedes Zustands  $y$ .  $\square$

Wir zeigen nun noch die im Beweis verwendete Gleichung (2.9) auf Seite 43:

Dazu werden wir folgende Beziehung verwenden:

$$\Pi^n(x, y) = \sum_{k=1}^{\infty} P^x(\tau_y = k) \cdot \Pi^{n-k}(y, y) \quad \forall n \in \mathbb{Z}_+. \quad (2.20)$$

(Das ist klar: Die Wahrscheinlichkeit, in  $n$  Schritten von  $x$  nach  $y$  zu gelangen, ist gleich der (über  $k$  summierten) Wahrscheinlichkeit, in  $k$  Schritten zum ersten Mal bei  $y$  zu landen und danach in  $n - k$  Schritten wieder bei  $y$  zu landen.) Wir betrachten das Produkt  $R_{xy}(r) \cdot P_{yy}(r)$ :

$$\begin{aligned} R_{xy}(r) \cdot P_{yy}(r) &= \sum_{n=1}^{\infty} P^x(\tau_y = n) \cdot r^n \cdot \sum_{l=0}^{\infty} \Pi^l(y, y) \cdot r^l \\ &= \sum_{n=1}^{\infty} \sum_{l=0}^{\infty} P^x(\tau_y = n) \cdot \Pi^l(y, y) \cdot r^{n+l} \\ &= \sum_{n=1}^{\infty} \sum_{n+l=1}^{\infty} P^x(\tau_y = n) \cdot \Pi^l(y, y) \cdot r^{n+l} \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P^x(\tau_y = n) \cdot \Pi^{m-n}(y, y) \cdot r^m \\ &= \sum_{m=1}^{\infty} \Pi^m(x, y) \cdot r^m = P_{xy} - \Pi^0(x, y) \cdot r^0 = P_{xy}(r) - \delta_{xy}. \end{aligned}$$

Im ersten Schritt wurde das zweite Summenzeichen vorgezogen. Dann wurde der Index der zweiten Summe in  $n + l$  geändert und die Summation bei 1 gestartet. (Das Ergebnis bleibt gleich). Im dritten Schritt ersetzt  $m = n + l$

und die Reihenfolge der Summen ist verändert. Jetzt erkennt man in der zweiten Summe Gleichung (2.20). Setzt man diese Gleichung ein, steht in der nächsten Zeile bis auf den fehlenden Summanden  $m = 0$  der Ausdruck für  $P_{xy}(r)$ . Somit ist die Gleichung gezeigt.

### 2.4.2 Positive Rekurrenz und Nullrekurrenz

Wie lange dauert es im Mittel, bis die Markov-Kette zu ihrem (rekurrenten) Startpunkt zurückkehrt? Dieser Frage wollen wir in diesem Abschnitt auf den Grund gehen. Man unterscheidet zwei Fälle:

**Definition 28.** Ein rekurrenter Zustand  $x \in E$  heißt *positiv rekurrent*, wenn die mittlere Rückkehrzeit endlich ist, d. h. wenn  $\mathbb{E}^x(\tau_x) < \infty$ . Andernfalls heißt  $x$  *nullrekurrent*.

Jetzt brauchen wir also nur noch ein Kriterium, um festzustellen, ob es sich um einen positiv rekurrenten oder nullrekurrenten Zustand handelt. Dazu dient der folgende Satz.

**Satz 15.** Existenz von stationären Verteilungen und positive Rekurrenz.

*Eine irreduzible Markov-Kette (mit Matrix  $\Pi$ ) besitzt genau dann eine stationäre Verteilung  $\alpha$ , wenn alle Zustände positiv rekurrent sind. Diese stationäre Verteilung ist dann eindeutig bestimmt mit  $\alpha(x) = \frac{1}{\mathbb{E}^x(\tau_x)} > 0$ . (vgl. [6])*

Bevor wir den Satz beweisen, bedürfen die Begriffe „stationäre Verteilung“ und „irreduzibel“ noch einer Erklärung.

**Definition 29.** *Stationäre Verteilung.*

Eine Verteilung  $\alpha$  heißt *stationäre Verteilung* für  $\Pi$  mit Zustandsraum  $E$ , wenn

$$\alpha(y) = \sum_{x \in E} \alpha(x) \cdot \Pi(x, y). \quad (2.21)$$

Fasst man  $\alpha$  als Zeilenvektor auf kann man auch  $\alpha \cdot \Pi = \alpha$  schreiben. Also ist  $\alpha$  ein linker Eigenvektor zu  $\Pi$ .

**Definition 30.** *Irreduzible Markov-Kette.*

Eine Markov-Kette heißt *irreduzibel*, wenn alle Zustände miteinander kommunizieren. Das heißt, für beliebige  $x, y \in E$  gibt es ein  $k = k(x, y) \geq 0$  mit  $\Pi^k(x, y) > 0$ .

Außerdem verwenden wir folgende Proposition:

**Proposition 11.** (vgl. [6])

*Miteinander kommunizierende rekurrente Zustände sind alle positiv rekurrent oder alle nullrekurrent.*

*Beweis.* Der Beweis schließt an den Beweis von Satz 14 auf Seite 42 an. Nach Definition 28 ist für Nullrekurrenz und positive Rekurrenz der Erwartungswert  $\mathbb{E}^x(\tau_x)$  ausschlaggebend. Wie in Satz 10 auf Seite 25 gezeigt, ist der Erwartungswert genau der Wert der ersten Ableitung der erzeugenden Funktion an der Stelle 1. Die erzeugende Funktion von  $\tau_x$  haben wir in Gleichung (2.8) auf Seite 42 schon definiert.

Unter Ausnutzung der Regel von de l'Hospital und mit Gleichung (2.10) auf Seite 43 können wir Folgendes schreiben:

$$\lim_{r \nearrow 1} (1-r) \cdot P_{xx}(r) = \lim_{r \nearrow 1} \frac{1-r}{1-R_{xx}(r)} = \frac{1}{R'_{xx}(1)} = \frac{1}{\mathbb{E}^x(\tau_x)} =: \alpha(x). \quad (2.22)$$

Damit ist ein Zustand  $x$  positiv rekurrent, wenn  $\alpha(x) > 0$  und nullrekurrent, wenn  $\alpha(x) = 0$ .

Aus den für kommunizierende Zustände geltenden Ungleichungen (2.18) und (2.19) auf Seite 44 folgern wir dann

$$\frac{1}{\nu\mu} \cdot \lim_{r \nearrow 1} (1-r)P_{yy} \geq \lim_{r \nearrow 1} (1-r)P_{xx} \geq \nu\mu \cdot \lim_{r \nearrow 1} (1-r)P_{yy}$$

und damit

$$\frac{1}{\nu\mu} \cdot \alpha(y) \geq \alpha(x) \geq \nu\mu \cdot \alpha(y). \quad (2.23)$$

Weil  $\nu > 0$  und  $\mu > 0$  ergibt sich daher als Resultat: Ist  $\alpha(x)$  positiv, so ist auch  $\alpha(y)$  positiv. Damit ist aber jeder mit  $x$  kommunizierende Zustand  $y$  positiv rekurrent, wenn  $x$  positiv rekurrent ist. Ebenso ist  $\alpha(y) = 0$  und damit  $y$  nullrekurrent, wenn  $\alpha(x) = 0$  ist.  $\square$

Nun zum Beweis von Satz 15 auf Seite 45:

*Beweis.* Wieder verwenden wir die erzeugenden Funktionen wie in Satz 14 auf Seite 42:

$$P_{xy}(r) := \sum_{n=0}^{\infty} \Pi^n(x, y) \cdot r^n, \quad (2.24)$$

$$R_{xy}(r) := \sum_{n=1}^{\infty} P^x(\tau_y = n) \cdot r^n. \quad (2.25)$$

mit  $0 \leq r < 1$  (und  $\Pi^0(x, y) := \delta_{xy}$ ).

Für den ersten Teil des Beweises nehmen wir an, dass es eine stationäre Verteilung  $a$  gibt, d. h.

$$a(y) = \sum_{x \in E} a(x) \cdot \Pi(x, y) \quad \forall y \in E \quad \text{bzw.} \quad a \cdot \Pi = a.$$

Damit gilt natürlich auch  $a \cdot \Pi^n = a$  und  $a(y) = \sum_{x \in E} a(x) \cdot \Pi^n(x, y)$  für alle  $n$ . Nun multiplizieren wir beide Seiten der zweiten Darstellung mit  $r^n$  und summieren dann über  $n$ :

$$\begin{aligned} a(y) \cdot r^n &= \sum_{x \in E} a(x) \cdot \Pi^n(x, y) \cdot r^n \\ a(y) \sum_{n=1}^{\infty} r^n &= \sum_{x \in E} a(x) \cdot \sum_{n=1}^{\infty} \Pi^n(x, y) \cdot r^n \\ &= \sum_{x \in E} (a(x) \cdot P_{xy}(r)) - a(y). \end{aligned}$$

Nun können wir die Summe auf der linken Seite, die ja bis auf den Summanden  $n = 0$  die geometrische Reihe darstellt, ersetzen durch  $\frac{1}{1-r} - 1 = \frac{r}{1-r}$  (denn  $|r| < 1$  ist erfüllt).

Für die rechte Seite verwenden wir Gleichung (2.9) auf Seite 43:

$$\begin{aligned} \sum_{x \in E} (a(x) \cdot P_{xy}(r)) - a(y) &= \sum_{x \in E} a(x) \cdot (\delta_{xy} + R_{xy}(r) \cdot P_{yy}(r)) - a(y) \\ &= a(y) + P_{yy}(r) \cdot \sum_{x \in E} (a(x) \cdot R_{xy}(r)) - a(y) = P_{yy}(r) \cdot \sum_{x \in E} a(x) \cdot R_{xy}(r). \end{aligned}$$

Insgesamt können wir also schreiben

$$r \cdot a(y) = (1 - r) \cdot P_{yy}(r) \cdot \sum_{x \in E} a(x) \cdot R_{xy}(r). \quad (2.26)$$

Nun bilden wir den linksseitigen Grenzwert  $r \nearrow 1$ . Dazu verwenden wir Gleichung (2.22) auf Seite 46 mit  $\frac{1}{\mathbb{E}^y(\tau_y)} = \alpha(y)$ .

$$a(y) = \alpha(y) \sum_{x \in E} a(x) \cdot R_{xy}(1) \quad (2.27)$$

Aus dieser Gleichung folgt  $a(y) \leq \alpha(y)$ , denn  $R_{xy}(1) \leq 1$  und  $\sum_{x \in E} a(x) = 1$ . Wäre  $y$  nullrekurrent, wäre  $\mathbb{E}^y(\tau_y) = \infty$  laut Definition und damit  $\alpha(y) = 0$ . Ebenso wäre bei transientem Zustand  $\alpha(y) = 0$ . In beiden Fällen wäre dann aber auch  $a(y) = 0$ , und zwar für alle  $y \in E$ . Damit wäre aber die Bedingung  $\sum_{x \in E} a(x) = 1$  für eine stationäre Verteilung nicht erfüllt. Also muss es mindestens ein  $y \in E$  geben mit  $a(y) > 0$ . Dann ist  $y$  positiv rekurrent. Wegen Proposition 11 auf Seite 45 sind dann aber alle Zustände positiv rekurrent.

Es ist  $R_{xy}(1)$  die Wahrscheinlichkeit dafür, von  $x$  aus  $y$  in endlicher Zeit zu erreichen. Damit ist

$$R_{xx}(1) \leq R_{xy}(1)R_{yx}(1) + (1 - R_{xy}(1)), \quad (2.28)$$

denn jeder Weg von  $x$  zurück zu  $x$  führt entweder über  $y$  oder nicht über  $y$ . Damit ist die Wahrscheinlichkeit, in endlicher Zeit zu  $x$  zurückzukehren, sicher kleiner oder gleich der Wahrscheinlichkeit, in endlicher Zeit von  $x$  zu  $y$  und wieder zurück  $x$  zu gelangen plus der Wahrscheinlichkeit, nicht zu  $y$  zu kommen (kleinergleich deshalb, weil auch alle Wege, die nicht zu  $y$ , aber auch nicht mehr zu  $x$  kommen, eingeschlossen sind). Weil  $x$  rekurrent ist, ist  $R_{xx}(1) = 1$ . Die Irreduzibilität bedingt außerdem, dass  $R_{xy}(1) > 0$ . Somit können wir obige Ungleichung umformen zu

$$R_{xy}(1) \leq R_{xy}(1)R_{yx}(1).$$

Daraus folgt  $R_{yx}(1) \geq 1$ . Weil aber auch  $R_{yx}(1) \leq 1$  gilt, folgt daraus

$$R_{yx}(1) = 1 \quad \forall x, y \in E. \quad (2.29)$$

Damit können wir Gleichung (2.27) auf Seite 47 vereinfachen zu

$$a(y) = \alpha(y) \quad \forall y \in E. \quad (2.30)$$

Die stationäre Verteilung ist also eindeutig bestimmt.

Jetzt bleibt noch die „Rückrichtung“ zu zeigen. Dazu nehmen wir an, dass alle Zustände positiv rekurrent sind, also

$$\frac{1}{\mathbb{E}^y(\tau_y)} = \alpha(y) > 0 \quad \forall y \in E.$$

Wir verwenden wieder die geometrische Reihe und schreiben

$$\sum_{n=1}^{\infty} r^n \sum_{y \in E} \Pi^n(x, y) = \frac{r}{1-r},$$

da  $\sum_{y \in E} \Pi^n(x, y) = 1$  gilt. Zieht man das zweite Summenzeichen vor, sieht man, dass auf der linken Seite (bis auf  $n = 0$ )  $P_{xy}(r)$  steht:

$$\begin{aligned} \sum_{y \in E} (P_{xy}(r) - \delta_{xy}) &= \frac{r}{1-r} \\ (1-r) \sum_{y \in E} (P_{xy}(r) - \delta_{xy}) &= r \\ (1-r) \sum_{y \in E} P_{yy}(r) \cdot R_{xy}(r) &= r < 1. \end{aligned}$$

Im letzten Schritt wurde wieder Gleichung (2.9) auf Seite 43 verwendet. Nun machen wir wieder den Grenzübergang  $r \nearrow 1$  und mit Gleichung (2.22)

auf Seite 46 erhalten wir zunächst für jede endliche Teilmenge  $E'$  von  $E$   $\sum_{y \in E'} \alpha(y) R_{xy}(1) \leq 1$  und dann für ganz  $E$

$$\sum_{y \in E} \alpha(y) R_{xy}(1) \leq 1.$$

Wegen Gleichung (2.29) auf Seite 48 ist sogar

$$\sum_{y \in E} \alpha(y) \leq 1. \quad (2.31)$$

Wir müssen nun zeigen, dass Gleichheit gilt und dass  $\alpha$  die stationäre Verteilung ist. Dazu starten wir mit  $\Pi^{n+1}(x, y) = \sum_{z \in E} \Pi^n(x, z) \cdot \Pi(z, y)$ . Diese Gleichung multiplizieren wir mit  $r^n$  und summieren anschließend über  $n$ :

$$\begin{aligned} \frac{1}{r} \cdot r^{n+1} \cdot \Pi^{n+1}(x, y) &= \sum_{z \in E} r^n \cdot \Pi^n(x, z) \cdot \Pi(z, y) \\ \frac{1}{r} \sum_{n=1}^{\infty} \Pi^{n+1}(x, y) \cdot r^{n+1} &= \sum_{z \in E} \Pi(z, y) \cdot \sum_{n=1}^{\infty} \Pi^n(x, z) \cdot r^n \\ \frac{1}{r} (P_{xy}(r) - \delta_{xy} - r \cdot \Pi(x, y)) &= \sum_{z \in E} \Pi(z, y) \cdot (P_{xz}(r) - \delta_{xz}). \end{aligned}$$

Auf beiden Seiten wurden die für das Einsetzen von  $P_{xy}(r)$  bzw.  $P_{xz}(r)$  fehlenden Summanden subtrahiert.

Nun verwenden wir wieder Gleichung (2.9) auf Seite 43 und erhalten

$$\begin{aligned} \frac{1}{r} (\delta_{xy} + R_{xy}(r) \cdot P_{yy}(r) - \delta_{xy}) - \Pi(x, y) &= \sum_{z \in E} \Pi(z, y) \cdot (\delta_{xz} + R_{xz}(r) \cdot P_{zz}(r) - \delta_{xz}) \\ \frac{1}{r} R_{xy}(r) \cdot P_{yy}(r) - \Pi(x, y) &= \sum_{z \in E} R_{xz}(r) \cdot P_{zz}(r) \cdot \Pi(z, y). \end{aligned}$$

Jetzt multiplizieren wir noch mit  $(1 - r)$ :

$$\frac{1}{r} \cdot R_{xy}(r) \cdot (1 - r) \cdot P_{yy}(r) - \Pi(x, y) = \sum_{z \in E} R_{xz}(r) \cdot (1 - r) \cdot P_{zz}(r) \cdot \Pi(z, y).$$

Wiederum betrachten wir den Grenzwert  $r \nearrow 1$  und erhalten mit Gleichung (2.22) auf Seite 46 und Gleichung (2.29) auf Seite 48 zunächst für endliche Teilsommen, dann für ganz  $E$

$$\alpha(y) \geq \sum_{z \in E} \alpha(z) \cdot \Pi(z, y) \quad \forall y \in E. \quad (2.32)$$

Nun summieren wir über  $y$  und verwenden (2.31):

$$1 \geq \sum_{y \in E} \alpha(y) \geq \sum_{z \in E} \alpha(z) \cdot \sum_{y \in E} \Pi(z, y) = \sum_{z \in E} \alpha(z) \quad (2.33)$$

weil  $\sum_{y \in E} \Pi(z, y) = 1$ . Weil die Summe endlich ist, muss schon

$$\alpha(y) = \sum_{z \in E} \alpha(z) \cdot \Pi(z, y) \quad (2.34)$$

gelten.

Jetzt setzen wir  $a(y) = \frac{\alpha(y)}{\sum_{x \in E} \alpha(x)}$ . Dieses  $a$  ist eine stationäre Verteilung, wie sich leicht überprüfen lässt. Weil wir aber schon wissen, dass die stationäre Verteilung eindeutig bestimmt ist, muss  $a(y) = \alpha(y)$  gelten. Deshalb ist  $\sum_{y \in E} \alpha(y) = 1$ . Wir haben also im Fall positiver Rekurrenz eine stationäre Verteilung gefunden.  $\square$

### Konvergenz gegen die stationäre Verteilung

Als Quelle für diesen Teil diene [6].

Als finalen Schritt wollen wir einen Satz über das Konvergenzverhalten von Markov-Ketten formulieren. Dazu benötigen wir allerdings noch eine Eigenschaft von Markovketten: Periodizität.

#### Definition 31. Periodizität.

Als *Periode*  $d(x)$  eines Zustands  $x$  bezeichnet man den größten gemeinsamen Teiler aller  $n \in \mathbb{Z}_+$  mit  $\Pi^n(x, x) > 0$ . Ist  $\Pi^n(x, x) = 0$  für alle  $n$ , setzen wir  $d(x) = \infty$ .

Zustände mit Periode 1 nennt man *aperiodisch*. Von einer aperiodischen Markov-Kette spricht man, wenn alle Zustände aperiodisch sind.

**Bemerkung.** Für aperiodische Zustände gibt es (ohne Beweis) ein  $n_0 \in \mathbb{Z}_+$  mit

$$\Pi^n(x, x) > 0 \quad \forall n \geq n_0.$$

Kommunizieren die Zustände  $x$  und  $y$  miteinander, gibt es ein  $m_0 \in \mathbb{Z}_+$  mit  $\Pi^{m_0}(y, x) > 0$ . Damit ist aber

$$\Pi^n(y, x) \geq \Pi^{m_0}(y, x) \cdot \Pi^{n-m_0}(x, x) > 0$$

für alle  $n \geq n_0 + m_0$ .

Ist eine Markov-Kette also irreduzibel und aperiodisch, dann gibt es für je zwei Zustände  $x$  und  $y$  ein  $N = N(x, y)$  mit  $\Pi^n > 0$  für alle  $n \geq N$ .

Nun zum angekündigten Satz:

**Satz 16.** Konvergenz gegen die stationäre Verteilung

Ist  $(X_n)_{n \geq 0}$  eine irreduzible, aperiodische Markov-Kette mit Zustandsraum  $E$  und Übergangsmatrix  $\Pi$  mit stationärer Verteilung  $\alpha$ , dann gilt

$$\lim_{n \rightarrow \infty} |P(X_n = x) - \alpha(x)| = 0$$

unabhängig von der Startverteilung.

*Beweis.* Der Beweis beruht auf der Kopplung zweier Markov-Ketten. Dazu sei  $(Y_n)_{n \geq 0}$  eine weitere, von  $(X_n)_{n \geq 0}$  unabhängige Markov-Kette mit derselben Übergangsmatrix  $\Pi$ . Dann ist auch  $(X_n, Y_n)_{n \geq 0}$  eine Markov-Kette mit Zustandsraum  $E^2$  und Übergangswahrscheinlichkeiten

$$\Psi(x, u)(y, v) := \Pi(x, y) \cdot \Pi(u, v) \quad \forall (x, u), (y, v) \in E^2.$$

Die so definierte *Produkt-Markov-Kette* ist irreduzibel, aperiodisch und alle Zustände sind positiv rekurrent:

Da die einzelnen Markov-Ketten irreduzibel und aperiodisch sind, gibt es ein  $N \in \mathbb{Z}_+$ , sodass

$$\Pi^n(x, y) > 0 \quad \text{und} \quad \Pi^n(u, v) > 0 \quad \forall n \geq N.$$

Daraus folgt aber auch, dass

$$\Psi^n(x, u)(y, v) = \Pi^n(x, y) \cdot \Pi^n(u, v) > 0 \quad \forall n \geq N.$$

Damit ist auch die Produkt-Markov-Kette irreduzibel und aperiodisch.

Die durch

$$\alpha^*(x, u) := \alpha(x) \cdot \alpha(u) \quad \forall (x, u) \in E^2$$

definierte Verteilung ist die stationäre Verteilung der Produkt-Markov-Kette:

$$\begin{aligned} \sum_{(x,u) \in E^2} \alpha^*(x, u) \cdot \Psi(x, u)(y, v) &= \sum_{x \in E} \sum_{u \in E} \alpha(x) \cdot \alpha(u) \cdot \Pi(x, y) \cdot \Pi(u, v) \\ &= \left( \sum_{x \in E} \alpha(x) \cdot \Pi(x, y) \right) \cdot \left( \sum_{u \in E} \alpha(u) \cdot \Pi(u, v) \right) = \alpha(y) \cdot \alpha(v) = \alpha^*(y, v). \end{aligned}$$

Daher ist die Produkt-Markov-Kette positiv rekurrent.

Wir definieren nun

$$\tau := \min\{n \in \mathbb{Z}_+ : X_n = Y_n\} \quad \text{mit } \tau := \infty \text{ falls } X_n \neq Y_n \forall n \in \mathbb{Z}_+.$$

Aufgrund der Rekurrenz ist für alle  $x \in E$  die Zufallsgröße

$$\begin{aligned} \tau_{(x,x)} &:= \min\{n \in \mathbb{Z}_+ : (X_n, Y_n) = (x, x)\} \\ \text{mit } \tau_{(x,x)} &:= \infty \text{ falls } (X_n, Y_n) \neq (x, x) \forall n \in \mathbb{Z}_+ \end{aligned}$$

fast sicher endlich. Da  $\tau \leq \tau_{(x,x)}$  gilt, muss auch  $\tau$  fast sicher endlich sein. Für alle  $n \geq 1$  und  $m < n$  gilt unabhängig von der Verteilung von  $(X_0, Y_0)$

$$\begin{aligned} P(X_n = x \cap \tau = m) &= P((X_n, Y_n) \in \{x\} \times E \cap \tau = m) \\ &= \sum_{y \in E} \sum_{v \in E} \Psi^{(n-m)}(v, v)(x, y) \cdot P((X_m, Y_m) = (v, v) \cap \tau = m) \\ &= \sum_{v \in E} \Pi^{(n-m)}(v, x) \cdot P((X_m, Y_m) = (v, v) \cap \tau = m) \cdot \underbrace{\sum_{y \in E} \Pi^{(n-m)}(v, y)}_{=1} \\ &= \sum_{v \in E} \Pi^{(n-m)}(v, x) \cdot P((X_m, Y_m) = (v, v) \cap \tau = m). \end{aligned}$$

Das Ereignis im ersten Schritt ist gleich dem Ereignis, dass  $X_m = Y_m = v$  und dann die Produkt-Markov-Kette in  $(n-m)$  Schritten von  $(v, v)$  zu  $(x, y)$  gelangt, wenn man dabei über die frei wählbaren Zustände  $y$  und  $v$  summiert. Auch für  $P(Y_n = x \cap \tau = n)$  für alle  $n \geq 1$  und  $m > n$  erhält man denselben Ausdruck. Das ergibt insgesamt

$$P(X_n = x \cap \tau < n) = P(Y_n = x \cap \tau < n).$$

Weil  $P(X_n = x) = P(X_n = x \cap \tau < n) + P(X_n = x \cap \tau \geq n)$  gilt, folgt daraus aber sofort

$$\begin{aligned} |P(X_n = x) - P(Y_n = x)| &= |P(X_n = x \cap \tau \geq n) - P(Y_n = x \cap \tau \geq n)| \\ &\leq P(X_n = x \cap \tau \geq n) + P(Y_n = x \cap \tau \geq n) \\ &\leq 2 \cdot P(\tau \geq n). \end{aligned}$$

Weil aber  $\tau$  fast sicher endlich ist, gilt  $\lim_{n \rightarrow \infty} P(\tau \geq n) = 0$  und daher auch

$$\lim_{n \rightarrow \infty} |P(X_n = x) - P(Y_n = x)| = 0.$$

Weil diese Betrachtungen unabhängig von der Startverteilung durchgeführt wurden, können wir  $P(Y_0 = x) = \alpha(x)$  setzen. Damit ist dann aber auch  $P(Y_n = x) = \alpha(x)$ , weil  $\alpha$  ja die stationäre Verteilung ist. Und dann haben wir schon

$$\lim_{n \rightarrow \infty} |P(X_n = x) - \alpha(x)| = 0.$$

□

Nach dieser eingehenden Betrachtung der Theorie hinter den Markov-Ketten wollen wir uns nun im nächsten Kapitel einem speziellen, etwas angewandteren Bereich zuwenden, den sogenannten Irrfahrten.



# Kapitel 3

## Irrfahrten

Dieses Kapitel ist eine Anwendung der Theorie der Markov-Ketten. Irrfahrten finden sich in vielen Bereichen. So kann zum Beispiel die Bewegung eines Teilchens in einer Flüssigkeit als Irrfahrt betrachtet werden – oder eben das Verhalten eines Betrunkenen am Heimweg und das eines Kindes in einem Klettergerüst. Die Betrachtungen schränken wir auf sogenannte einfache Irrfahrten auf  $\mathbb{Z}$  ein, die nur die Möglichkeit eines Schritts der Länge 1 in eine Richtung haben.

Die Großteil dieses Kapitels findet sich in ähnlicher Form in [10]. Die Quellen wurden in diesem Fall nicht explizit angegeben. Einige Abschnitte stammen aber auch aus [5] und [12]. Die Quellen [1], [4] und [9] wurden zum Abgleich der Inhalte verwendet.

### 3.1 Einfache Irrfahrten auf $\mathbb{Z}$

Wir stellen uns folgende Situation vor:

Ein betrunkenen Mensch steht auf einer schnurgeraden Straße, die auf beiden Seiten bis ins Unendliche geht. Der Mensch beginnt zu gehen. Da er die Orientierung verloren hat und nicht weiß, wo er hingehen soll bzw. will, entscheidet er immer wieder aufs Neue, in welche Richtung er gehen soll und wie weit er geht. Wo wird sich der Mensch wann befinden? Dazu muss man nur den gesamten zurückgelegten Weg in einer Richtung nehmen und den gesamten Weg in die Gegenrichtung abziehen. Doch was hat das mit Markov-Ketten zu tun?

Wir versuchen nun, ein mathematisches Modell für diese Situation aufzustellen. Dazu machen wir einige vereinfachende Annahmen. Der Betrunkene beginnt zum Zeitpunkt 0 mit seiner Wanderung und nach jeder ganzen Zeit-

einheit (z.B. nach jeder Minute) stellt er seine Überlegungen über Richtung und Schrittlänge im nächsten Zeitintervall an. Die Schrittlänge ist unabhängig von der Schrittlänge pro Zeiteinheit konstant. Wir nehmen eine Schrittlänge von 1 (z.B. Meter) an. Außerdem setzen wir voraus, dass die Anzahl der Schritte pro Zeiteinheit identisch verteilt ist. Schrittlänge und Schrittrichtung werden im Folgenden zur *Schrittzahl*  $Z \in \mathbb{Z}$  zusammengefasst, deren Betrag die Schrittlänge und deren Vorzeichen die Schrittrichtung angibt. (Die Schrittzahl unseres Menschen ist natürlich beschränkt. Für die allgemeine Betrachtung ist dies allerdings nicht Voraussetzung.)

Eigentlich ist die Straße der Zustandsraum. Wir ersetzen die Straße durch  $\mathbb{Z}$ , die Menge der ganzen Zahlen. Die Schrittlänge und -richtung  $(Z_i)_{i \geq 1}$  modellieren wir als  $\mathbb{Z}$ -wertige unabhängige Zufallsvariablen mit identischer Verteilung  $\varrho$ . Für die Zeitrechnung nehmen wir die natürlichen Zahlen. Da wir am Ort zum Zeitpunkt  $n$  interessiert sind, setzen wir  $X_0 = 0$  (Start bei 0) und  $X_n = \sum_{i=1}^n Z_i$  für  $n \geq 1$ . (Da die Richtung der Schrittzahlen im Modell durch das Vorzeichen gegeben ist, ist die Summe der Schritte logischerweise der Ort.)

Unter diesen Voraussetzungen ist  $(X_n)_{n \geq 0}$  eine Markov-Kette zur Übergangsmatrix  $\Pi(x, y) = \varrho(y - x)$ ,  $x, y \in \mathbb{Z}$ . Doch warum ist die Folge eine Markov-Kette?

Dazu ziehen wir die Bedingung für eine Markov-Kette, Gleichung (2.1) aus Definition 21 auf Seite 27, heran:

$$\Pi(x_n, x_{n+1}) = P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n).$$

Die Wahrscheinlichkeit für  $X_{n+1} = x_{n+1}$  (unter der Bedingung der bekannten „Vorgeschichte“  $X_0, \dots, X_n$ ) ist gleich der Wahrscheinlichkeit (unter selber Bedingung) dafür, dass der letzte Schritt  $Z_{n+1}$  den Wert  $x_{n+1} - x_n$  hat. Die Schrittlänge  $Z_{n+1}$  ist aber unabhängig von  $X_0, \dots, X_n$  mit  $\varrho$  verteilt. Insgesamt ergibt sich also:

$$\begin{aligned} \Pi(x_n, x_{n+1}) &= P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) \\ &= P(Z_{n+1} = x_{n+1} - x_n | X_0 = x_0, \dots, X_n = x_n) = \varrho(x_{n+1} - x_n). \end{aligned}$$

Markov-Ketten dieser Gestalt heißen *Irrfahrten*.

Im Folgenden betrachten wir sogenannte einfache Irrfahrten. Das sind Irrfahrten bei denen das Objekt nur einen Schritt nach vorne oder hinten machen kann. (Ein betrunkenen Mensch wird sein Tempo ja auch nicht groß variieren.)

Dazu sei  $p$  eine feste Zahl aus dem offenen Intervall  $(0, 1)$  und  $\{Z_n : n \in \mathbb{Z}^+\}$  eine Folge von Zufallsvariablen mit Werten aus  $\{-1, 1\}$ , wobei 1 mit Wahrscheinlichkeit  $p$  angenommen wird. (D. h. es geht mit Wahrscheinlichkeit  $p$  vorwärts. Der Betrunkene hat zumindest eine Ahnung, in welche Richtung er gehen sollte und bevorzugt daher eine Richtung.) Für die Verteilung  $\varrho$  der Zufallsvariablen erhält man

$$\begin{aligned}\varrho(+1) &= p, \\ \varrho(-1) &= 1 - p =: q.\end{aligned}$$

Die formale Darstellung der Zufallsvariablen  $(Z_n)_{n \geq 0}$  sieht folgendermaßen aus:

$$Z_k : E \rightarrow \{-1, 1\} \quad Z_k(\epsilon) = \epsilon_k \quad \text{für } \epsilon = (\epsilon_1, \dots),$$

wobei  $E = \{\epsilon = (\epsilon_1, \dots) : \epsilon_i \in \{-1, 1\}\}$ .

Für ein Ereignis  $\epsilon = (\epsilon_1, \dots, \epsilon_n) \in \{-1, 1\}^n$  und  $n \in \mathbb{Z}^n$  gilt dann:

$$\begin{aligned}P(Z_1 = \epsilon_1, \dots, Z_n = \epsilon_n) &= p^{N(\epsilon)} \cdot q^{n-N(\epsilon)}, \quad q = 1 - p \quad \text{und} \\ N(\epsilon) = |\{m : \epsilon_m = 1\}| &= \frac{n + S_n(\epsilon)}{2}, \quad \text{wobei } S_n(\epsilon) = \sum_{m=1}^n \epsilon_m.\end{aligned}$$

Die Formel für die Anzahl der Zufallsvariablen mit Wert  $+1$  lässt sich leicht verifizieren: Offensichtlich gilt:

$$\begin{aligned}N(\epsilon) - (n - N(\epsilon)) &= S_n(\epsilon) \quad | + n \\ 2N(\epsilon) &= n + S_n(\epsilon).\end{aligned}$$

Nun setzt man

$$X_0 = 0 \quad \text{und} \quad X_n = \sum_{m=1}^n Z_m \quad \text{für } n \in \mathbb{Z}^+. \quad (3.1)$$

(Ab hier werden wir für die Summe  $S_n(\epsilon)$  eines bestimmten Ereignisses  $X_n(\epsilon)$  schreiben.)

Die  $(X_n)_{n \geq 0}$  sind Zufallsvariablen mit

$$X_k : E \rightarrow \mathbb{Z} \quad X_k(\epsilon) = \sum_{k=1}^n Z_k(\epsilon).$$

Auf diese Weise kann eine Irrfahrt, die nur einen Schritt machen kann, beschrieben werden. Allerdings ist diese Beschreibung ziemlich „statisch“.

Hauptaugenmerk wird auf den gegenwärtigen Zustand gelegt, und nicht so sehr auf die Veränderung. Gerade bei Irrfahrten interessiert man sich aber für die dynamische Veränderung.

Um den dynamischen Aspekt hervorzuheben, könnte man (3.1) ersetzen durch

$$P(X_0 = 0) = 1 \quad \text{und} \\ P(X_n - X_{n-1} = \epsilon \mid X_0, \dots, X_{n-1}) = \begin{cases} p & \text{falls } \epsilon = 1 \\ q, & \text{falls } \epsilon = -1. \end{cases} \quad (3.2)$$

Die Aussage ist dieselbe, (3.2) ist aber die dynamischere Beschreibung. Speziell beschreibt (3.2) den sicheren Start bei 0 zum Zeitpunkt  $n = 0$  und die Verteilung der Differenzen zwischen zwei aufeinanderfolgenden Zufallsvariablen  $X_n$ . In diesem Fall geschieht zu jedem Zeitpunkt  $n \in \mathbb{Z}$  ein Schritt nach vorne mit Wahrscheinlichkeit  $p$  und ein Schritt zurück mit Wahrscheinlichkeit  $q$ , und zwar unabhängig von der „Vorgeschichte“.

### 3.1.1 Verteilung zum Zeitpunkt $n$

Im Folgenden wird  $P(X_n = m)$  auf zwei Arten berechnet. Die erste Art basiert auf (3.1) auf Seite 57.

Natürlich ist klar, dass  $P(|X_n| \leq n) = 1$  gilt. Die größte Entfernung vom Startpunkt  $X_0 = 0$  erreicht man, wenn man immer nur in eine Richtung geht. Mit  $n$  Schritten ist man maximal  $n$  Schritte vom Start entfernt. Außerdem weiß man:

$$\begin{aligned} n \text{ gerade} &\Rightarrow P(X_n \text{ ist gerade}) = 1 \quad \text{und} \\ n \text{ ungerade} &\Rightarrow P(X_n \text{ ist ungerade}) = 1. \end{aligned}$$

Nun wissen wir, welche Zahlen  $m$  überhaupt als „Standort zum Zeitpunkt  $n$ “ in Frage kommen. Eine Zahl  $m$  erfülle die Voraussetzungen  $m \in \{-n, \dots, n\}$  und außerdem sei  $m$  gerade, falls  $n$  gerade und sonst ungerade. Des Weiteren sei  $\epsilon$  ein Ereignis  $\epsilon = (\epsilon_1, \dots, \epsilon_n) \in \{-1, 1\}^n$  mit  $X_n(\epsilon) = m$ . Dann ist  $N(\epsilon) = \frac{n+m}{2}$  und daher

$$P(Z_1 = \epsilon_1, \dots, Z_n = \epsilon_n) = p^{\frac{n+m}{2}} \cdot q^{\frac{n-m}{2}}. \quad (3.3)$$

Das Ereignis  $\epsilon$  ist nur eines von vielen mit dem Wert  $m$ . Einzige Voraussetzung ist, dass die Gesamtanzahl der positiven Schritte gleich ist. Auf die Reihenfolge der  $\epsilon_i$  kommt es nicht an. Daher gibt es  $\binom{n}{\frac{n+m}{2}}$  Ereignisse, die zur

Summe  $m$  führen. Für die gesuchte Wahrscheinlichkeit, nach  $n$  Schritten bei der Summe  $m$  angekommen zu sein, erhalten wir daher:

$$P(X_n = m) = \binom{n}{\frac{n+m}{2}} \cdot p^{\frac{n+m}{2}} \cdot q^{\frac{n-m}{2}} \quad (3.4)$$

falls  $m \in \mathbb{Z}$ ,  $|m| \leq n$  und falls  $m$  gerade ist genau dann, wenn  $n$  gerade ist, und 0 sonst.

Nun zur zweiten Berechnung, basierend auf der dynamischeren Beschreibung (3.2).

Dazu führen wir eine neue Notation ein:  $(P^n)_m = P(X_n = m)$ . Offensichtlich ist  $(P^0)_m = \delta_{0,m}$ , wobei  $\delta_{k,l}$  das *Kroneckerdelta* mit Wert 1 bei  $k = l$  und 0 sonst ist. Aus (3.2) auf Seite 58 sieht man:

$$\begin{aligned} P(X_n = m) &= P(X_{n-1} = m - 1 \wedge X_n = m) + P(X_{n-1} = m + 1 \wedge X_n = m) \\ &= p \cdot P(X_{n-1} = m - 1) + q \cdot P(X_{n-1} = m + 1), \end{aligned}$$

denn um zu  $m$  zu gelangen muss man entweder von  $m - 1$  einen Schritt nach vorne oder von  $m + 1$  einen Schritt zurück machen. In der oben eingeführten Notation sieht folgendermaßen aus:

$$(P^0)_m = \delta_{0,m} \text{ und } (P^n)_m = p \cdot (P^{n-1})_{m-1} + q \cdot (P^{n-1})_{m+1}. \quad (3.5)$$

(3.5) ist eine rekursive Darstellung zur Berechnung der  $(P^n)_m$ . Kennt man  $(P^0)_m$ , könnte man alle  $(P^n)_m$  berechnen. Es lässt sich leicht überprüfen, dass (3.4) diese Darstellung erfüllt. Es gilt:

$$\begin{aligned} (P^{n-1})_{m-1} &= \binom{n-1}{\frac{n+m}{2}-1} \cdot p^{\frac{m+n}{2}-1} \cdot q^{\frac{n-m}{2}}, \\ (P^{n-1})_{m+1} &= \binom{n-1}{\frac{n+m}{2}} \cdot p^{\frac{m+n}{2}} \cdot q^{\frac{n-m}{2}-1}. \end{aligned}$$

Eingesetzt in (3.5) erhält man:

$$\begin{aligned} &p \cdot (P^{n-1})_{m-1} + q \cdot (P^{n-1})_{m+1} \\ &= \binom{n-1}{\frac{n+m}{2}-1} \cdot p^{\frac{n+m}{2}} \cdot q^{\frac{n-m}{2}} + \binom{n-1}{\frac{n+m}{2}} \cdot p^{\frac{n+m}{2}} \cdot q^{\frac{n-m}{2}} \\ &= p^{\frac{m+n}{2}} \cdot q^{\frac{n-m}{2}} \cdot \left[ \binom{n-1}{\frac{n+m}{2}-1} + \binom{n-1}{\frac{n+m}{2}} \right] \\ &= \binom{n}{\frac{n+m}{2}} \cdot p^{\frac{n+m}{2}} \cdot q^{\frac{n-m}{2}} = (P^n)_m. \end{aligned}$$

Die Beschreibungen sind also nach wie vor äquivalent. Während mit der ersten Art  $(P^n)_m$  relativ einfach ausgerechnet werden kann, sind mit der zweiten Darstellung viele Rechenschritte notwendig. Zumindest kann man feststellen, ob ein bestimmter Wert  $m$  mit  $n$  Schritten überhaupt möglich ist. Dazu folgende Proposition:

**Proposition 12.** *Für eine Irrfahrt, die durch (3.2) beschrieben wird, gilt  $(P^n)_m = 0$ , außer  $m = 2l - n$  für  $0 \leq l \leq n$ .*

Formt man  $m = 2l - n$  um, erhält man  $l = \frac{m+n}{2}$ . Das wiederum ist die Anzahl der positiven Schritte  $N(\epsilon)$  in der ersten Darstellung. Und die muss natürlich kleiner oder gleich  $n$  sein. die Äquivalenz der Darstellungen ist also auch hier sofort sichtbar. Trotzdem soll Proposition 12 aus der zweiten Darstellung heraus bewiesen werden.

*Beweis.* Der Beweis erfolgt mittels vollständiger Induktion nach  $n$ .  
*Induktionsanfang:*  $n = 0$

$$(P^0)_m = \delta_{0,m} = \begin{cases} 1 & \text{falls } m = 0, \\ 0 & \text{sonst.} \end{cases}$$

Da  $n = 0$  ist auch  $l = 0$  und daher  $m = 2l - n = 0$ .

*Induktionsvoraussetzung:*

Angenommen  $(P^n)_m \neq 0$  genau dann wenn  $m = 2l - n$  für  $0 \leq l \leq n$  ist erfüllt.

*Induktionsschritt:* Wegen (3.5) auf Seite 59 gilt für  $n + 1$ :

$$(P^{n+1})_m = p \cdot (P^n)_{m-1} + q \cdot (P^n)_{m+1}.$$

Man will nun wissen, unter welchen Voraussetzungen  $(P^{n+1})_m$  ungleich 0 ist: Diese Summe ist genau dann ungleich 0, wenn mindestens einer der Summanden ungleich 0 ist.

$(P^n)_{m-1}$  ist genau dann ungleich 0 wenn gilt:  $m - 1 = 2l - n$

$(P^n)_{m+1}$  ist genau dann ungleich 0 wenn gilt:  $m + 1 = 2l - n$

und in beiden Fällen  $0 \leq l \leq n$ . (Induktionsvoraussetzung)

Man formt um:

$$\begin{array}{ll} m - 1 = 2l - n & m + 1 = 2l - n \\ m = 2l - n + 1 & m = 2l - n - 1 \\ = 2(l + 1) - (n + 1) & = 2l - (n + 1) \end{array}$$

Damit ist alles gezeigt, denn:

$(P^{n+1})_m \neq 0$  genau dann wenn  $m = 2l - (n + 1)$  oder  $m = 2(l + 1) - (n + 1)$  oder beides.

Außerdem gilt:  $0 \leq l + 1 \leq n + 1$  und  $0 \leq l \leq n + 1$ . □

Auch ohne die für (3.4) auf Seite 59 angestellten Überlegungen kann man zeigen, dass die Koeffizienten die Binomialkoeffizienten sind. Dazu überlegen wir, welche Ereignisse  $\epsilon$  zur Summe  $2l - n$  (nur die sind wegen Proposition 12 interessant) führen. Die Summe von  $x$  Vorwärts- und  $n - x$  Rückwärtsschritten muss  $2l - n$  ergeben:  $x - (n - x) = 2l - n$ . Also ist  $x$  gleich  $l$ . Für ein Ereignis  $\epsilon$  gilt also  $P(\epsilon) = p^l \cdot q^{n-l}$ . Da das für alle Möglichkeiten gilt, die zu  $2l - n$  führen, müssen wir noch mit der Anzahl der Möglichkeiten multiplizieren. Wir setzen für den Koeffizienten  $(C^n)_l$  und zeigen:

**Proposition 13.** *Stellt man  $(P^n)_{2l-n}$  in der Form  $(P^n)_{2l-n} = (C^n)_l \cdot p^l \cdot q^{n-l}$  dar, so gilt für die Koeffizienten  $(C^n)_l = (C^{n-1})_{l-1} + (C^{n-1})_l$*

*Beweis.* Induktion nach  $n$ .

*Induktionsanfang:*  $n = 1$

$$\begin{aligned} (C^0)_{l-1} + (C^0)_l &= p^{-l+1} \cdot q^{0+l-1} \cdot (P^0)_{2(l-1)-0} + p^{-l} \cdot q^{0+l} \cdot (P^0)_{2l-0} \\ &= p^{-l} \cdot q^{-1+l} \cdot \underbrace{(p \cdot (P^0)_{(2l-1)-1} + q \cdot (P^0)_{(2l-1)+1})}_{=(P^1)_{2l-1} \text{ wegen (3.5) auf Seite 59}} \\ &= p^{-l} \cdot q^{-1+l} \cdot (P^1)_{2l-1} = (C^1)_l. \end{aligned}$$

*Induktionsvoraussetzung:* Angenommen es gilt

$$(C^n)_l = (C^{n-1})_{l-1} + (C^{n-1})_l \text{ wenn } (C^n)_l = p^{-l} \cdot q^{-n+l} \cdot (P^n)_{2l-n}$$

*Induktionsschritt:*

$$\begin{aligned} (C^n)_{l-1} + (C^n)_l &= p^{-l+1} \cdot q^{-n+l-1} \cdot (P^n)_{2(l-1)-n} + p^{-l} \cdot q^{-n+l} \cdot (P^n)_{2l-n} \\ &= p^{-l} \cdot q^{-n+l-1} \cdot \underbrace{(p \cdot (P^n)_{2l-(n+1)-1} + q \cdot (P^n)_{2l-(n+1)+1})}_{=(P^{n+1})_{2l-(n+1)} \text{ wegen (3.5) auf Seite 59}} \\ &= p^{-l} \cdot q^{-(n+1)+l} \cdot (P^{n+1})_{2l-(n+1)} = (C^{n+1})_l. \end{aligned}$$

□

### 3.1.2 Eintrittszeiten

Wir interessieren uns nun für die Verteilung der Eintrittszeiten in einen bestimmten Punkt  $a \in \mathbb{Z}$ , d. h. dafür, wann der Betrunke eine bestimmten Ort erreicht. Dazu erinnern wir uns an Definition 24 auf Seite 37:

$$\tau_a = \inf\{n \geq 1 : X_n = a\}$$

Wir suchen nun die Verteilung von  $\tau_a$ , also einen Ausdruck für  $P(\tau_a = n)$ . (genaugenommen suchen wir  $\tau_{0a}$ . Weil wir aber  $X_0 = 0$  voraussetzen, werden wir im Folgenden immer  $\tau_a$  schreiben.) Wie wir aus vorhergehenden Überlegungen wissen, müssen wir nur  $n$ 's berücksichtigen mit  $n \geq |a|$  und  $n$  gerade  $\Leftrightarrow a$  gerade.

Für das Folgende schränken wir  $a$  auf  $\mathbb{Z}^+$  ein. natürlich gilt

$$P(\tau_a = n) = P(X_n = a \wedge \tau_a > n - 1) = p \cdot P(\tau_a > n - 1 \wedge X_{n-1} = a - 1).$$

Damit suchen wir nun  $P(\tau_a > n - 1 \wedge X_{n-1} = a - 1)$ .

Wie für Gleichung (3.3) auf Seite 58 ist die Wahrscheinlichkeit für ein Ereignis  $\epsilon \in \{-1, 1\}^{n-1}$  mit  $X_{n-1}(\epsilon) = a - 1$  gesucht, die sich zu  $p^{\frac{n+a}{2}-1} \cdot q^{\frac{n-a}{2}}$  ergibt. Damit ist

$$P(\tau_a = n) = N(n, a) \cdot p^{\frac{n+a}{2}} \cdot q^{\frac{n-a}{2}} \quad (3.6)$$

und  $N(n, a)$  steht für die Anzahl der Ereignisse deren Summe  $X_{n-1}(\epsilon)$  bis zum Zeitpunkt  $n - 1$  kleiner als  $a$  ist. Diese Ereignisse sind jene, für die gilt:  $X_l(\epsilon) \leq a - 1$  für  $0 \leq l \leq n - 1$  und  $X_{n-1}(\epsilon) = a - 1$ .

Das Problem wurde also auf die Bestimmung von  $N(n, a)$  reduziert.

Ähnlich wie in Gleichung (3.4) auf Seite 59 ist die Anzahl aller Ereignisse mit  $X_{n-1}(\epsilon) = a - 1$  gleich  $\binom{n-1}{\frac{n+a}{2}-1}$  und wir können schreiben:

$$N(n, a) = \binom{n-1}{\frac{n+a}{2}-1} - N'(n, a). \quad (3.7)$$

Hier steht  $N'(n, a)$  für die Anzahl der Ereignisse  $\epsilon \in \{-1, 1\}^{n-1}$ , bei denen  $X_l(\epsilon) \geq a$  für ein oder mehrere  $l \leq n - 1$  vorkommt und  $X_n = a - 1$  ist. Für die Lösung des Problems brauchen wir also einen Ausdruck für  $N'(n, a)$ .

Dazu müssen wir etwas weiter ausholen:

### Reflexionsprinzip

Dieser Abschnitt wurde in Anlehnung an [12] erstellt.

Wir definieren für eine Irrfahrt  $(X_n)_{n \geq 0}$  die Zufallsvariablen  $M_n$ , die das Maximum dieser Irrfahrt bis zum Zeitpunkt  $n$  angeben:

$$M_n : E \rightarrow \mathbb{Z} \quad M_n := \max\{X_1(\epsilon), \dots, X_n(\epsilon)\}$$

für  $n \in \mathbb{Z}^+$ . Für alle  $a > 0$  gilt offensichtlich

$$P(M_n \geq a) = P(\tau_a \leq n).$$

Das Ereignis „Maximum bis zum Zeitpunkt  $n$  ist mindestens  $a$ “ ist gleich dem Ereignis „Die Irrfahrt ist spätestens zum Zeitpunkt  $n$  das erste Mal bei  $a$ “.

**Satz 17.** Reflexionsprinzip.

Für eine Irrfahrt  $(X_n)_{n \geq 0}$  und  $a \in \mathbb{Z}_+$  und  $b < a, b \in \mathbb{Z}$  gilt:

$$P(M_n \geq a, X_n = b) = P(X_n = 2a - b).$$

*Beweis.* Wir zeigen, dass die Ereignisse  $U(n, a) := \{M_n \geq a, X_n = b\}$  und  $O(n, a) := \{M_n \geq a, X_n = 2a - b\}$  gleich mächtig sind. Dann sind auch die Wahrscheinlichkeiten gleich groß. (Die Bedingung „ $M_n \geq a$ “ ist bei  $O(n, a)$  nicht notwendig, da  $X_n > a$  wegen  $2a - b > a$  ohnehin erfüllt ist.)

Dazu definieren wir für einen Weg  $X \in \{(X_n)_{n \geq 0} \text{ mit } M_n \geq a\}$  den Zeitpunkt des ersten Erreichens von  $a$ :  $m(X) = \min\{k : X_k \geq a\}$ .

Außerdem definieren wir die *Reflexions*-Abbildung  $r(X) = X' = (X'_n)_{n \geq 0}$ . Dabei ist  $X'_k = X_k$  für  $0 \leq k \leq m(X)$  und  $X'_k = 2a - X_k$  für  $k > m(X)$ . Ab  $X_{m(X)}$  wird also an  $a$  gespiegelt.

Die Reflexion  $r$  bildet  $U(n, a)$  auf  $O(n, a)$  ab: Zu jeder Kette  $X$  aus  $U(n, a)$  liegt die ab  $m(X)$  an  $a$  gespiegelte Kette  $X'$  in  $O(n, a)$ , da  $X_n = b$  auf  $X'_n = 2a - b$  abgebildet wird. Natürlich bildet  $r$  auch  $O(n, a)$  auf  $U(n, a)$  ab. Außerdem ist  $r \circ r = \text{id}$  die identische Abbildung. Daraus folgt, dass  $r : U(n, a) \rightarrow O(n, a)$  bijektiv ist. Daher sind  $U(n, a)$  und  $O(n, a)$  gleich mächtig.  $\square$

Nun kehren wir zurück zur Verteilung der Rückkehrzeiten. Wir suchen die Anzahl  $N'(n, a)$  der Ereignisse  $\epsilon$  in  $\{-1, 1\}^{n-1}$  mit  $M_{n-1}(\epsilon) \geq a$  und  $X_{n-1}(\epsilon) = a - 1$ . Jetzt können wir das Reflexionsprinzip anwenden, denn

$$\begin{aligned} N'(n, a) &= |\{M_{n-1} \geq a, X_{n-1} = a - 1\}| \\ &= |\{X_{n-1} = 2a - (a - 1)\}| = |\{X_{n-1} = a + 1\}|. \end{aligned}$$

Die Anzahl der Irrfahrten mit  $X_{n-1} = a + 1$  kennen wir aber:

$$N'(n, a) = |\{X_{n-1} = a + 1\}| = \binom{n-1}{\frac{n+a}{2}}.$$

Eingesetzt in Gleichung (3.7) auf Seite 62 erhalten wir

$$N(n, a) = \binom{n-1}{\frac{n+a}{2} - 1} - \binom{n-1}{\frac{n+a}{2}} = \frac{a}{n} \cdot \binom{n}{\frac{n+a}{2}}.$$

Die zweite Gleichung wird weiter unten noch begründet. Jetzt müssen wir dieses Ergebnis nur noch in Gleichung (3.6) auf Seite 62 einsetzen:

$$P(\tau_a = n) = \frac{a}{n} \cdot \binom{n}{\frac{n+a}{2}} \cdot p^{\frac{n+a}{2}} \cdot q^{\frac{n-a}{2}} = \frac{a}{n} \cdot P(X_n = a). \quad (3.8)$$

Die letzte Gleichung folgt aus Gleichung (3.4) auf Seite 59.

Bis jetzt haben wir unsere Überlegungen auf  $a > 0$  beschränkt. Wir sind aber auch an der Verteilung für  $a < 0$  interessiert. Man könnte analog wie oben jeden Schritt für  $a < 0$  durchgehen. Eleganter und schneller ist allerdings folgende Vorgehensweise:

Wir „spiegeln“ die ganze Irrfahrt an der Zeitachse. Das heißt wir vertauschen  $p$  und  $q$ . Die Wahrscheinlichkeit, mit der ursprünglichen Irrfahrt nach  $n$  Schritten erstmals bei  $a < 0$  zu landen ist genau gleich groß wie die Wahrscheinlichkeit, mit der gespiegelten Irrfahrt bei  $-a$  zu landen. Es gilt aber  $-a > 0$ ! Daher kann obige Gleichung verwenden, wenn wir  $p$  und  $q$  vertauschen und  $a$  durch  $-a$  ersetzen.

$$P(\tau_{-a} = n) = -\frac{a}{2} \cdot \binom{n}{\frac{n-a}{2}} \cdot q^{\frac{n-a}{2}} \cdot p^{\frac{n+a}{2}} \quad (3.9)$$

weil für  $a < 0$  gilt  $-a = |a|$  und weil  $\binom{n}{\frac{n-a}{2}} = \binom{n}{\frac{n+a}{2}}$  können wir Gleichung (3.8) auf Seite 63 und Gleichung (3.9) zu folgendem Satz zusammenführen:

**Satz 18.** Verteilung der Eintrittszeiten.

Für eine einfache Irrfahrt  $(X_n)_{n \geq 0}$  gilt:

$$a \neq 0 \quad \Rightarrow \quad P(\tau_a = n) = \frac{|a|}{n} \cdot P(X_n = a). \quad (3.10)$$

für  $n \geq |a|$  und  $n$  gerade  $\Leftrightarrow a$  gerade. Sonst ist  $P(\tau_a = n) = 0$ .

Nachzuliefern ist noch die Gleichung

$$\binom{n-1}{\frac{n+a}{2}-1} - \binom{n-1}{\frac{n+a}{2}} = \frac{a}{n} \cdot \binom{n}{\frac{n+a}{2}}.$$

Dazu setzen wir  $k = \frac{n+a}{2}$  und formen  $\binom{n-1}{k-1} - \binom{n-1}{k}$  geeignet um:

$$\begin{aligned} \binom{n-1}{k-1} - \binom{n-1}{k} &= \frac{(n-1)!}{(k-1)!(n-k)!} - \frac{(n-1)!}{k!(n-k-1)!} \\ &= \frac{k(n-1)!}{k!(n-k)!} - \frac{(n-k)(n-1)!}{k!(n-k)!} \\ &= \frac{(n-1)!}{k!(n-k)!} \cdot (k-n+k) \\ &= \frac{n!}{k!(n-k)!} \cdot \frac{2k-n}{n} = \binom{n}{k} \cdot \frac{a}{n}. \end{aligned}$$

### 3.1.3 Eintritt in endlicher Zeit, Absorption am Rand

Im letzten Abschnitt wurde die Wahrscheinlichkeit berechnet, *nach  $n$  Schritten* zum ersten Mal bei  $a$  zu landen. Oft ist man aber daran interessiert, mit welcher Wahrscheinlichkeit man *in endlicher Zeit* (zum ersten Mal) zu einer Stelle  $a$  kommt. Insbesondere möchte man wissen, ob  $a$  mit Sicherheit in endlicher Zeit erreicht wird, also  $P(\tau_a < \infty) = 1$ . Das Unterfangen des Betrunkenen, einen bestimmten Ort zu finden, sei es sein Zuhause oder eine Gaststätte, sollte ja zumindest theoretisch Aussicht auf Erfolg haben. Dafür leistet die obige Gleichung (3.10) keine Hilfe. Mit ein paar Tricks und geschickten Umformungen wollen wir aber  $P(\tau_a < \infty)$  trotzdem bestimmen.

Im Folgenden sei  $a$  wieder eine feste Zahl  $> 0$ . Die Eintrittszeit  $\tau_a$  können wir als Funktion  $\tau_a(Z_1, \dots, Z_n, \dots)$  sehen, die  $\{-1, 1\}^{\mathbb{Z}^+}$  auf  $\mathbb{Z}^+ \cup \infty$  abbildet und zwar derart, dass für alle  $n \in \mathbb{Z}_+$  gilt:

$$\tau_a(\epsilon_1, \dots, \epsilon_n, \dots) > n \iff \sum_{k=1}^m \epsilon_k < a \quad \text{für } 1 \leq m \leq n.$$

Wir definieren nun  $\tau_1 \circ \Sigma^m := \tau_1(Z_{m+1}, \dots, Z_{m+n}, \dots)$ .

Damit gilt  $\tau_a = m \Rightarrow \tau_{a+1} = m + \tau_1 \circ \Sigma^m$ . Natürlich hängt das Ereignis  $\{\tau_a = m\}$  nur von  $(Z_1, \dots, Z_m)$  ab und das Ereignis  $\{\tau_1 \circ \Sigma^m < \infty\}$  von  $(Z_{m+1}, \dots)$ . Insbesondere sind diese Ereignisse unabhängig voneinander.

Es gilt außerdem:  $\{\tau_a = m \wedge \tau_{a+1} < \infty\} = \{\tau_a = m\} \cap \{\tau_1 \circ \Sigma^m < \infty\}$ .  
Damit erhalten wir:

$$\begin{aligned} P(\tau_{a+1} < \infty) &= \sum_{m=1}^{\infty} P(\tau_a = m \wedge \tau_{a+1} < \infty) \\ &= \sum_{m=1}^{\infty} P(\tau_a = m) \cdot P(\tau_1 \circ \Sigma^m < \infty) \\ &= P(\tau_1 < \infty) \cdot \sum_{m=1}^{\infty} P(\tau_a = m) = P(\tau_1 < \infty) \cdot P(\tau_a < \infty) \end{aligned}$$

Verwendet wurde hier, dass  $(B_{m+1}, \dots)$  und  $(B_1, \dots)$  dieselbe Verteilung haben und damit auch  $\tau_1 \circ \Sigma^m$  und  $\tau_1$ .

Wir haben gezeigt, dass  $P(\tau_{a+1} < \infty) = P(\tau_1 < \infty) \cdot P(\tau_a < \infty)$  gilt. Setzt man dies fort, erhält man insgesamt

$$P(\tau_a < \infty) = P(\tau_1 < \infty)^a \quad \text{für } a > 0.$$

Ganz analog kann man argumentieren für  $a < 0$ , nur dass hier  $-1$  an die Stelle von  $1$  tritt. Fasst man zusammen, kann man schreiben:

$$P(\tau_a < \infty) = P(\tau_{\text{sgn}(a)} < \infty)^{|a|} \quad \text{für } a \in \mathbb{Z} \setminus \{0\}. \quad (3.11)$$

Hier ist  $\operatorname{sgn}(a)$  die *Signum-Funktion* mit Wert  $-1$  für negative und  $+1$  für positive Argumente  $a$ .

Das Problem der Bestimmung von  $P(\tau_a < \infty)$  wurde auf die Bestimmung von  $P(\tau_1 < \infty)$  reduziert. Insbesondere wird  $a$  (bzw.  $-a$ ) nur dann mit Sicherheit eingenommen, wenn 1 mit Sicherheit erreicht wird:

$$\begin{aligned} P(\tau_1 < \infty) = 1 &\Rightarrow P(\tau_a < \infty) = 1, \\ P(\tau_{-1} < \infty) = 1 &\Rightarrow P(\tau_{-a} < \infty) = 1 \\ &\text{für } a \in \mathbb{Z}^+. \end{aligned}$$

Um  $P(\tau_1 < \infty)$  zu berechnen sind jetzt allerdings noch einige teils trickreiche Schritte notwendig. Schon der erste ist ziemlich raffiniert: Wir verwenden die erzeugende Funktion  $g_{\tau_1}$  von  $\tau_1$ . Denn es gilt

$$P(\tau_1 < \infty) = \lim_{r \nearrow 1} g_{\tau_1}(r). \quad (3.12)$$

Um Gleichung (3.12) nachzuvollziehen, sehen wir uns die rechte Seite näher an:

$$g_{\tau_1}(r) = \mathbb{E}(r^{\tau_1}) = \sum_{i=0}^{\infty} r^i P(\tau_1 = i).$$

Nun machen wir den Grenzübergang  $r \nearrow 1$ . Die verbleibende Summe ist dann genau die gesuchte Wahrscheinlichkeit.

$$\lim_{r \nearrow 1} \mathbb{E}(r^{\tau_1}) = \lim_{r \nearrow 1} \sum_{i=0}^{\infty} r^i \cdot P(\tau_1 = i) = \sum_{i=0}^{\infty} P(\tau_1 = i) = P(\tau_1 < \infty).$$

Wir fahren nun bei Gleichung (3.12) fort. Wie weiter oben schon gezeigt, müssen wir nur die ungeraden Zeiten berücksichtigen, da 1 ungerade ist und zu geraden Zeitpunkten daher nicht erreicht werden kann.

$$\lim_{r \rightarrow 1} \mathbb{E}(r^{\tau_1}) = \lim_{r \rightarrow 1} \sum_{n=1}^{\infty} r^{2n-1} \cdot P(\tau_1 = 2n-1) \quad (3.13)$$

Wir benötigen nun  $P(\tau_1 = 2n-1)$ . Dazu verwenden wir Gleichung (3.10) auf Seite 64 für  $a = 1$  und erhalten

$$P(\tau_1 = 2n-1) = \frac{1}{2n-1} \cdot \binom{2n-1}{n} \cdot p^n \cdot q^{n-1}. \quad (3.14)$$

Die Faktoren  $\frac{1}{2n-1} \cdot \binom{2n-1}{n}$  formen wir weiter um (die Schritte werden im Anschluss erklärt):

$$\begin{aligned}
 \frac{1}{2n-1} \cdot \binom{2n-1}{n} &= \frac{1}{2n-1} \cdot \frac{(2n-1)!}{n!(n-1)!} = \frac{(2(n-1))!}{n!(n-1)!} \\
 &= \frac{1}{n!} \cdot \frac{2(n-1) \cdot (2n-3) \cdot 2(n-2) \cdot (2n-5) \cdots 3 \cdot 2(1)}{(n-1) \cdot (n-2) \cdots 1} \\
 &= \frac{1}{n!} \cdot 2^{n-1} \cdot (2n-3) \cdot (2n-5) \cdots 5 \cdot 3 \\
 &= \frac{2^{n-1}}{n!} \cdot \prod_{k=1}^{n-1} (2k-1) \\
 &= \frac{2^{n-1}}{n!} \cdot (-2)^{n-1} \cdot \prod_{k=1}^{n-1} \left(\frac{1}{2} - k\right) \\
 &= (-1)^{n-1} \cdot \frac{4^{n-1}}{n!} \cdot \prod_{k=1}^{n-1} \left(\frac{1}{2} - k\right) \\
 &= (-1)^{n-1} \cdot \frac{4^n}{2 \cdot n!} \cdot \frac{1}{2} \cdot \prod_{k=1}^{n-1} \left(\frac{1}{2} - k\right) \\
 &= (-1)^{n-1} \cdot \frac{4^n}{2 \cdot n!} \cdot \prod_{k=0}^{n-1} \left(\frac{1}{2} - k\right) = (-1)^{n-1} \cdot \frac{4^n}{2} \cdot \binom{\frac{1}{2}}{n}.
 \end{aligned}$$

Im ersten Schritt wurde der Binomialkoeffizient ausgeschrieben und der erste Faktor mit dem Nenner des ersten Bruchs gekürzt. Dann wurden beim verbleibenden Bruch die Fakultäten angeschrieben und, wie oben angedeutet, geeignet gekürzt. Der Faktor 2 bleibt  $n-1$ -mal übrig, außerdem die ungeraden Faktoren. Diese werden im nächsten Schritt als Produkt angeschrieben. Aus jedem Faktor im Produkt wird dann  $(-2)$  herausgehoben. Die Vorfaktoren werden vereinfacht dargestellt. Im nächsten Schritt wird  $4^{n-1}$  zu  $4^n$  und der übrigbleibende Faktor  $\frac{1}{4}$  aufgeteilt. Der Faktor  $\frac{1}{2}$  wurde dann ins Produkt einbezogen, da er der Faktor für  $k=0$  ist. Im letzten Schritt wurde die Definition des *allgemeinen Binomialkoeffizienten* verwendet, die folgendermaßen lautet:

$$\binom{\alpha}{n} = \begin{cases} 1 & \text{falls } n = 0, \\ \frac{1}{n!} \cdot \prod_{k=0}^{n-1} (\alpha - k) & \text{falls } n \in \mathbb{Z}^+ \end{cases} \quad \text{für } \alpha \in \mathbb{R}.$$

Den gesamten Faktor fügen wir nun wieder in Gleichung (3.14) auf Seite 66

ein:

$$P(\tau_1 = 2n - 1) = (-1)^{n-1} \cdot \frac{4^n}{2} \cdot \binom{\frac{1}{2}}{n} \cdot p^n \cdot q^{n-1}$$

Damit können wir nun die Summe in Gleichung (3.13) auf Seite 66 neu darstellen:

$$\sum_{n=1}^{\infty} r^{2n-1} \cdot P(\tau_1 = 2n - 1) = \sum_{n=1}^{\infty} r^{2n-1} \cdot (-1)^{n-1} \cdot \frac{4^n}{2} \cdot \binom{\frac{1}{2}}{n} \cdot p^n \cdot q^{n-1}.$$

Diesen Ausdruck wollen wir nun noch weiter vereinfachen. Dazu formen wir die Summe so um, dass wir eine binomische Reihe ( $\sum_{k=0}^{\infty} \binom{\alpha}{k} x^k = (1+x)^\alpha$  für  $|x| < 1$ ) erhalten.

$$\begin{aligned} \sum_{n=1}^{\infty} r^{2n-1} \cdot (-1)^{n-1} \cdot \frac{4^n}{2} \cdot \binom{\frac{1}{2}}{n} \cdot p^n \cdot q^{n-1} &= -\frac{1}{2} \cdot r^{-1} \cdot q^{-1} \sum_{n=1}^{\infty} \binom{\frac{1}{2}}{n} \cdot (-1)^n \cdot r^{2n} \cdot 4^n \cdot p^n \cdot q^n \\ &= -\frac{1}{2qr} \cdot \left[ \sum_{n=1}^{\infty} \binom{\frac{1}{2}}{n} (-4pqr^2)^n + \binom{\frac{1}{2}}{0} \cdot (-4pqr^2)^0 - 1 \right] \\ &= \frac{1}{2qr} \cdot \left[ 1 - \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} (-4pqr^2)^n \right] \\ &= \frac{1}{2qr} \cdot \left[ 1 - (1 - 4pqr^2)^{\frac{1}{2}} \right] = \mathbb{E}(r^{\tau_1}) \end{aligned} \quad (3.15)$$

für  $|r| < 1$ .

Um die Summanden auf die Form  $(\dots)^n$  zu bringen, wurden die „störenden“ Faktoren herausgehoben und vor die Summe gezogen. Weil für die binomische Reihe die Summe von 0 bis  $\infty$  gehen muss, wurde im zweiten Schritt der entsprechende Summand (der den Wert 1 hat) addiert und wieder abgezogen. Anschließend wurde der Summand  $n = 0$  zur Summe hinzugefügt und dann die binomische Reihe verwendet.

Wie oben schon erwähnt, muss für die binomische Reihe in unserem Fall  $|-4pqr^2| < 1$  gelten. Der Ausdruck  $pq = p(1-p)$  hat sein Maximum an der Stelle  $\frac{1}{2}$  mit  $\frac{1}{4}$ . Multipliziert mit  $-4$  ergibt das ein Extremum von  $-1$ . Somit ist der Ausdruck betragsmäßig kleiner 1 genau dann, wenn  $r^2$  kleiner 1 ist und damit  $|r| < 1$  gilt.

Für den Erwartungswert  $\mathbb{E}(r^{\tau-1})$  muss man in  $\mathbb{E}(r^{\tau_1})$  nur  $p$  und  $q$  vertauschen:

$$\mathbb{E}(r^{\tau-1}) = \frac{1}{2pr} \cdot \left[ 1 - (1 - 4qpr^2)^{\frac{1}{2}} \right] \quad \text{für } |r| < 1. \quad (3.16)$$

Um einen Ausdruck für  $P(\tau_1 < \infty)$  zu bekommen, ist noch, wie in Gleichung (3.12) auf Seite 66, der Grenzübergang  $r \nearrow 1$  notwendig:

$$\lim_{r \nearrow 1} \mathbb{E}(r^{\tau_1}) = \lim_{r \nearrow 1} \frac{1}{2qr} \cdot \left[ 1 - \sqrt{1 - 4pqr^2} \right] = \frac{1}{2q} \cdot \left[ 1 - \sqrt{1 - 4pq} \right].$$

Mit einem weiteren Trick können wir den nochmals vereinfachen. Dazu betrachten wir den Ausdruck unter der Wurzel:

$$1 - 4pq = (p + q)^2 - 4pq = p^2 + 2pq + q^2 - 4pq = (p - q)^2.$$

Damit erhalten wir

$$\lim_{r \nearrow 1} \mathbb{E}(r^{\tau_1}) = \frac{1 - |p - q|}{2q}. \quad (3.17)$$

Diese an sich schon sehr einfache Formel lässt sich noch vereinfachen, wenn man die Fälle  $p \geq q$  und  $p < q$  unterscheidet. Wir können schreiben  $p - q = p - (1 - p) = 2p - 1$ . Ist  $p \geq q$  dann ist  $|p - q| = p - q$ :

$$\frac{1 - |p - q|}{2q} = \frac{1 - (p - q)}{2q} = \frac{1 - (2p - 1)}{2q} = \frac{2 - 2p}{2q} = \frac{2q}{2q} = 1. \quad (3.18)$$

Im Fall  $p < q$  ist  $|p - q| = -(p - q)$  und wir erhalten:

$$\frac{1 - |p - q|}{2q} = \frac{1 + (p - q)}{2q} = \frac{1 + (2p - 1)}{2q} = \frac{2p}{2q} = \frac{p}{q}. \quad (3.19)$$

Insgesamt erhalten wir mit Gleichung (3.18) und Gleichung (3.19)

$$P(\tau_1 < \infty) = \begin{cases} 1 & \text{falls } p \geq q, \\ \frac{p}{q} & \text{falls } p < q. \end{cases} \quad (3.20)$$

Wie oben erhalten wir auch hier durch Vertauschen von  $p$  und  $q$   $P(\tau_{-1} < \infty)$

$$P(\tau_{-1} < \infty) = \begin{cases} 1 & \text{falls } p \leq q, \\ \frac{q}{p} & \text{falls } p > q. \end{cases} \quad (3.21)$$

Zusammen mit Gleichung (3.11) auf Seite 65 können wir somit schreiben:

$$P(\tau_a < \infty) = \begin{cases} 1 & \text{falls } a \in \mathbb{Z}^+ \text{ und } p \geq q \text{ oder } -a \in \mathbb{Z}^+ \text{ und } p \leq q, \\ \left(\frac{p}{q}\right)^a & \text{falls } a \in \mathbb{Z}^+ \text{ und } p < q \text{ oder } -a \in \mathbb{Z}^+ \text{ und } p > q. \end{cases} \quad (3.22)$$

Wir haben eine schnelle und einfache Methode gefunden, um zu berechnen, mit welcher Wahrscheinlichkeit eine eindimensionale einfache Irrfahrt endliche Eintrittswahrscheinlichkeiten hat.

Hat die Irrfahrt absorbierende Ränder, können wir damit auch berechnen, mit welcher Wahrscheinlichkeit die Irrfahrt am Rand absorbiert wird. Denn laut Satz 12 auf Seite 38 ist ja die Absorptionswahrscheinlichkeit gerade  $h_z(x) = P(\tau_{xz} < \infty)$ .

Wir betrachten jetzt noch einmal den Fall  $p > q$ , d. h. die Wahrscheinlichkeit eines Schritts nach vorne ist größer als die eines Schritts nach hinten. Aus Gleichung (3.22) folgt, dass positive Werte  $a$  fast sicher in endlicher Zeit erreicht werden. Die Wahrscheinlichkeit, dass negative Werte  $a$  in endlicher Zeit erreicht werden, ist auf jeden Fall kleiner als 1 und wird vor allem umso kleiner, je kleiner  $a$  wird, d. h. je weiter das Ziel vom Startpunkt entfernt ist. In der Situation des orientierungslosen Betrunkenen bedeutet das Folgendes: Hat er eine Vermutung, in welche Richtung sein Ziel liegt und macht er deshalb bevorzugt Schritte in diese Richtung, wird er mit Wahrscheinlichkeit 1 in endlicher Zeit dort anlangen. Kritisch wird es, wenn er sein Ziel in der falschen Richtung vermutet. Dann ist die Wahrscheinlichkeit kleiner 1, jemals anzukommen.

Das entspricht auch der intuitiven Vorstellung: Wenn die Wahrscheinlichkeit zur Vorwärtsbewegung größer ist, wird die Bewegung insgesamt gesehen vorwärts passieren. Daher werden positive Werte sicher irgendwann erreicht, während negative Werte immer unwahrscheinlicher erreicht werden, je weiter sie in der Gegenrichtung liegen.

### 3.1.4 Rekurrenz oder Transienz

In diesem Abschnitt wollen wir uns nun mit dem bisher ausgeklammerten Fall des Eintritts in den Startpunkt 0 beschäftigen. Wir wollen Kriterien finden, wann eine Irrfahrt rekurrent ist, d. h. mit Sicherheit zum Ausgangspunkt zurückkehrt. Für den Betrunkenen wäre das die Antwort auf die Frage, ob er bei einer zu Hause gestarteten Zechtour mit Wahrscheinlichkeit 1 wieder zurückfindet.

Am einfachsten wäre es, könnten wir die Berechnung von  $P(\tau_0 < \infty)$  auf etwas zurückführen, das uns die Verwendung von Gleichung (3.22) auf Seite 69 ermöglicht. Dazu machen wir folgende Überlegung:

Klarerweise gilt

$$P(\tau_0 < \infty) = P(X_1 = 1 \wedge \tau_0 < \infty) + P(X_1 = -1 \wedge \tau_0 < \infty), \quad (3.23)$$

denn die Irrfahrt hat zwei Möglichkeiten für eine Rückkehr in endlicher Zeit: Entweder sie geht vom Start weg zu 1 und dann in endlicher Zeit zu 0, oder sie geht zu  $-1$  und kehrt zurück.

Diese neuen Wahrscheinlichkeiten lassen sich jetzt aber auch anders ausdrücken:

$$P(X_1 = 1 \wedge \tau_0 < \infty) = p \cdot P(\tau_{-1} < \infty), \quad (3.24)$$

$$P(X_1 = 1 \wedge \tau_0 < \infty) = q \cdot P(\tau_1 < \infty). \quad (3.25)$$

Warum ist das so? Wir begründen den ersten Fall: Mit Wahrscheinlichkeit  $p$  geht die Irrfahrt von 0 zu 1. Dort beginnt die Irrfahrt wieder neu. Die „Vorgeschichte“ beeinflusst ja die Entscheidung nicht, in welche Richtung es weitergeht. Die Wahrscheinlichkeit, von 1 aus irgendwann bei 0 zu landen, ist daher die Wahrscheinlichkeit, beim „Restart“ in 1 in endlicher Zeit zu „-1 aus der Sicht von 1“ zu gelangen. Aus der Sicht von 1 ist -1 aber der Wert um 1 kleiner, also 0. Genauso kann man den zweiten Fall begründen.

Wir setzen nun in Gleichung (3.23) auf Seite 70 ein:

$$P(\tau_0 < \infty) = p \cdot P(\tau_{-1} < \infty) + q \cdot P(\tau_1 < \infty). \quad (3.26)$$

Nun erinnern wir uns an die Gleichungen (3.20) und (3.21) auf Seite 69 und setzen sie ein. Dabei unterscheiden wir drei Fälle:

$$P(\tau_0 < \infty) = \begin{cases} p \cdot \frac{q}{p} + q \cdot 1 = 2q & \text{falls } p > q, \\ p \cdot 1 + q \cdot \frac{p}{q} = 2p & \text{falls } p < q, \\ p \cdot 1 + q \cdot 1 = 1 & \text{falls } p = q. \end{cases} \quad (3.27)$$

Man könnte auch schreiben

$$P(\tau_0 < \infty) = 2 \cdot \min\{p, q\}. \quad (3.28)$$

Man sieht, dass die Irrfahrt  $(X_n)_{n \geq 0}$  genau dann mit Wahrscheinlichkeit 1 zum Startpunkt zurückkehrt, wenn  $p = q$  gilt. Die symmetrische Irrfahrt hat also den Startzustand 0 als rekurrenten Zustand. Ist die Irrfahrt asymmetrisch, ist 0 ein transienter Zustand. Jetzt erinnern wir uns an Satz 14 auf Seite 42. Die Zustände der einfachen Irrfahrt kommunizieren offensichtlich miteinander. Und Satz 14 besagt, dass kommunizierende Zustände alle rekurrent oder alle transient sind. Damit können wir nun folgende Proposition formulieren:

**Proposition 14.** *Nur die symmetrische einfache Irrfahrt auf  $\mathbb{Z}$ ,  $(X_n)_{n \geq 0}$  mit  $p = q$  ist rekurrent.*

Für den betrunkenen Menschen heißt das, er findet nur dann mit Wahrscheinlichkeit 1 zu seinem Startpunkt zurück, wenn er jeden Schritt völlig ahnungslos setzt.

Wir haben jetzt ein Kriterium für die Rekurrenz einer Irrfahrt auf  $\mathbb{Z}$ , können aber keine Aussage darüber machen, wie lange es dauert, bis die Irrfahrt zurückkehrt. Dem umherirrenden Betrunkenen helfen unsere Überlegungen ja nicht wirklich, wenn er nicht weiß, ob er in einem vernünftigen Zeitrahmen zurückkehrt. Daher suchen wir jetzt den Erwartungswert für  $\tau_0$  unter der Bedingung, dass  $\tau_0 < \infty$ . Dazu sind wieder einige Schritte notwendig. Zuerst „zerlegen“ wir den Erwartungswert mittels der Definition der bedingten Wahrscheinlichkeit:

$$\begin{aligned} \mathbb{E}(\tau_0 | \tau_0 < \infty) &= \sum_{k=0}^{\infty} k \cdot P(\tau_0 = k | \tau_0 < \infty) \\ &= \sum_{k=0}^{\infty} k \cdot P(\{\tau_0 = k\} \cap \{\tau_0 < \infty\}) \cdot \frac{1}{P(\tau_0 < \infty)} \\ &= \mathbb{E}(\tau_0 \cdot 1_{\{\tau_0 < \infty\}}) \cdot \frac{1}{P(\tau_0 < \infty)}. \end{aligned} \quad (3.29)$$

Die Wahrscheinlichkeit  $P(\tau_0 < \infty)$  kennen wir von oben. Zu bestimmen ist noch  $\mathbb{E}(\tau_0 \cdot 1_{\{\tau_0 < \infty\}})$ . Dafür machen wir uns zu Nutze, dass

$$\lim_{r \nearrow 1} \mathbb{E}(\tau_0 \cdot r^{\tau_0}) = \mathbb{E}(\tau_0 \cdot 1_{\{\tau_0 < \infty\}}) \quad (3.30)$$

gilt. Für  $\tau_0 < \infty$  geht nämlich  $r^{\tau_0}$  gegen 1. Für  $\tau_0 \rightarrow \infty$  ist aber  $r^{\tau_0} = 0$  wenn wir  $r < 1$  annehmen.

Wir haben jetzt das Ziel vor Augen, müssen den Weg dahin aber noch finden. Dazu beginnen wir mit der erzeugenden Funktion von  $\tau_0$ :

$$g_{\tau_0}(r) = \mathbb{E}(r^{\tau_0}) = \sum_{k=0}^{\infty} r^k \cdot P(\tau_0 = k). \quad (3.31)$$

Den Ausdruck  $P(\tau_0 = k)$  erhalten wir durch ähnliche Überlegungen wie für Gleichung (3.23) auf Seite 70. Wir müssen außerdem nur die geraden Zeitpunkte  $2n$  berücksichtigen.

$$\begin{aligned} P(\tau_0 = 2n) &= P(X_1 = 1 \wedge \tau_0 = 2n) + P(X_1 = -1 \wedge \tau_0 = 2n) \\ &= p \cdot P(\tau_{-1} = 2n - 1) + q \cdot P(\tau_1 = 2n - 1) \end{aligned}$$

Eingesetzt in Gleichung (3.31) ergibt das

$$\begin{aligned}\mathbb{E}(r^{\tau_0}) &= \sum_{n=0}^{\infty} r^{2n} \cdot [p \cdot P(\tau_{-1} = 2n - 1) + q \cdot P(\tau_1 = 2n - 1)] \\ &= p \cdot r \cdot \sum_{n=0}^{\infty} r^{2n-1} \cdot P(\tau_{-1} = 2n - 1) + q \cdot r \cdot \sum_{n=0}^{\infty} r^{2n-1} \cdot P(\tau_1 = 2n - 1) \\ &= p \cdot r \cdot \mathbb{E}(r^{\tau_{-1}}) + q \cdot r \cdot \mathbb{E}(r^{\tau_1}).\end{aligned}$$

Die Erwartungswerte in der letzten Zeile kennen wir aber schon aus (3.16) und (3.15) auf Seite 68 und können daher schreiben

$$\mathbb{E}(r^{\tau_0}) = \sqrt{1 - 4pqr^2} \quad \text{für } |r| < 1. \quad (3.32)$$

Jetzt müssen wir nur noch den Schritt von  $r^{\tau_0}$  zu  $\tau_0 \cdot r^{\tau_0}$  schaffen. Dazu bilden wir die Ableitung von  $r_0^\tau$ :

$$\frac{d}{dr} r^{\tau_0} = \tau_0 \cdot r^{\tau_0-1} \quad \iff \quad r \cdot \frac{d}{dr} r^{\tau_0} = \tau_0 \cdot r^{\tau_0}$$

Für die Erwartungswerte heißt das dann

$$\mathbb{E}(\tau_0 \cdot r^{\tau_0}) = r \cdot \frac{d}{dr} \mathbb{E}(r^{\tau_0}) = \frac{4pqr^2}{\sqrt{1 - 4pqr^2}} \quad \text{für } |r| < 1. \quad (3.33)$$

Jetzt müssen wir nur noch den Grenzübergang  $r \nearrow 1$  durchführen. Außerdem ersetzen wir die Wurzel dann wie oben wieder durch  $|p - q|$ .

$$\mathbb{E}(\tau_0 \cdot 1_{\{\tau_0 < \infty\}}) = \lim_{r \rightarrow 1} \mathbb{E}(\tau_0 \cdot r^{\tau_0}) = \frac{4pq}{|p - q|} \quad (3.34)$$

Jetzt verfügen wir über alle Komponenten, um aus Gleichung (3.29) auf Seite 72  $\mathbb{E}(\tau_0 | \tau_0 < \infty)$  zu berechnen:

$$\mathbb{E}(\tau_0 | \tau_0 < \infty) = \frac{4pq}{|p - q|} \cdot \frac{1}{2 \cdot \min\{p, q\}}. \quad (3.35)$$

Der kleinere Wert von  $p$  und  $q$  im Zähler kürzt sich mit dem Minimum im Nenner. Übrig bleibt also das Maximum im Zähler. Wir erhalten

$$\mathbb{E}(\tau_0 | \tau_0 < \infty) = \frac{2 \cdot \max\{p, q\}}{|p - q|} = \frac{1 + |p - q|}{|p - q|} = 1 + \frac{1}{|p - q|}, \quad (3.36)$$

denn angenommen  $p$  ist das Maximum. Dann ist  $2p = p + q + |p - q| = 1 + |p - q|$ .

In den letzten Schritten wurde verschwiegen, dass der Fall  $p = q$  ausgeschlossen werden muss, da sich sonst im Nenner 0 ergibt.

Fasst man die letzten Ergebnisse zusammen gibt es ein erstaunliches Resultat:

- Ist die Irrfahrt symmetrisch, kehrt sie mit Wahrscheinlichkeit 1 zum Startpunkt 0 zurück. Allerdings ist der Erwartungswert für die Rückkehrzeit unendlich.
- Ist die Irrfahrt nicht symmetrisch, gibt es eine positive Wahrscheinlichkeit, dass sie nicht zurückkehrt. Kehrt sie allerdings zurück, so passiert das relativ schnell.

Der Betrunkene steckt in einem Dilemma: Setzt er seine Schritte gleichverteilt, kommt er zwar sicher zurück, kann aber überhaupt nicht sagen wann. Gibt er seiner Wanderung eine tendenzielle Richtung vor, kommt er möglicherweise nie zurück. Wenn doch, braucht er nur eine relativ geringe Schrittzahl.

Nebenbei haben wir noch gezeigt, dass der Startpunkt der symmetrischen Irrfahrt nullrekurrent ist. Man kann sogar zeigen, dass alle Zustände der symmetrischen Irrfahrt auf  $\mathbb{Z}$  nullrekurrent sind. Dazu verwenden wir Satz 15 auf Seite 45 und zeigen, dass es keine stationäre Verteilung gibt.

**Proposition 15.** Nullrekurrenz der symmetrischen Irrfahrt auf  $\mathbb{Z}$ . (vgl. [5])  
*Jeder Zustand  $x$  der einfachen symmetrischen Irrfahrt auf  $\mathbb{Z}$  ist nullrekurrent.*

*Beweis.* Der Beweis erfolgt indirekt.

Angenommen es gibt einen positiv rekurrenten Zustand  $x$ . Dann gibt es eine stationäre Verteilung  $\alpha$ . Weil  $\sum_{y \in \mathbb{Z}} \alpha(y) = 1$  gilt, existiert dann ein  $z$  mit  $\alpha(z) = m := \max_{y \in \mathbb{Z}} \alpha(y) > 0$ . Wegen der Stationarität von  $\alpha$  gilt

$$m = \alpha(z) = \alpha \Pi(z) = \sum_{y \in \mathbb{Z}} \alpha(y) \cdot \Pi(y, z) = \frac{1}{2} \cdot (\alpha(z-1) + \alpha(z+1)). \quad (3.37)$$

Nur wenn  $\alpha(z-1) = \alpha(z+1) = \alpha(z)$  erfüllt ist, ist  $\alpha(z)$  das Maximum. Setzt man diese Überlegungen fort, dann erkennt man, dass für alle  $y \in \mathbb{Z}$  gilt  $\alpha(y) = m$ . Das ist also ein Widerspruch zu  $\sum_{y \in \mathbb{Z}} \alpha(y) = 1$ . Es gibt also keine stationäre Verteilung und damit auch keinen nullrekurrenten Zustand.  $\square$

### 3.2 einfache Irrfahrten auf $\mathbb{Z}^d$

Die eindimensionale Irrfahrt beschreibt die Situation des umherirrenden Betrunkenen eigentlich nicht so gut, da in den wenigsten Fällen nur die Bewegung in einer Dimension möglich ist. Besser passt das Modell einer zweidimensionalen Irrfahrt, wenn man zum Beispiel annimmt, der Betrunkene sucht in einer Stadt (möglichst mit rasterartigem Grundriss) den Heimweg. Ein Schritt entspricht dann natürlich dem Weg von einer Kreuzung zur nächsten.

Für eine Irrfahrt in drei Dimensionen kann man die Vorstellung eines in alle Richtungen ins Unendliche ausgedehnten Klettergerüsts für Kinder heranziehen. Hier ist es für Aufsichtspersonen durchaus von Interesse, ob Kinder wieder zurückfinden oder im Klettergerüst verloren gehen.

Für die Betrachtung von einfachen Irrfahrten auf  $\mathbb{Z}^d$  ersetzen wir die Zufallsvariablen  $Z_n$ , die bei den eindimensionalen Irrfahrten verwendet wurden, durch  $d$ -komponentige Zufallsvektoren  $\mathbf{Z}_n$  (fett gedruckte Buchstaben symbolisieren Vektoren). Alle Koordinaten bis auf eine zufällig ausgewählte sind 0 und in der ausgewählten Koordinate steht im Fall der einfachen Irrfahrt ein Wert aus  $\{-1, 1\}$ . Wir nennen die Menge dieser Vektoren  $M_d$  ( $M_1$  wäre dann  $\{-1, 1\}$ ). Jeder Zustand in  $\mathbb{Z}^d$  hat genau  $2d$  Nachbarn, die in einem Schritt, also mit einem Zufallsvektor  $\mathbf{Z}$  erreicht werden können. Natürlich gibt es auch  $2d$  verschiedene Möglichkeiten für den Zufallsvektor, denn für jede der  $d$  Koordinaten gibt es zwei Werte. Im Folgenden werden wir die Betrachtungen auf den symmetrischen Fall beschränken, d. h. auf gleichverteilte Zufallsvektoren.

Die einfache Irrfahrt auf  $\mathbb{Z}^d$  definieren wir analog zur eindimensionalen Irrfahrt:

$$(\mathbf{X}_n)_{n \geq 0} \quad \text{mit} \quad \mathbf{X}_0 = \mathbf{0} \quad \text{und} \quad \mathbf{X}_n = \sum_{i=1}^n \mathbf{Z}_i \quad n \geq 1.$$

Mit Fokus auf die Dynamik kann man auch definieren

$$P(\mathbf{X}_0 = \mathbf{0}) = 1 \quad \text{und} \tag{3.38}$$

$$P(\mathbf{X}_n - \mathbf{X}_{n-1} = \epsilon | \mathbf{X}_0, \dots, \mathbf{X}_{n-1}) = p_\epsilon = \frac{1}{2d}. \tag{3.39}$$

für  $n \geq 1$  und  $\epsilon \in M_d$ .

Die Übergangsmatrix sieht folgendermaßen aus:

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

Natürlich ist auch diese Irrfahrt eine Markov-Kette.

Wie in einer Dimension interessieren wir uns auch hier für das Langzeitverhalten, für Rekurrenz und Transienz.

Schreiben wir  $\tau_0$  für die erste und  $\tau_0^{(n)}$  für die  $n$ -te Rückkehr zum Startpunkt, so gilt

$$P(\tau_0^{(n+1)} < \infty) = P(\tau_0^{(n)} < \infty) \cdot P(\tau_0 < \infty) \quad (3.40)$$

und damit

$$P(\tau_0^{(n)} < \infty) = P(\tau_0 < \infty)^n \quad \text{bzw.} \quad (3.41)$$

$$F_n(\mathbf{0}, \mathbf{0}) = F_1(\mathbf{0}, \mathbf{0})^n \quad \forall n \geq 1. \quad (3.42)$$

Das lässt sich leicht begründen, denn die Irrfahrt startet jedesmal wieder neu, die Ereignisse  $\{\tau_0^{(n)} < \infty\}$  und  $\{\tau_0 < \infty\}$  sind daher unabhängig.

In dieser Gleichung steckt nun die Aussage von Satz 13 auf Seite 41. Kommt die Irrfahrt mit Sicherheit wieder zum Startpunkt zurück, ist sie also rekurrent, kommt sie mit Wahrscheinlichkeit 1 unendlich oft zurück. Ist im transienten Fall die Wahrscheinlichkeit der erstmaligen Rückkehr kleiner als 1, ist die Wahrscheinlichkeit, unendlich oft zurückzukehren gleich 0.

Wütende Ehefrauen und besorgte Eltern müssen sich also im Fall von Rekurrenz keine Sorgen machen, denn der Ehemann bzw. das Kind kommen dann unendlich oft zurück.<sup>1</sup> Bei Transienz empfiehlt es sich allerdings, die betroffene Person nicht mehr fortzulassen, da die Wahrscheinlichkeit einer mehrmaligen Rückkehr gegen 0 strebt. Vermutlich wäre die Ehefrau des Betrunkenen eher erfreut über Transienz, Eltern von Kindern im Klettergerüst hoffen wahrscheinlich auf Rekurrenz.

Auch für die in Definition 25 auf Seite 41 eingeführte erwartete Anzahl der Rückkehrzeiten  $G(\mathbf{0}, \mathbf{0})$  ergibt sich ein einfacher Ausdruck:

$$\begin{aligned} G(\mathbf{0}, \mathbf{0}) &= \mathbb{E}(T_0) = \sum_{n \geq 0} P(T_0 > n) \\ &= 1 + \sum_{n \geq 1} P(\tau_0^{(n)} < \infty) = 1 + \sum_{n \geq 1} P(\tau_0 < \infty)^n \\ &= \sum_{n \geq 0} P(\tau_0 < \infty)^n = \frac{1}{1 - P(\tau_0 < \infty)} = \frac{1}{P(\tau_0 = \infty)}. \end{aligned}$$

---

<sup>1</sup>Natürlich gibt es auch Fälle, in denen der wütende Ehemann auf die betrunkene Frau wartet, und neuerdings vermutlich auch wütende „eingetragene PartnerInnen“, die auf ihre betrunkenen gleichgeschlechtlichen „eingetragenen PartnerInnen“ warten. Wir wollen bei diesem (ohnehin nicht ganz ernst zu nehmenden) Vergleich aber bei der klassischen Situation bleiben.

Dabei wurde  $T_{\mathbf{0}} > n \Leftrightarrow \tau_{\mathbf{0}}^{(n)} < \infty$  verwendet. Weil man mit Sicherheit bei  $\mathbf{0}$  beginnt, ist  $P(T_{\mathbf{0}} > 0) = 1$ . Weil  $\tau_{\mathbf{0}}^{(0)}$  nicht definiert ist, muss 1 addiert werden. Mit der Summenformel der geometrischen Reihe erhält man dann den letzten Ausdruck.

Mit Definition 25 haben wir nun

$$\frac{1}{P(\tau_{\mathbf{0}} = \infty)} = \sum_{n=0}^{\infty} P(\mathbf{X}_n = \mathbf{0}). \quad (3.43)$$

Damit haben wir ein (in Satz 13 auf Seite 41 schon allgemein gezeigtes) Kriterium für Rekurrenz:

**Proposition 16.** *Eine einfache Irrfahrt auf  $\mathbb{Z}^d$  ist rekurrent genau dann, wenn  $\sum_{n=0}^{\infty} P(\mathbf{X}_n = \mathbf{0}) = \infty$ .*

Denn für Rekurrenz muss ja  $P(\tau_{\mathbf{0}} = \infty) = 0$  sein.

Da wir uns auf symmetrische Irrfahrten beschränkt haben, können wir zeigen, dass zu geraden Zeitpunkten der Aufenthalt im Startpunkt  $\mathbf{0}$  am wahrscheinlichsten ist. Dazu betrachten wir einen beliebigen Zustand  $\mathbf{x} \in \mathbb{Z}^d$  zum Zeitpunkt  $2n$ :

$$\begin{aligned} P(\mathbf{X}_{2n} = \mathbf{x}) &= \sum_{\mathbf{y} \in \mathbb{Z}^d} P(\mathbf{X}_n = \mathbf{y} \wedge \mathbf{X}_{2n} - \mathbf{X}_n = \mathbf{x} - \mathbf{y}) \\ &= \sum_{\mathbf{y} \in \mathbb{Z}^d} P(\mathbf{X}_n = \mathbf{y}) \cdot P(\mathbf{X}_{2n} - \mathbf{X}_n = \mathbf{x} - \mathbf{y}) \\ &= \sum_{\mathbf{y} \in \mathbb{Z}^d} P(\mathbf{X}_n = \mathbf{y}) \cdot P(\mathbf{X}_n = \mathbf{x} - \mathbf{y}) \\ &\leq \sqrt{\sum_{\mathbf{y} \in \mathbb{Z}^d} P(\mathbf{X}_n = \mathbf{y})^2} \cdot \sqrt{\sum_{\mathbf{y} \in \mathbb{Z}^d} P(\mathbf{X}_n = \mathbf{x} - \mathbf{y})^2} \\ &= \sum_{\mathbf{y} \in \mathbb{Z}^d} P(\mathbf{X}_n = \mathbf{y})^2. \end{aligned}$$

Im vierten Schritt wurde die *Cauchy-Schwarzsche Ungleichung*  $\sum_n |a_n \cdot b_n| \leq \sqrt{\sum_n a_n^2} \cdot \sqrt{\sum_n b_n^2}$  verwendet. Der letzte Schritt folgt daraus, dass die beiden Summen unter den Wurzeln gleich sind.

Jetzt kommt die geforderte Symmetrie ins Spiel. Denn für die symmetri-

sche Irrfahrt gilt  $P(\mathbf{X}_n = \mathbf{y}) = P(\mathbf{X}_n = -\mathbf{y})$ . Damit erhalten wir

$$\begin{aligned} \sum_{\mathbf{y} \in \mathbb{Z}^d} P(\mathbf{X}_n = \mathbf{y})^2 &= \sum_{\mathbf{y} \in \mathbb{Z}^d} P(\mathbf{X}_n = \mathbf{y}) \cdot P(\mathbf{X}_n = -\mathbf{y}) \\ &= \sum_{\mathbf{y} \in \mathbb{Z}^d} P(\mathbf{X}_n = \mathbf{y}) \cdot P(\mathbf{X}_{2n} - \mathbf{X}_n = -\mathbf{y}) = P(\mathbf{X}_{2n} = \mathbf{0}). \end{aligned}$$

Somit erhalten wir als Resultat:

**Proposition 17.** *Ist eine einfache Irrfahrt  $(\mathbf{X}_n)_{n \geq 0}$  auf  $\mathbb{Z}^d$  symmetrisch, so ist zu geraden Zeiten der Aufenthalt im Startpunkt  $\mathbf{0}$  am wahrscheinlichsten:  $P(\mathbf{X}_{2n} = \mathbf{0}) = \max_{\mathbf{x} \in \mathbb{Z}^d} P(\mathbf{X}_{2n} = \mathbf{x})$ .*

Aufgrund der Symmetrie ist dies nicht wirklich überraschend. Man könnte jetzt vermuten, dass die Wahrscheinlichkeit, irgendwann beim Startpunkt zu landen, 1 sein müsste, da der Aufenthalt dort am wahrscheinlichsten ist. Wie wir sehen werden, ist dies aber nicht zwangsläufig der Fall.

Betrachten wir zuerst den Fall  $\mathbb{Z}^2$ .

### Einfache Irrfahrt auf $\mathbb{Z}^2$

Dieser Abschnitt verwendet [5].

Um wieder zum Startpunkt zurückzukehren, muss der Betrunkene eine gerade Anzahl von Schritten machen. Außerdem muss er in jeder Koordinate die gleiche Anzahl an Schritten, die er in eine Richtung geht, in die Gegenrichtung machen. Bei einer Schrittzahl von  $2n$  torkelt er also  $k$  Schritte in der ersten Koordinate in eine Richtung und  $k$  Schritte zurück. Für die zweite Koordinate bleiben dann noch  $2n - 2k$  Schritte übrig, von denen jeweils  $n - k$  in eine Richtung gesetzt werden.

Damit ist die Wahrscheinlichkeit für  $k$  Schritte in zwei Richtungen einer Koordinate und  $n - k$  in der anderen

$$\left(\frac{1}{4}\right)^{2n} \cdot \frac{(2n)!}{(k!)^2 \cdot ((n-k)!)^2},$$

weil die Reihenfolge der Schritte egal ist. Da für eine Rückkehr auch egal ist, wie groß  $k$  ist, müssen wir noch über  $k$  summieren und erhalten

$$P(\mathbf{X}_{2n} = \mathbf{0}) = 4^{-2n} \cdot \sum_{k=0}^n \frac{(2n)!}{(k!)^2 \cdot ((n-k)!)^2}.$$

Wird der Bruch um  $(n!)^2$  erweitert, lässt sich der Ausdruck mit Binomialkoeffizienten anschreiben.

$$\begin{aligned} 4^{-2n} \cdot \sum_{k=0}^n \frac{(2n)!}{(k!)^2 \cdot ((n-k)!)^2} \\ = 4^{-2n} \cdot \binom{2n}{n} \cdot \sum_{k=0}^n \binom{n}{k} \cdot \binom{n}{n-k} = \left( 2^{-2n} \cdot \binom{2n}{n} \right)^2 \end{aligned}$$

Im letzten Schritt wurde die Identität  $\sum_{k=0}^n \binom{n}{k} \cdot \binom{n}{n-k} = \binom{2n}{n}$  verwendet. Diese kann so begründet werden: Der Wert  $\binom{2n}{n}$  gibt die Anzahl der Möglichkeiten, aus  $(2n)$  Elementen  $n$  auszuwählen. Dieselbe Auswahl kann man auch treffen, indem man  $k$  Elemente aus einer Hälfte von  $n$  (entspricht  $\binom{n}{k}$ ) auswählt und die verbleibenden  $n-k$  Elemente aus der anderen Hälfte (entspricht  $\binom{n}{n-k}$ ). Da  $k$  nicht festgelegt ist, muss noch über  $k$  summiert werden, um alle Möglichkeiten zu erfassen.

Mit der Stirling-Formel  $n! \sim \sqrt{2\pi n} \cdot n^n \cdot e^{-n}$  können wir den Ausdruck weiter vereinfachen:

$$\begin{aligned} 2^{-2n} \cdot \binom{2n}{n} &= 2^{-2n} \cdot \frac{(2n)!}{n!^2} \\ &\sim 2^{-2n} \cdot \frac{\sqrt{2\pi 2n} \cdot (2n)^{2n} \cdot e^{-2n}}{(\sqrt{2\pi n} \cdot n^n \cdot e^{-n})^2} \\ &= \frac{1}{\sqrt{\pi n}} \end{aligned}$$

und daher ist

$$P(\mathbf{X}_{2n} = \mathbf{0}) \sim \frac{1}{n\pi}.$$

Damit ist aber die Greenfunktion  $G(\mathbf{0}, \mathbf{0}) = \infty$ , weil die Summe  $\sum_{n \geq 1} \frac{1}{n}$  divergiert. Der Startpunkt ist also rekurrent. Weil alle Zustände miteinander kommunizieren, erhalten wir wegen Satz 14 auf Seite 42 daher

**Satz 19.** *Die einfache symmetrische Irrfahrt auf  $\mathbb{Z}^2$  ist rekurrent.*

Der Betrunkene findet in einer Stadt mit Wahrscheinlichkeit 1 nach Hause zurück, sofern er „gleichverteilte Richtungsentscheidungen“ trifft.

Die wartende Ehefrau weiß also, dass ihr Mann wiederkommt.

Doch wie sieht sie Situation am Kinderspielplatz mit dem unendlich großen Klettergerüst aus? Weil sich schon die Hoffnung der Ehefrau nicht

erfüllte, dass der Betrunkene nicht zurückkommt, stellen wir auch hier die Vermutung an, dass die Eltern enttäuscht werden und das Kind nicht zurückfindet. Um zu zeigen, dass die Irrfahrt auf  $\mathbb{Z}^d$  für  $d \geq 3$  transient ist, suchen wir eine obere Schranke für die Summanden der Greenfunktion, sodass die Summe dieser Schranken endlich ist.

### Einfache Irrfahrt auf $\mathbb{Z}^d$

Wir vermuten, dass das Verhalten für alle Irrfahrten mit  $d \geq 3$  gleich ist und zeigen daher gleich den allgemeinen Fall.

Ähnlich wie in zwei Dimensionen können wir durch einige Überlegungen zur Wahrscheinlichkeit  $P(\mathbf{X}_{2n} = \mathbf{0})$  kommen. Die Wahrscheinlichkeit für einen bestimmten Weg, der in  $2n$  Schritten zum Start zurückführt, ist  $\left(\frac{1}{2d}\right)^{2n}$ . Wiederum muss in jeder Koordinate in beide Richtungen die gleiche Schrittzahl gesetzt werden, um zurückzukommen. Unter der Annahme, dass diese Schrittzahlen fixiert werden, müssen immer noch die vielen Möglichkeiten berücksichtigt werden, die sich durch die freie Abfolge der Schritte ergeben. Da  $2n$  Schritte gemacht werden und davon  $k_i$  in der  $i$ -ten Koordinate in eine Richtung und wieder zurück, ergibt sich als Wahrscheinlichkeit bei festgelegter Schrittzahl  $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}_+^d$  mit  $\sum k_i = n$ :

$$\left(\frac{1}{2d}\right)^{2n} \cdot \frac{(2n)!}{k_1!^2 \cdots k_d!^2}.$$

Die bis jetzt als fixiert angenommene Schrittzahl in den Koordinaten kann für die Rückkehr zum Start frei variiert werden, solange die Gesamtsumme gleich der Gesamtschrittzahl  $2n$  ist. Also muss noch über alle  $\mathbf{k}$  summiert werden, die  $\sum k_i = n$  erfüllen. Wir erhalten daher

$$P(\mathbf{X}_{2n} = \mathbf{0}) = (2d)^{-2n} \sum_{\mathbf{k}} \frac{(2n)!}{k_1!^2 \cdots k_d!^2}.$$

Wieder erweitern wir den Bruch mit  $(n!)^2$  und ziehen den Binomialkoeffizienten vor die Summe:

$$= (2d)^{-2n} \cdot \binom{2n}{n} \cdot \sum_{\mathbf{k}} \frac{(n!)^2}{k_1!^2 \cdots k_d!^2}.$$

Um hier weiter zu kommen, benötigen wir den Multinomialkoeffizienten aus Definition 11 auf Seite 14. Außerdem ziehen wir  $d^{-2n}$  in die Summe und erhalten dann

$$= 2^{-2n} \cdot \binom{2n}{n} \cdot \sum_{\mathbf{k}} d^{-2n} \binom{n}{\mathbf{k}}^2 = 2^{-2n} \cdot \binom{2n}{n} \cdot \sum_{\mathbf{k}} \left[ d^{-n} \binom{n}{\mathbf{k}} \right]^2.$$

Der Ausdruck in der Summe ist genau die Zähl-dichte der Multinomialverteilung  $M_{n, \frac{1}{d}}$ . Daher liefert Summation über  $\mathbf{k}$ :  $\sum_{\mathbf{k}} d^{-n} \binom{n}{\mathbf{k}} = 1$ . Mit diesem Wissen können wir die Summe weiter vereinfachen:

$$P(\mathbf{X}_{2n} = \mathbf{0}) \leq 2^{-2n} \cdot \binom{2n}{n} \cdot \max_{\mathbf{k}} \left[ d^{-n} \binom{n}{\mathbf{k}} \right].$$

Um zum letzten Zwischenergebnis zu kommen benötigen wir folgende Zwischenrechnung:

$$\begin{aligned} \sum_{\mathbf{k}} \left[ d^{-n} \binom{n}{\mathbf{k}} \right]^2 &= \sum_{\mathbf{k}} \left[ d^{-n} \binom{n}{\mathbf{k}} \right] \cdot \underbrace{\left[ d^{-n} \binom{n}{\mathbf{k}} \right]}_{\leq \max_{\mathbf{k}} \left[ d^{-n} \binom{n}{\mathbf{k}} \right]} \\ &\leq \underbrace{\sum_{\mathbf{k}} \left[ d^{-n} \binom{n}{\mathbf{k}} \right]}_{=1} \cdot \max_{\mathbf{k}} \left[ d^{-n} \binom{n}{\mathbf{k}} \right] = \max_{\mathbf{k}} \left[ d^{-n} \binom{n}{\mathbf{k}} \right]. \end{aligned}$$

Nun bleibt das Maximum zu bestimmen. Der Multinomialkoeffizient ist aber gerade dann maximal, wenn gilt  $\lfloor \frac{n}{d} \rfloor \leq k_i \leq \lceil \frac{n}{d} \rceil$ , oder anders ausgedrückt wenn  $|k_i - \frac{n}{d}| \leq 1$  für alle  $k_i$ .

Das lässt sich folgendermaßen begründen:

Angenommen es gibt ein  $k_j$  mit  $|k_j - \frac{n}{d}| > 1$ . Dann ist  $k_j \geq k_i + 2$  für ein die Bedingung erfüllendes  $k_i$ . Ersetzt man jetzt  $k_j$  durch  $k_j - 1$  und  $k_i$  durch  $k_i + 1$ , bleibt zwar die Summe  $\sum k_i = n$ , aber der Multinomialkoeffizient wird größer. Der größere Faktor  $k_j$  im Nenner wird nämlich durch den kleineren Faktor  $k_i + 1$  ersetzt und der Nenner wird damit kleiner. Also gilt für das Maximum tatsächlich  $|k_i - \frac{n}{d}| \leq 1$  für alle  $k_i$ .

Nun verwenden wir wieder die Stirling-Formel  $n! = \sqrt{2\pi n} \cdot n^n \cdot e^{-n}$ :

$$\begin{aligned} d^{-n} \cdot \binom{n}{\mathbf{k}} &= d^{-n} \cdot \frac{n!}{k_1! \cdots k_d!} \\ &\sim d^{-n} \cdot \frac{\sqrt{2\pi n} \cdot n^n \cdot e^{-n}}{\sqrt{2\pi k_1} \cdot k_1^{k_1} \cdot e^{-k_1} \cdots \sqrt{2\pi k_d} \cdot k_d^{k_d} \cdot e^{-k_d}} \\ &= d^{-n} \cdot \frac{(2\pi)^{\frac{1}{2}} \cdot n^{(n+\frac{1}{2})} \cdot e^{-n}}{(2\pi)^{\frac{d}{2}} \cdot \prod_{i=1}^d \left( k_i^{(k_i+\frac{1}{2})} \right) \cdot e^{(k_1+\cdots+k_d)}} \\ &= d^{-n} \cdot (2\pi)^{\left(-\frac{d-1}{2}\right)} \cdot n^{(n+\frac{1}{2})} \cdot \prod_{i=1}^d \left( k_i^{-(k_i+\frac{1}{2})} \right). \end{aligned}$$

Weil sich für den maximalen Binomialkoeffizienten die  $k_i$  um weniger als 1 von  $\frac{n}{d}$  unterscheiden, ersetzen wir für das Maximum:

$$\begin{aligned} \max_{\mathbf{k}} \left[ d^{-n} \binom{n}{\mathbf{k}} \right] &\sim d^{-n} \cdot (2\pi)^{\left(-\frac{d-1}{2}\right)} \cdot n^{\left(n+\frac{1}{2}\right)} \cdot \prod_{i=1}^d \left( \left( \frac{n}{d} \right)^{-\left(\frac{n}{d}+\frac{1}{2}\right)} \right) \\ &= d^{-n} \cdot (2\pi)^{\left(-\frac{d-1}{2}\right)} \cdot n^{\left(n+\frac{1}{2}\right)} \cdot \left( \left( \frac{n}{d} \right)^{-\left(\frac{n}{d}+\frac{1}{2}\right)} \right)^d \\ &= d^{-n} \cdot (2\pi)^{\left(-\frac{d-1}{2}\right)} \cdot n^{\left(n+\frac{1}{2}\right)} \cdot n^{\left(-n-\frac{d}{2}\right)} \cdot d^{\left(n+\frac{d}{2}\right)} \\ &= d^{\frac{d}{2}} \cdot (2\pi n)^{\left(-\frac{d-1}{2}\right)}. \end{aligned}$$

Wie schon bei der zweidimensionalen Irrfahrt verwenden wir die Folgerung aus der Stirling-Formel

$$2^{-2n} \cdot \binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}}$$

und erhalten dann insgesamt

$$\begin{aligned} P(\mathbf{X}_{2\mathbf{n}} = \mathbf{0}) &\leq (\pi n)^{-\frac{1}{2}} \cdot d^{\frac{d}{2}} \cdot (2\pi n)^{\left(-\frac{d-1}{2}\right)} \\ &= d^{\frac{d}{2}} \cdot \pi^{\frac{d}{2}} \cdot 2^{\left(-\frac{d-1}{2}\right)} \cdot n^{-\frac{d}{2}}. \end{aligned}$$

Die nur von  $d$  abhängigen, aber von  $n$  unabhängigen Faktoren fassen wir zu einer Konstante  $C(d) < \infty$  zusammen. Als Endresultat erhalten wir somit

$$P(\mathbf{X}_{2\mathbf{n}} = \mathbf{0}) \leq C(d) \cdot n^{-\frac{d}{2}}$$

und für die Greenfunktion

$$G(\mathbf{0}, \mathbf{0}) \leq 1 + C(d) \cdot \sum_{n \geq 1} n^{-\frac{d}{2}}$$

Für  $d = 1$  und  $d = 2$  divergiert die Summe. Daher ist Rekurrenz möglich und tritt, wie wir schon gezeigt haben, auch tatsächlich auf.

Für  $d \geq 3$  allerdings konvergiert  $\sum_{n \geq 1} n^{-\frac{d}{2}}$ . Damit haben wir eine endliche obere Schranke für die Greenfunktion gefunden und folgedessen muss auch die Greenfunktion endlich sein. In drei oder mehr Dimensionen ist der Startpunkt der einfachen symmetrischen Irrfahrt auf  $\mathbb{Z}^d$  also transient!

Wir haben gezeigt, dass die Greenfunktion für die Rückkehr zum Startpunkt  $\mathbf{0}$  endlich ist, weil wir nur daran interessiert waren. Allerdings wurde in

der Argumentation nirgends verwendet, dass wir nur den Startpunkt betrachten. Man könnte jeden Zustand  $\mathbf{x}$  einsetzen und käme zum selben Ergebnis. Also ist nicht nur der Startpunkt sondern die gesamte einfache symmetrische Irrfahrt auf  $\mathbb{Z}^d$  für  $d \geq 3$  transient. Das folgt auch aus Satz 14 auf Seite 42.

**Satz 20.** *Die einfache symmetrische Irrfahrt auf  $\mathbb{Z}^d$  ist für  $d \geq 3$  transient.*

Für Eltern von Kindern, die gerne in Klettergerüsten herumturnen, heißt das: Kinder in (unendlich großen) Klettergerüsten niemals unbeaufsichtigt lassen! Es gibt eine positive Wahrscheinlichkeit, dass sie nicht mehr zum Startpunkt zurückkehren.

Damit ist die Frage beantwortet, warum Betrunkene zurückfinden, Kinder im Klettergerüst hingegen nicht.



# Kapitel 4

## Simulation von Irrfahrten auf $\mathbb{Z}^d$

Das abschließende Kapitel gibt eine Möglichkeit wieder, wie man einfache Irrfahrten auf  $\mathbb{Z}^d$  mit dem CAS *Mathematica* (verwendete Version: 7.0.0) simulieren kann. Dazu werden rekursive Funktionen und die Befehle `If` und `RandomInteger` bzw. `RandomReal` verwendet.

Solche Simulationen sind im Prinzip mit jedem Programm realisierbar, das über (Pseudo-)Zufallszahlen und Bedingungsbeefehle verfügt.

Ist im Folgenden nur von Irrfahrten die Rede, so sind damit immer einfache Irrfahrten gemeint.

### 4.1 Irrfahrten in einer Dimension

#### 4.1.1 symmetrische Irrfahrt

Für die symmetrische Irrfahrt in einer Dimension lautet die Eingabe in *Mathematica*:

```
Clear[X];
X[0] = 0;
X[n_Integer] := X[n]
  = X[n-1]+If[RandomInteger[]==0, -1, 1];
Xn = Table[{k,X[k]}, {k,0,500}];
```

Der Befehl `Clear[X]` sorgt dafür, dass bei neuerlicher Auswertung die Berechnungen erneut durchgeführt werden.

Die zweite Zeile legt den Startpunkt fest. Zeile drei definiert die Irrfahrt rekursiv als „Vorgängerwert plus zufällig ( $\pm 1$ )“. Die zufällige Addition bzw.

Subtraktion von 1 wird erreicht durch die Befehle `If[condition, t, f]` und `RandomInteger[]`.

Dabei erzeugt `RandomInteger[{min, max}]` gleichverteilte ganze (Pseudo-) Zufallszahlen aus dem angegebenen Intervall. Wird das Intervall nicht festgelegt, wird  $[0, 1]$  verwendet. Der Befehl `If[]` überprüft eine Bedingung (*condition*) und verwendet bei Erfüllung den Wert  $t$ , bei Nichterfüllung den Wert  $f$ . Hier wurde überprüft, ob die Zufallszahl gleich 0 ist. Wenn ja wurde 1 abgezogen, ansonsten addiert.

Die letzte Zeile erzeugt eine Liste, die die Wertpaare  $\{k, X_k\}$  für  $0 \leq k \leq 500$  enthält. Geplottet wird `Xn` mit `ListPlot[Xn...]`. Dieser Befehl stellt Wertpaare als Punkte im Koordinatensystem dar.

Abbildung 4.1 auf Seite 87 zeigt zwei verschiedene Plots, die mit demselben Befehl erzeugt wurden.

### 4.1.2 asymmetrische Irrfahrt

Die Eingabe sieht jetzt etwas anders aus, da die Richtung nicht mehr gleichverteilt ist:

```
p = 0.6; q = 1 - p;
Clear[X];
X[0] = 0;
X[n_Integer] := X[n]
    = X[n-1]+If[RandomReal[] < p, 1, +1];
Xn = Table[{k, X[k]}, {k, 0, 500};
```

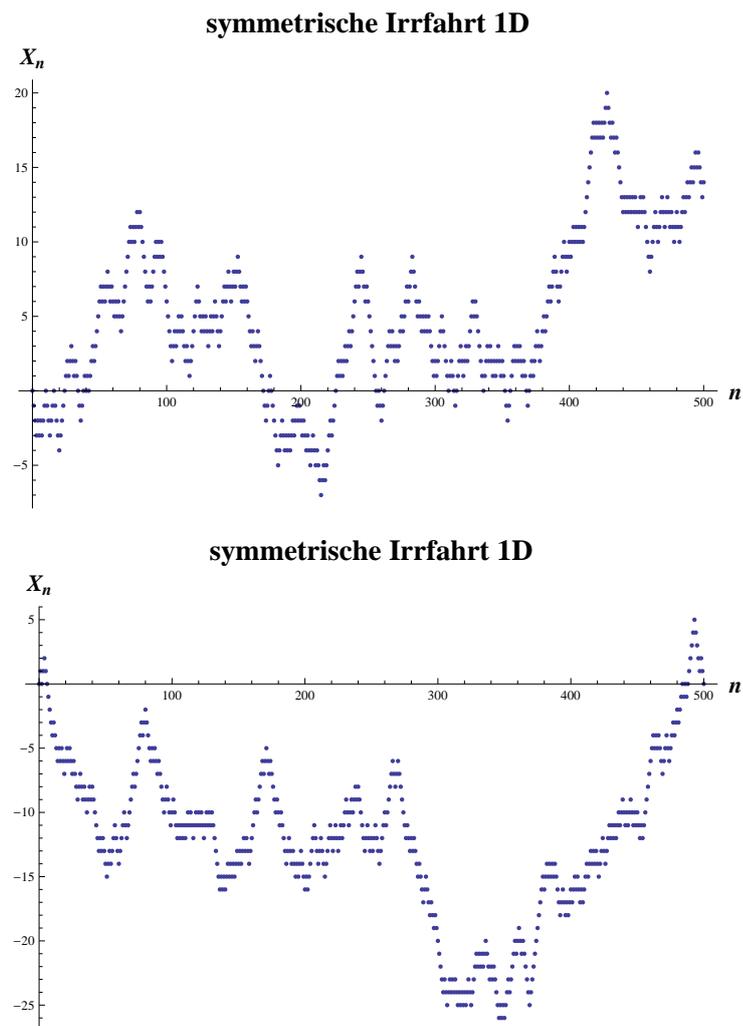
In der ersten Zeile wurde die Wahrscheinlichkeit  $p$  eines Schritts nach vor (+1) festgelegt. Der Unterschied zur symmetrischen Irrfahrt steckt in der Bedingung des `If`-Befehls: `RandomReal[]` erzeugt jetzt gleichverteilte *reelle* (Pseudo-) Zufallszahlen aus dem Intervall  $[0, 1]$ . Die Voraussetzung, dass 1 mit Wahrscheinlichkeit  $p$  eingenommen wird ist gleichbedeutend damit, 1 zu erhalten unter der Bedingung, dass eine Zufallszahl kleiner  $p$  erzeugt wird.

Die Abbildung 4.2 auf Seite 88 zeigt zwei verschiedene Plots einer asymmetrischen Irrfahrt mit  $p = 0,6$  und  $p = 0,55$ .

## 4.2 Irrfahrten in zwei Dimensionen

### 4.2.1 symmetrische Irrfahrt

Das Prinzip der Eingabe in *Mathematica* ändert sich bei der Einbeziehung einer zweiten Dimension nicht. Statt des Anfangswerts  $X_0 = 0$  wird jetzt ein

Abbildung 4.1: Graphen einer symmetrischen Irrfahrt auf  $\mathbb{Z}$

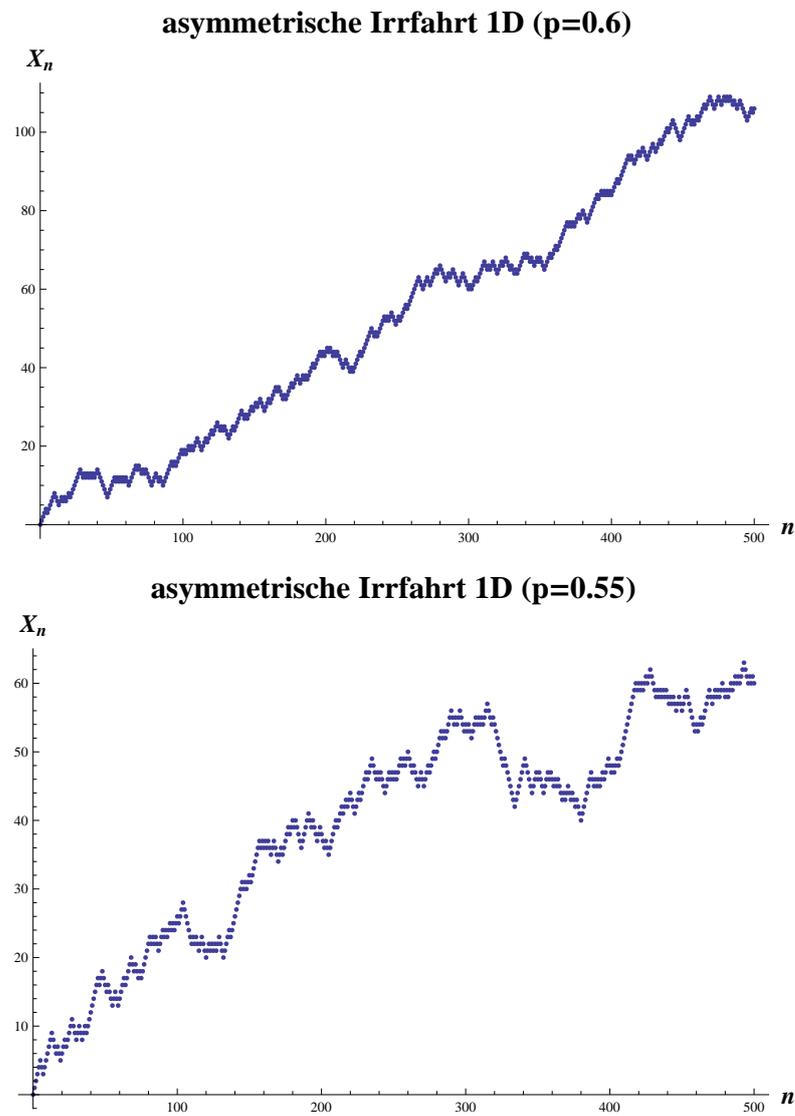


Abbildung 4.2: Graphen zweier asymmetrischer Irrfahrten mit  $p = 0.6$  bzw.  $p = 0.55$

Startpunkt  $(0, 0)$  vorgegeben und werden zufällige, ganzzahlige Vektoren der Länge 1 addiert. Nur die Ermittlung der Schrittrichtung ist komplizierter.

```
Clear[X];
X[0] = {0, 0};
X[n_Integer] := X[n]
  +If[= X[n-1]+RandomInteger[]==0,
      If[RandomInteger[]==0,{-1, 0},{1, 0}],
      If[RandomInteger[]==1,{0, 1},{0, -1}]];
Xn = Table[X[k], k, 0, 500];
```

Da die Änderung der Irrfahrt in beiden Koordinaten gleich wahrscheinlich ist, wird zuerst (erster If-Befehl) per Zufall bestimmt, welche Koordinate sich ändert (Zufallszahl 0 heißt Änderung in  $x$ -Richtung). Nach Festlegung der Koordinate muss analog wie im eindimensionalen Fall die Richtung ermittelt werden. Die letzte Zeile erzeugt eine Liste mit den Koordinaten als Einträge. Mittels `ListPlot[Xn...]` werden die Koordinaten graphisch dargestellt.

Eine zweite, einfachere Möglichkeit, die zweidimensionale symmetrische Irrfahrt zu simulieren, bieten komplexe Zahlen in der Polarform.

```
Clear[X];
X[0] = 0;
X[n_Integer] := X[n]
  = X[n-1]+e^(i RandomInteger[{0,3}]  $\pi/2$ );
Xn = Table[{Re[X[k]], Im[X[k]]}, {k, 0, 500}];
```

In jedem Schritt wird eine komplexe Zahl mit Betrag 1 und Winkel  $k \cdot \frac{\pi}{2}$  addiert, wobei  $k$  zufällig aus  $\{0, 1, 2, 3\}$  ermittelt wird. Für die Liste der Wertpaare werden dann einfach Real- und Imaginärteil als  $x$ - und  $y$ -Koordinate ausgegeben.

Abbildung 4.3 auf Seite 90 zeigt zwei Plots symmetrischer Irrfahrten in zwei Dimensionen.

Die zweidimensionale Irrfahrt kann, so wie auf  $\mathbb{Z}$ , gegen die Zeit aufgetragen werden. Am einfachsten funktiniert dies, wenn man den Einträgen der Tabelle `Xn` als erste Koordinate die Zeit zuordnet:

```
Xn = Table[{k, Re[X[k]], Im[X[k]]}, {k, 0, 500};
```

Dargestellt werden die Werte jetzt mit `ListPointPlot3d[]`, einem Befehl, welcher die Einträge einer Liste der Form  $\{\{x_1, y_1, z_y\}, \{x_2, y_2, z_2\}, \dots\}$  als Punkte in einem dreidimensionalen Koordinatensystem darstellt. So eine Darstellung zeigt Abbildung 4.4 auf Seite 91.

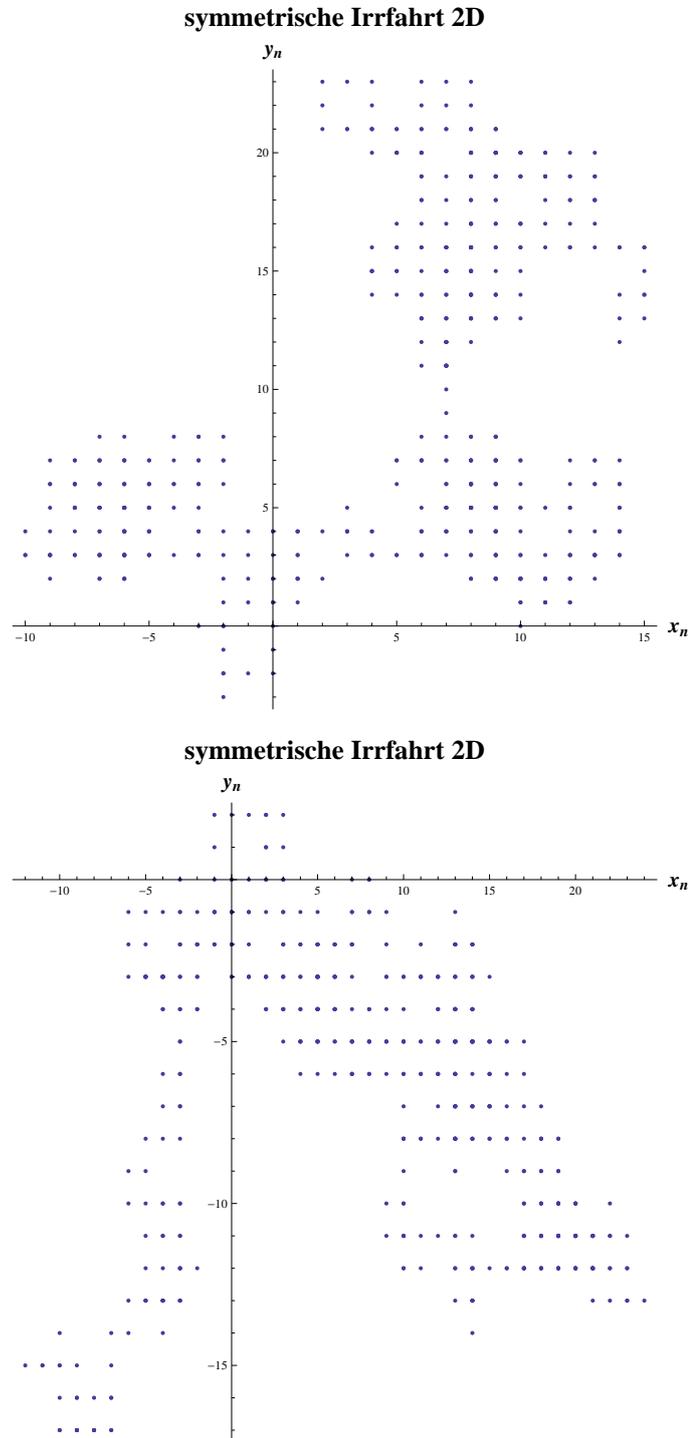
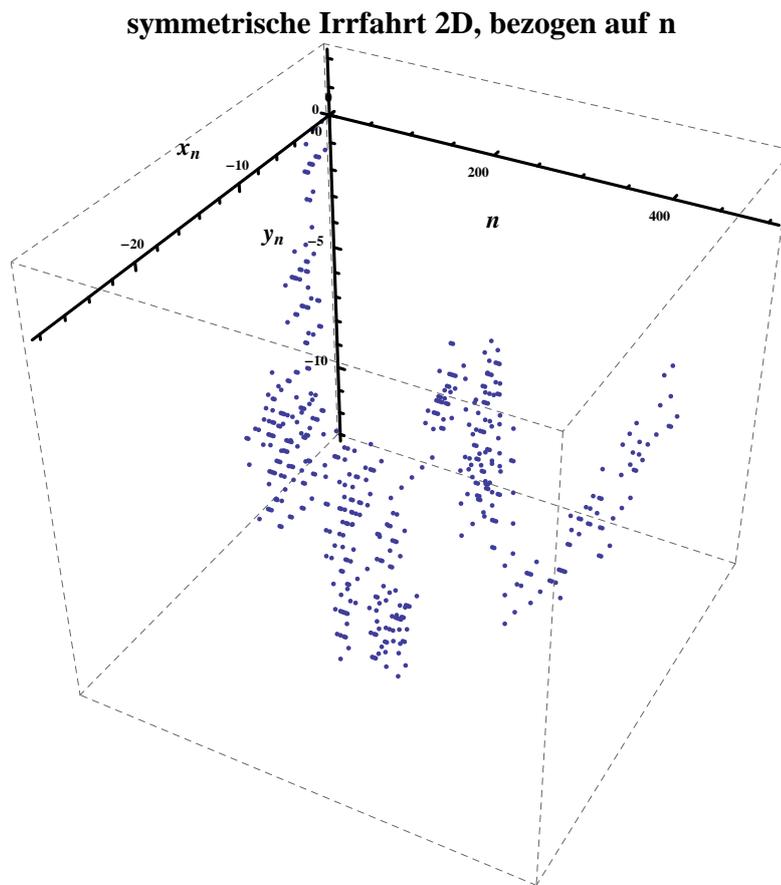


Abbildung 4.3: Graphen einer zweidimensionalen symmetrischen Irrfahrt

Abbildung 4.4: Graph einer symmetrischen 2D-Irrfahrt, bezogen auf  $n$

### 4.2.2 asymmetrische Irrfahrt

Wieder sind die Schrittrichtungen mit unterschiedlichen Wahrscheinlichkeiten behaftet. Allerdings wird jetzt auch den Koordinaten eine unterschiedliche Wahrscheinlichkeit zugeordnet:

```

px=0.4; py=0.4; pXy=0.7;
Clear[X];
X[0] = {0,0};
X[n_Integer] := X[n]
  = X[n-1]+If[RandomReal[]<pXy,
    If[RandomReal[]<px,{1,0},{-1,0}],
    If[RandomReal[]<py,{0,1},{0,-1}]];
Xn = Table[X[k],{k,0,500}];

```

Die Parameter  $px$ ,  $py$ ,  $pXy$  geben die Wahrscheinlichkeiten an, mit der ein Vorwärtsschritt in der ersten und in der zweiten Koordinate gemacht wird sowie die Wahrscheinlichkeit, dass die Bewegung in  $x$ -Richtung erfolgt. Die zufälligen Auswahlen von Koordinaten und Schrittrichtungen in den Koordinaten werden wie bei der symmetrischen zweidimensionalen Irrfahrt getroffen mit dem Unterschied, dass jetzt wieder `RandomReal[]` verwendet wird.

Abbildung 4.5 auf Seite 93 zeigt einen Graphen einer asymmetrischen Irrfahrt auf  $\mathbb{Z}^2$ .

### 4.2.3 „Irrfahrt“ mit Betrag 1

Obwohl die folgende Simulation keine Irrfahrt auf  $\mathbb{Z}^2$  mehr darstellt, wird sie aufgrund ihrer relativ einfachen Realisierung hinzugenommen. Voraussetzung ist jetzt nur noch, dass die Schrittgröße 1 beträgt. Die Richtung ist nicht mehr eingeschränkt. Die Simulation erfolgt wie schon bei der symmetrischen zweidimensionalen Irrfahrt mittels komplexer Zahlen. Jetzt wird der Winkel als gleichverteilte Zufallszahl aus  $[0, 2\pi]$  gewählt. Die Eingabe lautet folgendermaßen:

```

Clear[X];
X[0] = 0;
X[n_Integer] := X[n]
  = X[n-1]+e^(i RandomReal[{0,2π}]);
Xn = Table[{Re[X[k]], Im[X[k]]}, {k, 0, 500}];

```

Abbildung 4.6 auf Seite 94 zeigt so eine Simulation. vergleicht man mit Abbildung 4.4 auf Seite 91, erkennt man sofort den Unterschied: jetzt sind nicht mehr die Gitterpunkte in  $\mathbb{Z}^2$  besetzt sondern Punkte in  $\mathbb{R}^2$ .

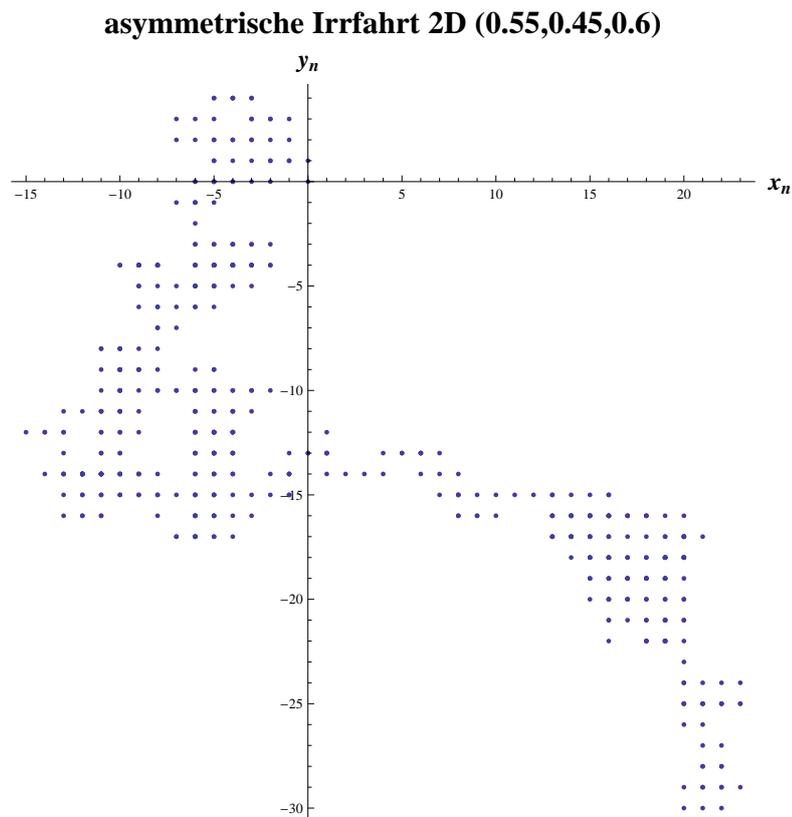


Abbildung 4.5: Graph einer asymmetrischen zweidimensionalen Irrfahrt mit  $p_x = 0.55$ ,  $p_y = 0.45$ ,  $p_{Xy} = 0.6$

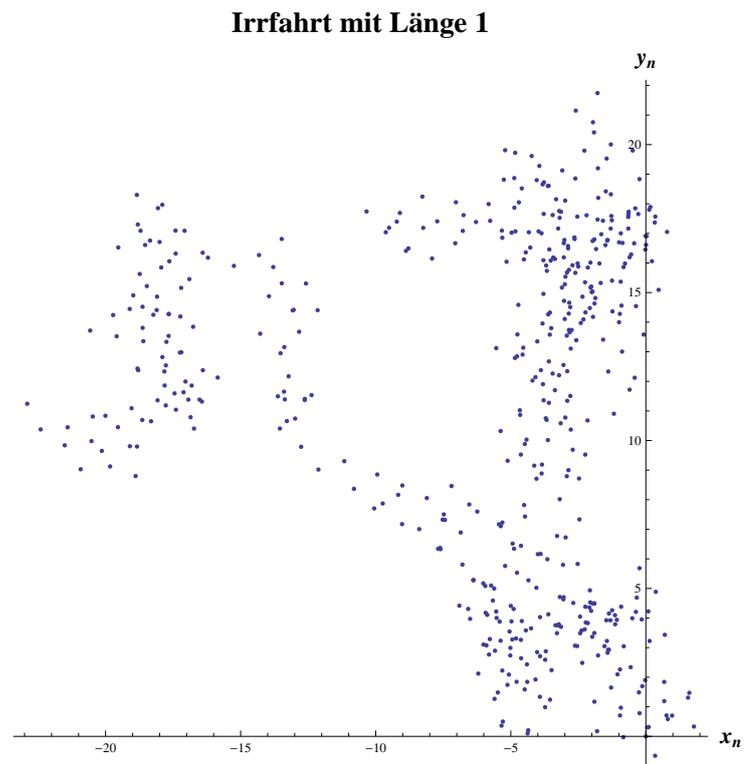


Abbildung 4.6: Graph einer Simulation mit Schrittlänge 1 und zufälliger Richtung

## 4.3 Irrfahrten in drei Dimensionen

Auch in drei Dimensionen kann eine Irrfahrt mit *Mathematica* simuliert werden. Das Prinzip bleibt gleich, die If-Befehle werden noch um eine Stufe verschachtelter.

### 4.3.1 symmetrische Irrfahrt

Die Eingabe in *Mathematica* sieht folgendermaßen aus:

```
Clear[X];
X[0] = 0, 0, 0;
X[n_Integer] := X[n]
  = X[n-1]+If[RandomInteger[{1,3}]==1,
    If[RandomInteger[]==0,{-1,0,0},{1,0,0}],
    If[RandomInteger[{2,3}]==2,
      If[RandomInteger[]==0,{0,-1,0},{0,1,0}],
      If[RandomInteger[]==0,{0,0,-1},{0,0,1}]]];
Xn = Table[X[k],{k, 0, 500}];
```

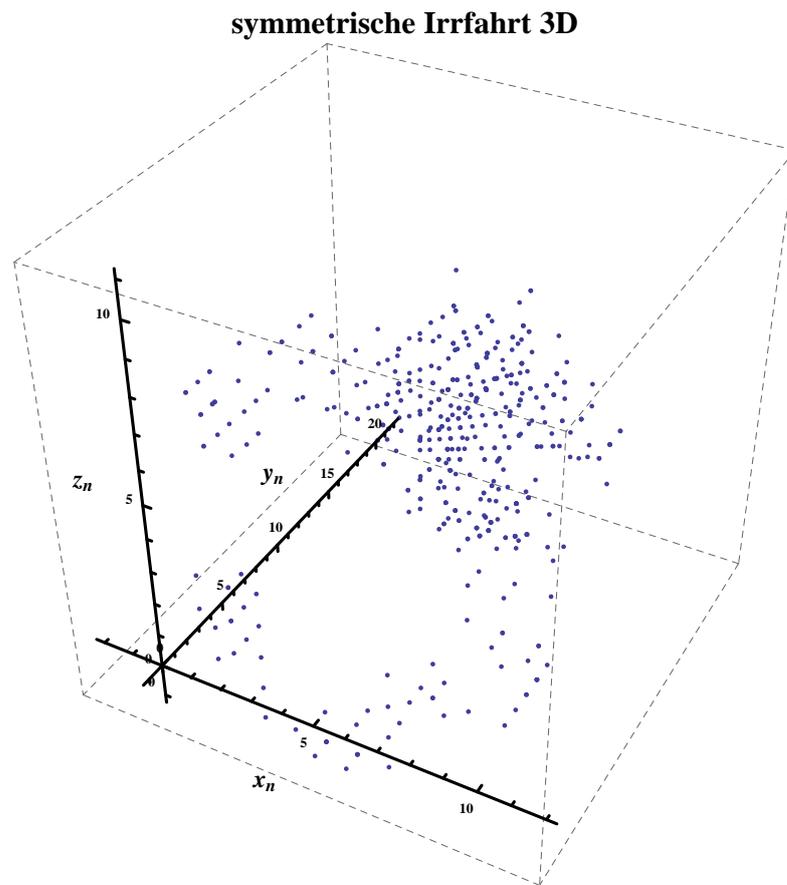
Da der Schritt mit Wahrscheinlichkeit  $\frac{1}{3}$  erfolgt, ist die erste Bedingung für eine Änderung in  $x$ -Richtung, dass aus der Menge  $\{1, 2, 3\}$  zufällig 1 ausgewählt wird. Bei erfüllter erster Bedingung wird wie im zweidimensionalen Fall die Richtung entschieden und der entsprechende Vektor  $(1, 0, 0)$  bzw.  $(-1, 0, 0)$  addiert. Bei nicht erfüllter erster Bedingung wird in den verbleibenden Koordinaten eine zweidimensionale Irrfahrt simuliert: mit Hilfe einer Zufallszahl aus  $\{2, 3\}$  wird eine Koordinate (2 für  $y$ , 3 für  $z$ ) ausgewählt und der entsprechende (zufällige) Vektor addiert. Wieder erzeugt die letzte Zeile eine Liste der Koordinaten, hier bis  $k = 500$ , die dann mit `ListPointPlot3D[Xn...]` geplottet werden.

Abbildung 4.7 auf Seite 96 zeigt einen entsprechenden Graphen.

### 4.3.2 asymmetrische Irrfahrt

Auch auf  $\mathbb{Z}^3$  lassen sich asymmetrische Irrfahrten simulieren. Allerdings müssen jetzt neben den Richtungswahrscheinlichkeiten in den drei Koordinaten noch die Wahrscheinlichkeiten für die Änderung in den einzelnen Koordinaten angegeben werden.

```
px=0.4; py=0.4; pz=0.6; pXyz=0.5; pxYz=0.2; pxyZ=1-pXyz-pxYz;
  pYz=pxYz/(pxYz+pxyZ);
Clear[X];
```

Abbildung 4.7: Graph einer symmetrischen Irrfahrt auf  $\mathbb{Z}^3$

```

X[0] = {0, 0, 0};
X[n_Integer] := X[n]
  = X[n-1]+If[RandomReal[]<pXyz,
    If[RandomReal[]<px, {1,0,0}, {-1,0,0}],
    If[RandomReal[]<pYz,
      If[RandomReal[]<py, {0,1,0}, {0,-1,0}],
      If[RandomReal[]<pz, {0,0,1}, {0,0,-1}]]];
Xn = Table[X[k], {k, 0, 500}];

```

Vom Prinzip funktioniert die Simulation wie der symmetrische Fall. Die Zufallsauswahlen erfolgen wie bei der asymmetrischen zweidimensionalen Irrfahrt. Einzig die Auswahl aus den verbleibenden Koordinaten, sofern die erste nicht gewählt wurde, bedarf einer kurzen Erklärung.

Wir brauchen die bedingte Wahrscheinlichkeit, dass die  $y$ -Koordinate ausgewählt wird unter der Bedingung, dass die  $y$ - oder die  $z$ -Koordinate gewählt werden muss (die  $x$ -Koordinate also nicht gewählt wurde), also  $pYz = P(y|y \cup z)$ . Im Folgenden bezeichnet „ $y$ “ das Ereignis „ $y$ -Koordinate ausgewählt“. Wegen Definition 13 auf Seite 15 und weil  $y$  und  $z$  disjunkt sind gilt:

$$pYz = P(y|y \cup z) = \frac{P(y \cap (y \cup z))}{P(y \cup z)} = \frac{P(y)}{P(y) + P(z)} = \frac{pxYz}{pxYz + pxyZ}$$

So wurde  $pYz$  in der zweiten Eingabezeile definiert. Eine Irrfahrt auf  $\mathbb{Z}^3$  zeigt Abbildung 4.8 auf Seite 98.

## 4.4 Irrfahrten in mehr als drei Dimensionen

Irrfahrten auf  $\mathbb{Z}^d$  für  $d > 3$  lassen sich in ähnlicher Weise mit *Mathematica* simulieren. Die Vorgehensweise ist immer dieselbe. Da solche Situationen jedoch weder vorstellbar sind noch die Möglichkeit besteht, sie zu plotten, wurde darauf verzichtet.

## 4.5 Irrfahrten mit absorbierendem Rand

Auch Irrfahrten mit absorbierendem Rand lassen sich mit *Mathematica* simulieren. Dazu verwendet man zwei zusätzliche If-Befehle, die bewirken, dass die Irrfahrt vom Rand nicht mehr wegkommt, weil die Schrittlänge 0 ist. Da sich das Prinzip in höheren Dimensionen nicht ändert, werden wir jetzt nur eindimensionale Irrfahrten betrachten.

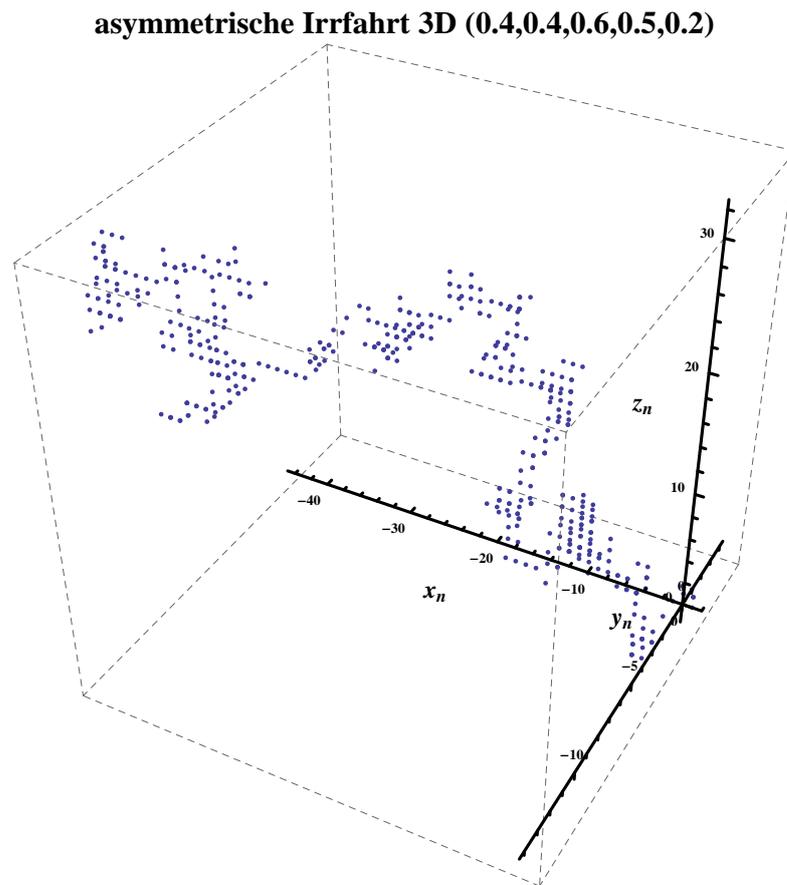


Abbildung 4.8: Graph einer 3D-Irrfahrt mit Parametern  $p_x = 0.4$ ,  $p_y = 0.4$ ,  $p_z = 0.6$ ,  $p_{Xyz} = 0.2$ ,  $p_{xYz} = 0.2$

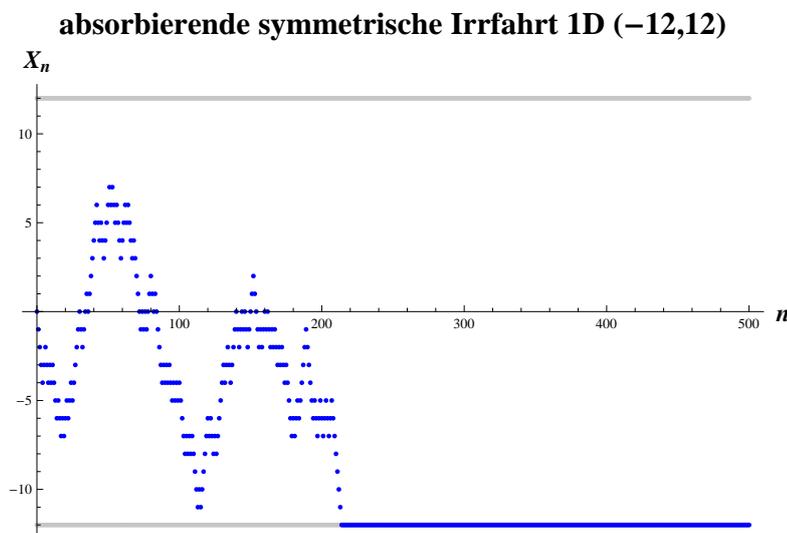


Abbildung 4.9: Irrfahrt mit absorbierendem Rand (-12,12)

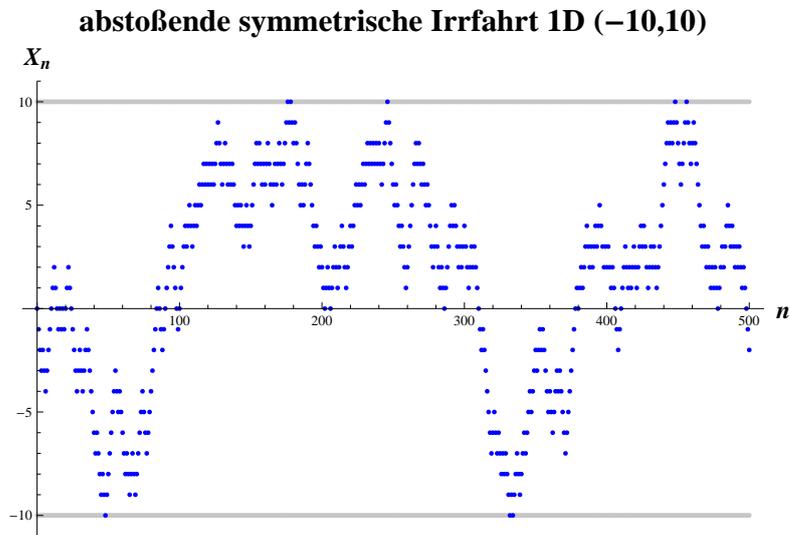
### Irrfahrt auf $\mathbb{Z}$ mit absorbierendem Rand

Die Eingabe in Mathematica sieht folgendermaßen aus:

```
Clear[X];
X[0] = 0; L = -12; R = 12;
X[n_Integer] := X[n]
  = X[n - 1] +
    If[X[n - 1] == L, 0,
      If[X[n - 1] == R, 0,
        If[RandomInteger[] == 0, -1, 1]]];
Xn = Table[{k, X[k]}, {k, 0, 500}]
```

In der zweiten Zeile werden Startpunkt, linker und rechter Rand festgelegt. Der erste If-Befehl bewirkt, dass im Fall  $X_{n-1} = L$  (die Irrfahrt ist am linken Rand) kein Schritt mehr gemacht wird. Der zweite If-Befehl bewirkt dasselbe für den rechten Rand. Sofern sich die Irrfahrt nicht am Rand befindet, kommt der dritte If-Befehl zum Tragen und die Irrfahrt entspricht der einfachen symmetrischen Irrfahrt auf  $\mathbb{Z}$  von oben.

Abbildung 4.9 auf Seite 99 zeigt so eine absorbierende Irrfahrt.

Abbildung 4.10: Irrfahrt mit absorbierendem Rand  $(-10,10)$ 

## 4.6 Irrfahrten mit abstoßendem Rand

Auch Irrfahrten mit abstoßendem Rand kann man relativ einfach simulieren. Die Eingabe sieht ähnlich aus wie bei der absorbierenden Irrfahrt, nur dass jetzt am Rand sicher ein Schritt weg vom Rand gemacht wird. Auch hier beschränken wir uns auf eine Dimension.

### Irrfahrt auf $\mathbb{Z}$ mit abstoßendem Rand

Die Eingabe sieht folgendermaßen aus:

```
Clear[X];
X[0] = 0; L = -10; R = 10;
X[n_Integer] := X[n]
  = X[n - 1] +
    If[X[n - 1] == L, +1,
      If[X[n - 1] == R, -1,
        If[RandomInteger[] == 0, -1, 1]]];
Xn = Table[{k, X[k]}, {k, 0, 500}]
```

Dabei sorgt der erste If-Befehl dafür, dass an der linken Grenze mit Sicherheit ein Schritt nach rechts bzw. vorne gemacht wird, der zweite verursacht einen Rückschritt von der rechten Grenze. Ist die Irrfahrt zwischen den Rändern, verhält sie sich wie die normale eindimensionale Irrfahrt, was durch den dritten If-Befehl realisiert wird.

# Literaturverzeichnis

- [1] M. N. Barber, B. W. Ninham, *Random and Restricted Walks*, Gordon and Breach, New York 1970.
- [2] H. Bauer, *Wahrscheinlichkeitstheorie* (4. Auflage), Walter de Gruyter, Berlin, 1990.
- [3] G. Bolch, S. Greiner, H. de Meer, K. Trivedi, *Queueing Networks and Markov Chains. Modeling and Performance Evaluation with Computer Science Applications*, John Wiley & Sons, New York, 1998.
- [4] K. L. Chung, J. B. Walsh, *Markov Processes, Brownian Motion, and Time Symmetry* (Second Edition), Springer, New York, 2005.
- [5] H. Georgii, *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik* (3. Auflage), Walter de Gruyter, Berlin, 2007.
- [6] C. Hesse, *Angewandte Wahrscheinlichkeitstheorie. Eine fundierte Einführung mit über 500 realitätsnahen Beispielen und Aufgaben*, Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden, 2003.
- [7] A. Irle, *Wahrscheinlichkeitstheorie und Statistik. Grundlagen - Resultate - Anwendungen*, B. G. Teubner, Stuttgart/Leipzig/Wiesbaden, 2001.
- [8] A. Klenke, *Wahrscheinlichkeitstheorie* (2. Auflage), Springer, Heidelberg, 2008.
- [9] G. F. Lawler, *Intersections of Random Walks*, Birkhäuser, Boston, 1991.
- [10] D. W. Stroock, *An Introduction to Markov Processes*, Springer-Verlag, Berlin Heidelberg, 2005.
- [11] [www.mathematik.hu-berlin.de/~riedle/winter06/stoch1.pdf](http://www.mathematik.hu-berlin.de/~riedle/winter06/stoch1.pdf)  
M. Riedle, *Script zu Wahrscheinlichkeitstheorie I*, Universität Mannheim, 2006. (abgerufen am 17.03.10 um 19:22)

- [12] [www.mathematik.hu-berlin.de/~riedle/winter06/stoch2.pdf](http://www.mathematik.hu-berlin.de/~riedle/winter06/stoch2.pdf)  
M. Riedle, *Script zu Wahrscheinlichkeitstheorie I*, Universität Mannheim, 2006. (abgerufen am 17.03.10 um 19:20)

# Anhang

## Zusammenfassung

Warum Betrunkene zurückfinden, Kinder im Klettergerüst hingegen nicht – Der Titel bezieht sich auf (nicht ganz ernst zu nehmende) praktische Anwendungen des eigentlichen Inhalts dieser Arbeit: Irrfahrten und Markov-Ketten. Da für die Behandlung eines wahrscheinlichkeitstheoretischen Themas ein Basiswissen zur Stochastik nötig ist, liefert das erste Kapitel diese Grundlagen und spezielle Inhalte, die für die weiteren Kapitel notwendig sind bzw. dort den logischen Aufbau unterbrechen würden. Außerdem dient das erste Kapitel dazu, mit den hier verwendeten Symbolen und Begriffen vertraut zu werden.

Das zweite Kapitel beschäftigt sich dann mit Markov-Ketten. Man könnte Markov-Ketten beschreiben als stochastische Prozesse mit extremem Kurzzeitgedächtnis, deren Verhalten im nächsten Schritt nur vom gegenwärtigen Zustand und nicht von der Vergangenheit abhängt. Betrachtet wird vor allem das Langzeitverhalten von Markov-Ketten: Die Wahrscheinlichkeit einer Absorption im Fall existierender absorbierender Zustände, die Wahrscheinlichkeit eines Eintritts in einen Zustand in endlicher Zeit, die Rückkehr zum Startpunkt in endlicher Zeit und stationäre Verteilungen. Den Abschluss bildet ein Satz über die Konvergenz gegen die stationäre Verteilung.

Den einfachen Irrfahrten auf  $\mathbb{Z}^d$  widmet sich das dritte Kapitel. Es handelt sich dabei um Markov-Ketten, die zu jedem Zeitpunkt die Möglichkeiten eines Schritts mit konstanter Länge in eine Richtung haben und dabei unabhängig davon entscheiden, woher sie gekommen sind. Für den nächsten Schritt ist nur der momentane Zustand ausschlaggebend. Nacheinander werden Irrfahrten in einer, zwei und in drei oder mehr Dimensionen behandelt. Insbesondere wird dem Langzeitverhalten Beachtung geschenkt. Man interessiert sich dafür, ob der Betrunkene (eine praktische Situation der ein- und zweidimensionalen Irrfahrt) zurückfindet und ob man Kinder im unendlichen Klettergerüst alleine lassen darf. Das Resultat steckt schon im Titel: Der Betrunkene kehrt mit Wahrscheinlichkeit 1 zurück, das Kind nicht unbedingt.

Im letzten Kapitel wird dann eine Möglichkeit beschrieben, die Irrfahrten des dritten Kapitels zu simulieren. Dazu wird das CAS *Mathematica* verwendet. Neben den Eingabebefehlen und ihren Erklärungen finden sich Plots verschiedener Irrfahrten.

## Abstract

Why drunks find back but children in a jungle gym don't – The title refers to real life-applications (not to be taken too seriously) of the thesis' content: Random walks and Markov chains.

In order to treat probabilistic problems some stochastic knowledge is necessary. Chapter one contains this basics and, additionally, special issues, which are necessary for following chapters and would probably interrupt the progress later on. Furthermore chapter one explains symbols, terms and definitions.

Chapter two deals with Markov chains. Markov chains can be described as stochastic processes, whose short time memory doesn't work very well. They only remember their present state, when they decide what to do in the next step. The past has no relevance. Special focus is laid on the long-term behavior of Markov chains: The probability of absorption in case of an existing absorbing element, the probability of finite first time passages, finite times of first return and stationary distributions. A proposition about the convergence to stationary distributions concludes the chapter.

Chapter three is dedicated to random walks, particularly to nearest neighbor random walks on  $\mathbb{Z}^d$ . Those are special Markov chains which always make just one step of fixed length. The decision where to go doesn't depend on the past. Only the present state causes the next state. First we examine the one-dimensional case, then two dimensions and finally random walks in dimensions higher than two. Again focus is laid on the long-term behavior. We want to find out whether drunks find back or not (a one- or two-dimensional random walk) and if it's wise not to supervise kids in a jungle gym. The result: Drunks find back, kids not necessarily.

The fourth chapter describes, how to simulate those random walks by using the CAS *Mathematica*. It contains the input commands used to get the output, their explanations and plots of random walks.

## Lebenslauf

### Persönliche Daten

Name: Dietmar Azesberger  
Adresse: 4122 Arnreit 46  
Geburtsdatum: 9. November 1986  
Staatsbürgerschaft: Österreich  
Eltern: Rupert und Veronika Azesberger  
Geschwister: David und Judith

### Ausbildung

1993-1997 Volksschule Arnreit  
1997-2005 BRG Rohrbach in OÖ  
2005 Matura am BRG Rohrbach in OÖ  
2005-2006 Präsenzdienst in Hörsching  
2006-2010 Studium an der Universität Wien  
UF Mathematik und UF Physik

### Sonstiges

Ehrenamtliche  
Tätigkeiten: aktives Mitglied der Freiwilligen Feuerwehr, des  
Musikvereins und der Landjugend Arnreit;  
Mitglied im Kirchenchor und Organist  
der Pfarre Arnreit