

# STOCHASTISCHE MODELLE

Michael Scheutzow

Vorlesungsskript

Technische Universität Berlin

Wintersemester 2019/20

VORLÄUFIGE VERSION  
FEBRUAR 2020



# Inhaltsverzeichnis

<b>1 Grundlagen</b>	<b>1</b>
1.1 Einleitung . . . . .	1
1.2 Warteschlangenmodelle . . . . .	3
1.3 Stochastische Prozesse und ihre Verteilung . . . . .	4
1.4 Endlich-dimensionale Verteilungen . . . . .	6
1.5 Simulation von Zufallsvariablen . . . . .	7
<b>2 Markovketten mit diskreter Zeit</b>	<b>11</b>
2.1 Einleitung . . . . .	11
2.2 Simulation einer Markovkette und Beispiele . . . . .	12
2.3 Definition und einfache Eigenschaften . . . . .	17
2.4 Rekurrenz und Transienz von Markovketten . . . . .	21
2.5 Klassifikation der Zustände . . . . .	25
2.6 Grenzwertsätze und invariante Verteilungen . . . . .	29
2.7 Beispiele . . . . .	37
2.8 Kartenmischen . . . . .	46
2.9 Mischzeiten . . . . .	49
2.10 Reversible Markovketten . . . . .	51
2.11 Perfekte Simulation: der Propp–Wilson Algorithmus . . . . .	55
<b>3 Markovketten mit stetiger Zeit</b>	<b>59</b>
3.1 Einleitung und Beispiele . . . . .	59
3.2 Definitionen und erste Ergebnisse . . . . .	62
3.3 Vorwärts- und Rückwärtsgleichungen . . . . .	70
3.4 Langzeitverhalten und invariante Maße . . . . .	76
3.5 Beispiele: Warteschlangen . . . . .	81
3.5.1 Warteschlangen . . . . .	81
3.5.2 Warteschlangennetze . . . . .	82
<b>4 Gaußsche Prozesse und Zeitreihenanalyse</b>	<b>87</b>
4.1 Einleitung . . . . .	87
4.2 Vorhersage stochastischer Prozesse . . . . .	90

4.3	Rekursive lineare Vorhersage . . . . .	93
4.4	Filterprobleme, der Kalmanfilter . . . . .	97
4.5	Schwach stationäre Prozesse . . . . .	100
4.6	Schätzung von Erwartungswert und Kovarianzfunktion . . . . .	101
4.7	Der nichtparametrische Ansatz; Spektraldichtenschätzung . . . . .	104
4.8	Datentransformation, Trendbereinigung, Saisonbereinigung . . . . .	108
4.9	Zusammenfassung . . . . .	110
	<b>Literatur</b>	<b>113</b>

# Kapitel 1

## Grundlagen

### 1.1 Einleitung

In diesem Kapitel wollen wir den Begriff des “Stochastischen Prozesses” anschaulich und formal einführen. Den Abschluss bildet ein Abschnitt über die Simulation von Zufallsvariablen mit vorgegebener Verteilung.

Ein *stochastischer Prozess* ist ein mathematisches Modell für einen realen Vorgang, der zufällig ist und von einem Parameter (meist der Zeit) abhängt. Beispiele für zufällige reale Vorgänge, auf die die Theorie der stochastischen Prozesse mit Erfolg angewendet wird, sind:

- Warteschlangen (auch Netze von Warteschlangen),
- Lagerhaltung,
- Ausbreitung von Epidemien,
- Ausbreitung von Genen,
- Populationsentwicklung,
- Wasserstand in einem Staudamm,
- Aktienkurse,
- Kapital eines Versicherungsunternehmens,
- Belastung eines Bauteils eines Fahrzeugs während der Fahrt,
- Temperaturverteilung auf der Erdoberfläche.

Alle diese Beispiele sind zeitabhängig und in der Regel nicht mit Sicherheit vorhersagbar, weswegen sich eine stochastische Modellierung anbietet. Zu stochastischen Prozessen werden diese Beispiele erst dann, wenn man festlegt, mit welchen Wahrscheinlichkeiten die einzelnen Realisierungen auftreten. Man kann einen stochastischen Prozess auffassen als eine Zufallsvariable mit Werten nicht in  $\mathbb{R}$ , sondern in einem geeigneten Raum von (Zeit-)funktionen. Ein stochastischer Prozess ist also eine zufällige Funktion (mit festem Definitions- und Wertebereich).

Zu einem stochastischen Prozess gehört immer ein *Zustandsraum*  $E$  und eine *Parametermenge*  $I$ . Der Zustandsraum ist der Wertebereich (oder eine Obermenge davon), die Parametermenge der Definitionsbereich der zufälligen Funktionen. Zustandsraum  $E$  und Parametermenge  $I$  können im Prinzip beliebige nichtleere Mengen sein; für  $E$  werden wir aber gleich eine zusätzliche Struktur fordern. Wichtig sind aber folgende Spezialfälle.

$$\begin{array}{l}
 E \\
 I
 \end{array}
 \left\{ \begin{array}{l}
 \text{endlich,} \\
 \text{abzählbar unendlich,} \\
 \mathbb{R}, \\
 \mathbb{R}^n, \\
 \text{Funktionenraum,} \\
 \text{diskret (endlich, } \mathbb{N}, \mathbb{N}_0, \mathbb{Z}), \\
 \text{kontinuierlich (} \mathbb{R}, \mathbb{R}^+, \mathbb{R}^n, \text{Kugeloberfläche, } \dots \text{).}
 \end{array} \right.$$

$I$  ist in den meisten Anwendungsfällen eine Menge von Zeitpunkten. Es gibt aber auch interessante Beispiele, für die  $I$  nicht als “Zeit”, sondern als “Ort” zu interpretieren ist. Interessiert man sich etwa für die Temperaturverteilung auf der gesamten Erdoberfläche zu einem *festen* Zeitpunkt, dann ist der Definitionsbereich  $I$  (z. B.) die Kugeloberfläche.

Bei der Beschreibung realer Phänomene ist die Trennung von  $I$  und  $E$  nicht immer eindeutig. Betrachtet man zum Beispiel die Temperaturverteilung auf der Erdoberfläche als Funktion der Zeit, so hat man unter anderen die folgenden Möglichkeiten:

- $I = \mathbb{R} = \text{“Zeit”}$  und  $E = \{f: \text{Kugeloberfläche} \rightarrow \mathbb{R}^+\}$ ,
- $I = \mathbb{R} \times \text{Kugeloberfläche} = \text{“Zeit} \times \text{Ort”}$  und  $E = \mathbb{R}^+ = \text{“Temperatur”}$ .

Der Übergang von einem realen Prozess zu dem mathematischen Modell eines stochastischen Prozesses beinhaltet in der Regel gewisse Hypothesen oder Annahmen (auch Vereinfachungen) über den realen Prozess sowie gegebenenfalls statistische Verfahren, z. B. zum Schätzen von Parametern, die bei der Modellierung zunächst nicht festgelegt wurden. Statistische Verfahren im Zusammenhang mit stochastischen Prozessen werden oft als *Zeitreihenanalyse* bezeichnet. Wir werden uns damit erst im letzten Kapitel beschäftigen.

Gelegentlich kann oder muss man auf statistische Verfahren verzichten, etwa dann, wenn das Modell so plausibel ist, dass es keiner statistischen Absicherung bedarf (z. B., dass bei einem fair aussehenden Würfel alle Seiten gleich wahrscheinlich und aufeinander folgende Würfe unabhängig sind) oder dann, wenn die Modellierung als stochastischer Prozess vorgenommen wird, bevor Realisierungen beobachtbar sind (etwa bei Zuverlässigkeitsuntersuchungen von erst in Betrieb zu nehmenden industriellen Anlagen).

Wenn man ein mathematisches Modell, d. h. einen stochastischen Prozess, festgelegt hat, dann kann man je nach Anwendung an verschiedenen Fragen interessiert sein. Zum Beispiel an

- der Berechnung gewisser Wahrscheinlichkeiten (analytisch oder numerisch),
- der Berechnung von (optimalen) Steuerungen,
- qualitativen Aussagen (z. B. Konvergenz gegen Gleichgewicht),
- der Simulation des Prozesses auf dem Rechner.

Unter Umständen erkennt man bei diesen Schritten, dass das Modell den realen Prozess doch nicht hinreichend gut beschreibt. Dann bietet es sich an, eine andere Modellierung zu versuchen und zu testen (etwa durch Simulation). Aber auch dann, wenn man hinreichendes Vertrauen zu der Modellierung hat, sind Berechnungen und Simulationen interessant, um gegebenenfalls entsprechend zu reagieren (z. B. Aktien verkaufen, Versicherungsprämien erhöhen, Kapazitätsausweitung).

## 1.2 Warteschlangenmodelle

Warteschlangensituationen sind in den Anwendungen sehr wichtig (Rechnernetze, Kunden in Geschäften, Bearbeitung von Anträgen) und gleichzeitig oft recht gut und einfach durch stochastische Prozesse modellierbar. In einer Reihe von interessanten Spezialfällen kann man analytische Resultate gewinnen. Selbst wenn dies nicht gelingt, kann man Warteschlangenmodelle meist ohne großen Aufwand auf dem Rechner simulieren. Wir werden an verschiedenen Stellen der Vorlesung auf Warteschlangenmodelle eingehen. Hier werden wir zunächst nur eine grobe Beschreibung einer Teilklasse von Warteschlangenmodellen angeben, die in der Literatur besonders ausführlich behandelt wird. Diese Modelle werden durch vier durch Schrägstriche getrennte Symbole gekennzeichnet:

$$V_A/V_B/A_B/K_{\max}, \quad (1.2.1)$$

mit den Bedeutungen:

$V_A$  Verteilung der Zwischenankunftszeiten,

$V_B$  Verteilung der Bedienungszeiten,

$A_B$  Anzahl der Bediener,

$K_{\max}$  maximale Kapazität.

Hierbei sind die folgenden Annahmen üblich:

- Die Kunden kommen einzeln an,
- die Zeitdauern zwischen zwei aufeinanderfolgenden Ankünften sind unabhängig und identisch verteilt,
- ebenso die Bedienungszeiten.

In den ersten beiden Positionen (also  $V_A$  und  $V_B$ ) werden die Verteilungen der Zwischenankunftszeiten und der Bedienungszeiten eingetragen. Entweder kann dort explizit die Verteilung notiert werden oder (was üblicher ist) durch folgende Symbole *Klassen* von Verteilungen (d. h. die *genaue* Verteilung wird offengelassen):

- $G$  = “general” (d. h. beliebig)
- $D$  = “deterministisch”
- $M$  = “Markovsch”, d. h. Exponentialverteilung

- $E_k$  = “Erlangverteilung mit Formparameter  $k$ ”, d. h. Verteilung der Summe von  $k$  unabhängigen exponentialverteilten Zufallsvariablen mit demselben Parameter ( $k \in \mathbb{N}$ ).

In der dritten Position (also  $A_B$ ) steht eine natürliche Zahl, die die Zahl der Bediener angibt, unter Umständen auch  $\infty$ .

In der vierten Position (also  $K_{\max}$ ) steht eine natürliche Zahl, die die maximale Zahl von Kunden im System angibt. Es wird angenommen, dass Ankünfte bei vollem System abgewiesen werden (und nicht warten). Gibt es keine Kapazitätsbeschränkung, so läßt man die vierte Position meist leer, anstatt  $\infty$  einzutragen.

Weitere Angaben über ein Warteschlangenmodell, die aus der obigen Notation nicht hervorgehen, werden bei Bedarf extra spezifiziert. Neben einer Festlegung von Parametern ist dies z. B. die Regelung, in welcher Reihenfolge die Kunden bedient werden. Gängige Regeln hierfür sind *FIFO* (“first in first out”), seltener *LIFO* (“last in first out”, Beispiel: Stapel auf Schreibtisch), zufällige Auswahl und Reihenfolge nach Prioritäten eventuell in Verbindung mit anderen Regeln bei gleicher Priorität. Dabei ist festzulegen, ob Bedienungen auch unterbrochen werden, wenn eine Anforderung mit höherer Priorität eintrifft.

Wir betonen noch einmal, dass häufig in der Realität Modelle, die mit der obigen Notation beschreibbar sind, zu primitiv sind. Das heisst nicht, dass kompliziertere Modelle einer mathematischen Beschreibung unzugänglich wären. In der Tat werden in der Literatur der letzten 20 Jahre vielfach recht komplizierte Warteschlangenmodelle, die z. B. Abhängigkeiten oder periodische Schwankungen berücksichtigen, untersucht.

Fragestellungen im Zusammenhang mit Warteschlangenmodellen sind zum Beispiel:

- Konvergiert die Verteilung der Anzahl der Kunden im System gegen eine Grenzverteilung, oder “explodiert” die Verteilung?
- Verteilung der Wartezeit eines Kunden?
- Verteilung der Zeit, bis das System wieder leer ist?
- Wie ändern sich diese Größen, wenn die Zahl der Bediener um 1 erhöht wird? (oder wenn sich durch bessere Schulung oder besseren Leistungsanreiz die Bedienungszeitverteilung verringert).

### 1.3 Stochastische Prozesse und ihre Verteilung

Wir wollen nun formal den Begriff eines stochastischen Prozesses definieren.

**Definition 1.3.1.** Ein *stochastischer Prozess* ist eine Familie  $(X_t)_{t \in I}$  von reellwertigen Zufallsvariablen  $X_t$ ,  $t \in I$ , die auf demselben Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  definiert sind. Dabei ist  $I$  eine beliebige (nichtleere) Indexmenge.

**Bemerkung 1.3.2.** Meist ist  $I$  eine Teilmenge von  $\mathbb{R}$  und wird als “Zeit” interpretiert.

Oft erlaubt man auch allgemeinere Zustandsräume  $E$ . Um von der Verteilung einer  $E$ -wertigen Zufallsvariablen reden zu können, muss auf  $E$  eine  $\sigma$ -Algebra  $\mathcal{E}$  definiert werden. Das Paar  $(E, \mathcal{E})$  nennt man einen *Messraum*. Wir erinnern daran, dass eine  $\sigma$ -Algebra  $\mathcal{E}$  über einer nichtleeren Menge  $E$  eine Teilmenge der Potenzmenge  $2^E$  ist, die die leere Menge enthält und abgeschlossen gegen Komplement- und abzählbarer Vereinigungsbildung ist.  $\sigma$ -Algebren sind die natürlichen Definitionsbereiche für Wahrscheinlichkeitsmaße.



**Definition 1.3.3.** Sei  $(E, \mathcal{E})$  ein Messraum. Ein *stochastischer Prozess mit Zustandsraum*  $(E, \mathcal{E})$  ist eine Familie  $(X_t)_{t \in I}$  von messbaren Abbildungen  $X_t, t \in I$ , von demselben Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  nach  $(E, \mathcal{E})$ . Für festes  $\omega \in \Omega$  heisst die Abbildung  $X(\omega): I \rightarrow E$  *Pfad, Trajektorie* oder *Realisierung* des stochastischen Prozesses. Falls  $I = \mathbb{N}_0$  oder  $I = [0, \infty)$ , so heisst die Verteilung von  $X_0$  die *Startverteilung* des Prozesses. (Falls  $E$  diskret ist, so spricht man auch vom *Startvektor*.)

Wir erinnern daran, dass eine Abbildung  $X: (E_1, \mathcal{E}_1) \rightarrow (E_2, \mathcal{E}_2)$  zwischen zwei Messräumen *messbar* heisst, wenn alle Urbilder unter  $X$  von Elementen in  $\mathcal{E}_2$  in  $\mathcal{E}_1$  liegen, also

$$B \in \mathcal{E}_2 \quad \implies \quad \{\omega \in E_1: X(\omega) \in B\} \in \mathcal{E}_1. \quad (1.3.1)$$

Man mache sich die Analogie zum Begriff der Stetigkeit einer Abbildung zwischen topologischen Räumen klar!

Für Zufallsvariable ist der Begriff der *Verteilung* sehr wichtig. Fast immer sind die Größen, die einen an einer Zufallsvariablen interessieren, eine Funktion der Verteilung der Zufallsvariablen, zum Beispiel Erwartungswert, Varianz und die Überschreitungswahrscheinlichkeit einer Grenze. Für stochastische Prozesse gilt Ähnliches. Daher ist es unser nächstes Ziel, die Verteilung eines stochastischen Prozesses zu definieren. Wir erinnern an die Definition der Verteilung einer reellwertigen Zufallsvariable, d. h. des Falls  $|I| = 1$ ,  $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ , wobei  $\mathcal{B}$  die *Borel- $\sigma$ -Algebra* über  $\mathbb{R}$  ist, nämlich die von den Intervallen erzeugte.

Eine Zufallsvariable  $X$  ist nun eine messbare Abbildung

$$X: (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}), \quad (1.3.2)$$

und die Verteilung  $\hat{P}$  von  $X$  ist das Bildmaß auf  $(\mathbb{R}, \mathcal{B})$  von  $\mathbb{P}$  unter  $X$  d. h.

$$\hat{P}(B) := \mathbb{P}(X^{-1}(B)) := \mathbb{P}(\{\omega \in \Omega: X(\omega) \in B\}), \quad B \in \mathcal{B}. \quad (1.3.3)$$

Man sieht hier, dass die Messbarkeit von  $X$  wichtig ist! Sie garantiert nämlich, dass  $X^{-1}(B)$  in  $\mathcal{F}$  liegt und somit  $\mathbb{P}(X^{-1}(B))$  überhaupt definiert ist.

In der allgemeinen Situation hat man statt  $\mathbb{R}$  Funktionen von  $I$  nach  $E$ , also Elemente aus  $E^I$ , d. h. zufällige Funktionen statt zufälliger Zahlen. Um wie im reellwertigen Fall wie oben eine Verteilung  $\hat{P}$  definieren zu können, braucht man auf  $E^I$  eine  $\sigma$ -Algebra. Die am meisten verwendete und üblichste ist die Produkt- $\sigma$ -Algebra<sup>1</sup>  $\mathcal{E}^I$ :

**Definition 1.3.4.**  $\mathcal{E}^I$  ist die kleinste  $\sigma$ -Algebra über  $E^I$ , die alle endlich-dimensionalen Rechtecke enthält, d. h. Mengen der Form

$$\{f \in E^I: f(i_1) \in E_1, \dots, f(i_k) \in E_k\}, \quad \text{mit } k \in \mathbb{N}, i_1, \dots, i_k \in I, E_1, \dots, E_k \in \mathcal{E}. \quad (1.3.4)$$

*Übungsaufgabe:* In Definition 1.3.4 und schon vorher bei der Definition der Borel- $\sigma$ -Algebra brauchten wir das folgende Resultat: Sei  $E$  eine nichtleere Menge und  $\mathcal{C}$  eine Familie von Teilmengen der Potenzmenge  $2^E$ . Dann gibt es eine kleinste  $\sigma$ -Algebra  $\mathcal{E}$  über  $E$ , die  $\mathcal{C}$  enthält.

**Lemma 1.3.5.** Sei  $(X_t)_{t \in I}$  ein stochastischer Prozess mit Zustandsraum  $(E, \mathcal{E})$ . Definiere eine Abbildung  $X: (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E^I, \mathcal{E}^I)$  durch

$$(X(\omega))(t) := X_t(\omega), \quad \omega \in \Omega, t \in I. \quad (1.3.5)$$

Dann ist  $X$  messbar, d. h.  $X^{-1}(B) \in \mathcal{F}$  für alle  $B \in \mathcal{E}^I$ .

<sup>1</sup>Wir erinnern daran, dass  $\mathcal{E}^I$  nicht die Menge aller Abbildungen  $I \rightarrow \mathcal{E}$  bezeichnet, sondern die von dieser Menge erzeugte  $\sigma$ -Algebra.

**Beweis.** Übung. □

Nach diesen Vorbereitungen können wir die Verteilung eines stochastischen Prozesses ganz analog zur Verteilung einer Zufallsvariablen definieren.

**Definition 1.3.6.** Die Verteilung  $\widehat{P}$  eines stochastischen Prozesses  $(X_t)_{t \in I}$  auf  $(\Omega, \mathcal{F}, P)$  mit Zustandsraum  $(E, \mathcal{E})$  ist das Bildmaß von  $\mathbb{P}$  unter der in (1.3.5) definierten Abbildung  $X$ ; in Formeln:

$$\widehat{P}(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega : X(\omega) \in B\}), \quad B \in \mathcal{E}^I. \quad (1.3.6)$$

**Bemerkung 1.3.7.** Aus der Verteilung  $\widehat{P}$  eines Prozesses  $(X_t)_{t \in I}$  ist im Allgemeinen nicht erkennbar, ob zum Beispiel im Fall  $I = \mathbb{R}, E = \mathbb{R}$  alle Pfade stetige Funktionen sind (vgl. Übungsaufgabe). Die frühere Bemerkung, dass für Zufallsvariablen nahezu alle interessanten Fragen nur durch die Kenntnis der Verteilung beantwortet werden können, ist daher nur mit Einschränkungen auf stochastische Prozesse übertragbar. Meist versucht man, zu gegebenem  $\widehat{P}$  einen stochastischen Prozess mit Verteilung  $\widehat{P}$  so zu konstruieren, dass die Pfade so “glatt” oder “regulär” wie möglich sind, zumindest rechtsseitig stetig mit linksseitigen Grenzwerten. Wir werden darauf in dieser Vorlesung aber nicht im Detail eingehen.

## 1.4 Endlich-dimensionale Verteilungen

Die Verteilung eines stochastischen Prozesses ist oft ein recht kompliziertes Objekt und ist meist nicht – oder nur mit großer Mühe – explizit angebar. Es stellt sich daher die Frage, ob man Verteilungen implizit charakterisieren kann, etwa dadurch, dass man lediglich die gemeinsamen Verteilungen zu endlich vielen Zeitpunkten angibt. Die Frage ist dann, ob dadurch die Verteilung eindeutig festgelegt wird. Dies ist eine Frage nach der eindeutigen Fortsetzbarkeit zu einem Wahrscheinlichkeitsmaß, dessen Werte nur auf einer Teilmenge von  $\mathcal{F}$  gegeben sind.

Wir schreiben  $J \sqsubset I$ , falls  $J$  eine endliche nichtleere Teilmenge von  $I$  ist.

**Definition 1.4.1.** Sei  $(X_t)_{t \in I}$  ein stochastischer Prozess. Die Familie  $(\widehat{P}_J)_{J \sqsubset I}$  aller (gemeinsamen) Verteilungen  $\widehat{P}_J$  von  $(X_t)_{t \in J}$  für alle  $J \sqsubset I$  heisst *Familie der endlich-dimensionalen Verteilungen* von  $(X_t)_{t \in I}$ , bzw. von  $\widehat{P}$ , wenn dies die Verteilung von  $(X_t)_{t \in I}$  ist.

Die endlich dimensionale Verteilung  $\widehat{P}_J$  ist ein Wahrscheinlichkeitsmaß auf  $(E^J, \mathcal{E}^J)$ . Offenbar ist  $\widehat{P}_J$  durch  $\widehat{P}$  eindeutig bestimmt. Man kann  $\widehat{P}_J$  als das Bildmaß von  $\widehat{P}$  unter der Projektion  $\pi_J: (E^I, \mathcal{E}^I) \rightarrow (E^J, \mathcal{E}^J)$  interpretieren, die  $f \in E^I$  auf  $(\pi_J f)(j) := f(j)$  für  $j \in J$  abbildet (diese Abbildung  $\pi_J$  ist messbar!).

Der folgende Satz 1.4.3 wird nicht nur die Eindeutigkeit einer Fortsetzung in vielen Fällen positiv beantworten, sondern auch notwendige und hinreichende Bedingungen an Familien von Wahrscheinlichkeitsmaßen  $\overline{P}_J$  mit  $J \sqsubset I$  bereitstellen dafür, dass die  $\overline{P}_J$  gerade die endlich-dimensionalen Verteilungen einer Verteilung  $\overline{P}$  sind. Dazu müssen die  $\overline{P}_J$  offensichtlich gewisse Konsistenzbedingungen erfüllen.

**Definition 1.4.2.** Sei  $I$  eine nichtleere Indexmenge,  $(E, \mathcal{E})$  ein Messraum, und für jedes  $J \sqsubset I$  sei ein Wahrscheinlichkeitsmaß  $\overline{P}_J$  auf  $(E^J, \mathcal{E}^J)$  vorgegeben. Die Familie  $(\overline{P}_J)_{J \sqsubset I}$  heisst *konsistent*, wenn für alle  $J_1 \sqsubset I$  und  $J_2 \sqsubset I$  mit  $J_1 \subset J_2$  jeweils gilt, dass  $\overline{P}_{J_1}$  gleich der Einschränkung (oder Projektion) von  $\overline{P}_{J_2}$  auf  $J_1$  ist.

Ein *polnischer Raum* ist ein vollständiger metrischer Raum mit abzählbarer Basis der Topologie. Falls nicht anders vermerkt, versehen wir einen polnischen Raum immer mit der  $\sigma$ -Algebra der Borelmengen, d. h. der kleinsten  $\sigma$ -Algebra, die alle offenen Mengen enthält. Polnische Räume haben sich als sehr natürliche Zustandsräume von stochastischen Prozessen erwiesen; siehe auch Bemerkung 1.4.4.

Der folgende (theoretisch sehr wichtige) Satz stellt klar, dass zu konsistenten Familien von Wahrscheinlichkeitsmaßen in natürlicher Weise ein stochastischer Prozess gehört.

**Satz 1.4.3** (Existenzsatz von Kolmogorov). *Wenn  $E$  ein polnischer Raum ist und  $I$  eine Indexmenge und  $(\bar{P}_J)_{J \sqsubset I}$  eine konsistente Familie von Wahrscheinlichkeitsmaßen ist, dann existiert genau ein Wahrscheinlichkeitsmaß  $\bar{P}$  auf  $(E^I, \mathcal{E}^I)$ , so dass die  $\bar{P}_J$  die endlich dimensionalen Verteilungen von  $\bar{P}$  sind. Weiter existiert ein Wahrscheinlichkeitsraum und darauf ein stochastischer Prozess  $(X_t)_{t \in I}$  mit Zustandsraum  $(E, \mathcal{E})$  und Verteilung  $\bar{P}$ .*

**Beweis.** Der Beweis ist zu aufwendig, um ihn hier zu präsentieren. Der interessierte Leser sei auf [Ba68, S. 288 ff] verwiesen.  $\square$

Eine Konsequenz von Satz 1.4.3 ist, dass, um einen Prozess  $(X_n)_{n \in \mathbb{N}_0}$  mit Werten in einem polnischen Raum zu definieren, es ausreicht, für jedes  $n \in \mathbb{N}_0$  nur die Verteilung des Vektors  $(X_0, \dots, X_n)$  anzugeben unter der Voraussetzung, dass die Folge der so definierten Verteilungen konsistent ist.

**Bemerkung 1.4.4.** Die meisten interessanten Zustandsräume  $(E, \mathcal{E})$  sind polnisch, z. B.  $\mathbb{R}$ ,  $\mathbb{R}^n$ ,  $\mathbb{C}^n$ ,  $\mathbb{R}^{\mathbb{N}}$ ,  $C([0, 1], \mathbb{R}^n)$ . Ohne die Voraussetzung “polnisch” ist Satz 1.4.3 falsch. Der letzte Teil der Aussage von Satz 1.4.3 ist recht leicht zu sehen. Man kann als Wahrscheinlichkeitsraum immer  $(E^I, \mathcal{E}^I, \bar{P})$  wählen und  $X_i(f) := f(i)$  für  $i \in I$  und  $f \in E^I$  wählen. Dann ist die in Lemma 1.3.5 definierte Abbildung  $X$  gerade die Identität und somit die Verteilung von  $X$  gleich  $\bar{P}$ .

## 1.5 Simulation von Zufallsvariablen

Gegeben sei eine Zufallsvariable  $U$  mit einer Gleichverteilung auf  $[0, 1]$ , d. h. die Verteilungsfunktion  $F_U$  von  $U$  ist gegeben durch

$$F_U(x) = \mathbb{P}(U \leq x) = \begin{cases} 0, & \text{falls } x \leq 0, \\ x, & \text{falls } 0 \leq x \leq 1 \\ 1, & \text{falls } x \geq 1. \end{cases} \quad (1.5.1)$$

Weiter sei  $F: \mathbb{R} \rightarrow [0, 1]$  eine vorgegebene Verteilungsfunktion. Man finde eine Funktion  $g: [0, 1] \rightarrow \overline{\mathbb{R}}$ , so dass die Zufallsvariable  $X := g(U)$  die vorgegebene Verteilungsfunktion  $F$  hat.

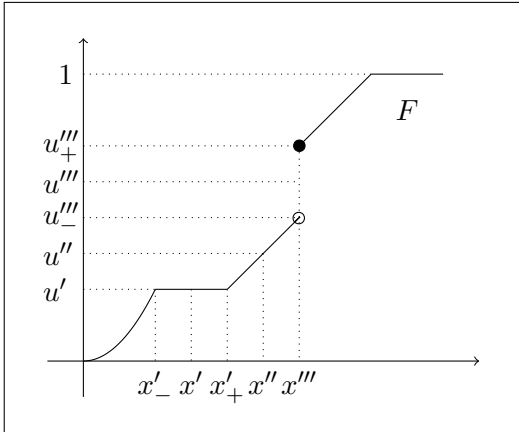
Da die meisten Rechner (Pseudo-)Zufallsgeneratoren besitzen, die Realisierungen von unabhängigen, auf  $[0, 1]$  gleichverteilten Zufallsvariablen (näherungsweise) erzeugen, braucht man den ersten Wert dieser Folge lediglich in  $g$  einzusetzen und erhält damit eine Zufallsvariable mit Verteilungsfunktion  $F$ . Entsprechend kann man Folgen von unabhängigen Zufallsvariablen (auch mit verschiedenen Verteilungen) simulieren.

**Lemma 1.5.1.** *Man kann*

$$g(u) := \inf\{y \in \mathbb{R}: F(y) \geq u\}, \quad u \in [0, 1] \quad (1.5.2)$$

wählen ( $\inf \emptyset = \infty$ ).

**Bemerkung 1.5.2.** Wenn  $F$  stetig und streng monoton wachsend ist, dann ist  $g(u) = F^{-1}(u)$  für alle  $u \in \mathbb{R}$ . Graphisch kann man sich  $g$  auch für allgemeines  $F$  wie folgt veranschaulichen:



Man trägt die Realisierung  $u$  (die der Rechner liefert) auf der Ordinate ab und läuft von  $-\infty$  nach rechts, bis man auf den Graph von  $F$  stößt. Der zugehörige Abszissenwert ist  $g(u)$ .

**Beweis von Lemma 1.5.1.** Es gilt für  $u \in [0, 1]$

$$u \leq F(x) \iff g(u) \leq x, \tag{1.5.3}$$

denn  $u \leq F(x) \Rightarrow g(u) \leq g(F(x)) \leq x$  und  $g(u) \leq x \Rightarrow u \leq F(g(u)) \leq F(x)$ . (Die Äquivalenz in (1.5.3) gilt nicht, wenn “ $\leq$ ” auf beiden Seiten durch “ $<$ ” ersetzt wird!)

Daraus folgt wegen  $\mathbb{P}(U \in (0, 1)) = 1$

$$\mathbb{P}(X \leq x) = \mathbb{P}(g(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x), \quad x \in \mathbb{R}. \tag{1.5.4}$$

Das heißt, dass  $X = g(U)$  wirklich die vorgegebene Verteilungsfunktion besitzt. □

**Bemerkung 1.5.3.** Die in (1.5.1) angegebene Funktion  $g$  ist nicht die einzige, so dass  $g(U)$  die vorgegebene Verteilung hat.

**Beispiel 1.5.4.** Wie simuliert man eine exponentialverteilte Zufallsvariable?

Sei also  $X \sim \text{Exp}(\lambda)$  für  $\lambda > 0$ , dann hat  $X$  die Verteilungsfunktion

$$F(x) = \begin{cases} 0, & \text{falls } x \leq 0, \\ 1 - e^{-\lambda x}, & \text{falls } x \geq 0. \end{cases} \tag{1.5.5}$$

Wir halten ein  $u \in (0, 1)$  fest. Für die in (1.5.1) definierte Funktion  $g$  gilt  $g(u) = F^{-1}(u) =: x$ , also  $u = F(x) = 1 - e^{-\lambda x}$ .

Löst man diese Gleichung nach  $x$  auf, dann erhält man  $x = g(u) = -1/\lambda \log(1 - u)$ .

Also ist  $X := -1/\lambda \log(1 - U)$   $\text{Exp}(\lambda)$ -verteilt.

**Bemerkung 1.5.5.** Zur Simulation von normalverteilten Zufallsvariablen ist das obige Verfahren nicht sehr gut geeignet, da die Umkehrfunktion der Verteilungsfunktion  $F$  sich nicht einfach darstellen läßt. Ohne Beweis nennen wir eine wesentlich bessere Möglichkeit:

Seien  $U_1$  und  $U_2$  unabhängige, auf  $[0, 1]$  gleichverteilte Zufallsvariablen und

$$\begin{aligned} X_1 &:= (-2 \log U_1)^{1/2} \cos(2\pi U_2), \\ X_2 &:= (-2 \log U_1)^{1/2} \sin(2\pi U_2) \end{aligned}$$

Dann sind  $X_1$  und  $X_2$  unabhängig und beide  $\mathcal{N}(0, 1)$ -verteilt. Will man nur *eine*  $\mathcal{N}(0, 1)$ -verteilte Zufallsvariable, so ignoriert man  $X_2$ .



# Kapitel 2

## Markovketten mit diskreter Zeit

### 2.1 Einleitung

*Markovketten* modellieren zufällige Vorgänge mit endlich oder abzählbar unendlich vielen Zuständen mit diskreter Zeit ( $I = \mathbb{N}_0$ ), bei denen Übergänge zwischen den Zuständen mit vorgegebenen Wahrscheinlichkeiten unabhängig von der Vorgeschichte – d. h. auf welchem Pfad man in den gegenwärtigen Zustand kam – stattfinden. Den Zustandsraum bezeichnen wir wieder mit  $E$ . Man kann o.B.d.A. immer  $E = \{1, \dots, n\}$  für ein  $n \in \mathbb{N}$  oder  $E = \mathbb{N}$  wählen. Als  $\sigma$ -Algebra  $\mathcal{E}$  auf  $E$  nehmen wir – wie bei abzählbaren Mengen üblich – immer die Potenzmenge von  $E$ . Anschaulich kann man sich eine Markovkette als gerichteten Graphen vorstellen, wobei die Ecken die Elemente von  $E$  darstellen und gerichtete Kanten solche Übergänge, die positive Wahrscheinlichkeit haben. Jede Kante beschriftet man mit der zugehörigen Übergangswahrscheinlichkeit, zum Beispiel

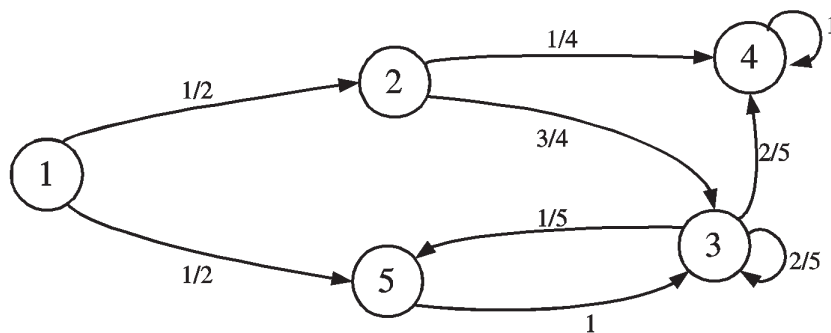


Abbildung 2.1.1: Graphische Darstellung einer Markovkette

Eine Markovkette ist charakterisiert durch eine Funktion  $P: E \times E \rightarrow [0, 1]$ , wobei wir  $P = (p_{ij})_{i,j \in E}$  schreiben und  $p_{ij}$  die Übergangswahrscheinlichkeit nach  $j$  angibt, wenn man sich gerade in  $i$  befindet.  $P$  lässt sich als (eventuell unendliche) Matrix interpretieren, wobei jedem Zustand eine Zeile und eine Spalte entspricht.

**Definition 2.1.1.** Eine Matrix  $P = (p_{ij})_{i,j \in E}$  heißt *Übergangsmatrix* oder *stochastische Matrix*,

falls gelten:

- (i)  $p_{ij} \geq 0$  für alle  $i, j \in E$ ,
- (ii)  $\sum_{j \in E} p_{ij} = 1$  für alle  $i \in E$ .

**Bemerkung 2.1.2.** Wir setzen immer voraus, dass die Übergangsmatrix sich zeitlich nicht ändert (*zeitliche Homogenität*) und machen dies zu einem Bestandteil der Definition einer Markovkette. Manche Autoren lassen eine zeitliche Inhomogenität bei der Definition einer Markovkette zu (d. h.  $P$  hängt noch von der Zeit  $n$  ab), behandeln dann aber meistens doch nur den zeitlich homogenen Fall ausführlicher.

Bislang haben wir den Begriff einer Markovkette noch nicht mit dem eines stochastischen Prozesses in Verbindung gebracht. Da wir unter einer Markovkette nur den Übergangsmechanismus, der durch  $P$  gegeben ist, verstehen wollen, ist eine Markovkette nicht apriori ein stochastischer Prozess. Durch  $P$  wird nicht einmal die Verteilung eines stochastischen Prozesses eindeutig festgelegt, da  $P$  nichts darüber aussagt, mit welcher Verteilung auf  $E$  der Prozess starten soll. Erst eine Startverteilung und eine Übergangsmatrix  $P$  zusammen legen eindeutig eine Verteilung auf  $E^I$  fest. Wir werden später darauf genauer eingehen.

## 2.2 Simulation einer Markovkette und Beispiele

Es sei ein höchstens abzählbarer Zustandsraum  $E$ , eine stochastische Matrix  $P$  und ein Wahrscheinlichkeitsvektor  $a$  gegeben. Weiter stehe eine Folge  $U_1, U_2, \dots$  von unabhängigen, auf  $[0, 1]$  gleichverteilten Zufallsvariablen zur Verfügung. Man simuliere damit eine Markovkette mit Startverteilung  $a$  und Übergangsmatrix  $P$ . Wir nehmen hier (o.B.d.A.) an, dass  $E = \{1, \dots, n\}$  oder  $E = \mathbb{N}$  ist.

Zunächst simuliert man die Startposition, d. h. eine  $E$ -wertige Zufallsvariable  $X_0$  mit der Startverteilung  $a$ . Wir haben schon gesehen, wie man dies macht. Um die Lesbarkeit zu verbessern, schreiben wir einige Schleifen explizit aus.

### 1. Simulation des Startwertes $X_0$

$k = 0$ .

$j = 1$ .

Wenn  $U_1 \leq a_1$ , setze  $X_0 = j$ , gehe nach 2.

$j = 2$ .

Wenn  $U_1 \leq a_1 + a_2$ , setze  $X_0 = j$ , gehe nach 2.

$\vdots$

### 2. Simulation von $X_{k+1}$

$i := j$

$k \rightarrow k + 1$ .

$j = 1$ .

Wenn  $U_{k+1} \leq p_{i1}$ , setze  $X_k = j$ , gehe nach 2.

$j = 2$ .

Wenn  $U_{k+1} \leq p_{i1} + p_{i2}$ , setze  $X_k = j$ , gehe nach 2.



⋮

Als Abbruchkriterium wird man meist die Überschreitung einer gewissen Grenze von  $k$  wählen. Alternativ könnte man so lange simulieren, bis ein bestimmter Zustand eine definierte Anzahl von Besuchen erfahren hat.

Bevor wir die Theorie der Markovketten weiter behandeln, betrachten wir zunächst einige Beispiele.

**Beispiel 2.2.1** (Symmetrische Irrfahrt auf  $\mathbb{Z}$ ). Hier ist  $E = \mathbb{Z}$  und

$$p_{ij} = \begin{cases} \frac{1}{2}, & \text{falls } |i - j| = 1, \\ 0 & \text{sonst.} \end{cases} \tag{2.2.1}$$

Diese Markovkette beschreibt ein Teilchen, das pro Zeiteinheit auf  $\mathbb{Z}$  um eins nach rechts oder links springt, und zwar immer mit Wahrscheinlichkeit  $1/2$ . Die Übergangsmatrix  $P = (p_{ij})_{i,j \in \mathbb{Z}}$  ist dann eine unendlich große Dreibandmatrix, die auf der Hauptdiagonalen ausschliesslich Nullen hat und auf den beiden Nebendiagonalen immer den Wert  $1/2$ .

Das Anfangsstück eines Pfades der symmetrischen Irrfahrt (mit Start in Null) kann zum Beispiel so aussehen (linear interpoliert):

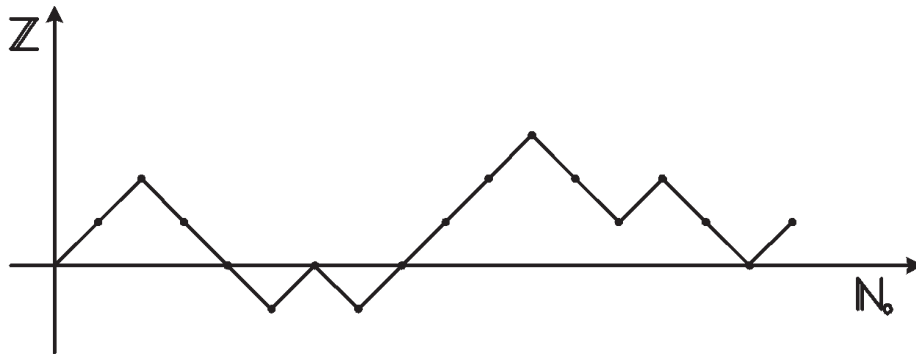


Abbildung 2.2.1: Realisierung einer symmetrischen Irrfahrt

**Beispiel 2.2.2** (Symmetrische Irrfahrt auf  $\mathbb{Z}^d$ ,  $d \in \mathbb{N}$ ). Hier ist  $E = \mathbb{Z}^d$  und

$$p_{ij} = \begin{cases} \frac{1}{2d}, & \text{falls } |i - j| = 1, \\ 0 & \text{sonst.} \end{cases} \tag{2.2.2}$$

(Mit  $|\cdot|$  meinen wir die Summe der Beträge der Koeffizienten.) Man sieht leicht, dass  $\sum_{j \in E} p_{ij} = 1$  für alle  $i$  ist, denn jeder Punkt im  $d$ -dimensionalen Gitter  $\mathbb{Z}^d$  hat  $2d$  Nachbarn (d.h. Elemente mit euklidischem Abstand 1).

Zu diesen (und vielen anderen) Beispielen kann man die folgenden Fragen stellen:

- Was ist die Verteilung von  $X_n$  (d. h. des Ortes nach  $n$  Schritten) bei bekannter Startverteilung?
- Wie groß ist die Wahrscheinlichkeit, jemals zum Ausgangspunkt zurückzukehren?
- Wie groß ist die Wahrscheinlichkeit, unendlich oft zum Ausgangspunkt zurückzukehren?
- Wie groß ist die erwartete Anzahl von Besuchen des Ausgangspunktes (wenn der Prozess unendlich lange läuft)?

**Beispiel 2.2.3** (uiv Folgen). Folgen von unabhängigen, identisch verteilten Zufallsgrößen sind insbesondere Markovketten: Sei  $X_0, X_1, X_2, \dots$  eine Folge von u.i.v. Zufallsvariable mit abzählbarem Zustandsraum  $E$ . Sei  $b_i = \mathbb{P}(X_0 = i)$  für alle  $i \in E$ . Dann gilt  $\sum_{i \in E} b_i = 1$ . Offenbar ist  $(X_0, X_1, X_2, \dots)$  eine Markovkette mit Startverteilung  $b = (b_i)_{i \in E}$  und Übergangsmatrix  $P = (b_j)_{i,j \in E}$ . Insbesondere sind alle Spalten  $P$  konstant, da die Zufallsvariablen  $X_0, X_1, X_2, \dots$  unabhängig sind.

**Beispiel 2.2.4** (Irrfahrten). Sei  $X_1, X_2, \dots$  eine Folge von u.i.v. Zufallsvariablen mit Zustandsraum  $\mathbb{Z}$  und Verteilung gegeben durch  $b_i = \mathbb{P}(X_1 = i)$ . Wir setzen  $S_0 := 0$  und  $S_n = \sum_{k=1}^n X_k$ . Dann ist  $(S_n)_{n \in \mathbb{N}_0}$  eine Markovkette mit Startverteilung  $a_i = \delta_{i0}$  (d. h.  $a_0 = 1$  und  $a_i = 0$  für alle  $i \neq 0$ ). Die Übergangsmatrix ist  $P = (b_{j-i})_{i,j \in \mathbb{Z}}$ . Die Markovkette  $(S_n)_{n \in \mathbb{N}_0}$  heißt *Irrfahrt* (*random walk*) auf  $\mathbb{Z}$ . Für  $b_1 = b_{-1} = 1/2$  und  $b_i = 0$  für  $|i| \neq 1$  erhält man als Spezialfall die symmetrische Irrfahrt.

**Beispiel 2.2.5** (Asymmetrische Irrfahrt mit absorbierenden Rändern). Zwei Spieler  $A$  und  $B$  haben zusammen  $N$  Euro. In jeder Spielrunde verliert  $A$  mit Wahrscheinlichkeit  $p \in (0, 1)$  einen Euro an  $B$  und gewinnt von  $B$  mit Wahrscheinlichkeit  $q = 1 - p$  einen Euro. Ist einer der Spieler pleite, so endet das Spiel.

Betrachtet man das Vermögen von  $A$  nach  $n$  Runden, so wird dies durch eine Markovkette mit Zustandsraum  $E = \{0, \dots, N\}$  und Übergangsmatrix

$$P = \begin{pmatrix} 1 & 0 & 0 & \dots & & & 0 \\ p & 0 & q & 0 & \dots & & 0 \\ 0 & p & 0 & q & 0 & \dots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & p & 0 & q & 0 \\ & & & 0 & p & 0 & q \\ 0 & \dots & & & 0 & 0 & 1 \end{pmatrix} \quad (2.2.3)$$

beschrieben. Nun sind folgende Fragen von Interesse:

- Wenn das Startkapital von  $A$   $i$  Euro ist, wie groß ist seine Chance (irgendwann) alles zu gewinnen?
- Wie lange dauert das Spiel? (Verteilung, Erwartungswert).

**Beispiel 2.2.6** (Asymmetrische Irrfahrt mit reflektierenden Rändern). Die Spielregeln sind wie im Beispiel 2.2.5 mit der einzigen Ausnahme, dass ein Spieler dann, wenn er pleite ist,

in der nächsten Runde auf jeden Fall 1 Euro vom anderen Spieler erhält. Die entsprechende Markovkette hat als Übergangsmatrix

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & & & 0 \\ p & 0 & q & 0 & \dots & & 0 \\ 0 & p & 0 & q & 0 & \dots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & p & 0 & q & 0 \\ & & & 0 & p & 0 & q \\ 0 & \dots & & & 0 & 1 & 0 \end{pmatrix} \tag{2.2.4}$$

Fragen:

- Wie oft ist  $A$  bis zur Zeit  $n$  pleite gewesen, d. h. wie oft wurde der Zustand 0 besucht? (Verteilung, Erwartungswert, Asymptotik für große  $n$ ).
- Existiert der Grenzwert

$$\lim_{n \rightarrow \infty} \frac{\#\text{Pleiten von } A \text{ bis } n}{\#\text{Pleiten von } B \text{ bis } n}$$

fast sicher? Wenn ja, ist der Grenzwert deterministisch? Wie lässt er sich gegebenenfalls berechnen?

**Beispiel 2.2.7** (Warenlagermodell). Die tägliche Nachfrage nach Waren in einem Lager sei gegeben durch u.i.v. Zufallsvariable  $X_1, X_2, X_3, \dots$  mit  $b_i = \mathbb{P}(X_1 = i)$  für  $i \in \mathbb{N}_0$  und  $\sum_{i \in \mathbb{N}_0} b_i = 1$ . Seien natürliche Zahlen  $s, S \in \mathbb{N}_0$  mit  $0 \leq s < S$  vorgegeben. Am Abend von Tag  $n$  wird eine Bestellung aufgegeben, wenn im Lager weniger als  $s$  Einheiten vorhanden sind. Die Bestellung trifft bis zum nächsten Morgen ein. Bestellt wird soviel, dass am nächsten Morgen  $S$  Einheiten im Lager sind. Wir definieren den Lagerinhalt am Abend als die Differenz vom Lagerinhalt am Morgen und der Nachfrage am Tag, auch wenn diese Zahl negativ ist, weil die Nachfrage nicht befriedigt wurde. Man kann wahlweise annehmen, dass nicht befriedigte Nachfrage endgültig verloren ist, oder aber, dass über Nacht entsprechend mehr bestellt und den Kunden nachgeliefert wird. Man bezeichnet diese – von den zwei Parametern  $s$  und  $S$  abhängige – Bestellstrategie als  $(s, S)$ -Warenlagerpolitik.

Seien  $Y_n$  bzw.  $Z_n$  der Lagerinhalt am Morgen bzw. am Abend von Tag  $n$ . Dann gilt

$$Y_{n+1} = \begin{cases} Y_n - X_n, & \text{wenn } Y_n - X_n \geq s, \\ S, & \text{wenn } Y_n - X_n < s. \end{cases} \tag{2.2.5}$$

Offenbar ist  $(Y_n)_{n \in \mathbb{N}_0}$  eine Markovkette mit Zustandsraum  $\{s, \dots, S\}$ . Die zugehörige Übergangsmatrix ist

$$P_Y = \begin{pmatrix} b_0 & 0 & 0 & \dots & 0 & \sum_{i=1}^{\infty} b_i \\ b_1 & b_0 & 0 & \dots & 0 & \sum_{i=2}^{\infty} b_i \\ b_2 & b_1 & b_0 & 0 & \dots & \sum_{i=3}^{\infty} b_i \\ \vdots & \vdots & & & & \vdots \\ b_{S-s-1} & b_{S-s-2} & \dots & & b_0 & \sum_{i=S-s}^{\infty} b_i \\ b_{S-s} & b_{S-s-1} & \dots & & b_1 & b_0 + \sum_{i=S-s+1}^{\infty} b_i \end{pmatrix} \tag{2.2.6}$$

Für den Lagerinhalt am Abend gilt

$$Z_{n+1} = \begin{cases} Z_n - X_{n+1}, & \text{wenn } Z_n \geq s, \\ S - X_{n+1}, & \text{wenn } Z_n < s. \end{cases} \quad (2.2.7)$$

Auch  $(Z_n)_{n \in \mathbb{N}_0}$  ist eine Markovkette. Der Zustandsraum ist  $\{S, S-1, S-2, \dots\}$ . (Wenn die Nachfrage durch eine Konstante beschränkt ist, dann kann man auch einen endlichen Zustandsraum wählen.) Die Übergangsmatrix  $P_Z = (p_{ij})_{i,j \leq S}$  ist gegeben durch:

$$p_{ij} = \begin{cases} b_{i-j}, & \text{falls } s \leq i \leq S \text{ und } i \geq j, \\ 0, & \text{falls } s \leq i \leq S \text{ und } i < j, \\ b_{S-j}, & \text{falls } i < s. \end{cases} \quad (2.2.8)$$

Eine wichtige Frage ist die nach den Kosten der  $(s, S)$ -Bestellpolitik mit dem Ziel der Optimierung der Parameter  $s$  und  $S$ . Wir machen hierzu folgende Annahmen:

- Pro Bestellung fallen Fixkosten  $K_B$  an.
- Für die Lagerhaltung fallen am Tag  $n$  die Kosten  $f_1(Y_n)$  an, wobei  $f_1$  eine nichtnegative monoton wachsende Funktion ist (Energiekosten, Versicherungsprämien, Zinsen).  $f_1$  wird man meist als linear oder jedenfalls konkav ansetzen.
- Kosten für nicht sofort befriedigte Nachfrage: Wenn  $Z_n < 0$  ist, dann fallen Kosten in Höhe von  $f_2(Z_n^-)$  an, wobei

$$Z_n^- = \begin{cases} -Z_n, & \text{falls } Z_n < 0, \\ 0 & \text{sonst,} \end{cases} \quad (2.2.9)$$

und  $f_2: \mathbb{N}_0 \rightarrow \mathbb{R}$  monoton wachsend ist (z. B. linear). Diese Kosten fallen entweder wirklich an dadurch, dass die Ware z. B. als Ausgleich für die nicht sofortige Verfügbarkeit direkt dem Kunden zugeleitet wird oder "fiktiv" dadurch, dass Kunden verärgert sind und später nicht mehr als Nachfrager auftreten.

- Nur von  $S$  abhängige Lagerinvestitions- oder Mietkosten pro Tag. Diese seien durch eine monotone Funktion  $f_3: \mathbb{N} \rightarrow \mathbb{R}$  gegeben.

Die Durchschnittsgesamtkosten der ersten  $n$  Tage sind daher

$$K^{(n)} = \frac{1}{n} \sum_{k=1}^n \left( K_B \mathbb{1}_{\{s-1, s-2, \dots\}}(Z_k) + f_1(Y_k) + f_2(Z_k^-) + f_3(S) \right).$$

Die variablen Bestellkosten pro Einheit haben wir hier nicht berücksichtigt. Im Hinblick auf eine Optimierung von  $s$  und  $S$  ist dies dann zulässig, wenn erstens diese Kosten pro Einheit nicht von der Bestellmenge abhängen (es also keinen Mengenrabatt gibt) und zweitens die Nachfrage nicht von der Bestellstrategie abhängt. Letzteres wollen wir hier annehmen. Damit dies realistisch ist, können wir annehmen, dass Kunden als Ausgleich für nicht sofort lieferbare Ware (die am nächsten Morgen nachgeliefert wird) soviel pro Einheit erhalten, dass dies keinen Einfluss auf die künftige Nachfrage hat. Diese Ausgleichszahlungen werden unter  $f_2$  berücksichtigt.

Selbstverständlich kann man auch andere Annahmen machen, ohne dass die Markoveigenschaft zerstört wird.

Interessant ist die Frage der Konvergenz der Zufallsvariablen  $K^{(n)}$  (die eigentlich erst durch die Festlegung einer Startverteilung zu Zufallsvariablen werden) für  $n \rightarrow \infty$ . Wir werden in Beispiel 2.7.2 zeigen, dass die  $K^{(n)}$  nicht nur mit Wahrscheinlichkeit 1 konvergieren, sondern dass dieser Grenzwert sogar deterministisch ist. Wir werden auch sehen, wie man ihn berechnet. Diesen Grenzwert, der noch von  $s$  und  $S$  abhängt, kann man dann optimieren, indem man  $s$  und  $S$  so wählt, dass er ein globales Minimum annimmt.

**Beispiel 2.2.8** (Verbreitung von Gerüchten). In einem von  $N$  Dörfern ( $N \geq 2$ ) sei ein Gerücht entstanden. Aus den  $N$  Dörfern werden zu jeder Zeiteinheit  $n = 1, 2, \dots$  zwei verschiedene zufällig ausgewählt, die telefonisch Kontakt aufnehmen. Ist einem der beiden Dörfer das Gerücht bekannt, so teilt es dies dem anderen mit. Kennen beide Dörfer das Gerücht, dann ist der Anrufer enttäuscht über den Misserfolg und erzählt es fortan nie mehr. (Wir nehmen wohl nicht ganz unrealistisch an, dass innerhalb eines Dorfes das Gerücht sofort allen bekannt ist, wenn einer es kennt. Statt "Dörfer" könnte man auch "Personen" wählen.)

Sei  $X_n = (S_n, I_n, R_n)$ , wobei

$S_n$  = Anzahl der Orte zur Zeit  $n$ , die das Gerücht nicht kennen,

$I_n$  = Anzahl der Orte zur Zeit  $n$ , die das Gerücht kennen und noch weitererzählen,

$R_n$  = Anzahl der Orte zur Zeit  $n$ , die das Gerücht kennen, aber nicht mehr erzählen.

Das Modell und die Bezeichnungen ( $S = susceptible$ ,  $I = infected$ ,  $R = removed$ ) sind an Infektionsmodelle angelehnt.  $I$  sind dabei die Infizierten, die in  $S$  können noch angesteckt werden, die in  $R$  sind immun (oder gestorben). Während Ansteckungen mit dem obigen Modell halbwegs realistisch modelliert werden können, ist es weniger plausibel, dass beim Zusammentreffen von zwei Infizierten einer immun wird. Bei Infektionsmodellen werden an dieser Stelle deswegen andere Modellannahmen gemacht.

Da  $S_n + I_n + R_n = N$  für alle  $n$  gilt, enthält  $X_n$  redundante Information. Wir betrachten daher  $Y_n = (S_n, I_n)$ . Die Folge  $(Y_n)_{n \in \mathbb{N}_0}$  ist offenbar eine Markovkette mit Zustandsraum

$$E = \{(m_1, m_2) \in \mathbb{N}_0^2 : m_1 + m_2 \leq N\}.$$

Die Übergangswahrscheinlichkeiten sind (wir setzen  $r := N - s - i$ ):

$$p_{(s,i),(\tilde{s},\tilde{i})} = \frac{1}{N(N-1)} \times \begin{cases} s(s-1) + 2sr + r(r-1), & \text{falls } \tilde{s} = s \text{ und } \tilde{i} = i, \\ 2si, & \text{falls } \tilde{s} = s-1 \text{ und } \tilde{i} = i+1, \\ i(i-1) + 2ir, & \text{falls } \tilde{s} = s \text{ und } \tilde{i} = i-1, \\ 0 & \text{sonst.} \end{cases} \quad (2.2.10)$$

Interessant ist die Frage nach der Verteilung der Anzahl der Dörfer, die niemals das Gerücht erfahren, wie diese Verteilung von  $N$  abhängt, und ob sie (bei geeigneter Skalierung) für  $N \rightarrow \infty$  konvergiert und gegebenenfalls wogegen. Das Modell geht übrigens auf Daley und Kendall zurück. Eine interessante Arbeit dazu (mit zahlreichen Literaturhinweisen) ist [Pi90].

## 2.3 Definition und einfache Eigenschaften

Wir kehren nun zur Theorie der Markovketten zurück. Zunächst definieren wir formal, was eine Markovkette zu einer Übergangsmatrix  $P$  und Startverteilung  $a$  ist. Dann untersuchen wir, wie

man die endlich dimensionalen Verteilungen berechnet. Im folgenden ist immer  $E$  eine abzählbare (d.h. nichtleere endliche oder abzählbar unendliche) Menge.

**Definition 2.3.1.** Sei  $P: E \times E \rightarrow [0, 1]$  eine stochastische Matrix und  $a: E \rightarrow [0, 1]$  ein Wahrscheinlichkeitsvektor. Ein stochastischer Prozess  $(X_n)_{n \in \mathbb{N}_0}$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  mit Werten in  $E$  heißt *Markovkette mit Übergangsmatrix  $P$  und Startverteilung  $a$* , wenn

- (i)  $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) = p_{i_n i_{n+1}}$   
für alle  $n \in \mathbb{N}_0$  und  $i_0, \dots, i_{n+1} \in E$  mit  $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) > 0$  ist und
- (ii)  $\mathbb{P}(X_0 = i_0) = a_{i_0}$  für alle  $i_0 \in E$ .

In der Situation der letzten Definition sprechen wir auch kurz von einer Markovkette zu  $(P, a)$ . Die folgende Aussage zeigt insbesondere, dass die endlich dimensionalen Verteilungen einer Markovkette zu  $(P, a)$  eindeutig durch  $P$  und  $a$  festgelegt sind.

**Proposition 2.3.2.** Sei  $P: E \times E \rightarrow [0, 1]$  eine stochastische Matrix und  $a: E \rightarrow [0, 1]$  ein Wahrscheinlichkeitsvektor. Weiter sei  $X_0, X_1, \dots$  eine Folge von  $E$ -wertigen Zufallsvariablen auf  $(\Omega, \mathcal{F}, \mathbb{P})$ . Dann sind äquivalent:

- a)  $(X_n)_{n \in \mathbb{N}_0}$  ist eine Markovkette zu  $(P, a)$ .
- b) Für jedes  $n \in \mathbb{N}_0$  und  $i_0, i_1, \dots, i_n \in E$  gilt

$$\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = a_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}.$$

**Beweis.** b)  $\Rightarrow$  a): Wir zeigen, dass aus b) die Eigenschaften (i) und (ii) von Definition 2.3.1 folgen (sofern die bedingten Wahrscheinlichkeiten definiert sind).

- (i)  $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) = \frac{\mathbb{P}(X_0 = i_0, \dots, X_{n+1} = i_{n+1})}{\mathbb{P}(X_0 = i_0, \dots, X_n = i_n)} = p_{i_n i_{n+1}}$ .
- (ii)  $\mathbb{P}(X_0 = i_0) = a_{i_0}$  folgt direkt aus b), indem man  $n = 0$  setzt.

a)  $\Rightarrow$  b): Dies folgt recht einfach mittels vollständiger Induktion nach  $n \in \mathbb{N}_0$ , wobei man beim Induktionsschritt von  $n$  auf  $n + 1$  unterscheiden muss, ob  $\mathbb{P}(X_{i_0} = i_0, \dots, X_n = i_n)$  gleich oder ungleich Null ist.  $\square$

**Proposition 2.3.3.** Für eine Markovkette  $(X_n)_{n \in \mathbb{N}_0}$  zu  $(P, a)$  gilt

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_n i_{n+1}}, \quad n \in \mathbb{N}_0, i_n, i_{n+1} \in E,$$

sofern  $\mathbb{P}(X_n = i_n) > 0$ .

**Beweis.** Sei  $\mathbb{P}(X_n = i_n) > 0$ . Dann gilt mit Proposition 2.3.2

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n) = \frac{\mathbb{P}(X_{n+1} = i_{n+1}, X_n = i_n)}{\mathbb{P}(X_n = i_n)} = \frac{\sum_{i_{n-1}, \dots, i_0} \mathbb{P}(X_{n+1} = i_{n+1}, \dots, X_0 = i_0)}{\sum_{i_{n-1}, \dots, i_0} \mathbb{P}(X_n = i_n, \dots, X_0 = i_0)} = p_{i_n i_{n+1}}.$$

$\square$

Bislang haben wir nicht gezeigt, dass zu jedem  $P$  und  $a$  auch eine Markovkette existiert (obwohl dies anschaulich sicher klar ist – immerhin haben wir ja ein Simulationsverfahren angegeben). Die Existenz folgt aber aus dem folgenden Satz, dessen Beweis leicht aus dem Satz von Kolmogorov (Satz 1.4.3), indem man zeigt, dass die durch Proposition 2.3.2 festgelegten endlich-dimensionalen Verteilungen konsistent sind.

**Satz 2.3.4** (Existenz von Markovketten). *Zu jeder stochastischen Matrix  $P: E \times E \rightarrow [0, 1]$  und jedem Wahrscheinlichkeitsvektor  $a: E \rightarrow [0, 1]$  existiert eine bzgl. Verteilung eindeutige Markovkette zu  $(P, a)$ .*

Der folgende Satz zeigt, wie man die Verteilung von  $X_k$  erhält.

**Satz 2.3.5.** *Für eine Markovkette mit Übergangsmatrix  $P$  und Startvektor  $a$  gilt*

$$\mathbb{P}(X_k = j) = \sum_{i_0, i_1, \dots, i_{k-1} \in E} a_{i_0} p_{i_0 i_1} \cdots p_{i_{k-1} j} = (aP^k)_j,$$

wenn man  $a$  als Zeilenvektor schreibt und den letzten Ausdruck im Sinne der Multiplikation von Matrizen auffasst.

**Beweis.** leicht. □

**Bemerkung 2.3.6.** Man sieht leicht, dass mit zwei stochastischen Matrizen über derselben Menge  $E$  auch deren Produkt eine stochastische Matrix ist. Insbesondere gilt dies für  $P^k$ .

Im Folgenden bezeichnen wir die Koeffizienten der  $k$ -ten Potenz von  $P$  mit  $p_{ij}^{(k)}$ , also insbesondere  $p_{ij}^{(0)} = \delta_{ij}$  und  $p_{ij}^{(1)} = p_{ij}$ .

**Satz 2.3.7.** *Sei  $\mathbb{P}(X_0 = i) > 0$ . Dann ist für jedes  $j \in E$  und jedes  $k \in \mathbb{N}_0$*

$$p_{ij}^{(k)} = \mathbb{P}(X_k = j \mid X_0 = i)$$

die Wahrscheinlichkeit, in  $k$  Schritten von  $i$  nach  $j$  zu kommen.

**Beweis.** Dies ist ein Spezialfall von Satz 2.3.5 mit  $a_i = 1$  und  $a_m = 0$  für  $m \neq i$ . □

**Bemerkung 2.3.8.** Für große  $k$  ist es nicht sinnvoll,  $P^k$  explizit durch Matrixmultiplikation zu berechnen, sondern z. B. im Fall  $|E| < \infty$  über die Jordanzerlegung  $P = AJA^{-1}$ . Dann ist  $P^k = AJ^k A^{-1}$ , wobei  $J^k$  wesentlich einfacher zu berechnen ist als  $P^k$ . Da der Aufwand zur Berechnung von  $A$  unabhängig von  $k$  ist, lohnt er sich für große  $k$ .

**Satz 2.3.9** (Chapman-Kolmogorov-Gleichungen). *Für jede stochastische Matrix  $P = (p_{ij})_{i,j \in E}$  gilt*

$$p_{ij}^{(n+m)} = \sum_{l \in E} p_{il}^{(n)} p_{lj}^{(m)}, \quad i, j \in E, m, n \in \mathbb{N}_0. \quad (2.3.1)$$

**Beweis.** Dies folgt aus der Beziehung  $P^{n+m} = P^n P^m$  durch Ausschreiben der Koeffizienten. □

Wir definieren nun für jedes  $n \in \mathbb{N}_0$  die  $\sigma$ -Algebra  $\mathcal{F}_n$  auf  $\Omega$  als die Familie aller (notwendigerweise abzählbaren) Vereinigungen der Form  $\{\omega \in \Omega : X_0(\omega) = i_0, \dots, X_n(\omega) = i_n\}$  mit  $i_0, \dots, i_n \in E$ . Es ist klar, dass  $\mathcal{F}_n$  eine  $\sigma$ -Algebra ist und in  $\mathcal{F}$  enthalten ist. Weiter gilt  $\mathcal{F}_m \subset \mathcal{F}_n$  für  $m < n$ . Anschaulich beschreibt  $\mathcal{F}_n$  die Information, die ein Beobachter der Markovkette bis zur Zeit  $n$  hat.

**Satz 2.3.10.** Sei  $(X_n)_{n \in \mathbb{N}_0}$  eine Markovkette in  $E$  mit Übergangsmatrix  $P = (p_{ij})_{i,j \in E}$ . Dann gilt für alle  $n, k \in \mathbb{N}_0$ ,  $i_0, \dots, i_k \in E$  und  $F \in \mathcal{F}_n$  mit  $\mathbb{P}(\{X_n = i_0\} \cap F) > 0$

$$\begin{aligned} & \mathbb{P}(X_{n+k} = i_k, \dots, X_{n+1} = i_1 \mid \{X_n = i_0\} \cap F) \\ &= \mathbb{P}(X_{n+k} = i_k, \dots, X_{n+1} = i_1 \mid X_n = i_0) \\ &= p_{i_0 i_1} \cdots p_{i_{k-1} i_k}. \end{aligned} \quad (2.3.2)$$

**Beweis.** Der einfache Beweis folgt ähnlich wie der von Proposition 2.3.2.  $\square$

**Bemerkung 2.3.11.** Satz 2.3.10, dessen Aussage auch als (einfache oder schwache) *Markov-eigenschaft* bezeichnet wird, besagt, dass die durch  $X_n = i_n, \dots, X_0 = i_0$  bedingten endlich dimensionalen Verteilungen (und damit nach Satz 1.4.3 die Verteilung) des Prozesses ab Zeit  $n$  nur von  $i_n$ , nicht aber von  $n$  oder  $i_0$  bis  $i_{n-1}$  abhängen.

Für die folgenden Betrachtungen ist das sogenannte *Borel-Cantelli-Lemma* nützlich. In der Formulierung verwenden wir die folgende Schreibweise: Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $A_1, A_2, \dots \in \mathcal{F}$ . Dann sei

$$\begin{aligned} \{A_n \text{ u. o.}\} &:= \{\omega \in \Omega: \omega \in A_n \text{ für unendlich viele } n \in \mathbb{N}\} \\ &= \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n \\ &=: \limsup_{n \rightarrow \infty} A_n. \end{aligned} \quad (2.3.3)$$

Die Bezeichnung *lim sup* kommt daher, dass für eine reelle Zahlenfolge  $(a_n)_{n \in \mathbb{N}}$  gilt:

$$\left(-\infty, \limsup_{n \rightarrow \infty} a_n\right) \subset \limsup_{n \rightarrow \infty} (-\infty, a_n) \subset \limsup_{n \rightarrow \infty} (-\infty, a_n] \subset \left(-\infty, \limsup_{n \rightarrow \infty} a_n\right]. \quad (2.3.4)$$

Die Bezeichnung

$$\liminf_{n \rightarrow \infty} A_n := \{\omega: \omega \in A_n \text{ für alle bis auf endlich viele } n\} \quad (2.3.5)$$

werden wir nicht benötigen.

**Satz 2.3.12** (Borel-Cantelli-Lemmata). Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum, und seien  $A_1, A_2, \dots \in \mathcal{F}$ .

(i) Wenn  $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ , dann gilt  $\mathbb{P}(\{A_n \text{ u. o.}\}) = 0$ .

(ii) Wenn die  $A_n$  unabhängig sind und  $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$  gilt, dann gilt  $\mathbb{P}(\{A_n \text{ u. o.}\}) = 1$ .

**Beweis.**

(i)

$$\mathbb{P}(\{A_n \text{ u. o.}\}) = \mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\right) = \lim_{m \rightarrow \infty} \mathbb{P}\left(\bigcup_{n=m}^{\infty} A_n\right) \leq \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \mathbb{P}(A_n) = 0. \quad (2.3.6)$$



(ii) Es genügt zu zeigen, dass

$$\mathbb{P}\left(\left(\bigcup_{n=m}^{\infty} A_n\right)^c\right) = 0 \quad (2.3.7)$$

für alle  $m \in \mathbb{N}$  gilt, denn dann ist  $\mathbb{P}(\bigcup_{n=m}^{\infty} A_n) = 1$  für alle  $m \in \mathbb{N}$ , und somit haben wir  $\mathbb{P}(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n) = 1$ .

Nun gilt

$$\begin{aligned} \mathbb{P}\left(\left(\bigcup_{n=m}^{\infty} A_n\right)^c\right) &= \left(\bigcap_{n=m}^{\infty} A_n^c\right) = \lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcap_{n=m}^N A_n^c\right) = \lim_{N \rightarrow \infty} \prod_{n=m}^N \mathbb{P}(A_n^c) = \prod_{n=m}^{\infty} \mathbb{P}(A_n^c) \\ &= \prod_{n=m}^{\infty} [1 - \mathbb{P}(A_n)] \leq \prod_{n=m}^{\infty} \exp\{-\mathbb{P}(A_n)\} = \exp\left\{-\sum_{n=m}^{\infty} \mathbb{P}(A_n)\right\} = 0, \end{aligned} \quad (2.3.8)$$

wobei die Unabhängigkeit beim dritten Gleichheitszeichen benutzt wurde.

□

**Bemerkung 2.3.13.** Wir werden im folgenden nur den (einfachen) Teil (i) von Satz 2.3.12 benutzen. Man beachte, dass Satz 2.3.12 besagt, dass für unabhängige Ereignisse ein Null-Eins-Gesetz der folgenden Art gilt:  $\mathbb{P}(\{A_n \text{ u. o.}\})$  ist entweder Null oder Eins – andere Werte sind nicht möglich.

## 2.4 Rekurrenz und Transienz von Markovketten

Wir werden nun Rekurrenz(=Wiederkehr-)eigenschaften von Markovketten studieren. Wie zuvor sei  $P$  die Übergangsmatrix einer Markovkette mit Zustandsraum  $E$ . Für  $j \in E$  sei  $(\Omega, \mathcal{F}, \mathbb{P}_j)$  ein Wahrscheinlichkeitsraum und  $X_0, X_1, \dots$  eine darauf definierte Markovkette mit Zustandsraum  $E$ , Übergangsmatrix  $P$  und Startvektor  $a = \delta_j$ , wobei  $\delta_j(j) = 1$  und  $\delta_j(i) = 0$  für  $i \neq j$ . Mit anderen Worten,  $\mathbb{P}_j(X_0 = j) = 1$ , d. h., die Kette startet unter  $\mathbb{P}_j$  im Zustand  $j$ . Weiter sei

$$f_{ji} = \mathbb{P}_j\left(\bigcup_{n=1}^{\infty} \{X_n = i\}\right) \quad (2.4.1)$$

die Wahrscheinlichkeit, dass bei Start in  $j$  der Zustand  $i$  mindestens einmal irgendwann besucht wird.

**Satz 2.4.1** (Rekurrenz und Transienz).

(i) Wenn  $f_{jj} = 1$ , dann gelten

$$\mathbb{P}_j(\{X_n = j \text{ u. o.}\}) = 1 \quad \text{und} \quad \sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty. \quad (2.4.2)$$

In diesem Fall heißt  $j$  rekurrent.

(ii) Wenn  $f_{jj} < 1$ , dann gelten

$$\mathbb{P}_j(\{X_n = j \text{ u. o.}\}) = 0 \quad \text{und} \quad \sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty. \quad (2.4.3)$$

In diesem Fall heißt  $j$  transient.

**Bemerkung 2.4.2.** Die Zahl  $\sum_{n=1}^{\infty} p_{jj}^{(n)}$  ist die erwartete Anzahl von Besuchen in  $j$  nach dem Start, denn

$$\mathbb{E}_j \sum_{n=1}^{\infty} \mathbb{1}\{X_n = j\} = \sum_{n=1}^{\infty} \mathbb{E}_j \mathbb{1}\{X_n = j\} = \sum_{n=1}^{\infty} p_{jj}^{(n)}, \quad (2.4.4)$$

wobei  $\mathbb{E}_j$  den Erwartungswert bezüglich  $\mathbb{P}_j$  bezeichnet.

**Beweis von Satz 2.4.1.** Sei  $F_n := \{X_n = j, X_{n+k} \neq j \text{ für alle } k \in \mathbb{N}\}$  für  $n \in \mathbb{N}$ , und sei  $F_0 := \{X_k \neq j \text{ für alle } k \in \mathbb{N}\}$ .  $F_n$  ist also das Ereignis, dass zur Zeit  $n$  der Zustand  $j$  zum letzten Mal besucht wird. Es gilt  $\{X_n = j \text{ u. o.}\}^c = \bigcup_{n=0}^{\infty} F_n$ , also

$$1 - \mathbb{P}_j(\{X_n = j \text{ u. o.}\}) = \sum_{n=0}^{\infty} \mathbb{P}_j(F_n), \quad (2.4.5)$$

da die  $F_n$  disjunkt sind. Weiter gilt wegen der Markoveigenschaft

$$\mathbb{P}_j(F_n) = \mathbb{P}_j(X_{n+k} \neq j \text{ für alle } k \geq 1 \mid X_n = j) \mathbb{P}_j(X_n = j) = \mathbb{P}_j(F_0) p_{jj}^{(n)}, \quad (2.4.6)$$

und

$$\mathbb{P}_j(F_0) = 1 - f_{jj}. \quad (2.4.7)$$

(i) Für  $f_{jj} = 1$  folgt aus (2.4.7)  $\mathbb{P}_j(F_0) = 0$  und damit aus (2.4.6)  $\mathbb{P}_j(F_n) = 0$  für alle  $n \in \mathbb{N}$ . Aus (2.4.5) folgt nun  $\mathbb{P}_j(\{X_n = j \text{ u. o.}\}) = 1$ . Weiter gilt  $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$ , denn anderenfalls folgte aus Satz 2.3.12(i), dass  $\mathbb{P}_j(\{X_n = j \text{ u. o.}\}) = 0$ .

(ii) Für  $f_{jj} < 1$  folgt aus (2.4.7)  $\mathbb{P}_j(F_0) > 0$ , und mit (2.4.5) und (2.4.6) folgt

$$1 - \mathbb{P}_j(\{X_n = j \text{ u. o.}\}) = \sum_{n=0}^{\infty} \mathbb{P}_j(F_n) = \mathbb{P}_j(F_0) \left(1 + \sum_{n=1}^{\infty} p_{jj}^{(n)}\right).$$

Da die linke Seite durch Eins beschränkt ist und  $\mathbb{P}_j(F_0) > 0$  gilt, muss  $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$  sein. Aus Satz 2.3.12(i) folgt dann  $\mathbb{P}_j(\{X_n = j \text{ u. o.}\}) = 0$ .  $\square$

**Bemerkung 2.4.3.** Aus dem Beweis von Teil (ii) sieht man, dass im Fall  $f_{jj} < 1$  die Formel

$$1 = (1 - f_{jj}) \left(1 + \sum_{n=1}^{\infty} p_{jj}^{(n)}\right) \quad (2.4.8)$$

gilt. Die erwartete Anzahl von Besuchen in  $j$  ab dem Startzeitpunkt ist daher

$$1 + \sum_{n=1}^{\infty} p_{jj}^{(n)} = \frac{1}{1 - f_{jj}}. \quad (2.4.9)$$

**Bemerkung 2.4.4.** Man könnte die Resultate von Satz 2.4.1 anschaulich auch wie folgt begründen.

Wenn  $f_{jj} = 1$  ist, dann kehrt man sicher nach  $j$  zurück. Sobald man nach  $j$  zurückgekehrt ist, hat man wieder dieselbe Situation wie am Anfang, d. h. man wird  $j$  auch ein drittes Mal sicher besuchen usw. Also wird man  $j$  sogar unendlich oft besuchen. Die erwartete Anzahl von Besuchen in  $j$  ist natürlich erst recht unendlich. Somit hat man (i) gezeigt.

Ist andererseits  $f_{jj} < 1$ , so hat man bei jedem Besuch eine gewisse Chance – nämlich  $1 - f_{jj}$  – niemals mehr nach  $j$  zurückzukehren. Da man beliebig viele Versuche hat, wird dies sicher irgendwann passieren. Daher gilt  $\mathbb{P}_j(\{X_n = j \text{ u. o.}\}) = 0$ . Es ist klar, dass die Anzahl der Besuche in  $j$  geometrisch verteilt ist:

$$\mathbb{P}_j(k = \text{Anzahl der Besuche ohne Start in } j) = f_{jj}^k(1 - f_{jj}), \quad k \in \mathbb{N}_0.$$

Da der Erwartungswert einer solchen geometrischen Verteilung gleich  $f_{jj}(1 - f_{jj})^{-1}$  ist, ist die erwartete Anzahl von Besuchen in  $j$  (mit Start)  $1 + f_{jj}(1 - f_{jj})^{-1} = (1 - f_{jj})^{-1}$ , also insbesondere endlich, womit (ii) und die Formel (2.4.9) gezeigt ist.

Diese Argumente sind durchaus richtig, aber an einer Stelle lückenhaft, nämlich dort, wo argumentiert wird, dass nach der ersten Rückkehr in  $j$  der Prozess so weiterläuft als würde er (unabhängig) neu gestartet. Die Markoveigenschaft alleine sagt nicht, dass dies stimmt, denn sie ist nur für feste und nicht für zufällige Zeiten formuliert. Die erste Rückkehrzeit ist aber zufällig. Wir wollen nun zeigen, dass die Markoveigenschaft auch für gewisse solcher zufälligen Zeiten (“Stoppzeiten”) gilt. Diese Eigenschaft wird als *starke Markoveigenschaft* bezeichnet. Wir betonen, dass die starke Markoveigenschaft zwar für Markovketten gilt, nicht aber für beliebige “Markovprozesse”. Wir definieren zunächst den Begriff der Stoppzeit.

**Definition 2.4.5.** Sei  $(X_n)_{n \in \mathbb{N}_0}$  eine Markovkette mit Startverteilung  $a$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$ . Eine Abbildung  $\tau: \Omega \rightarrow \mathbb{N}_0 \cup \{\infty\}$  heißt *Stoppzeit* (bezüglich  $(X_n)_{n \in \mathbb{N}_0}$ ) wenn gilt:

$$\{\tau = n\} \in \sigma(X_0, \dots, X_n), \quad n \in \mathbb{N}_0. \quad (2.4.10)$$

Wir erinnern daran, dass die  $\sigma$ -Algebra  $\sigma(X_0, \dots, X_n)$  weiter oben unter dem Namen  $\mathcal{F}_n$  eingeführt worden ist. Die Eigenschaft in (2.4.10) bedeutet also, dass sich das Ereignis  $\{\tau = n\}$  als Vereinigung über Ereignisse der Form  $\{X_0 = i_0, \dots, X_n = i_n\}$  darstellen lässt oder – anschaulich – dass die Tatsache, ob  $\tau = n$  gilt, d. h. zur Zeit  $n$  gestoppt wird, nur von den Realisierungen von  $X_0, \dots, X_n$  (und nicht von zukünftigen  $X_{n+1}, X_{n+2}, \dots$ ) abhängt. Stoppzeiten (oder Stopppregeln) sind also solche, für deren Anwendung man keiner hellseherischen Fähigkeiten bedarf.

Beispiele für Stoppzeiten sind (mit einem Zustand  $i \in E$ )

- $\tau =$  Zeitpunkt des ersten Besuchs in  $i$
- $\tau =$  Zeitpunkt des fünften Besuchs in  $i$ .

Abgesehen von Spezialfällen ist aber

- $\tau =$  Zeitpunkt des letzten Besuchs in  $i$

keine Stoppzeit, da man ohne hellseherische Fähigkeiten nicht sicher sein kann, dass man nicht doch noch einmal nach  $i$  zurückkehrt (es sei denn,  $f_{ii}$  wäre Null).

Offensichtlich sind Stoppzeiten eine Verallgemeinerung fester (deterministischer) Zeiten, d. h. die konstante Abbildung  $\tau := n \in \mathbb{N}_0$  ist eine Stoppzeit. Zur Formulierung der starken Markoveigenschaft ist es nützlich, die  $\sigma$ -Algebra  $\mathcal{F}_\tau$  – die sogenannte  $\tau$ -Vergangenheit – als natürliche Verallgemeinerung von  $\mathcal{F}_n$  im Spezialfall  $\tau = n$  zu definieren. Dabei soll ein Ereignis  $F$  in  $\mathcal{F}_\tau$  liegen genau dann, wenn das Eintreffen oder Nichteintreffen von  $F$  aufgrund der Beobachtung  $X_0, \dots, X_\tau$  erkennbar ist. Wir definieren  $\mathcal{F}_\tau$  als die Menge der Teilmengen von  $\Omega$ , die sich als (notwendigerweise abzählbare) Vereinigung von Mengen der Form

$$\{\omega: \tau(\omega) = n, X_0(\omega) = i_0, \dots, X_n(\omega) = i_n\}$$

für  $n \in \mathbb{N}_0$ ,  $i_0, \dots, i_n \in E$  und einer Teilmenge von  $\{\omega: \tau(\omega) = \infty\}$  in  $\mathcal{F}$  schreiben lassen. Man beachte, dass insbesondere  $\{\omega: \tau(\omega) = n\} \in \mathcal{F}_\tau$  für alle  $n \in \mathbb{N}_0$  gilt. Man rechnet leicht nach, dass  $\mathcal{F}_\tau$  eine  $\sigma$ -Algebra ist.

**Satz 2.4.6** (Starke Markoveigenschaft). *Sei  $(X_n)_{n \in \mathbb{N}_0}$  eine Markovkette mit Übergangsmatrix  $P$  und Startverteilung  $a$ . Weiter sei  $\tau$  eine Stoppzeit für  $(X_n)_{n \in \mathbb{N}_0}$ . Dann gilt die starke Markoveigenschaft, d. h.*

$$\mathbb{P}(X_{\tau+1} = i_1, \dots, X_{\tau+k} = i_k \mid \{X_\tau = i_0\} \cap F \cap \{\tau < \infty\}) = p_{i_0 i_1} \dots p_{i_{k-1} i_k} \quad (2.4.11)$$

für alle  $F \in \mathcal{F}_\tau$ ,  $k \in \mathbb{N}$ ,  $i_0, \dots, i_k \in E$ , für die  $\mathbb{P}(\{X_\tau = i_0\} \cap F \cap \{\tau < \infty\}) > 0$ .

**Beweis.** Sei zunächst  $F_n \in \mathcal{F}_n$  mit den Eigenschaften  $F_n \subset \{\tau = n\}$  für ein  $n \in \mathbb{N}_0$  und  $\mathbb{P}(\{X_\tau = i_0\} \cap F_n) > 0$ . Dann folgt aus Satz 2.3.10

$$\begin{aligned} & \mathbb{P}(X_{\tau+1} = i_1, \dots, X_{\tau+k} = i_k \mid \{X_\tau = i_0\} \cap F_n) \\ &= \mathbb{P}(X_{n+1} = i_1, \dots, X_{n+k} = i_k \mid \{X_n = i_0\} \cap F_n) \\ &= p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k}. \end{aligned} \quad (2.4.12)$$

Für beliebiges  $F \in \mathcal{F}_\tau$  schreibe man

$$F = \bigcup_{n=0}^{\infty} (\{\tau = n\} \cap F) \cup (\{\tau = \infty\} \cap F). \quad (2.4.13)$$

Dann gilt  $F_n := \{\tau = n\} \cap F \in \mathcal{F}_n$  nach Definition von  $\mathcal{F}_\tau$  und die Behauptung folgt leicht.  $\square$

**Bemerkung 2.4.7.** Satz 2.4.6 besagt, dass die beiden bedingten endlich dimensionalen Verteilungen der Prozesse  $(X_{\tau+k})_{k \in \mathbb{N}_0}$  gegeben  $\tau = n$ ,  $X_0 = i_0, \dots, X_n = i_n$  und gegeben  $X_\tau = i_n$  gleich sind. Mit Satz 1.4.3 sind dann auch ihre Verteilungen gleich.

**Beispiel 2.4.8.** Wir wollen untersuchen, ob die symmetrische Irrfahrt auf  $\mathbb{Z}^d$  rekurrent ist, genauer, ob die Null ein rekurrenter Zustand ist.

Zunächst behandeln wir den Fall  $d = 1$ ; wir betrachten allgemeiner die asymmetrische Irrfahrt auf  $\mathbb{Z}$ , d. h., mit einem Parameter  $p \in (0, 1)$  und  $q = 1 - p$  ist die Übergangsmatrix  $P = (p_{ij})_{i,j \in \mathbb{Z}}$  gegeben durch

$$p_{ij} = \begin{cases} p, & \text{falls } j = i + 1, \\ q, & \text{falls } j = i - 1, \\ 0 & \text{sonst.} \end{cases} \quad (2.4.14)$$

Es gilt für alle  $n \in \mathbb{N}$

$$\mathbb{P}_0(X_{2n} = 0) = \binom{2n}{n} p^n q^n \quad \text{und} \quad \mathbb{P}_0(X_{2n+1} = 0) = 0, \quad (2.4.15)$$

also

$$\sum_{k=1}^{\infty} p_{00}^{(k)} = \sum_{n=1}^{\infty} \binom{2n}{n} p^n q^n = \sum_{n=1}^{\infty} \frac{(2n)!}{n!^2} p^n q^n. \quad (2.4.16)$$

Wir wollen untersuchen, ob die Reihe konvergiert oder divergiert, um mit Satz 2.4.1 zu entscheiden, ob 0 rekurrent oder transient ist. Hierzu benutzen wir die bekannte *Stirlingformel* (Beweis z. B. in [Fe68] oder [No97])

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad (2.4.17)$$

wobei  $\sim$  bedeutet, dass der Quotient beider Seiten mit  $n \rightarrow \infty$  gegen 1 konvergiert. Somit gilt

$$\frac{(2n)!}{n!^2} p^n q^n \sim \frac{\sqrt{2\pi 2n} (2n)^{2n} e^{-2n} p^n q^n}{2\pi n^{2n+1} e^{-2n}} = \frac{(4pq)^n}{\sqrt{\pi n}}. \quad (2.4.18)$$

Setzt man dies in die Reihe in (2.4.16) ein, so sieht man, dass sie genau dann divergiert, wenn  $4pq = 1$  ist, d. h. wenn  $p = q = 1/2$  ist. (Man mache sich klar, dass das Ersetzen von asymptotisch gleichen Ausdrücken an der Konvergenz oder Divergenz der Reihe nichts ändert!) Somit haben wir gesehen:

- Die eindimensionale symmetrische Irrfahrt ist rekurrent.
- Die eindimensionale asymmetrische Irrfahrt (d. h. mit  $p \neq 1/2$ ) ist transient.

Genau genommen müssten wir sagen: Der Zustand 0 ist rekurrent bzw. transient. Es ist aber offensichtlich, dass wegen der räumlichen Homogenität diese Eigenschaft allen Zuständen zukommt, was die obige Sprechweise rechtfertigt.

Im Fall  $d \geq 2$  zeigt eine ähnliche, aber aufwändigere Rechnung: Die symmetrische Irrfahrt ist

- rekurrent für  $d = 2$ ,
- transient für  $d \geq 3$ .

(vgl. [Fe68, S. 360 f] oder [KT75, S. 67 ff]).

## 2.5 Klassifikation der Zustände

Wir kehren nun wieder zu allgemeinen Fragen über Markovketten zurück. Ist bei einer Markovkette ein Zustand rekurrent, dann ist immer noch die Frage nach der Verteilung der ersten Rückkehrzeit interessant und dabei speziell die Frage, ob diese Zeit einen endlichen oder unendlichen Erwartungswert hat. Da das Nachprüfen von Kriterien für Rekurrenz oder auch für die Endlichkeit des Erwartungswertes der Rückkehrzeit mit einem gewissen Aufwand verbunden ist, stellt sich die Frage, ob man wirklich für jeden Zustand einer Markovkette diese Kriterien getrennt abprüfen muss. Glücklicherweise ist dies nicht so. Man kann nämlich die Zustände einer Markovkette recht einfach so in Klassen einteilen, dass alle Zustände einer Klasse dieselben Rekurrenzeigenschaften haben. Wir werden dies im Folgenden präzisieren und beweisen.

Im gesamten Rest dieses Kapitels sei  $E$  eine endliche oder höchstens abzählbar unendliche Menge, und  $P = (p_{ij})_{i,j \in E}$  sei eine stochastische Matrix auf  $E$ . Mit  $\mathbb{X} = (X_k)_{k \in \mathbb{N}_0}$  bezeichnen wir eine Markovkette auf  $E$  mit Übergangsmatrix  $P$ . Die Kette sei auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  definiert. Falls die Markovkette im Zustand  $i \in E$  startet, schreiben wir  $\mathbb{P}_i$  statt  $\mathbb{P}$ .

Für einen rekurrenten Zustand  $j \in E$  sei  $R_k^{(j)}$  der (zufällige) Zeitpunkt der  $k$ -ten Rückkehr in den Zustand  $j$ , wobei wir  $R_0^{(j)} = 0$  setzen. Ferner sei  $T_k^{(j)} = R_k^{(j)} - R_{k-1}^{(j)}$  für  $k \in \mathbb{N}$  die

Zeitdauer zwischen dem  $(k-1)$ -ten und  $k$ -ten Besuch in  $j$ . Wegen der Rekurrenz von  $j$  sind die  $R_k^{(j)}$  alle fast sicher endlich und damit die  $T_k^{(j)}$  wohldefiniert. Es ist plausibel, dass bei Start in  $j$  (d. h. unter  $\mathbb{P}_j$ ) die Variablen  $T_1^{(j)}, T_2^{(j)}, \dots$  u. i. v. sind, denn nach der ersten Rückkehr nach  $j$  verhält sich der Prozess bezüglich seiner Verteilung ja genauso wie am Anfang bei Start in  $j$ , und was nach  $R_1^{(j)}$  passiert, ist unabhängig von dem, was vor  $R_1^{(j)}$  passierte. Dies beweisen wir nun formal.

**Satz 2.5.1.** *Wenn  $j$  rekurrent ist, dann ist  $(T_k^{(j)})_{k \in \mathbb{N}}$  eine Folge von u. i. v. Zufallsvariablen auf  $(\Omega, \mathcal{F}, \mathbb{P}_j)$ .*

**Beweis.** Zunächst zeigen wir, dass  $R_1^{(j)}, R_2^{(j)}, \dots$  Stoppzeiten sind. Dann folgt die Behauptung aus der starken Markoveigenschaft.

Nun gilt

$$\begin{aligned} \{R_1^{(j)} = n\} &= \{X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j\}, \\ \{R_2^{(j)} = n\} &= \{X_n = j, \text{ es gibt genau ein } k \in \{1, \dots, n-1\} \text{ mit } X_k = j\}, \end{aligned} \quad (2.5.1)$$

und so weiter. Diese Ereignisse liegen offensichtlich in  $\sigma(X_1, \dots, X_n)$ , denn ihr Eintreten oder Nichteintreten kann man aufgrund der Realisierung von  $X_0$  bis  $X_n$  erkennen. Daher sind  $R_1^{(j)}, R_2^{(j)}, \dots$  Stoppzeiten.  $\square$

Mit  $\mathbb{E}_j$  bezeichnen wir den Erwartungswert bezüglich  $\mathbb{P}_j$ , d. h. bei Start in  $j$ .

**Definition 2.5.2.** Seien  $i, j \in E$ .

- a) Sei  $j$  rekurrent. Dann heißt  $j$  *positiv rekurrent*, wenn  $\mathbb{E}_j T_1^{(j)} < \infty$  und *nullrekurrent*, wenn  $\mathbb{E}_j T_1^{(j)} = \infty$ . (Die Begriffe “positiv” und “null” beziehen sich auf den Kehrwert von  $\mathbb{E}_j T_1^{(j)}$ .)
- b) Der Zustand  $j$  heißt *erreichbar* von  $i$ , und wir schreiben  $i \rightarrow j$ , wenn ein  $n \in \mathbb{N}$  existiert mit  $p_{ij}^{(n)} > 0$ .
- c) Wenn  $i \rightarrow j$  und  $j \rightarrow i$ , dann sagt man, dass  $i$  und  $j$  *kommunizieren*, und wir schreiben  $i \leftrightarrow j$ .
- d) Der Zustand  $j$  heißt *absorbierend*, wenn aus  $j \rightarrow i$  folgt, dass  $i = j$ . Dies ist äquivalent zu  $p_{jj} = 1$  und zu  $p_{jj}^{(n)} = 1$  für alle  $n \in \mathbb{N}$ .
- e) Eine nichtleere Menge  $M \subset E$  heißt *kommunizierende Klasse*, wenn
  - (i)  $i \leftrightarrow j$  für alle  $i, j \in M$ ,
  - (ii) wenn  $i \in M, j \in E$  und  $i \rightarrow j$ , dann folgt  $j \in M$ .
- f) Ist  $j$  Element einer kommunizierenden Klasse, dann heißt  $j$  *wesentlich*, sonst *unwesentlich*.
- g) Ist  $E$  selbst eine kommunizierende Klasse, dann heißt die Markovkette *irreduzibel*.

**Bemerkung 2.5.3.** Jeder Zustand  $j$  liegt in höchstens einer kommunizierenden Klasse. Der einzige Kandidat ist offenbar  $M = \{i \in E : j \rightarrow i\}$ .

Man sieht leicht, dass eine Markovkette (oder ihre Übergangsmatrix  $P$ ) genau dann irreduzibel ist, falls für jedes Paar  $i, j \in E$  ein  $n \in \mathbb{N}$  existiert mit  $p_{ij}^{(n)} > 0$ .

**Satz 2.5.4.** “ $\leftrightarrow$ ” ist eine Äquivalenzrelation auf der Menge der wesentlichen Zustände. Die zugehörigen Äquivalenzklassen sind die kommunizierenden Klassen.

**Beweis.** “ $\leftrightarrow$ ” ist offensichtlich reflexiv und symmetrisch auf der Menge der wesentlichen Zustände. Wir zeigen, dass “ $\leftrightarrow$ ” auch transitiv ist:

Es gelte  $i \leftrightarrow j$  und  $j \leftrightarrow k$ . Dann existieren  $n_1, n_2 \in \mathbb{N}$  mit  $p_{ij}^{(n_1)} > 0$  und  $p_{jk}^{(n_2)} > 0$ . Daher gilt

$$p_{ik}^{(n_1+n_2)} = \sum_{l \in E} p_{il}^{(n_1)} p_{lk}^{(n_2)} \geq p_{ij}^{(n_1)} p_{jk}^{(n_2)} > 0, \tag{2.5.2}$$

d. h.  $i \rightarrow k$ . Ebenso folgt  $k \rightarrow i$ . Die zweite Aussage des Satzes ist klar.

**Bemerkung 2.5.5.** Für unwesentliche Zustände gilt nicht immer  $i \leftrightarrow i$ , z. B. dann nicht, wenn  $p_{ji} = 0$  für alle  $j \in E$  gilt. Gelegentlich wird auch auf der Menge *aller* Zustände eine Äquivalenzrelation eingeführt. Um die Reflexivität  $i \leftrightarrow i$  auch für unwesentliche Zustände  $i$  zu garantieren, definiert man dann “ $i \rightarrow j$ ”, wenn  $p_{ij}^{(n)} > 0$  für ein  $n \in \mathbb{N}_0$  (statt nur  $n \in \mathbb{N}$ ) ist.

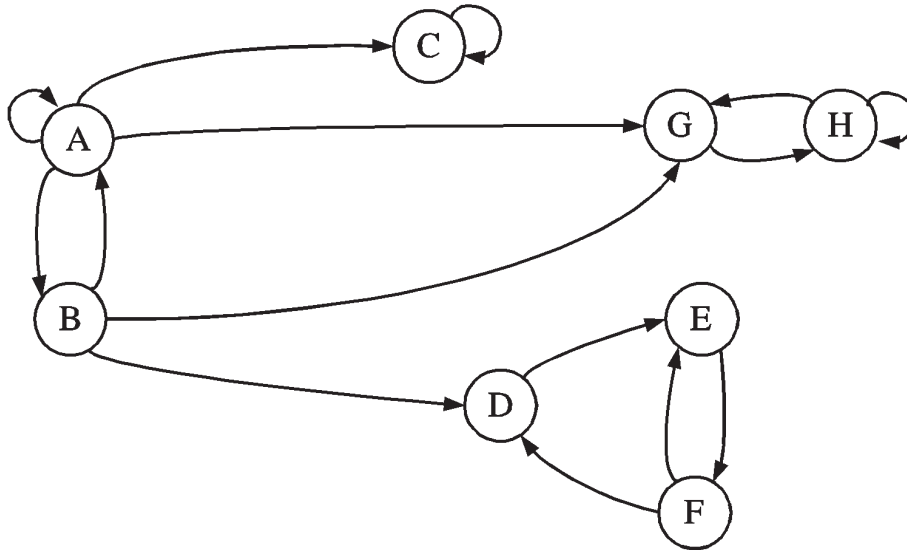


Abbildung 2.5.1: Beispiel einer Markovkette

**Beispiel 2.5.6.**  $\{C\}$ ,  $\{G, H\}$ ,  $\{D, E, F\}$  sind kommunizierende Klassen.  $A$  und  $B$  sind unwesentlich, aber  $A \leftrightarrow B$ .  $C$  ist absorbierend.

Wir zeigen im weiteren, dass Rekurrenz, Transienz, positive Rekurrenz und die noch zu definierende *Periode* von Zuständen Klasseeigenschaften sind, d. h., wenn  $i$  und  $j$  in derselben (kommunizierenden) Klasse sind, dann sind beide rekurrent oder beide transient usw.

**Satz 2.5.7.** Sei  $i \in E$  rekurrent und  $i \rightarrow j$ , dann gilt  $j \rightarrow i$  und  $j$  ist rekurrent.

**Beweis.** Um  $j \rightarrow i$  zu zeigen, führen wir einen Widerspruchsbeweis. Angenommen, es gilt nicht  $j \rightarrow i$ , d. h.  $p_{ji}^{(n)} = 0$  für alle  $n \in \mathbb{N}$ . Wähle  $n_0 \in \mathbb{N}$  mit  $p_{ij}^{(n_0)} > 0$ . Dann gilt

$$\begin{aligned} 0 &= \mathbb{P}_i(\{X_n = i \text{ nur endlich oft}\}) \\ &\geq \mathbb{P}_i(X_{n_0} = j, X_{n_0+1} \neq i, X_{n_0+2} \neq i, \dots) \\ &\geq p_{ij}^{(n_0)} \mathbb{P}_j(X_1 \neq i, X_2 \neq i, \dots) \\ &= p_{ij}^{(n_0)} > 0, \end{aligned} \tag{2.5.3}$$

was ein Widerspruch ist. Also gilt  $j \rightarrow i$ .

Nun zeigen wir die Rekurrenz von  $j$ . Seien  $r$  und  $s \in \mathbb{N}$  gewählt mit  $p_{ji}^{(r)} > 0$  und  $p_{ij}^{(s)} > 0$ . Dann gilt  $p_{jj}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ii}^{(n)} p_{ij}^{(s)}$  für jedes  $n \in \mathbb{N}$ . Also folgt

$$\sum_{n=1}^{\infty} p_{jj}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ij}^{(s)} \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty, \tag{2.5.4}$$

d. h.  $j$  ist rekurrent. □

**Korollar 2.5.8.** *Unwesentliche Zustände sind transient. Äquivalent: rekurrente Zustände sind wesentlich.*

**Beweis.** Sei  $i$  rekurrent, und setze  $C_i = \{j \in E : i \rightarrow j\}$ . Nach Satz 2.5.7 ist  $C_i$  eine kommunizierende Klasse, d. h.  $i$  ist wesentlich. □

**Korollar 2.5.9.** *Rekurrenz und Transienz sind Klasseneigenschaften.*

**Beweis.** Folgt sofort aus Satz 2.5.7. □

**Bemerkung 2.5.10.** Es gibt Markovketten ohne wesentliche Zustände, z. B.:  $E = \mathbb{N}$ ,  $p_{i,i+1} = 1$  für  $i \in \mathbb{N}$ ,  $p_{ij} = 0$  sonst.

**Definition 2.5.11.** Sei  $j \in E$  mit  $f_{jj} > 0$ , und sei  $d$  die größte natürliche Zahl, so dass

$$\sum_{k=1}^{\infty} \mathbb{P}_j(T_1^{(j)} = kd) + \mathbb{P}_j(T_1^{(j)} = \infty) = 1, \tag{2.5.5}$$

also  $d = \text{g.g.T.}\{n : p_{jj}^{(n)} > 0\}$ . Dann heißt  $d$  *Periode* von  $j$ .  $j$  heißt *periodisch*, wenn  $d > 1$  und *aperiodisch*, wenn  $d = 1$  ist.

**Beispiel 2.5.12.** Bei der eindimensionalen symmetrischen Irrfahrt haben alle Zustände Periode 2.

**Satz 2.5.13.** *Wenn  $i \leftrightarrow j$ , dann sind  $i$  und  $j$  beide transient oder beide rekurrent und  $i$  und  $j$  haben dieselbe Periode. Insbesondere sind Rekurrenz, Transienz und die Periode Klasseneigenschaften.*



**Beweis.** Die erste Aussage folgt aus Satz 2.5.7. Wir zeigen die zweite Aussage: Sei  $i \leftrightarrow j$ . Dann gilt  $f_{ii} > 0$  und  $f_{jj} > 0$ , und die Perioden  $d_i$  bzw.  $d_j$  von  $i$  bzw.  $j$  sind daher definiert. Seien  $n_1$  und  $n_2 \in \mathbb{N}$  gewählt mit  $p_{ij}^{(n_1)} > 0$  und  $p_{ji}^{(n_2)} > 0$ . Nun ist  $n_1 + n_2$  offenbar ein ganzzahliges Vielfaches von  $d_i$  und

$$p_{ii}^{(n_1+n_2)} \geq p_{ij}^{(n_1)} p_{jj}^{(n_2)} p_{ji}^{(n_2)}, \quad (2.5.6)$$

d. h.  $p_{jj}^{(n)} = 0$ , wenn  $n$  kein Vielfaches von  $d_i$  ist. Somit gilt  $d_i \leq d_j$ . Entsprechend folgt  $d_i \geq d_j$  und somit  $d_i = d_j$ .  $\square$

**Bemerkung 2.5.14.** Die Tatsache, dass auch die positive Rekurrenz eine Klasseneigenschaft ist, formulieren wir in Satz 2.6.14.

## 2.6 Grenzwertsätze und invariante Verteilungen

Wir betrachten zwei Arten von Grenzwertsätzen für Markovketten: solche für die  $n$ -schrittigen Übergangswahrscheinlichkeiten und solche für die Aufenthaltshäufigkeiten  $N_n(j)$  im Zustand  $j$  bis zur Zeit  $n$ , jeweils für  $n \rightarrow \infty$ . Zunächst erledigen wir die erste Frage für transiente Zustände.

**Satz 2.6.1.** *Ist  $j \in E$  transient, dann gilt  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$  für alle  $i \in E$ .*

**Beweis.** Wenn  $j \in E$  transient ist, gilt  $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$  nach Satz 2.4.1 und daher insbesondere  $\lim_{n \rightarrow \infty} p_{jj}^{(n)} = 0$ . Sei  $i \in E$  und

$$f_{ij}^{(k)} = \mathbb{P}_i(X_1 \neq j, \dots, X_{k-1} \neq j, X_k = j) \quad (2.6.1)$$

die Wahrscheinlichkeit,  $j$  von  $i$  aus nach  $k$  Schritten erstmalig zu besuchen. Es gilt  $\sum_{k=1}^{\infty} f_{ij}^{(k)} = f_{ij} \leq 1$ , vergleiche (2.4.1). Sei  $i \neq j$ . Dann gilt

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = \sum_{n=1}^{\infty} \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} \sum_{n=k}^{\infty} p_{jj}^{(n-k)} = f_{ij} \sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty. \quad (2.6.2)$$

Somit gilt  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$ .  $\square$

**Lemma 2.6.2.** *Sei  $C$  eine aperiodische Klasse, und seien  $i, j \in C$ . Dann existiert ein  $n_0 = n_0(i, j) \in \mathbb{N}$ , so dass  $p_{ij}^{(n)} > 0$  für alle  $n \geq n_0$  gilt.*

**Beweis.** Sei  $j \in C$  und  $A := \{n \in \mathbb{N} : p_{jj}^{(n)} > 0\}$ . Da  $j$  wesentlich ist, folgt  $A \neq \emptyset$ . Wegen der Aperiodizität von  $j$  folgt die Existenz von  $r \in \mathbb{N}$  und  $a_1, \dots, a_r \in A$ , so dass  $\text{ggT}(a_1, \dots, a_r) = 1$ . Mit dem euklidischen Algorithmus lassen sich  $c_1, \dots, c_r \in \mathbb{Z}$  bestimmen, so dass  $c_1 a_1 + \dots + c_r a_r = 1$ . Sei  $s := \sum_{k=1}^r a_k$  und  $n = sx + y$  mit  $n \in \mathbb{N}$ ,  $x \in \mathbb{N}_0$ , und  $0 \leq y < s$ . Dann gilt  $n = \sum_{k=1}^r (x + yc_k) a_k$ . Sei  $x_0 := (s-1) \max\{|c_1|, \dots, |c_r|\}$ . Für alle  $n \geq \tilde{n}_0 := sx_0$  gilt dann  $s_k := x + yc_k \geq 0$  für alle  $k$  und

$$p_{jj}^{(n)} = p_{jj}^{(\sum_{k=1}^r s_k a_k)} \geq \prod_{k=1}^r p_{jj}^{(s_k a_k)} \geq \prod_{k=1}^r (p_{jj}^{(a_k)})^{s_k} > 0, \quad (2.6.3)$$

da  $a_1, \dots, a_r \in A$ . Ist  $i \in C$ , dann existiert  $m \in \mathbb{N}$ , so dass  $p_{ij}^{(m)} > 0$  und somit

$$p_{ij}^{(m+n)} \geq p_{ij}^{(m)} p_{jj}^{(n)} > 0 \quad (2.6.4)$$

für alle  $n \geq \tilde{n}_0$ , also  $p_{ij}^{(n)} > 0$  für alle  $n \geq n_0 := m + \tilde{n}_0$ .  $\square$

**Satz 2.6.3.** Sei  $C$  eine rekurrente Klasse,  $i, k \in C$  und

$$\bar{\pi}_k^{(i)} = \mathbb{E}_i \left( \sum_{n=1}^{\infty} \mathbb{1}\{X_n = k, X_{n-1} \neq i, \dots, X_1 \neq i\} \right) \quad (2.6.5)$$

die erwartete Zahl von Besuchen in  $k$  vor der Rückkehr nach  $i$  bei Start in  $i$ . Dann gelten

$$0 < \bar{\pi}_k^{(i)} < \infty \quad \text{und} \quad \bar{\pi}_k^{(i)} = \sum_{j \in C} \bar{\pi}_j^{(i)} p_{jk}. \quad (2.6.6)$$

**Bemerkung 2.6.4.** Schreibt man  $\bar{\pi}^{(i)} = (\bar{\pi}_j^{(i)})_{j \in C}$  als Zeilenvektor und setzt  $\bar{\pi}_j^{(i)} := 0$  für  $j \in E \setminus C$ , so lautet (2.6.6) in Matrixform:  $\bar{\pi}^{(i)} = \bar{\pi}^{(i)} P$ .

**Beweis von Satz 2.6.3.** Wir zeigen zuerst die zweite Behauptung in (2.6.6). Man beachte, dass  $\bar{\pi}_i^{(i)} = 1$  gilt und definiere  $\infty \times 0 = 0$ . Es gilt

$$\begin{aligned} \sum_{j \in C} \bar{\pi}_j^{(i)} p_{jk} &= \sum_{n=1}^{\infty} \sum_{j \in C \setminus \{i\}} \mathbb{P}_i(X_n = j, X_{n-1} \neq i, \dots, X_1 \neq i) p_{jk} + p_{ik} \\ &= \sum_{n=1}^{\infty} \mathbb{P}_i(X_{n+1} = k, X_n \neq i, \dots, X_1 \neq i) + \mathbb{P}_i(X_1 = k) \\ &= \bar{\pi}_k^{(i)}, \end{aligned} \quad (2.6.7)$$

wobei wir beim vorletzten Gleichheitszeichen die Markov-Eigenschaft benutzt haben.

Nun zeigen wir  $0 < \bar{\pi}_k^{(i)} < \infty$  für  $k \neq i$ . Wegen der Rekurrenz existiert ein  $n \in \mathbb{N}$  mit  $p_{ki}^{(n)} > 0$ . Nach dem eben Gezeigten gilt  $\bar{\pi}^{(i)} = \bar{\pi}^{(i)} P$ , also auch  $\bar{\pi}^{(i)} = \bar{\pi}^{(i)} P^n$  für alle  $n \in \mathbb{N}$ . Insbesondere gilt

$$1 = \bar{\pi}_i^{(i)} = \sum_{j \in C} \bar{\pi}_j^{(i)} p_{ji}^{(n)}, \quad (2.6.8)$$

also folgt  $\bar{\pi}_k^{(i)} < \infty$ , da alle Summanden  $\geq 0$  sind.

Weiter gibt es ein  $m \in \mathbb{N}$  mit  $p_{ik}^{(m)} > 0$  und somit

$$\bar{\pi}_k^{(i)} = \sum_{j \in C} \bar{\pi}_j^{(i)} p_{jk}^{(m)} \geq \bar{\pi}_i^{(i)} p_{ik}^{(m)} = p_{ik}^{(m)} > 0. \quad (2.6.9)$$

$\square$

**Bemerkung 2.6.5.** Satz 2.6.3 dient uns hauptsächlich als Hilfssatz zum Beweis von Grenzwertsätzen, ist aber auch für sich genommen interessant. Um die  $\bar{\pi}_k^{(i)}$  zu berechnen, muss man

das lineare Gleichungssystem in (2.6.6) mit der Normierungsbedingung  $\bar{\pi}_i^{(i)} = 1$  und Nebenbedingungen  $0 < \bar{\pi}_k^{(i)} < \infty$  für  $k \in C$  lösen. Wir werden in Satz 2.6.8 sehen, dass die Lösung eindeutig ist. Man beachte, dass sich (2.6.6) auch durch Simulation (approximativ) lösen lässt: Man simuliere eine Markovkette mit Start in  $i$  bis zur ersten Rückkehr nach  $i$  und merke sich die Zahl der Besuche in allen anderen Zuständen. Dies wiederhole man oft (z. B. 1000 Mal) mit unabhängigen Zufallszahlen. Die Mittelwerte der Zahl der Besuche in  $k \in C$  über die Simulationsläufe ist eine Näherung für  $\bar{\pi}_k^{(i)}$ .

Man beachte weiterhin, dass

$$\sum_{k \in C} \bar{\pi}_k^{(i)} = \mathbb{E}_i T_1^{(i)} \tag{2.6.10}$$

gilt, insbesondere also  $i$  positiv rekurrent ist genau dann, wenn  $\sum_{k \in C} \bar{\pi}_k^{(i)} < \infty$ .

**Definition 2.6.6.** Eine Lösung  $\pi$  von

$$\pi = \pi P, \quad \pi = (\pi_i)_{i \in E} \in [0, \infty]^E, \tag{2.6.11}$$

heißt *invariantes Maß* von  $P$  (oder der Markovkette). Gilt zusätzlich  $\sum_{i \in E} \pi_i = 1$ , dann heißt  $\pi$  *invariantes Wahrscheinlichkeitsmaß* oder *invariante Verteilung* von  $P$ .

**Bemerkung 2.6.7.** (i) Ein invariantes Maß  $\pi \neq 0$  mit  $\pi_k < \infty$  für alle  $k \in E$  ist ein Linkseigenvektor von  $P$  zum Eigenwert 1.

(ii) Der Vektor  $(\bar{\pi}_k^{(i)})_{k \in E}$  aus Satz 2.6.3 ist für jedes  $i \in C$  ein invariantes Maß von  $P$ , wenn man  $\bar{\pi}_k^{(i)} = 0$  für  $k \notin C$  definiert.

(iii) Ist  $\pi$  ein invariantes Wahrscheinlichkeitsmaß, dann gilt für die speziell gewählte Startverteilung  $a = \pi$

$$\mathbb{P}(X_n = j) = (\pi P^n)_j = \pi_j, \quad j \in E, n \in \mathbb{N}_0. \tag{2.6.12}$$

Von dieser Eigenschaft her kommt die Bezeichnung “invariant”. Man zeigt leicht, dass dann automatisch sogar

$$\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = \mathbb{P}(X_1 = i_0, \dots, X_{n+1} = i_n), \quad n \in \mathbb{N}_0, i_0, \dots, i_n \in E, \tag{2.6.13}$$

gilt. Man sagt,  $(X_n)_{n \in \mathbb{N}_0}$  ist ein *stationärer Prozess*.

**Satz 2.6.8.** Ist  $C$  eine rekurrente Klasse, dann existiert bis auf konstante Vielfache ( $\in [0, \infty]$ ) genau ein invariantes Maß  $\pi$  auf  $C$ , genauer: es existiert ein  $\pi \in [0, \infty]^E$  mit  $\pi_k = \sum_{j \in C} \pi_j p_{jk}$  für alle  $k \in C$  und  $\pi_k = 0$  für alle  $k \in E \setminus C$  und  $\pi_k > 0$  für alle  $k \in C$  und für jedes invariante Maß  $\tilde{\pi}$  mit  $\tilde{\pi}_k = 0$  für alle  $k \in E \setminus C$  existiert ein  $c \in [0, \infty]$ , so dass  $\tilde{\pi}_k = c\pi_k$  für alle  $k \in E$  gilt (mit der üblichen Konvention  $\infty \times 0 = 0$ ).

**Beweis.** Die Existenz von  $\pi$  folgt aus Satz 2.6.3. Zu zeigen bleibt die Eindeutigkeit. Sei also  $\pi$  invariant auf  $C$ . Wenn  $\pi \equiv \infty$  auf  $C$  gilt, so ist  $\pi$  ein konstantes (nämlich unendliches) Vielfaches jedes  $\bar{\pi}^{(i)}$ , also können wir annehmen, dass ein  $i \in C$  existiert mit  $\pi_i < \infty$ . Dann gilt für alle  $k \in C$

$$\begin{aligned} \pi_k &= \sum_{j \in C} \pi_j p_{jk} = \pi_i p_{ik} + \sum_{j \in C \setminus \{i\}} \pi_j p_{jk} = \pi_i p_{ik} + \sum_{j \in C \setminus \{i\}} \left( \pi_i p_{ij} + \sum_{j_1 \in C \setminus \{i\}} \pi_{j_1} p_{j_1 j} \right) p_{jk} \\ &= \pi_i p_{ik} + \pi_i \sum_{j \in C \setminus \{i\}} p_{ij} p_{jk} + \sum_{j \in C \setminus \{i\}} \sum_{j_1 \in C \setminus \{i\}} \pi_{j_1} p_{j_1 j} p_{jk} = \dots \end{aligned} \tag{2.6.14}$$

Also gilt

$$\pi_k \geq \pi_i \sum_{m=1}^{\infty} \mathbb{P}_i(X_1 \neq i, X_2 \neq i, \dots, X_{m-1} \neq i, X_m = k) = \pi_i \bar{\pi}_k^{(i)}, \quad (2.6.15)$$

und somit

$$\pi_i = \sum_{k \in C} \pi_k p_{ki}^{(n)} \geq \sum_{k \in C} \pi_i \bar{\pi}_k^{(i)} p_{ki}^{(n)} = \pi_i \bar{\pi}_i^{(i)} = \pi_i, \quad n \in \mathbb{N}. \quad (2.6.16)$$

Da zu jedem  $k \in C$  ein  $n_k \in \mathbb{N}$  existiert mit  $p_{ki}^{(n_k)} > 0$  und da  $\pi_i < \infty$  ist, gilt also in (2.6.15) Gleichheit für alle  $k \in C$ , also ist

$$\pi_k = \pi_i \bar{\pi}_k^{(i)}, \quad k \in C, \quad (2.6.17)$$

und somit ist  $\pi$  konstantes Vielfaches (nämlich  $\pi_i$ -faches) von  $\bar{\pi}^{(i)}$ .  $\square$

**Bemerkung 2.6.9.** Aus Satz 2.6.8 folgt insbesondere, dass sich die invarianten Maße  $\bar{\pi}^{(i)}$  aus Satz 2.6.3 für verschiedene  $i$  nur um konstante Vielfache unterscheiden.

**Satz 2.6.10.** *Wenn es eine invariante Verteilung  $\pi$  gibt, dann sind alle  $j \in E$  mit  $\pi_j > 0$  rekurrent.*

**Beweis.**

$$\begin{aligned} \infty &= \sum_{n=0}^{\infty} \pi_j = \sum_{n=0}^{\infty} \sum_{k \in E} \pi_k p_{kj}^{(n)} = \sum_{k \in E \setminus \{j\}} \pi_k f_{kj} \sum_{n=0}^{\infty} p_{jj}^{(n)} + \pi_j \sum_{n=0}^{\infty} p_{jj}^{(n)} \\ &\leq \left( \sum_{k \in E} \pi_k \right) \sum_{n=0}^{\infty} p_{jj}^{(n)} = \sum_{n=0}^{\infty} p_{jj}^{(n)}, \end{aligned} \quad (2.6.18)$$

wobei das dritte Gleichheitszeichen wie im Beweis von Satz 2.6.1 folgt. Also ist  $j$  rekurrent nach Satz 2.4.1.  $\square$

**Korollar 2.6.11.** *Auf einer transienten Klasse  $C$  gibt es keine invariante Verteilung.*

**Bemerkung 2.6.12.** Es kann aber auf einer transienten Klasse  $C$  durchaus invariante Maße geben, wie das Beispiel der asymmetrischen Irrfahrt auf  $\mathbb{Z}$  zeigt. Hier ist  $\pi_i = 1$ ,  $i \in \mathbb{Z}$  ein invariantes Maß, aber auch  $\pi_i = (p/q)^i$ ,  $i \in \mathbb{Z}$ , d. h. auf transienten Klassen sind nicht notwendigerweise alle invarianten Maße proportional.

**Satz 2.6.13.** *Sei  $\pi$  eine invariante Verteilung auf einer (notwendigerweise rekurrenten) Klasse  $C$ . Dann gilt*

$$\pi_i = \frac{1}{\mathbb{E}_i T_1^{(i)}} \in (0, 1], \quad i \in C. \quad (2.6.19)$$

**Beweis.** Wegen Satz 2.6.8 ist  $\pi$  eindeutig, und wegen Satz 2.6.3 gilt  $\pi_k = \alpha \bar{\pi}_k^{(i)}$  für alle  $k \in C$  mit einem festen  $i \in C$  und einem  $\alpha > 0$ . Da  $\sum_{k \in C} \bar{\pi}_k^{(i)} = \mathbb{E}_i T_1^{(i)}$  und  $\sum_{k \in C} \pi_k = 1$ , folgt  $\alpha = (\mathbb{E}_i T_1^{(i)})^{-1}$ , und damit die Behauptung.  $\square$

Nach diesen Vorbereitungen erhalten wir die folgenden wichtigen Aussagen 2.6.14 bis 2.6.16.

**Satz 2.6.14.** *Sei  $C$  eine rekurrente Klasse. Dann sind äquivalent:*

- (i) Es existiert ein  $i \in C$ , das positiv rekurrent ist.
- (ii) Es gibt auf  $C$  eine invariante Verteilung.
- (iii) Alle  $i \in C$  sind positiv rekurrent.

Insbesondere ist positive Rekurrenz eine Klasseeigenschaft.

**Beweis.**

(iii) $\Rightarrow$ (i) Das ist trivial, da  $C \neq \emptyset$ .

(i) $\Rightarrow$ (ii) Nach (i) gilt  $\sum_{k \in C} \bar{\pi}_k^{(i)} = \mathbb{E}_i T_1^{(i)} < \infty$ . Nach Satz 2.6.3 definiert  $\pi_k := (\mathbb{E}_i T_1^{(i)})^{-1} \bar{\pi}_k^{(i)}$  für  $k \in C$  und  $\pi_k = 0$  für  $k \in E \setminus C$  eine invariante Verteilung  $\pi$ .

(ii) $\Rightarrow$ (iii) Dies folgt unmittelbar aus Satz 2.6.13.

□

**Satz 2.6.15.** Die Markovkette mit Übergangsmatrix  $P$  sei irreduzibel, aperiodisch und positiv rekurrent mit (nach Satz 2.6.14 existierender) invarianter Verteilung  $\pi$ . Dann gilt

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j, \quad i, j \in E. \tag{2.6.20}$$

**Beweis.** Diese Aussage wurde früher (z. B. [Fe68]) rein analytisch bewiesen. Sie besagt insbesondere, dass unter den Voraussetzungen des Satzes die Verteilung von  $X_n$  im Grenzwert  $n \rightarrow \infty$  nicht vom Startpunkt  $X_0 = i$  abhängt, die Verteilung der Markovkette also unabhängig vom Start gegen die Gleichgewichtsverteilung  $\pi$  konvergiert. Heute bevorzugt man den folgenden “stochastischen” Beweis, der zwar kaum kürzer, aber anschaulicher ist. Die darin verwendete Kopplungstechnik (*coupling technique*) findet auch bei anderen Beweisen nutzbringend Anwendung. “Gekoppelt” werden dabei zwei Markovketten mit unterschiedlichem Startpunkt.

Sei  $\mathbb{W} = (W_k)_{k \in \mathbb{N}_0} = (X_k, Y_k)_{k \in \mathbb{N}_0}$  eine Markovkette mit Zustandsraum  $E \times E$  und Übergangswahrscheinlichkeiten  $p_{(i,j),(k,l)} = p_{ik} p_{jl}$  für alle  $i, j, k, l \in E$ . Mit anderen Worten: Die Komponenten  $\mathbb{X} = (X_k)_k$  und  $\mathbb{Y} = (Y_k)_k$  sind unabhängige Markovketten mit derselben Übergangsmatrix  $P$ . Wir betrachten einige Eigenschaften von  $\mathbb{W} = (W_k)_k$ .

(i) Für jedes Quadrupel  $i, j, k, l$  existiert nach Lemma 2.6.2 ein  $n_0 = n_0(i, j, k, l)$ , so dass

$$p_{(i,j),(k,l)}^{(n)} = p_{ik}^{(n)} p_{jl}^{(n)} > 0, \quad n \geq n_0, \tag{2.6.21}$$

da  $P$  aperiodisch ist. Also ist auch die Markovkette  $\mathbb{W}$  irreduzibel und aperiodisch.

(ii) Man prüft leicht nach, dass  $\tilde{\pi}_{(i,j)} := \pi_i \pi_j$ ,  $i, j \in E$  eine invariante Verteilung  $\tilde{\pi}$  von  $\mathbb{W}$  definiert. Also ist nach Satz 2.6.10 die Markovkette  $(W_k)_k$  rekurrent und nach Satz 2.6.14 sogar positiv rekurrent.

Sei nun  $i_0 \in E$  beliebig. Dann gilt wegen (ii) für  $i, j \in E$

$$\mathbb{P}_{(i,j)}(W_n = (i_0, i_0) \text{ u. o.}) = 1. \quad (\text{Warum?}) \tag{2.6.22}$$

Sei  $\tau$  der erste Zeitpunkt, an dem die Markovkette  $\mathbb{W}$  den Zustand  $(i_0, i_0)$  erreicht. Dann gilt (undefinierte bedingte Wahrscheinlichkeiten setze man Null):

$$\begin{aligned} |p_{ik}^{(n)} - p_{jk}^{(n)}| &= |\mathbb{P}_{(i,j)}(X_n = k, \tau \leq n) + \mathbb{P}_{(i,j)}(X_n = k \mid \tau > n)\mathbb{P}_{(i,j)}(\tau > n) \\ &\quad - \mathbb{P}_{(i,j)}(Y_n = k, \tau \leq n) - \mathbb{P}_{(i,j)}(Y_n = k \mid \tau > n)\mathbb{P}_{(i,j)}(\tau > n)| \\ &\leq |\mathbb{P}_{(i,j)}(X_n = k, \tau \leq n) - \mathbb{P}_{(i,j)}(Y_n = k, \tau \leq n)| + \mathbb{P}_{(i,j)}(\tau > n). \end{aligned} \quad (2.6.23)$$

Für  $n \rightarrow \infty$  geht  $\mathbb{P}_{(i,j)}(\tau > n)$  gegen Null, da  $\mathbb{P}_{(i,j)}(\tau < \infty) = 1$ . Der letzte Ausdruck in den Betragsstrichen ist für alle  $i, j, k, n$  identisch Null, wie wir gleich formal zeigen werden. Anschaulich ist dies klar, denn wenn  $\tau \leq n$  ist,  $\mathbb{X}$  und  $\mathbb{Y}$  sich also schon in  $i_0$  getroffen haben, dann haben sie danach dieselbe Verteilung.

Nun folgt der formale Beweis.

$$\begin{aligned} \mathbb{P}_{(i,j)}(X_n = k, \tau \leq n) &= \sum_{m=0}^n \mathbb{P}_{(i,j)}(X_n = k, \tau = m) \\ &= \sum_{m=0}^n \mathbb{P}_{(i,j)}(X_n = k \mid \tau = m, W_m = (i_0, i_0))\mathbb{P}_{(i,j)}(\tau = m) \\ &= \sum_{m=0}^n p_{i_0 k}^{(n-m)} \mathbb{P}_{(i,j)}(\tau = m) \\ &= \sum_{m=0}^n \mathbb{P}_{(i,j)}(Y_n = k \mid \tau = m, W_m = (i_0, i_0))\mathbb{P}_{(i,j)}(\tau = m) \\ &= \mathbb{P}_{(i,j)}(Y_n = k, \tau \leq n), \end{aligned} \quad (2.6.24)$$

wobei wir bei der dritten und vierten Identität die starke Markoveigenschaft (Satz 2.4.6) verwendet haben.

Wenn wir (2.6.24) in (2.6.23) einsetzen, erhalten wir

$$\lim_{n \rightarrow \infty} |p_{ik}^{(n)} - p_{jk}^{(n)}| = 0. \quad (2.6.25)$$

Sei nun  $\varepsilon > 0$  beliebig und  $E_0 \subset E$  endlich, so dass  $\sum_{k \in E \setminus E_0} \pi_k < \varepsilon$ . Dann folgt

$$|\pi_j - p_{ij}^{(n)}| = \left| \sum_{k \in E} \pi_k (p_{kj}^{(n)} - p_{ij}^{(n)}) \right| \leq \sum_{k \in E_0} \pi_k |p_{kj}^{(n)} - p_{ij}^{(n)}| + \varepsilon, \quad (2.6.26)$$

also mit (2.6.25)

$$\limsup_{n \rightarrow \infty} |\pi_j - p_{ij}^{(n)}| \leq \varepsilon. \quad (2.6.27)$$

Da  $\varepsilon > 0$  beliebig war, folgt die Behauptung.  $\square$

Wir wollen nun das Analogon von Satz 2.6.15 im nullrekurrenten Fall zeigen.

**Satz 2.6.16.** *Die Markovkette mit Übergangsmatrix  $P$  sei irreduzibel, aperiodisch und nullrekurrent. Dann gilt*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0, \quad i, j \in E. \quad (2.6.28)$$

**Beweis.** Wie im Beweis von Satz 2.6.15 definieren wir die Markovkette  $\mathbb{W}$ , die wie in Satz 2.6.15 irreduzibel und aperiodisch ist. Wir unterscheiden zwei Fälle je nachdem ob  $\mathbb{W}$  transient oder rekurrent ist (beides ist möglich):

Falls  $\mathbb{W}$  transient ist, gilt für  $i, j \in E$  nach Satz 2.6.1:  $(p_{ij}^{(n)})^2 = p_{(i,i),(j,j)}^{(n)} \rightarrow 0$  für  $n \rightarrow \infty$ , also folgt die Behauptung.

Falls  $\mathbb{W}$  rekurrent ist, dann definieren wir  $\tau$  wie im Beweis von Satz 2.6.15, woraus (2.6.25) folgt. Angenommen, es gäbe ein Paar  $(i, j) \in E \times E$ , so dass  $\alpha := \limsup_{n \rightarrow \infty} p_{ij}^{(n)} > 0$ . Dann existiert eine Teilfolge  $(n_m)_m$ , so dass  $\lim_{m \rightarrow \infty} p_{ij}^{(n_m)} = \alpha$ . Wegen  $\sum_{k \in E} \bar{\pi}_k^{(i)} = \mathbb{E}_i T_1 = \infty$  existiert eine endliche Menge  $M \subset E$  mit  $\sum_{k \in M} \bar{\pi}_k^{(i)} > \frac{2}{\alpha} \bar{\pi}_j^{(i)}$ . Wähle  $m_0$  so groß, dass  $|p_{kj}^{(n_m)} - \alpha| < \alpha/2$  für alle  $m \geq m_0$  und  $k \in M$  (benutze (2.6.25)). Dann gilt für  $m \geq m_0$

$$\bar{\pi}_j^{(i)} = \sum_{k \in E} \bar{\pi}_k^{(i)} p_{kj}^{(n_m)} \geq \sum_{k \in M} \bar{\pi}_k^{(i)} \frac{\alpha}{2} > \bar{\pi}_j^{(i)}, \quad (2.6.29)$$

was unmöglich ist. □

**Bemerkung 2.6.17.** Natürlich kann man die Sätze 2.6.15 und 2.6.16 auch für nicht irreduzible Markovketten formulieren, wenn man annimmt, dass  $i, j$  in einer aperiodischen positiv bzw. nullrekurrenten Klasse liegen.

Wir wollen nun noch sehen, was man im periodischen Fall sagen kann und gleichzeitig die wichtigsten Resultate der Sätze 2.6.1 bis 2.6.16 zusammenfassen.

**Satz 2.6.18.** a) Ist  $j \in E$  aperiodisch und positiv rekurrent, so gilt

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \frac{f_{ij}}{\mathbb{E}_j T_1^{(j)}} = f_{ij} \pi_j, \quad i \in E, \quad (2.6.30)$$

wobei  $\pi$  die auf der Klasse von  $j$  konzentrierte invariante Verteilung ist.

b) Ist  $j \in E$  nullrekurrent, so gilt  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$  für alle  $i \in E$ .

c) Ist  $j \in E$  positiv rekurrent mit Periode  $d > 1$ , dann gilt

$$\lim_{n \rightarrow \infty} p_{jj}^{(nd)} = \frac{d}{\mathbb{E}_j T_1^{(j)}} = \pi_j d, \quad (2.6.31)$$

wobei  $\pi$  die auf der Klasse von  $j$  konzentrierte invariante Verteilung ist.

d) Ist  $j \in E$  transient, so gilt  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$  für alle  $i \in E$ .

e) (Verallgemeinerung von c)). Ist  $j \in E$  positiv rekurrent mit Periode  $d > 1$ , dann gilt

$$\lim_{n \rightarrow \infty} p_{ij}^{(nd+r)} = \frac{d}{\mathbb{E}_j T_1^{(j)}} f_{ij}^{*,r} = \pi_j d f_{ij}^{*,r}, \quad (2.6.32)$$

wobei  $r \in \{1, \dots, d\}$  und  $f_{ij}^{*,r} := \sum_{m=0}^{\infty} f_{ij}^{(md+r)}$ .

**Beweis.**

- a) Ist  $i$  in derselben Klasse wie  $j$ , so ist  $f_{ij} = 1$ , und die Behauptung folgt aus Satz 2.6.15 und Satz 2.6.13. Anderenfalls gilt (mit  $f_{ij}^{(k)}$  aus Satz 2.6.1):

$$p_{ij}^{(n)} = \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}. \quad (2.6.33)$$

Wegen  $1 \geq p_{jj}^{(n-k)} \rightarrow (\mathbb{E}_j T_1^{(j)})^{-1} = \pi_j$  und  $\sum_{k=1}^{\infty} f_{ij}^{(k)} = f_{ij}$  folgt die Behauptung allgemein. (Man beachte, dass hier die Vertauschung von  $\lim$  und unendlicher Summation zulässig ist.)

- b) Wir zeigen zunächst  $\lim_{n \rightarrow \infty} p_{jj}^{(n)} = 0$ . Im aperiodischen Fall folgt dies aus Satz 2.6.16. Im periodischen Fall mit Periode  $d$  ist die Markovkette  $(X_{dn})_n$  aperiodisch und  $j$  ebenfalls nullrekurrent, also  $\lim_{n \rightarrow \infty} p_{jj}^{(nd)} = 0$ . Wenn  $m$  kein ganzzahliges Vielfaches von  $d$  ist, gilt  $p_{jj}^{(m)} = 0$ , also ist  $\lim_{n \rightarrow \infty} p_{jj}^{(n)} = 0$ . Für  $i \neq j$  folgt die Behauptung dann wie in a) (mit  $\pi_j = (\mathbb{E}_j T_1^{(j)})^{-1} = 0$ ).
- c) Folgt analog zum periodischen Fall in b). Dabei beachte man, dass  $(X_{dn})_n$  ebenfalls die invariante Verteilung  $\pi$  hat und die erwartete Rückkehrzeit von  $(X_{dn})_n$  nach  $i$  gleich  $1/d$  mal derjenigen von  $(X_n)_n$  ist.
- d) Dies ist Satz 2.6.1.
- e) Einfache Verallgemeinerung von c).

□

Wir formulieren noch zwei einfache Folgerungen.

**Korollar 2.6.19.** a) Jede Markovkette auf einem endlichen Zustandsraum  $E$  hat mindestens einen positiv rekurrenten Zustand.

b) Jede endliche Klasse  $C$  ist positiv rekurrent.

**Beweis.**

- a) Wäre dies nicht so, dann gälte  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$  für alle  $i, j \in E$  nach Satz 2.6.18b) und d), was aber wegen  $1 = \sum_{j \in E} p_{ij}^{(n)} \rightarrow 0$  unmöglich ist.
- b) folgt ebenso (ersetze oben  $E$  durch  $C$ ).

□

Die folgende Aussage folgt unschwer aus Satz 2.6.18.

**Satz 2.6.20.** a) Es existiert eine eindeutige invariante Verteilung genau dann, wenn es genau eine positiv rekurrente Klasse gibt.

- b) Es existiert ein  $\pi \in [0, \infty)^E$  mit  $\sum_{i \in E} \pi_i = 1$  und  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j$  für alle  $i, j \in E$  genau dann, wenn es genau eine aperiodische positiv rekurrente Klasse  $C$  gibt und  $f_{ij} = 1$  für alle  $i \in E, j \in C$  gilt.



**Beweis.** Übungsaufgabe. □

Ohne Beweis präsentieren wir noch zwei sogenannte “starke” Grenzwertsätze, die fast sichere Aussagen über das Verhalten für  $n \rightarrow \infty$  machen. Wir erinnern daran, dass  $N_n(j) = \sum_{k=0}^n \mathbb{1}\{X_k = j\}$  die Zahl der Besuche im Zustand  $j$  bis zum Zeitpunkt  $n$  bezeichnet.

**Satz 2.6.21.** Sei  $C$  eine rekurrente Klasse,  $\bar{\pi}_k^{(i)}$  für  $i, k \in C$  wie in Satz 2.6.3 und  $N_n(j)$  die Anzahl der Besuche der Markovkette  $(X_k)_{k=1, \dots, n}$  in  $j$ . Dann gilt

$$\mathbb{P}_i \left( \lim_{n \rightarrow \infty} \frac{N_n(j)}{N_n(k)} = \frac{\bar{\pi}_j^{(i)}}{\bar{\pi}_k^{(i)}} \right) = 1 \quad i, j, k \in C. \tag{2.6.34}$$

**Beweis.** [Br68, S. 143 f]. Wir werden den Beweis in der Vorlesung präsentieren. □

**Satz 2.6.22.** Sei  $C$  eine positiv rekurrente Klasse und  $\pi \in [0, \infty)^E$  die zugehörige invariante Verteilung, d. h. die eindeutige Lösung von

$$\pi = \pi P, \quad \sum_{i \in C} \pi_i = 1, \quad \pi_j = 0, \quad j \in E \setminus C. \tag{2.6.35}$$

Weiter gelte für ein  $f: E \rightarrow \mathbb{R}$ :

$$\sum_{j \in C} |f(j)| \pi_j < \infty. \tag{2.6.36}$$

Dann gilt

$$\mathbb{P}_i \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(X_k) = \sum_{j \in E} f(j) \pi_j \right) = 1, \quad i \in C. \tag{2.6.37}$$

**Beweis.** Dies ist ein Spezialfall des *Ergodensatzes* für stationäre Prozesse (siehe z. B. [Br68, S. 118 ff]). □

**Bemerkung 2.6.23.** Wählt man speziell  $f(k) = \delta_{jk}$  für ein  $j \in C$ , dann folgt

$$\mathbb{P}_i \left( \lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \pi_j \right) = 1, \quad i \in C. \tag{2.6.38}$$

Alternativ kann man die Aussage ohne Verwendung des Ergodensatzes direkt aus Satz 2.6.21 herleiten. Wir werden dies in der Vorlesung tun (nicht vergessen!).

## 2.7 Beispiele

**Beispiel 2.7.1.** Ist die symmetrische Irrfahrt auf  $\mathbb{Z}$  (siehe Beispiel 2.2.1) positiv rekurrent?

Es gilt  $p_{00}^{(2n)} \sim (\pi n)^{-1/2}$  nach Satz 2.4.8. Da  $p_{00}^{(2n-1)} = 0$  für alle  $n \in \mathbb{N}$  ist, gilt also  $\lim_{n \rightarrow \infty} p_{00}^{(n)} = 0$ . Da 0 ein rekurrenter Zustand ist, folgt nach Satz 2.6.18c), dass 0 nullrekurrent sein muss. Da die symmetrische Irrfahrt irreduzibel ist, ist sie (oder  $\mathbb{Z}$ ) nullrekurrent. Alternativ folgt die Nullrekurrenz aus Satz 2.6.14 und der Tatsache, dass  $\pi$  definiert als  $\pi_i = 1$  für alle  $i \in \mathbb{Z}$  ein unendliches invariantes Maß der Kette ist.

**Beispiel 2.7.2** ( $(s, S)$ -Warenlagermodell, Fortsetzung von Beispiel 2.2.7). Betrachte zunächst

$$Y_n := \text{Lagerinhalt am Morgen des Tages } n. \quad (2.7.1)$$

Wir nehmen an, dass  $\mathbb{P}(X_1 > 0) > 0$  gilt.

Die Menge  $C := \{j \in \{s, \dots, S\} : S \rightarrow j\}$  ist eine kommunizierende Klasse, die den Zustand  $S$  enthält. Sie ist endlich und daher nach Korollar 2.6.19 positiv rekurrent. Alle anderen Zustände (sofern vorhanden) sind unwesentlich und damit transient.  $C$  kann periodisch oder aperiodisch sein; dies hängt von der Verteilung von  $X$  ab. Es gilt offensichtlich  $f_{ij} = 1$  für alle  $i \in \{s, \dots, S\}$ ,  $j \in C$ . Die Markovkette  $(Z_n)_n$  mit

$$Z_n := \text{Lagerinhalt am Abend des Tages } n \quad (2.7.2)$$

besteht ebenfalls aus einer positiv rekurrenten Klasse  $\tilde{C}$  und eventuell weiteren unwesentlichen Zuständen. Man kann zeigen, dass auch hier  $f_{ij} = 1$  für alle  $j \in \tilde{C}$  und alle  $i \in \{S, S-1, \dots\}$  gilt.

Wir betrachten nun die durchschnittlichen Kosten pro Tag,

$$K^{(n)} := \frac{1}{n} \sum_{k=1}^n \text{Kosten am Tag } k, \quad (2.7.3)$$

und stellen die Frage, ob  $K^{(n)}$  konvergiert, und wenn ja, wogegen.

Wie in Beispiel 2.2.7 erläutert wurde, kann man die Kosten aufspalten wie folgt:

$$K^{(n)} = \frac{1}{n} \sum_{k=1}^n (K_B \mathbb{1}_{\{s-1, s-2, \dots\}}(Z_k) + f_1(Y_k) + f_2(Z_k^-) + f_3(S)), \quad (2.7.4)$$

wobei die vier Summanden die Bestellkosten, die Lagerhaltungskosten, die Kosten für nichtbefristete Nachfrage und die Lagerinvestitions- und Mietkosten modellieren.

Seien  $\pi^{(Z)}$  bzw.  $\pi^{(Y)}$  die (eindeutigen) invarianten Verteilungen der Markovketten  $(Z_k)_k$  bzw.  $(Y_k)_k$ . Nach Satz 2.6.22 gelten die drei Grenzwertaussagen

$$\frac{1}{n} \sum_{k=1}^n K_B \mathbb{1}_{\{s-1, s-2, \dots\}}(Z_k) \longrightarrow K_B \sum_{j=-\infty}^{s-1} \pi_j^{(Z)}, \quad (2.7.5)$$

$$\frac{1}{n} \sum_{k=1}^n f_1(Y_k) \longrightarrow \sum_{j=s}^S f_1(j) \pi_j^{(Y)}, \quad (2.7.6)$$

$$\frac{1}{n} \sum_{k=1}^n f_2(Z_k^-) \longrightarrow \sum_{j=-\infty}^{-1} f_2(-j) \pi_j^{(Z)} \quad (2.7.7)$$

jeweils mit Wahrscheinlichkeit 1, wobei wir bei der letzten Konvergenz annehmen, dass die rechte Seite endlich ist (ansonsten würden bei nichtnegativem  $f_2$  die durchschnittlichen Kosten gegen unendlich konvergieren). Genaugenommen gelten die Grenzwertaussagen zunächst nur unter der Voraussetzung, dass man in den positiv rekurrenten Klassen von  $(Y_n)_n$  und  $(Z_n)_n$  startet. Man kann aber leicht einsehen, dass diese Zusatzvoraussetzung nicht nötig ist.

Somit haben wir gezeigt, dass die durchschnittlichen Kosten pro Tag für  $n \rightarrow \infty$  einen deterministischen Grenzwert haben. Dieser lässt sich durch zwei invariante Verteilungen ausdrücken,

die sich wiederum als Lösung von zwei linearen Gleichungssystemen berechnen lassen. Dieser Grenzwert der Kosten hängt noch (auf im allgemeinen komplizierte Weise) von  $s$  und  $S$  ab. In vielen Fällen ist es interessant, diese Funktion bezüglich  $s$  und  $S$  zu minimieren. Wir gehen auf dieses Problem nicht näher ein.

In vielen Fällen dürfte die (auf einem Rechner in der Regel schnelle) Auswertung der Grenzfunktion an z. B. 100 Wertepaaren  $(s, S)$  bereits einen guten Überblick über vernünftige Wahlen von  $s$  und  $S$  liefern.

**Beispiel 2.7.3** (M/G/1-Warteschlange). Sei  $X(t)$  die Anzahl der Kunden im System zur Zeit  $t \geq 0$  bei einer M/G/1-Warteschlange. Der stochastische Prozess  $(X(t))_{t \geq 0}$  hat in der Regel nicht die Markoveigenschaft (nur dann, wenn auch die Bedienungszeiten exponentialverteilt und damit gedächtnislos sind). Wenn aber  $\tau_1 < \tau_2 < \dots$  die (zufälligen) Zeitpunkte sind, an denen die Bedienungszeit eines Kunden endet, dann definiert  $Y_n := X(\tau_n+)$ ,  $n \in \mathbb{N}$ , eine Markovkette  $(Y_n)_n$ , da die Zwischenankunftszeiten gedächtnislos sind. Dabei definieren wir  $X(\tau_n+) = \lim_{s \downarrow \tau_n} X(s)$ . Man bezeichnet  $(Y_n)_n$  als eine in  $(X(t))_{t \geq 0}$  eingebettete Markovkette.

Sei  $G$  die Verteilungsfunktion der Bedienungszeit. Wir nehmen an, dass  $G(0) = 0$  ist und der Erwartungswert  $\nu$  der Bedienungszeit endlich ist. Den Parameter der (exponentialverteilten) Zwischenankunftszeiten nennen wir  $\lambda$ . Der Zustandsraum von  $(Y_n)_n$  ist offenbar  $\mathbb{N}_0$ . Wir berechnen zunächst die Übergangsmatrix  $P$  von  $(Y_n)_n$ . Danach beschäftigen wir uns mit der Frage, ob die Markovkette positiv rekurrent ist. Diese Eigenschaft ist sicher wünschenswert, da nach Satz 2.6.18 ansonsten die Warteschlange “explodiert” in dem Sinn, dass z. B.  $\lim_{n \rightarrow \infty} \mathbb{P}(Y_n \leq 10^{10}) = 0$  gilt.

Wir bestimmen nun  $p_{ij}$  für  $i, j \in \mathbb{N}_0$ . Sei  $Y_0$  die Zahl der Kunden, die am Ende einer Bedienungszeit noch im System sind. Den nächsten Kunden, der bedient wird, bezeichnen wir mit der Nummer 1 usw. Sei zunächst  $i \geq 1$  und  $K$  die Zahl der Kunden, die während der Bedienungszeit  $B$  des ersten Kunden eintreffen. Dann gilt

$$\begin{aligned} p_{ij} &= \mathbb{P}(Y_1 = j \mid Y_0 = i) \\ &= \mathbb{P}(K = j - i + 1 \mid Y_0 = i) \\ &= \begin{cases} \int_0^\infty \mathbb{P}_i(K = j - i + 1 \mid B = x) dG(x), & \text{falls } j - i + 1 \geq 0, \\ 0 & \text{sonst.} \end{cases} \end{aligned} \tag{2.7.8}$$

Wir berechnen nun den Integranden: Im Fall  $j - i + 1 \geq 0$  gilt

$$\mathbb{P}_i(K = j - i + 1 \mid B = x) = \mathbb{P}_i(K \geq j - i + 1 \mid B = x) - \mathbb{P}_i(K \geq j - i + 2 \mid B = x). \tag{2.7.9}$$

Die Wahrscheinlichkeit, dass innerhalb der Zeit  $x$  mindestens  $k$  Kunden eintreffen, ist gleich der Wahrscheinlichkeit, dass die Summe von  $k$  unabhängigen  $\text{Exp}(\lambda)$ -verteilten Zufallsvariablen einen Wert kleiner oder gleich  $x$  annimmt. Diese ist gegeben durch

$$\mathbb{P}(K \geq k \mid B = x) = \int_0^x \frac{\lambda^k y^{k-1}}{(k-1)!} e^{-\lambda y} dy, \quad k \in \mathbb{N}. \tag{2.7.10}$$

Setzt man dies mit  $k = j - i + 1$  bzw.  $k = j - i + 2$  in (2.7.9) ein, so ergibt sich

$$P(K = j - i + 1 \mid B = x) = \int_0^x \frac{\lambda^{j-i+1} y^{j-i}}{(j-i)!} e^{-\lambda y} dy - \int_0^x \frac{\lambda^{j-i+2} y^{j-i+1}}{(j-i+1)!} e^{-\lambda y} dy. \tag{2.7.11}$$

Indem man das erste Integral partiell integriert, sieht man, dass die rechte Seite von (2.7.11) gleich  $(\lambda x)^{j-i+1} e^{-\lambda x} / (j-i+1)!$  ist. Die Zahl der im Zeitraum  $x$  ankommenden Kunden ist daher Poisson-verteilt mit Parameter  $\lambda x$ . Somit gilt für  $i \geq 1$  und  $j-i+1 \geq 0$ :

$$p_{ij} = \int_0^\infty \frac{(\lambda x)^{j-i+1}}{(j-i+1)!} e^{-\lambda x} dG(x). \quad (2.7.12)$$

Weiter gilt (wie man leicht einsieht)  $p_{0j} = \mathbb{P}(K = j) = p_{1j}$ . Mit der Abkürzung

$$c_r = \mathbb{P}(K = r) = \int_0^\infty \frac{(\lambda x)^r}{r!} e^{-\lambda x} dG(x), \quad r \in \mathbb{N}_0, \quad (2.7.13)$$

folgt

$$P = (p_{ij})_{i,j \in \mathbb{N}_0} = \begin{pmatrix} c_0 & c_1 & c_2 & c_3 & c_4 & \dots \\ c_0 & c_1 & c_2 & c_3 & c_4 & \dots \\ 0 & c_0 & c_1 & c_2 & c_3 & \dots \\ 0 & 0 & c_0 & c_1 & c_2 & \dots \\ 0 & 0 & 0 & c_0 & c_1 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \quad (2.7.14)$$

Offenbar sind alle  $c_r$  positiv. Daher ist die Markovkette irreduzibel und aperiodisch. Für die positive Rekurrenz der Markovkette ist nach Satz 2.6.14 notwendig und hinreichend, dass eine invariante Verteilung existiert (die dann nach Satz 2.6.8 automatisch eindeutig ist).

Zur Abkürzung definieren wir

$$\bar{c}_i = \sum_{j=i+1}^\infty c_j = \mathbb{P}(K \geq i+1), \quad i \in \mathbb{N}_0. \quad (2.7.15)$$

Dann gilt

$$\mathbb{E}K = \sum_{i=0}^\infty i c_i = \sum_{i=0}^\infty i \int_0^\infty \frac{(\lambda x)^i}{i!} e^{-\lambda x} dG(x) = \int_0^\infty \lambda x dG(x) = \lambda \nu =: \rho. \quad (2.7.16)$$

und

$$\sum_{i=0}^\infty \bar{c}_i = \sum_{i=0}^\infty \mathbb{P}(K \geq i+1) = \mathbb{E}K = \rho. \quad (2.7.17)$$

Alternativ kann man  $\mathbb{E}K$  auch durch die Formel

$$\mathbb{E}K = \lim_{z \uparrow 1} \frac{d}{dz} \hat{c}(z) \quad \text{mit} \quad \hat{c}(z) = \sum_{i=0}^\infty c_i z^i \quad (2.7.18)$$

(vgl. WT I) berechnen.

Die Zahl  $\rho$  heißt *Verkehrsdichte* der Warteschlange. Wegen  $\nu < \infty$  ist sie endlich. Große Werte von  $\rho$  entstehen durch große  $\lambda$ , d. h. durch kurze mittlere Kundenabstände  $1/\lambda$  oder durch langsame Bedienung (große  $\nu$ ). Es ist daher zu erwarten, dass große  $\rho$  zu langen Warteschlangen führen. Wir werden gleich sehen, dass der Wert  $\rho = 1$  kritisch ist in dem Sinne, dass für kleinere Werte die Markovkette positiv rekurrent ist und für größere nicht. Dies ist nicht verwunderlich, denn  $\rho = 1$  bedeutet gerade, dass die erwartete Bedienungszeit  $\nu$  gleich der erwarteten Zwischenankunftszeit  $\lambda^{-1}$  ist.

Um dies alles zu zeigen, lösen wir das lineare Gleichungssystem  $\pi = \pi P$ . Schreibt man die einzelnen Gleichungen untereinander und addiert die ersten  $k$  für alle  $k \in \mathbb{N}$ , dann erhält man das äquivalente Gleichungssystem

$$\begin{aligned} \pi_1 c_0 &= \pi_0 \bar{c}_0 \\ \pi_2 c_0 &= \pi_0 \bar{c}_1 + \pi_1 \bar{c}_1 \\ \pi_3 c_0 &= \pi_0 \bar{c}_2 + \pi_1 \bar{c}_2 + \pi_2 \bar{c}_1 \\ &\vdots \end{aligned} \tag{2.7.19}$$

Wenn nun  $\sum_{i=0}^{\infty} \pi_i = 1$  ist, dann kann man alle Gleichungen addieren und erhält

$$(1 - \pi_0)c_0 = \pi_0 \rho + \sum_{i=1}^{\infty} \pi_i \sum_{j=1}^{\infty} \bar{c}_j = \pi_0 \rho + (1 - \pi_0)(\rho - \bar{c}_0). \tag{2.7.20}$$

Auflösen nach  $\pi_0$  ergibt  $\pi_0 = c_0 - \rho + \bar{c}_0 = 1 - \rho$ . Durch sukzessives Einsetzen in (2.7.19) erhält man rekursiv alle  $\pi_i$  für  $i \in \mathbb{N}$ . Offenbar muss  $\rho < 1$  sein, damit eine invariante Verteilung existiert, denn sonst wäre  $\pi_0 \leq 0$ . Also ist die Markovkette im Fall  $\rho \geq 1$  nicht positiv rekurrent. Sei umgekehrt  $0 < \rho < 1$ . Setzt man  $\pi_0 = 1 - \rho$  und berechnet die  $\pi_i$  aus (2.7.19) rekursiv, dann sind offenbar alle  $\pi_i$  positiv. Da (2.7.19) äquivalent zu  $\pi = \pi P$  ist, bleibt nur zu zeigen, dass  $\sum_{i=0}^{\infty} \pi_i = 1$  ist. Setzt man  $s_k = \sum_{i=1}^k \pi_i$  und addiert die ersten  $k$  Gleichungen von (2.7.19), dann erhält man

$$\begin{aligned} s_k c_0 &= (1 - \rho) \sum_{j=0}^{k-1} \bar{c}_j + \sum_{i=1}^{k-1} \pi_i \sum_{j=1}^{k-i} \bar{c}_j \\ &\leq (1 - \rho)\rho + s_{k-1}(\rho - \bar{c}_0) \\ &= (1 - \rho)\rho + s_{k-1}(c_0 - (1 - \rho)). \end{aligned} \tag{2.7.21}$$

Hieraus folgt

$$0 < \pi_k = s_k - s_{k-1} \leq \frac{1 - \rho}{c_0}(\rho - s_{k-1}), \tag{2.7.22}$$

woraus  $s_{k-1} \leq \rho$  für alle  $k$  und somit die Endlichkeit von  $\sum_{i=0}^{\infty} \pi_i$  folgt. Falls  $\sum_{i=0}^{\infty} \pi_i \neq 1$  ist, normiert man die  $\pi_i$  so, dass die Summe 1 ist, und sieht wie vorher, dass das normierte  $\pi_0$  gleich  $1 - \rho$  ist. D. h. es gilt  $\sum_{i=0}^{\infty} \pi_i = 1$  (man musste also gar nicht normieren!). Dies zeigt, dass die Markovkette im Fall  $\rho < 1$  positiv rekurrent ist.

Wir werden nun noch eine Formel für die erzeugende Funktion von  $\pi$  im Fall  $\rho < 1$  herleiten und damit den Erwartungswert der Verteilung  $\pi$  berechnen. Seien

$$\hat{\pi}(z) = \sum_{i=0}^{\infty} \pi_i z^i, \quad \hat{c}(z) = \sum_{i=0}^{\infty} c_i z^i, \quad \text{für } z \in \mathbb{C}. \tag{2.7.23}$$

Beide Reihen konvergieren für alle  $|z| < 1$  absolut. Multipliziert man die  $i$ -te Gleichung von  $\pi = \pi P$  mit  $z^i$  und summiert über  $i \in \mathbb{N}_0$ , so erhält man für  $0 < |z| < 1$  die Beziehung

$$\hat{\pi}(z) = \hat{c}(z)(\pi_0 + \pi_1 + \pi_2 z + \pi_3 z^2 + \dots) = \frac{\hat{c}(z)}{z}(\pi_0 z - \pi_0 + \hat{\pi}(z)). \tag{2.7.24}$$

Löst man die Gleichung nach  $\hat{\pi}(z)$  auf, dann erhält man

$$\hat{\pi}(z) = \frac{\hat{c}(z)\pi_0(z-1)}{z - \hat{c}(z)} = \frac{(1-\rho)(z-1)\hat{c}(z)}{z - \hat{c}(z)}. \tag{2.7.25}$$

Diese Gleichung gilt für alle  $z \in \mathbb{C}$  mit  $0 \leq |z| < 1$ , denn aus  $\rho < 1$  folgt  $z \neq \widehat{c}(z)$  für  $0 < |z| < 1$  aus (2.7.24) und für  $z = 0$  folgt die Gleichung, da beide Seiten stetig in  $z = 0$  sind. Die Formel in (2.7.25) ist als *Pollaczek-Khintchin-Formel* bekannt. Aus ihr lassen sich durch mehrfache Differentiation nach  $z$  und Auswerten für  $z = 0$  die  $\pi_i$  (etwas mühsam) berechnen. Einfacher ist die Berechnung des Erwartungswertes  $L$  von  $\pi$  (die mittlere Warteschlangenlänge am Ende einer Bedienungszeit):

$$L = \sum_{i=0}^{\infty} i\pi_i = \lim_{z \uparrow 1} \frac{d}{dz} \sum_{i=0}^{\infty} \pi_i z^i = \lim_{z \uparrow 1} \frac{d}{dz} \widehat{\pi}(z). \quad (2.7.26)$$

Nun ist für  $0 < z < 1$

$$\frac{d}{dz} \widehat{\pi}(z) = (1 - \rho) \frac{(z - \widehat{c}(z))(\widehat{c}(z) + (z - 1)\widehat{c}'(z)) - (z - 1)\widehat{c}(z)(1 - \widehat{c}'(z))}{(z - \widehat{c}(z))^2}, \quad (2.7.27)$$

wobei  $\widehat{c}'(z)$  als  $\frac{d}{dz} \widehat{c}(z)$  zu verstehen ist. Nun konvergieren Zähler und Nenner für  $z \uparrow 1$  gegen Null. Wendet man zweimal die Regel von De l'Hôpital an, so erhält man

$$L = \lim_{z \uparrow 1} \frac{d}{dz} \widehat{\pi}(z) = \rho + \frac{\widehat{c}''(1)}{2(1 - \rho)} = \rho + \frac{\rho^2(\sigma^2/\nu^2 + 1)}{2(1 - \rho)}, \quad (2.7.28)$$

wobei

$$\sigma^2 = \int_0^{\infty} t^2 dG(t) - \nu^2 \quad (2.7.29)$$

die Varianz der Bedienungszeit ist, denn

$$\widehat{c}''(z) = \sum_{i=2}^{\infty} c_i i(i-1)z^i = \sum_{i=2}^{\infty} \int_0^{\infty} i(i-1) \frac{(\lambda x z)^i}{i!} e^{-\lambda x} dG(x) = z^2 \int_0^{\infty} (\lambda x)^2 e^{-\lambda x} e^{\lambda x z} dG(x) \quad (2.7.30)$$

also

$$\widehat{c}''(1) = \lambda^2 \int_0^{\infty} x^2 dG(x) = \lambda^2 \mathbb{E}B^2 = \lambda^2(\nu^2 + \sigma^2). \quad (2.7.31)$$

Im Fall  $\sigma^2 = \infty$  (aber  $\nu < \infty$  und  $\rho < 1$ ) existiert zwar eine invariante Verteilung  $\pi$ , ihr Erwartungswert ist aber unendlich.

**Bemerkung 2.7.4.** Mit denselben Methoden kann man auch die mittlere Wartezeit eines Kunden berechnen. Es erscheint plausibel (und ist richtig), dass die mittlere Zeit  $W$ , die ein Kunde (wartend oder bedient werdend) im System verbringt, gleich  $L/\lambda$  ist, denn wenn Kunden im durchschnittlichen Abstand  $1/\lambda$  eintreffen und die Durchschnittszeit  $L/\lambda$  bleiben, dann sind im Mittel gerade  $L$  Kunden im System. Für weitere Diskussionen hierzu verweisen wir auf die Literatur, z. B. [HS82, S. 251] oder [KT81, S. 502 ff].

**Beispiel 2.7.5** (Verzweigungsprozesse). Wir nehmen an, dass zur Zeit  $n = 0$  ein Individuum existiert, das mit Wahrscheinlichkeiten  $p_k$  genau  $k \in \mathbb{N}_0$  Nachkommen produziert. Jeder der Nachkommen produziert wieder unabhängig voneinander Nachkommen mit derselben Verteilung usw. Anwendungsbeispiele sind Zellteilungsvorgänge, die Ausbreitung von Familiennamen usw. Sei  $S_n$  die Zahl der Individuen in der  $n$ -ten Generation. Offenbar ist  $(S_n)_{n \in \mathbb{N}_0}$  eine Markovkette mit Zustandsraum  $\mathbb{N}_0$  und Startwert 1. Ein solcher Prozess heißt *Galton-Watson-Prozess*.

Wir interessieren uns zunächst für die Wahrscheinlichkeit, dass die Population irgendwann ausstirbt. Wir setzen für  $n \in \mathbb{N}_0$

$$d_n := \mathbb{P}(S_n = 0) \quad \text{und} \quad \widehat{G}_n(z) := \sum_{k=0}^{\infty} z^k \mathbb{P}(S_n = k) = \mathbb{E}z^{S_n}. \quad (2.7.32)$$

Wir nummerieren die Individuen der ersten Generation mit  $i = 1$  bis  $S_1$  durch und bezeichnen die Zahl der Nachkommen des  $i$ -ten Individuums der ersten Generation in der  $n$ -ten Generation mit  $S_{n-1}^{(i)}$ . Das Baugesetz des Verzweigungsprozesses kann mit der Formel

$$S_n = \sum_{i=1}^{S_1} S_{n-1}^{(i)} \quad (2.7.33)$$

beschrieben werden. Im Folgenden schreiben wir kurz  $G$  statt  $\widehat{G}_1$ . Nun folgt (vgl. WT I)

$$\begin{aligned} \widehat{G}_n(z) &= \mathbb{E}z^{S_n} = \mathbb{E}z^{\sum_{i=1}^{S_1} S_{n-1}^{(i)}} = \sum_{k=0}^{\infty} \mathbb{E}\left(z^{\sum_{i=1}^k S_{n-1}^{(i)}} \mid S_1 = k\right) p_k \\ &= \sum_{k=0}^{\infty} \mathbb{E}\left(\prod_{i=1}^k z^{S_{n-1}^{(i)}} \mid S_1 = k\right) p_k \\ &= \sum_{k=0}^{\infty} (\mathbb{E}z^{S_{n-1}})^k p_k \\ &= G(\widehat{G}_{n-1}(z)), \end{aligned} \quad (2.7.34)$$

wobei beim fünften Gleichheitszeichen die Unabhängigkeit der Nachkommensanzahlen verschiedener Individuen verwendet wurde. Nun ist

$$d_n = \widehat{G}_n(0) = G(\widehat{G}_{n-1}(0)) = G(d_{n-1}), \quad n \in \mathbb{N}, \quad \text{und} \quad d_0 = 0. \quad (2.7.35)$$

Mit anderen Worten, die Folge  $(d_n)_n$  entsteht durch Iteration der Funktion  $G$  mit Startwert Null.

Offenbar ist die Folge  $(d_n)_{n \in \mathbb{N}}$  monoton nicht fallend und durch 1 beschränkt. Der Grenzwert  $d$  ist die Wahrscheinlichkeit, dass die Population irgendwann ausstirbt. Da  $G$  stetig auf  $[0, 1]$  ist, folgt  $d = G(d)$ , d. h.  $d$  ist ein *Fixpunkt* von  $G$ .

Im Fall  $p_0 = 0$  folgt  $d_n = 0$  für alle  $n \in \mathbb{N}_0$ , also  $d = 0$ , d. h., die Population kann nicht aussterben.

Ist  $p_0 > 0$  und  $p_0 + p_1 = 1$ , so ist  $d = 1$ , d. h., die Population stirbt fast sicher aus.

Im Fall  $p_0 + p_1 < 1$  ist die Funktion  $G = \widehat{G}_1$  offensichtlich strikt konvex auf  $(0, 1)$  und monoton nichtfallend auf  $[0, 1]$ , und es gilt  $G(1) = 1$ . Daher hat  $G$  außer dem Wert 1 höchstens einen weiteren Fixpunkt in  $[0, 1]$ . Dies ist genau dann der Fall, wenn  $\nu = G'(1) > 1$  ist (eventuell auch  $\infty$ ), d. h., wenn die erwartete Zahl  $n$  von Nachkommen größer als 1 ist.

Ist also  $\nu \leq 1$ , dann folgt  $d = 1$ , d. h. die Population stirbt mit Sicherheit irgendwann aus. Im Fall  $\nu > 1$  bleibt zu untersuchen, ob  $d = 1$  ist oder ob  $d$  die eindeutige Lösung von  $G(x) = x$  mit  $x < 1$  ist. Wir zeigen, dass Letzteres der Fall ist. Aus dem linken Bild erkennt man dies leicht: Offenbar gilt  $d_n < 1$  für alle  $n \in \mathbb{N}$  (Induktion). Wenn  $\lim_{n \rightarrow \infty} d_n = 1$  wäre, dann gäbe es ein  $n$  mit  $x < d_n < 1$  für alle genügend großen  $n$  und somit  $d_{n+1} = G(d_n) < d_n$ , was falsch ist. Also muss die Folge  $(d_n)_n$  gegen die Lösung  $x \in [0, 1)$  von  $G(x) = x$  mit konvergieren. Damit haben wir den folgenden Satz gezeigt.

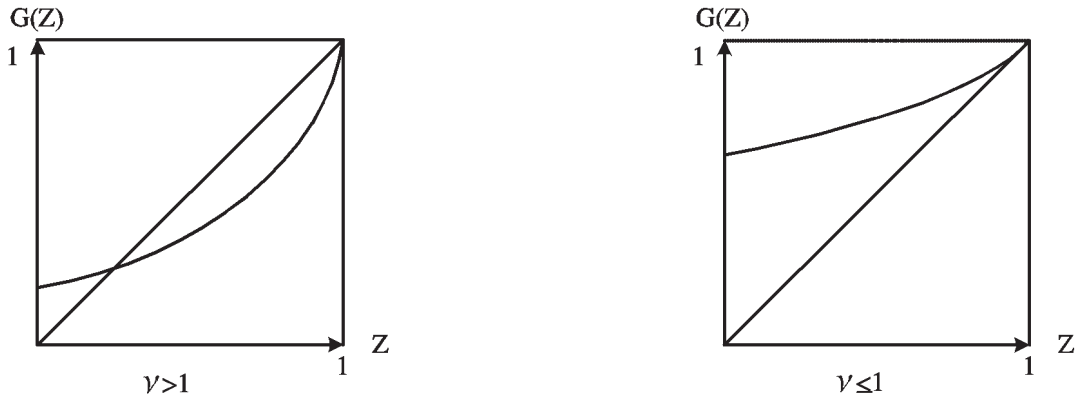


Abbildung 2.7.1: Erzeugende Funktion der Nachkommenverteilung

**Satz 2.7.6.** Sei  $\nu = \sum_{j=0}^{\infty} jp_j \in [0, \infty]$  die erwartete Anzahl von Nachkommen. Wenn  $0 < p_0 \leq 1$  und  $\nu \leq 1$ , dann gilt  $d = 1$ . Wenn  $0 < p_0 \leq 1$  und  $\nu > 1$ , dann ist  $d$  die eindeutige Lösung  $x \in (0, 1)$  von  $G(x) = x$ . Wenn  $p_0 = 0$ , dann ist  $d = 0$ .

Es ist interessant, die Verteilung der Größe  $C$  der gesamten Nachkommenschaft zu bestimmen. Sei  $C_n := 1 + S_1 + \dots + S_n$  für  $n \in \mathbb{N}_0$  die Zahl aller Individuen bis zur  $n$ -ten Generation und  $C_{n-1}^{(i)}$  die Größe der gesamten Nachkommenschaft des  $i$ -ten Individuums der ersten Generation (es selbst eingeschlossen) bis zur  $n$ -ten Generation. Weiter sei

$$\hat{H}_n(z) = \sum_{k=0}^{\infty} z^k \mathbb{P}(C_n = k) \tag{2.7.36}$$

die erzeugende Funktion von  $C_n$ . Dann gilt für  $0 \leq z < 1$

$$\begin{aligned} \hat{H}_n(z) &= \mathbb{E}z^{C_n} = \mathbb{E}z^{1+\sum_{i=1}^{S_1} C_{n-1}^{(i)}} = \sum_{k=0}^{\infty} \mathbb{E}\left(z^{1+\sum_{i=1}^k C_{n-1}^{(i)}} \mid S_1 = k\right) p_k \\ &= z \sum_{k=0}^{\infty} p_k (\hat{H}_{n-1}(z))^k = zG(\hat{H}_{n-1}(z)). \end{aligned} \tag{2.7.37}$$

Es gilt für  $0 \leq z < 1$  und  $n \in \mathbb{N}_0$

$$\hat{H}_{n+1}(z) = \mathbb{E}z^{C_{n+1}} \leq \mathbb{E}z^{C_n} = \hat{H}_n(z) \tag{2.7.38}$$

Da für  $0 \leq z < 1$  die Folge  $(\hat{H}_n(z))_n$  fallend und durch Null nach unten beschränkt ist, besitzt sie einen Grenzwert  $\hat{H}(z)$ . Aus der Stetigkeit von  $G$  folgt

$$\hat{H}(z) = zG(\hat{H}(z)). \tag{2.7.39}$$

Diese Fixpunktgleichung hat für alle  $0 \leq z < 1$  eine eindeutige Lösung in  $[0, 1]$  (auch z. B. im Fall  $p_1 = 1$ ). Wir zeigen, dass (wie zu erwarten)

$$\hat{H}(z) = \sum_{k=0}^{\infty} \mathbb{P}(C = k) z^k \tag{2.7.40}$$



gilt, wenn  $C = \lim_{n \rightarrow \infty} C_n$  ist. (Dieser Grenzwert existiert wegen der Monotonie der Folge  $(C_k)_k$ , ist aber eventuell  $\infty$ .) Es gilt für  $k \in \mathbb{N}_0$ , (aber nicht unbedingt für  $k = \infty$ !)

$$\lim_{n \rightarrow \infty} \mathbb{P}(C_n = k) = \mathbb{P}(C = k) \tag{2.7.41}$$

(man zeige dies!). Aus dem Konvergenzsatz für erzeugende Funktionen (WT I) folgt (2.7.40).

Wir können  $\widehat{H}(z)$  durch (2.7.40) auch für  $z = 1$  definieren. Dann ist  $\widehat{H}(1) = \mathbb{P}(C < \infty) = d$ . Man beachte, dass im Fall  $d < 1$  zwar  $\lim_{n \rightarrow \infty} \widehat{H}_n(1) = 1$  existiert, aber *ungleich*  $\widehat{H}(1)$  ist.

Wir fassen noch einmal zusammen:

**Satz 2.7.7.**  $\widehat{H}(z)$  ist für  $0 \leq z < 1$  die eindeutige Lösung von  $x = zG(x)$ . Weiter gilt  $d = \widehat{H}(1) = \mathbb{P}(C < \infty)$ .

**Korollar 2.7.8.** Wenn  $\nu \leq 1$ , dann gilt

$$\mathbb{E}C = \frac{1}{1 - \nu}, \tag{2.7.42}$$

wobei die rechte Seite im Fall  $\nu = 1$  als  $\infty$  zu lesen ist.

**Beweis.** Im Fall  $p_0 = 0$  ist  $C \equiv \infty$  und  $\nu = 1$  und damit die Behauptung klar. Sei also  $p_0 > 0$ . Dann folgt  $\mathbb{P}(C < \infty) = 1$ , und für  $0 \leq z < 1$  gilt

$$\frac{d\widehat{H}}{dz}(z) = \frac{d}{dz}(zG(\widehat{H}(z))) = G(\widehat{H}(z)) + zG'(\widehat{H}(z))\frac{d\widehat{H}}{dz}(z). \tag{2.7.43}$$

Auflösen nach  $\frac{d\widehat{H}}{dz}(z)$  liefert

$$\mathbb{E}C = \lim_{z \uparrow 1} \frac{d\widehat{H}}{dz}(z) = \lim_{z \uparrow 1} \frac{G(\widehat{H}(z))}{1 - zG'(\widehat{H}(z))} = \frac{1}{1 - \nu}. \tag{2.7.44}$$

□

**Bemerkung 2.7.9.** Alternativ kann man Korollar 2.7.8 wie folgt zeigen: differenziert man die linke und rechte Seite von Gleichung (2.7.34) nach  $z$  und bildet beidseitig den Grenzwert  $z \uparrow 1$ , so folgt  $\mathbb{E}S_n = \nu \mathbb{E}S_{n-1}$  und wegen  $\mathbb{E}S_0 = 1$  per vollständiger Induktion  $\mathbb{E}S_n = \nu^n$  für alle  $n \in \mathbb{N}_0$ . Im Fall  $\nu \leq 1$  folgt also

$$\mathbb{E}C = \sum_{n=0}^{\infty} \mathbb{E}S_n = \sum_{n=0}^{\infty} \nu^n = \frac{1}{1 - \nu} \tag{2.7.45}$$

**Beispiel 2.7.10** (M/G/1- Warteschlange als Verzweigungsprozess). Wenn ein Kunde bei einer M/G/1-Warteschlange das leere System betritt, so wird er als Stammvater eines Verzweigungsprozesses angesehen. Die Nachkommen eines Kunden sind die Kunden, die während seiner Bedienungszeit eintreffen. Die Gedächtnislosigkeit der Zwischenankunftszeiten garantiert, dass die Nachkommenzahlen unabhängig sind. Die erwartete Zahl von Nachkommen ist die Verkehrsdichte  $\rho$ . Also stirbt der Prozess für  $\rho \leq 1$  mit Wahrscheinlichkeit 1 aus, für  $\rho > 1$  dagegen mit geringerer Wahrscheinlichkeit. Dies bedeutet, dass die in Satz 2.7.3 betrachtete Markovkette  $(Y_n)_{n \in \mathbb{N}_0}$  im Fall  $\rho \leq 1$  rekurrent und im Fall  $\rho > 1$  transient ist. Zusammen mit dem Ergebnis aus Satz 2.7.3 sehen wir, dass die Markovkette positiv rekurrent bzw. nullrekurrent bzw. transient ist, je nachdem, ob  $\rho < 1$ ,  $\rho = 1$  oder  $\rho > 1$  ist. Im Fall  $\rho \leq 1$  ist  $(1 - \rho)^{-1}$  die erwartete Zahl der in einer Bedienungsperiode (*busy period*) bedienten Kunden.

**Bemerkung 2.7.11.** Mit ähnlichen Methoden kann man auch die Verteilung der Länge der Bedienungsperiode berechnen [HS82, S. 198 ff].

Wir erwähnen noch ein Paradoxon der Warteschlangentheorie.

**Beispiel 2.7.12** (Ein Paradoxon). Es seien  $\rho < 1$ ,  $L$  die erwartete Warteschlangenlänge (nachdem ein Kunde bedient wurde),  $K = (1 - \rho)^{-1}$  die erwartete Zahl der in einer Bedienungsperiode bedienten Kunden und  $0 < \sigma^2 \leq \infty$  die Varianz der Bedienungszeit. Dann gilt

$$\begin{aligned} L < \infty, K < \infty, & \text{ falls } \sigma^2 < \infty, \\ L = \infty, K < \infty, & \text{ falls } \sigma^2 = \infty. \end{aligned} \quad (2.7.46)$$

Dies scheint im Widerspruch zu der folgenden Überlegung zu stehen: Seien  $\rho < 1$  und  $\sigma^2 = \infty$ . Dann gilt offenbar im Gleichgewicht  $\mathbb{E}Y_n = L = \infty$ , aber

$$Y_n \leq \text{Anzahl der in der zugehörigen Bedienungsperiode bedienten Kunden}. \quad (2.7.47)$$

Bildet man beidseitig Erwartungswerte, so steht aber links  $L = \infty$  und rechts  $K < \infty$ , was offenbar nicht sein kann. Wo liegt der Fehlschluss?

## 2.8 Kartenmischen

In diesem Abschnitt beschäftigt uns die folgende Frage: gegeben sei ein perfekt geordneter Kartenstapel mit  $N$  Spielkarten (nummeriert von 1 bis  $N$ ). Nun wird das Blatt mit einem noch festzulegenden Verfahren  $n$  Mischungsschritten unterworfen. Wie groß muss  $n$  im Vergleich zu  $N$  gewählt werden, damit das Blatt am Ende gut gemischt ist? Zunächst präzisieren wir, was wir unter ‘gut gemischt’ verstehen wollen. Vor und nach jedem Mischungsschritt ist die Reihenfolge der Karten durch eine Permutation von  $N$  Elementen beschrieben, also einem Element der Permutationsgruppe  $S_N$ . Da die Mischungsschritte zufällig sind, haben wir also bei spezialisiertem Mischverfahren nach  $n$  Schritten eine Wahrscheinlichkeitsverteilung  $\mu_n$  auf  $S_N$  gegeben. Vor dem ersten Mischungsschritt ist die Verteilung ein Punktmaß auf der Identität  $e$  von  $S_N$ . Perfekt gemischt ist das Blatt dann, wenn  $\mu_n$  die Gleichverteilung  $\pi$  auf  $S_N$  ist. Da wir kaum darauf hoffen können, dass das Blatt nach einer festen Anzahl von Schritten perfekt gemischt ist, begnügen wir uns damit, dass die Wahrscheinlichkeitsmaße  $\mu_n$  und  $\pi$  nah beieinander liegen. Dazu brauchen wir einen Abstandsbegriff für Wahrscheinlichkeitsmaße auf einer endlichen Menge. Ein geeigneter Abstandsbegriff ist die Totalvariation.

**Definition 2.8.1.** Seien  $\mu$  und  $\nu$  Wahrscheinlichkeitsmaße auf der endlichen Menge  $E$  (ausgestattet mit der Potenzmenge als  $\sigma$ -Algebra). Dann heißt

$$\|\mu - \nu\| := \sup_{A \subseteq E} |\mu(A) - \nu(A)|$$

*Totalvariationsabstand* von  $\mu$  und  $\nu$ .

Es ist sehr leicht zu sehen, dass  $0 \leq \|\mu - \nu\| \leq 1$  gilt, sowie  $\|\mu - \nu\| = \frac{1}{2} \sum_{j \in E} |\mu_j - \nu_j|$ .

Wir betrachten nun wieder einen Kartenstapel mit  $N$  Karten und analysieren das Mischverfahren *Top to Random*. Dabei wird in jedem Mischungsschritt die oberste Karte des Stapels abgehoben und rein zufällig an eine der  $N$  möglichen Positionen in das restliche Kartenblatt

geschoben (dabei ist auch zugelassen, dass die Karte wieder oben auf den Stapel gelegt wird). Danach wird das Verfahren wiederholt, wobei die gewählten Positionen, in die die abgehobenen Karten gesteckt werden, unabhängig voneinander sind. Die einzelnen Mischungsschritte sind also unabhängige und identisch verteilte Zufallsvariable  $X_i$ ,  $i = 1, 2, \dots$  und  $Y_n := X_n \circ \dots \circ X_1$  ist die Kartenreihenfolge nach  $n$  Schritten. Dabei ist  $\circ$  die Hintereinanderausführung von Elementen der Permutationsgruppe  $S_N$ .  $Y_1, Y_2, \dots$  ist eine *Irrfahrt* auf der Gruppe  $S_N$  und daher insbesondere eine Markovkette mit Zustandsraum  $S_N$ . Diese ist offensichtlich irreduzibel und aperiodisch. Die eindeutige invariante Verteilung ist die Gleichverteilung  $\pi$  auf  $S_N$  und nach Satz 2.6.15 gilt  $\lim_{n \rightarrow \infty} p_{ej}^{(n)} = \pi_j = 1/N!$  für alle  $j \in S_N$ . Hieraus folgt insbesondere  $\lim_{n \rightarrow \infty} \|\mu_n - \pi\| = 0$ , wobei  $\mu_n(j) := p_{ej}^{(n)}$ ,  $j \in S_N$ . Damit wollen wir uns aber nicht begnügen, sondern herausfinden, wie  $n$  als Funktion von  $N$  wachsen muss, damit z.B.  $\|\mu_n - \pi\| \approx 1/100$  gilt.

Zunächst stellen wir uns die Frage, ob es eine Stoppzeit  $T$  gibt, so dass nach  $T$  Mischungsschritten “Top to Random” das Blatt perfekt gemischt ist, also jede Kartenreihenfolge mit genau derselben Wahrscheinlichkeit eintritt. Sei  $T'$  der erste Zeitpunkt, zu dem die Karte Nummer  $N$ , die zu Beginn ganz unten im Stapel liegt, erstmalig die Position 1 im Stapel erreicht. Man mache sich klar, dass  $T := T' + 1$  wirklich eine Stoppzeit ist und überlege sich, dass das Blatt nach  $T$  Schritten perfekt gemischt ist. Tatsächlich gilt sogar  $\mathbb{P}_e\{Y_T = j | T = k\} = 1/N!$  für alle  $j \in S_N$  und alle  $k \in \mathbb{N}$ , für die die Bedingung positive Wahrscheinlichkeit hat. Man sagt in diesem Fall, dass  $T$  eine *stark gleichverteilte Stoppregel* ist. Es ist leicht einzusehen, dass man statt  $T'$  auch die erste Zeit wählen kann, nachdem die Karte Nummer  $N - 1$  die Position 1 erreicht hat.

Wir werden nun die Asymptotik sowohl von  $\mathbb{E}T$  als auch der Varianz von  $T$  bestimmen und dann zeigen, dass für (deterministische)  $n$ , die deutlich unterhalb von  $\mathbb{E}T$  liegen, das Blatt nach  $n$  Schritten schlecht gemischt ist, während es für  $n$ , die deutlich oberhalb von  $\mathbb{E}T$  liegen, sehr gut gemischt ist.

In der Vorlesung zeigen wir, dass

$$\mathbb{E}T = N\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{N}\right) = N \log N + \gamma N + o(N)$$

gilt, wobei  $\gamma \approx 0.57721566490153286060651209008240243104215933593992$  (laut Wikipedia) die Euler-Mascheroni Konstante ist, und

$$\text{var}(T) = N^2 \sum_{j=1}^N \frac{1}{j^2} - N \sum_{j=1}^N \frac{1}{j} \sim cN^2, \quad \text{wobei } c := \sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}.$$

Nun wollen wir zeigen, dass für jedes vorgegebene  $\varepsilon > 0$ , das Blatt nach  $n \leq (1 - \varepsilon)N \log N$  Schritten für große  $N$  sehr schlecht gemischt ist, für  $n \geq (1 + \varepsilon)N \log N$  dagegen sehr gut (wobei  $n$  deterministisch ist). Die erste Aussage ist ziemlich offensichtlich, da nach  $(1 - \varepsilon)N \log N$  Schritten mit sehr großer Wahrscheinlichkeit zum Beispiel die Karte Nummer  $N - 10$  noch nicht die Spitze des Stapels erreicht hat und daher die letzten 10 Karten des Stapels immer noch in der ursprünglichen Reihenfolge sind, was zur Folge hat, dass der Totalvariationsabstand  $\|\mu_n - \pi\|$  noch fast 1 und damit das Blatt sehr schlecht gemischt ist. Umgekehrt gilt mit der

Tschebychevungleichung

$$\begin{aligned}
\mathbb{P}\{T \geq (1 + \varepsilon)N \log N\} &= \mathbb{P}\{T - \mathbb{E}T \geq (1 + \varepsilon)N \log N - \mathbb{E}T\} \\
&= \mathbb{P}\{T - \mathbb{E}T \geq \varepsilon N \log N + O(N)\} \\
&\leq \frac{\text{var}T}{(\varepsilon N \log N + O(N))^2} \\
&\sim \frac{c}{\varepsilon^2 (\log N)^2} \rightarrow 0.
\end{aligned}$$

Es ist also für große  $N$  sehr wahrscheinlich, dass die Stoppzeit  $T$  nach  $(1 + \varepsilon)N \log N$  Mischschritten bereits eingetreten ist und daher sollte das Blatt zur Zeit  $n$  sehr gut gemischt sein.

Wir wollen nun zuerst  $\mathbb{P}\{T > m\}$  noch genauer nach oben abschätzen und dann eine präzise obere Abschätzung für den Totalvariationsabstand  $\|\mu_k - \pi\|$  herleiten. Dafür ist es nützlich, das zum Mischverfahren Top to Random analoge *Gutscheinsammelproblem* oder *Coupon Collector's Problem* zu betrachten: in jeder Schokoladenpackung befindet sich eines von  $N$  verschiedenen Bildern und zwar jedes mit derselben Wahrscheinlichkeit. Wieviele Schokoladenpackungen muss man kaufen, bis man von jedem Bild mindestens ein Exemplar besitzt? Ist  $S$  diese Anzahl, so überzeugt man sich sofort davon, dass  $S$  dieselbe Verteilung wie  $T$  hat. Sei nun  $N \in \mathbb{N}$ ,  $\beta \geq 0$ ,  $m := \lceil N \log N + \beta N \rceil$  und  $A_i$  das Ereignis, dass sich das Bild Nummer  $i$  nicht in einer der ersten  $m$  Schokoladenpackungen befindet, so gilt

$$\mathbb{P}\{T > m\} = \mathbb{P}\{S > m\} = \mathbb{P}\left(\bigcup_{i=1}^N A_i\right) \leq N \left(1 - \frac{1}{N}\right)^m \leq N e^{-(\log N + \beta)} = e^{-\beta}. \quad (2.8.1)$$

Um rigoros zeigen zu können, dass das Blatt nach  $n \approx (1 + \varepsilon)N \log N$  Schritten sehr gut gemischt ist, brauchen wir folgendes Lemma:

**Lemma 2.8.2.** *Sei  $Y_k$ ,  $k \geq 0$  wie oben definiert (mit Startwert  $Y_0 = e$ ) und  $S$  eine stark gleichverteilte Stoppregel, d.h.  $S$  ist eine Stoppzeit mit  $\mathbb{P}\{Y_k \in A | S = j\} = \pi(A)$  für alle  $0 \leq j \leq k$  und  $A \subseteq S_N$  sofern die Bedingung strikt positive Wahrscheinlichkeit hat, wobei  $\pi$  die Gleichverteilung auf  $S_N$  ist. Weiter sei wie oben  $\mu_k$  die Verteilung von  $Y_k$ . Dann gilt*

$$\|\mu_k - \pi\| \leq \mathbb{P}\{S > k\}.$$

**Beweis.** Sei  $A \subseteq S_N$ . Dann gilt

$$\begin{aligned}
\mu_k(A) = \mathbb{P}\{Y_k \in A\} &= \sum_{j=0}^k \mathbb{P}\{Y_k \in A, S = j\} + \mathbb{P}\{Y_k \in A, S > k\} \\
&= \pi(A) \sum_{j=0}^k \mathbb{P}\{S = j\} + \mathbb{P}\{Y_k \in A, S > k\} \\
&\leq \pi(A) + \mathbb{P}\{S > k\}.
\end{aligned}$$

Da  $A \subseteq S_N$  beliebig war, folgt die Behauptung.  $\square$

Wendet man das Lemma auf die stark gleichverteilte Stoppregel  $T$  an, so folgt unter Verwendung von Gleichung (2.8.1) die folgende Aussage.

**Korollar 2.8.3.** *Für  $\beta > 0$  und  $k := \lceil N \log N + \beta N \rceil$  gilt  $\|\mu_k - \pi\| \leq e^{-\beta}$ .*

Das Korollar besagt, dass das Blatt “schon” nach  $N \log N + 10N$  Mischschritten sehr gut gemischt ist.

Nachdem wir das einfache aber kaum praxisrelevante Mischverfahren *Top to Random* mathematisch analysiert haben, stellt sich die Frage, ob dies auch für realitätsnähere Mischverfahren möglich ist. Ein solches Verfahren ist *Riffle Shuffle*, bei dem der Kartenstapel zufällig in zwei Teile aufgespaltet wird (das heißt abgehoben wird) und dann der eine Teil zufällig in den anderen geschoben wird. Wir werden dieses Verfahren in der Vorlesung näher kennenlernen und analysieren. Es wird sich dabei zeigen, dass die erforderliche Anzahl von Mischungsschritten von der Ordnung  $\log N$  (statt  $N \log N$  bei *Top to Random*) ist. Eine gut lesbare Literaturquelle zu diesen Mischverfahren findet sich in der Monographie von Aigner und Ziegler, [Ai04], Seite 177ff.

## 2.9 Mischzeiten

Wir abstrahieren nun von der speziellen Situation im letzten Abschnitt und wollen allgemeiner für eine irreduzible und aperiodische Markovkette mit *endlichem* Zustandsraum  $E$  die Konvergenzgeschwindigkeit der Übergangswahrscheinlichkeiten gegen die invariante Verteilung  $\pi$  untersuchen. Eine gute Literaturquelle für diesen Abschnitt ist die Monographie von Levin, Peres und Wilmer [Le09]. Zunächst definieren wir, was wir unter einem *Coupling* verstehen wollen.

**Definition 2.9.1.** Seien  $\mu$  und  $\nu$  Wahrscheinlichkeitsmaße auf  $E$  (immer mit der Potenzmenge als  $\sigma$ -Algebra). Ein Wahrscheinlichkeitsmaß  $\xi$  auf  $E \times E$  heißt *Coupling* von  $\mu$  und  $\nu$ , wenn  $\mu$  bzw.  $\nu$  das Bildmaß von  $\xi$  unter der Projektionsabbildung von  $E \times E$  auf die erste bzw. zweite Komponente ist, d.h.  $\mu_i = \xi\{(i, j) : j \in E\}$  für  $i \in E$  und analog für  $\nu$ .

Gelegentlich nennt man auch ein Paar von  $E$ -wertigen Zufallsgrößen  $(X, Y)$  ein *Coupling* von  $\mu$  und  $\nu$ , wenn die gemeinsame Verteilung  $\xi$  von  $(X, Y)$  ein *Coupling* von  $\mu$  und  $\nu$  ist. Ein wichtiges Resultat ist das folgende:

**Proposition 2.9.2.** *Seien  $\mu$  und  $\nu$  Wahrscheinlichkeitsmaße auf der endlichen Menge  $E$  und  $\Delta := \{(i, i) : i \in E\}$  die Diagonale in  $E \times E$ . Dann gilt*

$$\|\mu - \nu\| = \inf\{\xi(\Delta^c)\},$$

wobei das Infimum über alle *Couplings*  $\xi$  von  $\mu$  und  $\nu$  gebildet wird. Weiterhin existiert ein *Coupling*  $\xi$ , für welches das Infimum angenommen wird. Dieses heißt *optimales Coupling*.

**Beweis.** Sei  $\xi$  ein *Coupling* von  $\mu$  und  $\nu$  und  $(X, Y)$  ein Zufallsvektor auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  mit Verteilung  $\xi$ . Dann gilt für die Menge  $A \subseteq E$ :

$$\mu(A) - \nu(A) = \mathbb{P}\{X \in A\} - \mathbb{P}\{Y \in A\} \leq \mathbb{P}\{X \in A, Y \notin A\} \leq \mathbb{P}\{X \neq Y\} = \xi(\Delta^c)$$

und daher folgt  $\|\mu - \nu\| \leq \inf\{\xi(\Delta^c)\}$ . Es bleibt zu zeigen, dass ein *Coupling*  $\xi$  existiert, so dass  $\|\mu - \nu\| = \xi(\Delta^c)$  gilt. Die Konstruktion dieses *Couplings* findet man in Kapitel 4 der Monographie [Le09].  $\square$

Sei nun bis zum Ende dieses Abschnitts  $P$  eine irreduzible stochastische Matrix auf dem endlichen Raum  $E$  mit (eindeutiger) invarianter Verteilung  $\pi$  und sei  $P_x^n$  die Verteilung der zugehörigen Markovkette nach  $n$  Schritten bei Start in  $x$ . Wir definieren

$$d(n) := \max_{x \in E} \|P_x^n - \pi\|, \quad \bar{d}(n) := \max_{x, y \in E} \|P_x^n - P_y^n\|.$$

**Definition 2.9.3.** Die Größen

$$t_{\text{mix}}(\varepsilon) := \min\{n \in \mathbb{N}_0 : d(n) \leq \varepsilon\}, \quad \varepsilon > 0 \quad \text{und} \quad t_{\text{mix}} := t_{\text{mix}}(1/4)$$

heißen *Mischzeiten* der Kette.

Der spezielle Wert  $1/4$  ist etwas willkürlich aber für einige Anwendungen bequem. Man sieht relativ leicht, dass sowohl  $n \mapsto d(n)$  als auch  $n \mapsto \bar{d}(n)$  nichtwachsend sind. Wir benötigen folgendes einfache Resultat.

**Proposition 2.9.4.**

$$d(n) \leq \bar{d}(n) \leq 2d(n).$$

**Beweis.** Die zweite Ungleichung folgt aus der Dreiecksungleichung für die Totalvariation. Wir zeigen nun die erste Ungleichung:

$$\begin{aligned} \|P_x^n - \pi\| &= \max_{ACE} |P_x^n(A) - \pi(A)| \\ &= \max_{ACE} \left| \sum_{y \in E} \pi_y (P_x^n(A) - P_y^n(A)) \right| \\ &\leq \max_{ACE} \sum_{y \in E} \pi_y |P_x^n(A) - P_y^n(A)| \\ &\leq \sum_{y \in E} \pi_y \max_{ACE} |P_x^n(A) - P_y^n(A)| \\ &= \sum_{y \in E} \pi_y \|P_x^n - P_y^n\| \leq \max_{y \in E} \|P_x^n - P_y^n\|, \end{aligned}$$

woraus sofort die Behauptung folgt. □

Wir wollen nun Proposition 2.9.2 auf Markovketten anwenden. Dazu seien  $(X_n)$  und  $(Y_n)$  Markovketten (nicht unbedingt unabhängig!) mit derselben Übergangsmatrix  $P$  auf demselben Wahrscheinlichkeitsraum. Man nennt die gemeinsame Verteilung auf dem Raum  $E^{\mathbb{N}_0} \times E^{\mathbb{N}_0}$  dann auch oft ein *coupling* (auch wenn hier der Zustandsraum anders als in Definition 2.9.1 unendlich (sogar überabzählbar) ist). Wir nehmen an, dass dabei aus  $X_n = Y_n$  folgt, dass auch  $X_k = Y_k$  für alle  $k \geq n$  gilt.

**Satz 2.9.5.** Seien  $X_0 = x$  und  $Y_0 = y$  und

$$\tau_c := \inf\{n : X_n = Y_n\},$$

dann folgt

$$\|P_x^n - P_y^n\| \leq \mathbb{P}_{x,y}\{\tau_c > n\} \quad \text{und} \quad d(n) \leq \max_{x,y \in E} \mathbb{P}_{x,y}\{\tau_c > n\}.$$

**Beweis.** Die erste Aussage folgt aus 2.9.2, die zweite aus der ersten zusammen mit Proposition 2.9.4. □

In der Vorlesung werden wir mehrere Beispiele kennenlernen, für die wir die Asymptotik der Mischzeiten abschätzen, unter anderen die Irrfahrt auf einem Zykel.

## 2.10 Reversible Markovketten

Eine gute Referenz für diesen Abschnitt ist die Monographie von Olle Häggström [Ha02] (außer dem Abschnitt über die Spektrallücke, den man besser in [Le09] findet). Eine wichtige Klasse von Markovketten sind die *reversiblen* Markovketten.

**Definition 2.10.1.** Sei  $P : E \times E \rightarrow [0, 1]$  eine stochastische Matrix.  $\pi \in [0, \infty)^E$  heißt *reversibles Maß* von  $P$ , falls

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{für alle } i, j \in E$$

gilt. Man sagt dann, dass  $(P, \pi)$  eine *reversible Markovkette* ist.

Ein reversibles Maß  $\pi$  ist immer ein invariantes Maß, denn für  $j \in E$  gilt

$$\sum_{i \in E} \pi_i p_{ij} = \sum_{i \in E} \pi_j p_{ji} = \pi_j \sum_{i \in E} p_{ji} = \pi_j.$$

Umgekehrt ist aber keineswegs jedes invariante Maß reversibel (Beispiel in Vorlesung).

Wir zeigen zuerst eine Anwendung reversibler Markovketten auf Irrfahrten auf Graphen. Sei  $(V, E)$  ein zusammenhängender ungerichteter endlicher Graph ohne Mehrfachkanten und ohne Schleifen. Mit  $d_i$  bezeichnen wir den Grad von  $i \in V$ , d.h. die Zahl der mit  $i$  inzidenten Kanten. Wir definieren eine Übergangsmatrix auf  $V$  durch

$$p_{ij} = \begin{cases} 1/d_i & \text{falls } (i, j) \in E, \\ 0 & \text{sonst.} \end{cases}$$

Die durch  $P$  beschriebene Markovkette heißt (symmetrische) Irrfahrt auf  $(V, E)$ . Wir zeigen, dass  $\pi_j := d_j$ ,  $j \in V$  ein reversibles Maß für  $P$  ist. Im Fall  $(i, j) \notin E$  gilt  $\pi_i p_{ij} = 0 = \pi_j p_{ji}$ , während im Fall  $(i, j) \in E$  gilt:  $\pi_i p_{ij} = d_i \frac{1}{d_i} = 1 = d_j \frac{1}{d_j} = \pi_j p_{ji}$ . Also ist  $\pi$  ein reversibles Maß von  $P$ . Weiterhin ist  $\pi_i := d_i \left( \sum_j d_j \right)^{-1}$  das eindeutige invariante Wahrscheinlichkeitsmaß von  $P$  (die Eindeutigkeit folgt aus der Irreduzibilität und diese aus der Annahme, dass der Graph zusammenhängend ist).

### Spektrallücke und Relaxationszeit

Zunächst sei  $P$  die Übergangsmatrix einer Markovkette mit endlichem Zustandsraum  $E$  und  $n := |E|$ .

**Lemma 2.10.2.**

- (i)  $\lambda = 1$  ist ein Eigenwert von  $P$ .
- (ii) Für jeden (reellen oder komplexen) Eigenwert  $\lambda$  von  $P$  gilt  $|\lambda| \leq 1$ .
- (iii) Ist  $P$  irreduzibel, dann hat der Eigenwert  $\lambda = 1$  algebraische Vielfachheit 1.
- (iv) Ist  $P$  irreduzibel und aperiodisch, dann ist  $\lambda = 1$  der einzige Eigenwert mit  $|\lambda| = 1$ .

**Beweis.**

- (i)  $f = (1, \dots, 1)^T$  ist ein Eigenvektor zum Eigenwert 1, da  $P$  stochastisch ist.
- (ii) Sei  $Pf = \lambda f$  für ein  $f \in \mathbb{C}^n \setminus \{0\}$  und sei  $i \in \{1, \dots, n\}$  so gewählt, dass  $f_i \neq 0$ . Dann folgt für jedes  $N \in \mathbb{N}$  zunächst  $P^N f = \lambda^N f$  und daraus

$$|\lambda|^N |f_i| = |(P^N f)_i| \leq \max_{1 \leq j \leq n} |f_j|,$$

woraus (man betrachte  $N \rightarrow \infty$ )  $|\lambda| \leq 1$  folgt.

- (iii) Wir zeigen zuerst, dass die geometrische Vielfachheit des Eigenwerts 1 1 ist. Dazu sei  $f$  ein Eigenvektor von  $P$  zum Eigenwert 1. Wir zeigen, dass  $f$  notwendig ein (komplexes) Vielfaches von  $(1, \dots, 1)^T$  ist. Wir nehmen zuerst zusätzlich an, dass  $f$  reelle Komponenten hat. Wir nehmen an, die Komponenten von  $f$  seien nicht alle gleich und  $f_k > \min_{j \in E} f_j = f_i$ . Da  $P$  irreduzibel ist, existiert ein  $N$  mit  $p_{i,k}^{(N)} > 0$  und aus  $\sum_{j \in E} p_{i,j}^{(N)} f_j = f_i$  folgt  $f_k = f_i$  was der Annahme widerspricht. Ist nun  $g + ih$  ein komplexer Eigenvektor zum Eigenwert 1 mit reellen Vektoren  $g$  und  $h$ , so folgt  $g + ih = P(g + ih) = Pg + iPh$  und daher  $Pg = g$  und  $Ph = h$ , also sind nach dem ersten Schritt  $g$  und  $h$  konstant und die Behauptung folgt.

Die Tatsache, dass auch die algebraische Vielfachheit 1 ist zeigt man mit der Jordanzerlegung von  $P$ , wobei man benutzt, dass die  $N$ -ten Potenzen der Matrix  $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  unbeschränkt sind (Details: Übungsaufgabe).

- (iv) Sei  $\lambda$  ein Eigenwert von  $P$  mit  $|\lambda| = 1$  und  $f \in \mathbb{C}^n \setminus \{0\}$  ein zugehöriger Eigenvektor. Nach dem Hauptsatz 2.6.15 folgt

$$\lambda^N f = P^N f \rightarrow \Pi f,$$

für eine gewisse Matrix  $\Pi$ , also muss auch die linke Seite für  $N \rightarrow \infty$  konvergieren, was aber nur im Fall  $\lambda = 1$  möglich ist.

□

Wir nehmen nun (bis zum Ende des Beweises von Theorem 2.10.5) an, dass  $(P, \pi)$  eine irreduzible reversible Markovkette mit endlichem Zustandsraum  $E$  mit  $|E| = n$  ist und dabei  $\pi$  sogar ein Wahrscheinlichkeitsmaß ist. Dann definiert

$$\langle f, g \rangle_\pi := \sum_{i \in E} f_i g_i \pi_i$$

ein Skalarprodukt auf  $\mathbb{R}^E$  (man beachte, dass  $\pi_i > 0$  für alle  $i \in E$  gilt!). Wir definieren die Matrix  $A = (a_{ij})$  durch  $a_{ij} := \pi_i^{1/2} \pi_j^{-1/2} p_{ij}$ . Dann ist  $A$  eine (reelle) symmetrische Matrix, zu der daher eine Orthonormalbasis  $\varphi^j$ ,  $j = 1, \dots, n$  aus reellen Eigenvektoren mit zugehörigen reellen Eigenwerten  $\lambda_j$  existiert. Offenbar ist  $\varphi := \sqrt{\pi}$  (komponentenweise) ein Eigenvektor von  $A$  zum Eigenwert 1 und wir nehmen ohne Einschränkung an, dass  $\varphi^1 = \varphi$  gilt. Sei nun  $D_\pi$  die Diagonalmatrix, bei der an der Stelle  $i, i$  der Wert  $\pi_i$  steht. Dann gilt – wie man leicht sieht –  $A = D_\pi^{1/2} P D_\pi^{-1/2}$  und  $f^j := D_\pi^{-1/2} \varphi^j$ ,  $j = 1, \dots, n$  bilden eine Orthonormalbasis von  $P$  mit den Eigenwerten  $\lambda_j$ ,  $j = 1, \dots, n$  bezüglich des Skalarprodukts  $\langle \cdot, \cdot \rangle_\pi$ . Man beachte, dass  $f^1 = (1, \dots, 1)^T$  gilt. Wir nehmen ohne Einschränkung an, dass die Eigenwerte absteigend der Größe nach geordnet sind, also  $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_n \geq -1$  gilt.



**Definition 2.10.3.** Sei

$$\lambda_* := \max\{|\lambda_2|, |\lambda_n|\}.$$

Dann heißt die Zahl  $\gamma_* := 1 - \lambda_*$  *absolute Spektrallücke* und  $\gamma := 1 - \lambda_2$  *Spektrallücke* von  $P$ . Weiter heißt

$$t_{\text{rel}} := \frac{1}{\gamma_*}$$

*Relaxationszeit* der Kette (oder von  $P$ ).

Man beachte, dass  $t_{\text{rel}} = \infty$  gilt genau dann wenn die Kette periodisch ist genau dann wenn die Kette Periode 2 hat genau dann wenn  $\lambda_n = -1$  ist.

Wir formulieren nun Abschätzungen zwischen der Relaxationszeit und der Mischzeit einer reversiblen Markovkette. Die Beweise beider Aussagen finden sich im Kapitel 12 der Monographie [Le09] (wobei das “insbesondere” im folgenden Satz direkt aus der Definition folgt).

**Satz 2.10.4.** Sei  $\pi_{\min} := \min_{i \in E} \pi_i$ . Dann gilt für alle  $\varepsilon > 0$

$$t_{\text{mix}}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon \pi_{\min}}\right) t_{\text{rel}}.$$

Insbesondere gilt für alle  $n \in \mathbb{N}$

$$\|P_x^n - \pi\| \leq C e^{-(1-\lambda_*)n},$$

wobei  $C := 1/\pi_{\min}$ .

**Satz 2.10.5.** Für alle  $\varepsilon > 0$  gilt

$$t_{\text{mix}}(\varepsilon) \geq (t_{\text{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right).$$

**Beweis.** Sei  $f$  ein Eigenvektor von  $P$  zum Eigenwert  $\lambda \neq 1$ . Nach obiger Überlegung ist  $f$  bezüglich  $\langle \cdot, \cdot \rangle_\pi$  orthogonal zum Eigenvektor  $(1, \dots, 1)^T$  von  $P$  zum Eigenwert 1 und daher gilt für jedes  $i \in E$  und  $N \in \mathbb{N}$

$$|\lambda|^N |f_i| = |(P^N f)_i| = \left| \sum_j (p_{i,j}^{(N)} f_j - \pi_j f_j) \right| \leq \|f\|_\infty \sum_j |p_{i,j}^{(N)} - \pi_j| \leq \|f\|_\infty 2d(N).$$

Wähle nun  $i$  so dass  $|f_i| = \|f\|_\infty$ . Dann folgt  $|\lambda|^N \leq 2d(N)$ . Für beliebig vorgegebenes  $\varepsilon > 0$  setze  $N := t_{\text{mix}}(\varepsilon)$ . Dann gilt  $|\lambda|^{t_{\text{mix}}(\varepsilon)} \leq 2\varepsilon$ . Da  $\lambda$  ein beliebiger von 1 verschiedener Eigenwert ist, folgt die Ungleichung für  $\lambda_*$  statt  $|\lambda|$ . Durch Logarithmieren erhält man die Behauptung unter Benutzung der Ungleichung  $\log x \leq x - 1$  für  $x > 0$ .  $\square$

Wir betrachten nun die Anwendung reversibler Markovketten auf das *Hardcore Modell*.

## Das Hardcore Modell

Sei  $G = (V, E)$  ein ungerichteter Graph ohne Schleifen mit  $|V| = k < \infty$ . Ein Element von  $\{0, 1\}^V$  heißt *Konfiguration*. Ein Element  $\xi \in \{0, 1\}^V$  heißt *zulässige (Hardcore)konfiguration*, wenn aus  $\xi_i = 1$  folgt, dass  $\xi_j = 0$  gilt für alle Nachbarn  $j$  von  $i$ , d.h. für alle  $j$  mit  $(i, j) \in E$ .

Wir interpretieren dies wie folgt:  $\xi_i = 1$  bedeutet, dass an dem Knoten  $i$  ein Teilchen sitzt und  $\xi_i = 0$  bedeutet, dass sich an dem Knoten  $i$  kein Teilchen befindet. Die Konfiguration  $\xi$  ist also zulässig genau dann wenn keine zwei benachbarten Positionen von Teilchen belegt sind. Eine interessante Aufgabe ist die, bei gegebenem  $G = (V, E)$  zufällig eine zulässige Konfiguration zu simulieren. Selbst für moderat große Graphen erweist sich dies als keineswegs trivial, da man nicht einmal mit vernünftigem Aufwand berechnen kann, wieviele zulässige Konfigurationen es überhaupt gibt. Sei  $\pi$  die Gleichverteilung auf der (endlichen) Menge  $M$  aller zulässigen Konfigurationen. Eine Idee ist die, zu dem Graphen eine irreduzible und aperiodische Markovkette  $X$  zu basteln, die als Zustandsraum die Menge aller zulässigen Konfigurationen hat und die als invariante Verteilung  $\pi$  hat. Man kann dann die Kette ausgehend von einer beliebigen zulässigen Konfiguration (zum Beispiel alles Null) sehr lange simulieren und weiß dann, dass die Verteilung gegen  $\pi$  konvergiert. Wenn man Glück hat und wie beim Kartenmischen eine Stoppzeit  $T$  findet, so dass die Verteilung von  $X_T$  genau  $\pi$  ist, dann ist das natürlich noch besser. Wir konstruieren nun eine solche Markovkette. Man beachte, dass wir bei der Konstruktion nicht wissen müssen, wie groß ihr Zustandsraum ist.

### Konstruktion einer Markovkette mit invarianter Verteilung $\pi$

1. Starte mit  $X_0 \equiv 0$ ,  $n = 0$ .
2. Wähle zufällig ein  $v \in V$ .
3. Wirf eine faire Münze.
4. Wenn die Münze 1 zeigt und alle Nachbarn von  $v$  in  $X_n$  Null sind, dann setze  $X_{n+1}(v) = 1$ , ansonsten setze  $X_{n+1}(v) = 0$ . Setze  $X_{n+1}(w) = X_n(w)$  für alle  $w \neq v$ .
5. Erhöhe  $n$  um 1.
6. Gehe nach 2.

Die durch diesen Algorithmus beschriebene Markovkette mit Übergangsmatrix  $P$  ist offensichtlich irreduzibel und aperiodisch. Wir zeigen, dass die Gleichverteilung  $\pi$  nicht nur eine invariante, sondern sogar eine reversible Verteilung der Markovkette ist, das heißt, dass

$$\pi_\xi p_{\xi\zeta} = \pi_\zeta p_{\zeta\xi} \quad (2.10.1)$$

für alle  $\xi, \zeta \in M$  ist. Um dies zu zeigen, definieren wir  $d := d(\xi, \zeta)$  als die Anzahl der Knoten, an denen sich  $\xi$  und  $\zeta$  unterscheiden.

1. Fall:  $d = 0$ : dann ist  $\xi = \zeta$  und damit (2.10.1) trivialerweise richtig.
2. Fall:  $d \geq 2$ : dann ist  $p_{\zeta\xi} = p_{\xi\zeta} = 0$  und damit gilt (2.10.1).
3. Fall:  $d = 1$ : in diesem Fall unterscheiden sich  $\xi$  und  $\zeta$  an genau einer Position  $v$ . Daher sind alle Nachbarn von  $v$  Null und es gilt:

$$\pi_\xi p_{\xi\zeta} = \frac{1}{|M|} \frac{1}{k} \frac{1}{2} = \pi_\zeta p_{\zeta\xi}.$$

Damit ist die Behauptung gezeigt.

Unbefriedigend bleibt dabei, dass man nicht weiß, wie lange man simulieren muss, um eine gute Approximation von  $\pi$  zu erhalten.

### Der Gibbssampler

Obiges Verfahren ist ein Spezialfall des *Gibbssamplers*, den wir jetzt beschreiben. Seien  $S$  und  $V$  endliche Mengen und  $\pi$  ein Wahrscheinlichkeitsmaß auf  $S^V$ . Die im folgenden Algorithmus beschriebene Markovkette mit Zustandsraum  $E := \{i \in S^V : \pi_i > 0\}$  nennt man den Gibbssampler für  $\pi$ .

1. Starte in irgendeinem Zustand  $i \in E$ :  $X_0 = i$ ,  $n = 0$ .
2. Wähle zufällig ein  $v \in V$ .
3. Wähle  $X_{n+1}(v)$  gemäß der bedingten Verteilung (bezüglich  $\pi$ ) von  $v$  gegeben alle anderen Werte von  $X_n$ .
4. Setze  $X_{n+1}(w) = X_n(w)$  für alle  $w \neq v$ .
5. Erhöhe  $n$  um 1.
6. Gehe nach 2.

Man zeigt wieder, dass  $\pi$  ein reversibles Maß für die Kette ist. Wenn diese Markovkette irreduzibel und aperiodisch ist (was nicht immer der Fall ist), dann konvergiert die Verteilung der Kette gegen  $\pi$ .

## 2.11 Perfekte Simulation: der Propp–Wilson Algorithmus

In diesem Abschnitt behandeln wir die folgende Frage:

Sei  $\pi$  ein Wahrscheinlichkeitsmaß auf der endlichen Menge  $E := \{s_1, \dots, s_k\}$  (mit typischerweise sehr großem  $k$ ), zum Beispiel die Gleichverteilung auf der Menge aller zulässigen Konfigurationen im Hardcoremodell. Wir nehmen an, dass  $\pi_i := \pi(\{i\}) > 0$  für alle  $i \in E$  gilt. Man finde eine *perfekte Simulation* von  $\pi$ , d.h. man bestimme eine Zufallsvariable  $Y$  mit Verteilung  $\pi$  und ein praktikables *exaktes* Simulationsverfahren für  $Y$ .

Eine Möglichkeit ist die folgende: man wählt sich zunächst eine  $E$ -wertige irreduzible und aperiodische Markovkette mit Übergangsmatrix  $P$ , deren invariante Verteilung  $\pi$  ist (das ist immer möglich). Der im folgenden angegebene *Propp–Wilson Algorithmus* (1996) ergibt als Output eine Zufallsvariable  $Y$  mit Verteilung  $\pi$ .

Zunächst konstruiert man sich eine zulässige *Update-Funktion* für die Übergangsmatrix  $P$ , das heißt eine Abbildung  $\phi : E \times [0, 1] \rightarrow E$  mit der Eigenschaft, dass für jede  $U[0, 1]$ -verteilte Zufallsvariable  $U$  gilt:

$$\mathbb{P}\{\phi(i, U) = j\} = p_{ij} \quad \text{für alle } i, j \in E.$$

Eine solche Funktion existiert immer, ist aber keineswegs eindeutig. Die Funktion  $\phi$  definiert eine *Kopplung* zwischen den Übergängen aus verschiedenen Zuständen.

**Propp–Wilson Algorithmus**

Sei  $N_0 := 0, N_1, N_2, \dots$  eine streng wachsende Folge in  $\mathbb{N}_0$  (z.B. 0, 1, 2, 4, 8, ...).

1.  $m := 1$
2. Simuliere unabhängige  $U[0, 1]$  verteilte Zufallsvariable  $U_{-N_m}, \dots, U_{-N_{m-1}-1}$  und bestimme die (zufällige) Menge  $M_m := \phi(\cdot, U_{-1}) \circ \dots \circ \phi(\cdot, U_{-N_m})(E)$ .
3. Wenn  $M_m = \{s\}$  einelementig ist, dann setze  $Y := s$  und der Algorithmus endet, sonst gehe nach 4.
4. Erhöhe  $m$  um 1. Gehe nach 2.

**Satz 2.11.1.** *Die Wahrscheinlichkeit, dass der Propp–Wilson Algorithmus terminiert, ist entweder 0 oder 1. Wenn er terminiert, dann gilt  $\mathbb{P}\{Y = i\} = \pi_i$  für alle  $i \in E$ .*

**Beweis.** Angenommen, der Algorithmus terminiert mit Wahrscheinlichkeit  $p > 0$ , dann existiert ein  $m_0 \in \mathbb{N}$ , so dass die Wahrscheinlichkeit, dass er spätestens nach  $m_0$  Schritten terminiert, mindestens  $p/2$  ist. Nun hat man aber für  $m \rightarrow \infty$  beliebig viele unabhängige Versuche, so dass die Wahrscheinlichkeit, dass der Algorithmus (nach endlich vielen Schritten) terminiert, 1 ist.

Wir nehmen nun an, dass der Algorithmus terminiert. Sei  $i \in E$  und  $\varepsilon > 0$ . Wir werden zeigen, dass

$$|\mathbb{P}\{Y = i\} - \pi_i| \leq \varepsilon.$$

Da der Algorithmus mit Wahrscheinlichkeit 1 terminiert, existiert ein  $m_0 \in \mathbb{N}$ , so dass

$$\mathbb{P}\{\text{Algor. bricht spätestens bei } m = m_0 \text{ ab}\} \geq 1 - \varepsilon.$$

Nun konstruieren wir uns eine Zufallsvariable  $\tilde{Y}$  wie folgt: sei  $Z$  unabhängig von allen  $U_k$  mit Verteilung  $\pi$  und  $\tilde{Y} := \phi(\cdot, U_{-1}) \circ \dots \circ \phi(\cdot, U_{-N_{m_0}})(Z)$ . Da  $\pi$  eine invariante Verteilung der Markovkette ist, ist die Verteilung von  $\tilde{Y}$  gerade  $\pi$ . Weiter gilt  $\tilde{Y} = Y$  falls der Algorithmus spätestens bei  $m = m_0$  abbricht, also  $\mathbb{P}\{Y \neq \tilde{Y}\} \leq \varepsilon$ . Es folgt

$$|\mathbb{P}\{Y = i\} - \pi_i| = |\mathbb{P}\{Y = i\} - \mathbb{P}\{\tilde{Y} = i\}| \leq \mathbb{P}\{Y = i, \tilde{Y} \neq i\} + \mathbb{P}\{Y \neq i, \tilde{Y} = i\} \leq \mathbb{P}\{Y \neq \tilde{Y}\} \leq \varepsilon.$$

Damit ist die Aussage gezeigt. □

**Beispiel 2.11.2.** Wir betrachten die Markovkette mit Zustandsraum  $E = \{1, 2\}$  und Übergangsmatrix  $P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ . Man findet leicht zwei verschiedene Updatefunktionen, so dass der Propp–Wilson Algorithmus im einen Fall terminiert und im anderen nicht!

Es ist sehr wichtig, die folgenden zwei Warnungen zu beherzigen.

**Warnung 2.11.3.** Man könnte vermuten, dass man  $Y$  auch durch eine Vorwärtsiteration wie folgt generieren kann: Für eine Updatefunktion  $\phi$  und unabhängige  $U[0, 1]$ -verteilte Zufallsgrößen  $U_1, U_2, \dots$  definiert man die Stoppzeit  $T$  durch

$$T := \inf\{n \in \mathbb{N} \mid |\phi(\cdot, U_n) \circ \dots \circ \phi(\cdot, U_1)(E)| = 1\},$$

d.h.  $T$  ist der erste Zeitpunkt, zu dem die Kardinalität des Bildes von  $E$  unter Vorwärtsiterationen der Updatefunktion einelementig ist. Wenn  $T$  fast sicher endlich ist und wir das eine

Element mit  $\hat{Y}$  bezeichnen, dann folgt im allgemeinen NICHT, dass  $\pi$  die Verteilung von  $\hat{Y}$  ist. Dazu betrachten wir folgendes Beispiel.

$$E = \{1, 2\}, P = \begin{pmatrix} .5 & .5 \\ 1 & 0 \end{pmatrix}, \quad (2.11.1)$$

$$\phi(i, x) = \begin{cases} 1 & \text{falls } i = 2, \\ 1 & \text{falls } i = 1, x < 1/2, \\ 2 & \text{falls } i = 1, x \geq 1/2. \end{cases}$$

Dann gilt  $\hat{Y} = 1$  fast sicher, aber  $\pi_1 = 2/3!$

**Warnung 2.11.4.** Es ist wichtig, dass man im zweiten Schritt des Propp–Wilson Algorithmus die schon vorher generierten Zufallsgrößen  $U_0, \dots, U_{-N_m-1}$  wieder recyclet und nicht neu generiert (auch wenn das Speicherplatz kostet und eine Neugenerierung billiger wäre). Wir zeigen dafür in der Vorlesung ein Beispiel (man kann das Beispiel in der vorherigen Warnung nehmen mit  $N_m := m$  für alle  $m \in \mathbb{N}_0$ ).

Ein ernsthaftes Problem bei der praktischen Anwendung des Propp–Wilson Algorithmus ist der oft enorme Speicherplatzbedarf und Rechenaufwand. Im folgenden interessanten Spezialfall ist dieser allerdings recht gering. Wenn man auf der Menge  $E$  eine partielle Ordnung definieren kann, die ein maximales Element  $\bar{M}$  und ein minimales Element  $\underline{M}$  besitzt – das heißt für alle  $i \in E$  gilt  $\underline{M} \leq i \leq \bar{M}$  – und die Markovkette so gebaut ist, dass eine zugehörige *monotone* Updatefunktion existiert, das heißt, dass aus  $i > j$  folgt, dass  $\phi(i, u) \geq \phi(j, u)$  für alle  $u \in [0, 1]$  gilt, dann genügt es, im zweiten Schritt des Propp–Wilson Algorithmus lediglich zu testen, ob  $\phi(\cdot, U_{-1}) \circ \dots \circ \phi(\cdot, U_{-N_m})(\bar{M}) = \phi(\cdot, U_{-1}) \circ \dots \circ \phi(\cdot, U_{-N_m})(\underline{M})$  gilt, da in diesem Fall wegen der Monotonie der Updatefunktion dann automatisch  $\phi(\cdot, U_{-1}) \circ \dots \circ \phi(\cdot, U_{-N_m})(E)$  einelementig ist.



# Kapitel 3

## Markovketten mit stetiger Zeit

### 3.1 Einleitung und Beispiele

*Markovketten mit stetiger Zeit (MKSZ)* sind Prozesse mit höchstens abzählbarem Zustandsraum  $E$ , die im Gegensatz zu Markovketten mit diskreter Zeit (MKDZ), die wir im vorigen Kapitel ausgiebig studiert haben, mit  $t \in [0, \infty)$  indiziert sind und die sich ebenfalls der Markoveigenschaft erfreuen (die wir weiter unten formulieren werden). Für die Modellierung zeitkontinuierlicher Systeme sind sie oft praktikabler als die Approximation durch MKDZ.<sup>1</sup> Anwendungen finden sich in vielen Bereichen. Wir betonen wieder die Anwendung auf Warteschlangen unter Einschluss von Netzen von Warteschlangen; siehe Abschnitt 3.5. Zunächst wollen wir anschaulich verstehen, was eine MKSZ ist und durch welche Größen sie charakterisiert ist. Wie im diskreten Fall wollen wir unter einer MKSZ nur den Übergangsmechanismus verstehen. Einen stochastischen Prozess – oder genauer: die Verteilung eines solchen – erhält man erst, wenn man noch eine Startverteilung auf  $E$  festlegt. In diesem Fall soll die Markoveigenschaft gelten, d. h., für jedes  $n \in \mathbb{N}$  und jede  $t_1 \leq t_2 \leq \dots \leq t_n \leq t$  sowie jede  $i_1, i_2, \dots, i_n, i \in E$  gilt

$$\begin{aligned} \mathbb{P}(X(t) = i \mid X(t_1) = i_1, \dots, X(t_n) = i_n) \\ &= \mathbb{P}(X(t) = i \mid X(t_n) = i_n) \\ &= \mathbb{P}_{i_n}(X(t - t_n) = i), \end{aligned} \tag{3.1.1}$$

wobei der Index  $i_n$  von  $\mathbb{P}$  bedeutet, dass der Prozess im Zustand  $i_n$  gestartet wird. Um diese Eigenschaft zu garantieren, müssen die Aufenthaltszeiten in einem Zustand bis zum nächsten Sprung gedächtnislos, d. h. exponentialverteilt sein (siehe Übungsaufgabe) und die Wahrscheinlichkeit, dann in einen bestimmten anderen Zustand zu springen, muss unabhängig von allem Vorherigen sein. Wenn die Aufenthaltsdauer in  $i \in E$   $\text{Exp}(c_i)$ -verteilt ist und die stochastische Matrix  $P = (p_{ij})_{i,j \in E}$  die Übergangswahrscheinlichkeiten nach einem Sprung beschreibt, wobei  $p_{ii} = 0$  für alle  $i \in E$  ist (man muss in einen anderen Zustand springen), so ist durch die Parameter  $c_i$  und die Matrix  $P$  offenbar der Übergangsmechanismus vollständig beschrieben. Zusammen mit einer Startverteilung  $a$  auf  $E$  ist damit ein stochastischer Prozess beschrieben. Wir erlauben den Fall  $c_i = 0$  für ein  $i \in E$ . In diesem Fall ist  $i$  absorbierend. Dagegen ignorieren wir zunächst den möglichen Fall  $c_i = \infty$ , in welchem man sofort springt. Wir zeigen, wie man den Prozess simuliert:

1. Man simuliere die Startverteilung  $a$  auf  $E$  (wie bei MKDZ). Sei  $X(0) = i_0$ ,  $n = 0$ .

---

<sup>1</sup>In diesem Kapitel werden wir Markovketten im Sinne der Definition 2.3.1 immer mit “MKDZ” bezeichnen.

2. Man simuliere die Aufenthaltszeit  $T_n$  in  $i_n$ . Sie ist  $\text{Exp}(c_{i_n})$ -verteilt.
3. Man simuliere den nächsten Zustand  $i_{n+1} \in E$  gemäß der Verteilung  $p_{i_n i_{n+1}}$ .
4. Setze  $n = n + 1$ .
5. Gehe nach 2.

Bei jeder Simulation verwende man eine neue (von den vorherigen unabhängige) auf  $[0, 1]$  gleichverteilte Zufallsvariable und transformiere sie entsprechend. Es bietet sich an, jeweils nach der Simulation in 2. den Plot der Realisierung zu aktualisieren. Nach Belieben baue man ein Abbruchkriterium ein (entweder als Funktion von  $n$  oder der Zeitdauer des Prozesses). Man kann dabei bei gewissen  $c = (c_i)_{i \in E}$  und  $P$  in die Situation kommen, dass selbst für  $n \rightarrow \infty$  der Prozess “steckenbleibt”, d. h.  $\sum_{n=0}^{\infty} T_n < \infty$  mit positiver Wahrscheinlichkeit passiert. Dieses Verhalten entspricht einer Explosion in endlicher Zeit. Hier stellt sich die Frage, ob man den Prozess auch noch nach der Zeit  $\sum_{n=0}^{\infty} T_n$  definieren kann oder will. Offen bleibt bei der obigen Konstruktion, ob die Pfade rechtsstetig oder linksstetig sind. Die Verteilung einer MKSZ legt dies nicht fest. Es hat sich aber gezeigt, dass es günstig ist, alle Pfade rechtsstetig zu wählen. Oft definiert man

$$q_{ij} = \begin{cases} c_i p_{ij} & \text{falls } i \neq j, \\ -c_i & \text{falls } i = j, \end{cases} \quad (3.1.2)$$

und bezeichnet  $Q = (q_{ij})_{i,j \in E}$  als die  $Q$ -Matrix der MKSZ. Offenbar kann man aus  $Q$  die  $c_i$  und jene Zeilen von  $P$ , für die  $c_i \neq 0$  ist, zurückgewinnen (die anderen Zeilen von  $P$  sind ohnehin bedeutungslos).

Bei der oben beschriebenen Simulation genügt es offensichtlich,  $a$  und die Matrix  $Q$  zu kennen.  $Q$  muss lediglich die Bedingungen  $q_{ij} \geq 0$  für alle  $i, j \in E$  mit  $i \neq j$  und  $\sum_{j \in E} q_{ij} = 0$  erfüllen.  $a$  und  $Q$  legen dann die Verteilung des Prozesses eindeutig fest – jedenfalls bis zur “Explosion”, sofern diese in endlicher Zeit passiert.

Um bereits in Kürze konkrete Beispiele für MKSZ behandeln zu können, erwähnen wir (Beweis später; siehe Satz 3.2.7), dass

$$q_{ij} = \lim_{t \downarrow 0} \frac{\mathbb{P}_i(X(t) = j)}{t}, \quad i, j \in E, i \neq j, \quad (3.1.3)$$

gilt. Die  $q_{ij}$  heißen daher auch *Übergangsraten*. Graphisch kann man eine MKSZ ähnlich wie im Fall einer MKDZ darstellen: man malt einen Pfeil von  $i$  nach  $j$ , wenn  $q_{ij} > 0$  ist und beschriftet ihn mit  $q_{ij}$ .

Die bisherigen Betrachtungen sind recht oberflächlich und dienen lediglich dazu, eine intuitive Vorstellung von MKSZ zu vermitteln. Präzise Definitionen, Annahmen, Aussagen und Beweise folgen im Abschnitt 3.2.

**Beispiel 3.1.1** (Geburts- und Todesprozesse). Geburts- und Todesprozesse nennt man solche MKSZ, deren Zustandsraum  $E = \mathbb{N}_0$  ist und für die die  $Q$ -Matrix die Form

$$q_{ij} = \begin{cases} \lambda_i, & \text{falls } j = i + 1, \\ \mu_i, & \text{falls } j = i - 1, i \geq 1, \\ 0 & \text{sonst, wenn } i \neq j, \end{cases} \quad (3.1.4)$$

besitzt. Die  $\lambda_i$  werden als *Geburtsraten*, die  $\mu_i$  als *Sterberaten* bezeichnet. Sind alle  $\mu_i = 0$  (bzw. alle  $\lambda_i = 0$ ), dann heisst die MKSZ (reiner) Geburtsprozess (bzw. (reiner) Todesprozess).



Geburts- und Todesprozesse werden zum Beispiel als (sehr einfache) Modelle zur Beschreibung der zeitlichen Evolution der Größe einer Population verwendet. Sie treten aber auch bei bestimmten Warteschlangenmodellen auf, wie wir im nächsten Beispiel sehen werden. Ein wichtiger Spezialfall ist der *Poissonprozess*, der als reiner Geburtsprozess mit  $\lambda_i = \lambda$  für alle  $i \in \mathbb{N}_0$  definiert ist, wobei  $\lambda > 0$  ein Parameter ist. Wir werden den Poissonprozess noch näher studieren.

**Beispiel 3.1.2** (M/M/c-Warteschlange). Die Zahl der Kunden in einer M/M/c-Warteschlange ist offenbar wegen der Gedächtnislosigkeit der Exponentialverteilung eine MKSZ. Seien  $\lambda$  bzw.  $\mu$  die Parameter der Exponentialverteilungen der Zwischenankunfts- bzw. Bedienungszeiten (für jeden der  $c$  Bediener). Da Sprünge immer nur zu benachbarten Zuständen möglich sind, ist die Zahl der Kunden sogar ein Geburts- und Todesprozess. Die Geburts- und Sterberaten sind gegeben durch

$$\lambda_i = \lambda \quad \text{für alle } i \in \mathbb{N}_0, \quad \mu_i = \begin{cases} i\mu & \text{falls } 0 \leq i \leq c, \\ c\mu & \text{falls } i \geq c. \end{cases} \quad (3.1.5)$$

Man beachte, dass die Sterberate proportional zur Zahl der Kunden ist, die gerade bedient werden.

**Beispiel 3.1.3** (M/E<sub>2</sub>/1-Warteschlange). Die Zahl der Kunden in einer M/E<sub>2</sub>/1-Warteschlange ist keine MKSZ, da die E<sub>2</sub>-Verteilung nicht gedächtnislos ist. Es lässt sich aber eine MKSZ mit größerem Zustandsraum so definieren, dass die Zahl der Kunden in einer M/E<sub>2</sub>/1-Warteschlange eine Funktion jener MKSZ ist. Wenn man für diese MKSZ alle Übergangswahrscheinlichkeiten (oder auch die invariante Wahrscheinlichkeitsverteilung – sofern sie existiert) berechnet, so kennt man damit automatisch auch die (invariante) Verteilung der Zahl der Kunden in der M/E<sub>2</sub>/1-Schlange.

Der “Trick” besteht darin, dass man die E<sub>2</sub>-verteilte Bedienungszeit (mit Erwartungswert  $\mu^{-1}$ ) künstlich in zwei Phasen zerlegt, die jeweils Exp( $2\mu$ )-verteilt und unabhängig sind. Merkt man sich nun zusätzlich zur Zahl der Kunden noch, in welcher Phase der Bedienung das System sich befindet, so hat man eine MKSZ vorliegen, da die Zeitdauern der einzelnen Phasen gedächtnislos sind. Der Zustandsraum dieser MKSZ ist

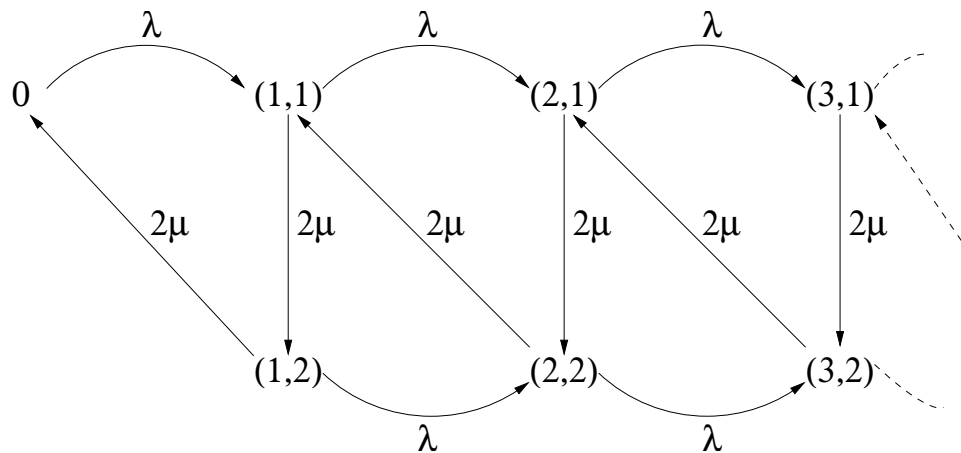
$$E = \{0\} \cup \{(n, i) : n \in \mathbb{N}, i \in \{1, 2\}\} = \{0\} \cup \mathbb{N} \times \{1, 2\}, \quad (3.1.6)$$

wobei “0” bedeutet, dass kein Kunde im System ist, und  $(n, i)$ , dass  $n \in \mathbb{N}$  Kunden im System sind und der Kunde, der gerade bedient wird, sich in der Bedienungsphase  $i \in \{1, 2\}$  befindet. Es spielt dabei keine Rolle, ob sich solche Bedienungsphasen real beobachten lassen, oder ob sie nur künstlich eingeführt wurden. Die  $Q$ -Matrix ist aus der folgenden Abbildung ablesbar.

Übergänge von Phase 1 zur Phase 2 finden immer mit Rate  $2\mu$  statt, da die Phase 1 Exp( $2\mu$ )-verteilt ist. Mit derselben Rate  $2\mu$  endet die Bedienungszeit eines Kunden, wenn er sich in Phase 2 befindet, worauf sich die Zahl der Kunden um Eins verringert und der nächste Kunde in Phase 1 bedient wird. Die Ankunftsrate ist immer  $\lambda$ .

**Beispiel 3.1.4** (M/H<sub>2</sub>/1-Warteschlange). Eine Mischung aus  $k \in \mathbb{N}$  Exponentialverteilungen wird gelegentlich als *Hyper-Exponentialverteilung* bezeichnet und mit  $H_k$  abgekürzt. Eine Zufallsvariable ist also genau dann  $H_k$ -verteilt, wenn ihre Verteilungsfunktion  $F$  die Gestalt

$$F(x) = \begin{cases} 0, & \text{falls } x \leq 0, \\ \sum_{i=1}^k q_i(1 - e^{-r_i x}), & \text{falls } x \geq 0 \end{cases} \quad (3.1.7)$$

Abbildung 3.1.1: M/E<sub>2</sub>/1-Warteschlange

hat, wobei  $r_1, \dots, r_k > 0$ ,  $q_1, \dots, q_k > 0$  und  $\sum_{i=1}^k q_i = 1$  gilt.  $H_k$ -Verteilungen (oder allgemeinere Mischungsverteilungen) können zum Beispiel auftreten, wenn Waren aus verschiedenen Produktionen gemischt werden. Die  $q_i$  sind dann die Anteile der Waren aus Produktion  $i$ .  $H_2$ -Verteilungen bei Bedienzeiten können dadurch entstehen, dass die Kunden eine von zwei Serviceleistungen in Anspruch nehmen, die jeweils exponentialverteilt mit verschiedenen Parametern sind.

Wie bei der M/E<sub>2</sub>/1-Schlange ist die Zahl der Kunden auch bei der M/H<sub>2</sub>/1-Schlange keine MKSZ, kann aber als Funktion einer MKSZ geschrieben werden. Bei dieser MKSZ merkt man sich neben der Zahl der Kunden noch den “Typ” der Bedienung (1 oder 2). In der Regel lässt sich ein solcher “Typ” (wie die “Phase” bei M/E<sub>2</sub>/1) nicht real beobachten.

Als Zustandsraum kann man wiederum die Menge  $E$  in (3.1.6) wählen, wobei “0” bedeutet, dass kein Kunde im System ist und  $(n, i)$ , dass  $n$  Kunden im System sind und der Kunde, der gerade bedient wird, vom Typ  $i \in \{1, 2\}$  ist. Die  $Q$ -Matrix ist aus dem folgenden Graph ablesbar, wobei  $q_1, q_2 = 1 - q_1$ , sowie  $r_1$  und  $r_2$  die Parameter der  $H_2$ -Verteilung seien.

**Bemerkung 3.1.5.** Mit ähnlichen Methoden (nur etwas komplizierter) kann man solche Warteschlangen behandeln, für die sowohl die Zwischenankunftszeiten als auch die Bedienzeiten jeweils (endliche) Mischungen von Erlangverteilungen (mit verschiedenen Parametern) sind. Dabei kann die Zahl der Bediener auch größer als 1 sein und eine Kapazitätsbeschränkung vorliegen. In einem bestimmten Sinn lässt sich damit jede G/G/c/K und G/G/c-Warteschlange approximieren. Der Zustandsraum ist für die MKSZ so zu wählen, dass er für jeden der  $c$  Bediener und für die Zwischenankunftszeit die volle Information über den “Typ” (d. h. die Komponente der Mischung) und die Phase enthält.

## 3.2 Definitionen und erste Ergebnisse

Nun zurück zur Theorie. Gegeben sei als Zustandsraum eine abzählbare (nichtleere) Menge  $E$ . Gelegentlich will man die Möglichkeit einer Explosion in endlicher Zeit nicht a priori ausschließen. Im Falle einer Explosion gibt es vielfältige Möglichkeiten, den Prozess als MKSZ weiterlaufen

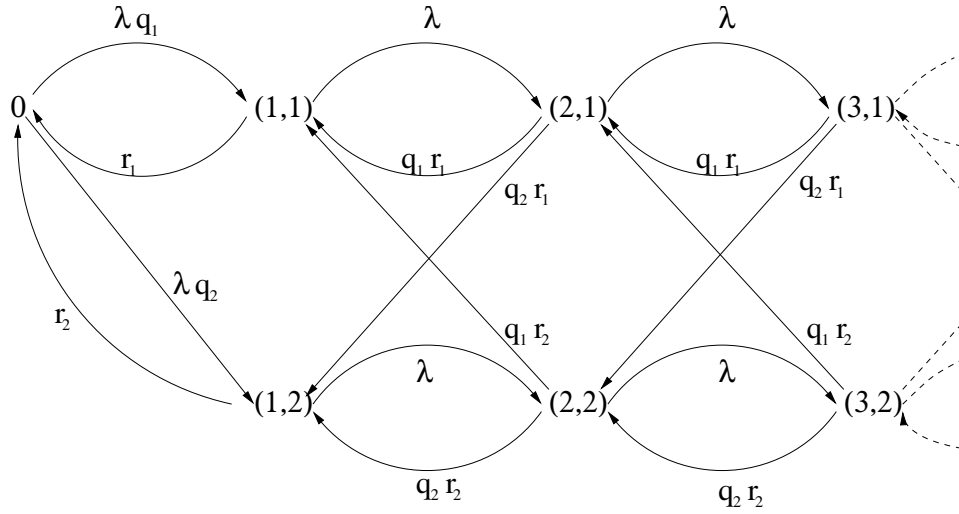


Abbildung 3.1.2: M/H<sub>2</sub>/1-Warteschlange

zu lassen, z. B. durch unmittelbares Springen in einen bestimmten Zustand mit “Neustart” der Kette. Oft will man aber auch den Fall betrachten, dass der Prozess im Zustand der Explosion verharret. Dies erreicht man dadurch, dass man einen zusätzlichen Zustand  $\partial$  (*Grab, Sarg*) zu  $E$  hinzufügt und vereinbart, dass sich der Prozess nach der Explosion für immer in  $\partial$  befindet. Dies hat zur Folge, dass  $\sum_{j \in E} \mathbb{P}_i(X(t) = j)$  kleiner als 1 sein kann.

Wir werden nun ähnlich wie im Fall von MKDZ vorgehen und das Analogon der Übergangsmatrix  $P$  definieren. Während im diskreten Fall durch eine Startverteilung  $a$  und die stochastische Matrix  $P$  die Verteilung der MKDZ zu  $a$  und  $P$  festgelegt ist, ist nicht klar, ob dies im stetigen Fall mit  $a$  und der Matrix  $P(1) := (\mathbb{P}_i(X(1) = j))_{i,j \in E}$  auch so ist, denn aus  $P(1)$  kann man zwar  $P(n)$  für  $n \in \mathbb{N}$  durch  $P(n) = P(1)^n$  berechnen, aber wie berechnet man z. B.  $P(1/2)$  aus  $P(1)$ ? Um diese Probleme zu umgehen, definieren wir, was eine (*sub-*)*markovsche Halbgruppe*  $(P(t))_{t \geq 0}$  ist. Wir sehen dann in Satz 3.2.4, dass zu einer Startverteilung und einer submarkovschen Halbgruppe eine (bezüglich Verteilung eindeutige) MKSZ existiert. In Satz 3.2.7 zeigen wir die schon in (3.1.3) behauptete Existenz der Grenzwerte  $q_{ij} = \lim_{t \downarrow 0} t^{-1} P_{ij}(t)$ . Danach folgen Aussagen (zum Teil ohne Beweis), die zeigen, dass die vorher beschriebene Simulation wirklich die zu  $a$  und  $(P(t))_{t \geq 0}$  gehörige MKSZ simuliert. Wie im zeitdiskreten Fall werden wir uns dann mit der Existenz und Berechnung der Grenzwerte von  $P_{ij}(t)$  für  $t \rightarrow \infty$  beschäftigen.

**Definition 3.2.1.** Sei  $E$  eine abzählbare, nichtleere Menge. Eine Familie  $(P(t))_{t > 0}$  von Matrizen  $(P_{ij}(t))_{i,j \in E}$  heißt *submarkovsche Halbgruppe* auf  $E$ , wenn für alle  $i, j \in E$  und alle  $t, s > 0$  gilt:

- (i)  $P_{ij}(t) \geq 0$ ,
- (ii)  $\sum_{k \in E} P_{ik}(t) \leq 1$ ,
- (iii)  $P_{ij}(t + s) = \sum_{k \in E} P_{ik}(t) P_{kj}(s)$  (*Chapman-Kolmogorov-Gleichung*)

Die Familie  $(P(t))_{t > 0}$  heißt *markovsch*, wenn zusätzlich in (ii) das Gleichheitszeichen für alle  $i \in E$  gilt, und sie heißt *standard*, wenn zusätzlich zu (i) - (iii) gilt:

(iv)  $\lim_{t \downarrow 0} P_{ij}(t) = \delta_{ij} (= P_{ij}(0))$  für alle  $i, j \in E$ .

**Definition 3.2.2.** Sei  $(P(t))_{t>0}$  eine submarkovsche Halbgruppe auf  $E$ . Weiter sei  $\partial \notin E$  und  $a$  eine Wahrscheinlichkeitsverteilung auf  $E \cup \{\partial\}$ . Dann heißt ein  $E \cup \{\partial\}$ -wertiger stochastischer Prozess  $(X(t))_{t \geq 0}$  eine *Markovkette in stetiger Zeit (MKSZ) zu  $a$  und  $(P(t))_{t>0}$* , wenn für alle  $n \in \mathbb{N}$ , und alle  $0 \leq t_1 < t_2 < \dots < t_n < t$  und alle  $i_1, \dots, i_n, i \in E$  gelten:

(i)  $\mathbb{P}(X(t) = i \mid X(t_1) = i_1, \dots, X(t_n) = i_n) = P_{i_n i}(t - t_n)$ , sofern die linke Seite definiert ist,

(ii)  $\mathbb{P}(X(0) = i) = a_i$  für alle  $i \in E \cup \{\partial\}$ ,

(iii)  $\mathbb{P}(X(t) = \partial \mid X(t_1) = i_1, \dots, X(t_{n-1}) = i_{n-1}, X(t_n) = \partial) = 1$ .

**Proposition 3.2.3.** Eine submarkovsche Halbgruppe auf  $E$  lässt sich zu einer markovschen auf  $E \cup \{\partial\}$  fortsetzen durch die Definition

$$P_{i\partial}(t) = \begin{cases} 1 & \text{falls } i = \partial \text{ und } t > 0, \\ 1 - \sum_{j \in E} P_{ij}(t) & \text{falls } i \in E \text{ und } t > 0. \end{cases} \quad (3.2.1)$$

In diesem Fall gilt die Bedingung (i) aus Definition 3.2.2 automatisch auch für  $i_1, \dots, i_n, i \in E \cup \{\partial\}$  (Übungsaufgabe). Die Fortsetzung ist die einzig mögliche genau dann, wenn die Halbgruppe nicht markovsch auf  $E$  ist (Übungsaufgabe).

**Satz 3.2.4.** Sei  $a$  eine Startverteilung auf  $E \cup \{\partial\}$ , und sei  $(P(t))_{t>0}$  submarkovsch. Dann existiert eine MKSZ  $\mathbb{X} = (X(t))_{t \geq 0}$  zu  $a$  und  $(P(t))_{t>0}$  im Sinne von Definition 3.2.2. Für diese MKSZ gilt für alle  $n \in \mathbb{N}_0$ , alle  $0 = t_0 < t_1 < \dots < t_n$  und alle  $i_0, \dots, i_n \in E$ :

$$\mathbb{P}(X(t_0) = i_0, X(t_1) = i_1, \dots, X(t_n) = i_n) = a_{i_0} P_{i_0 i_1}(t_1 - t_0) \cdots P_{i_{n-1} i_n}(t_n - t_{n-1}). \quad (3.2.2)$$

Insbesondere ist die Verteilung von  $\mathbb{X}$  eindeutig durch  $a$  und  $(P(t))_{t>0}$  bestimmt.

**Beweis.** Dies folgt wie im Fall von MKDZ mit Hilfe des Satzes von Kolmogorov 1.4.3. Wir verzichten auf eine genauere Begründung.  $\square$

**Bemerkung 3.2.5.** Wie im Fall einer MKDZ kann man (3.2.2) auch als Definition einer MKSZ zu  $a$  und  $(P(t))_{t>0}$  wählen.

Die bisherigen Aussagen lassen vermuten, dass sich MKSZ im wesentlichen wie MKDZ behandeln lassen. Dies ist insofern richtig, als zum Beispiel  $X(0), X(s), X(2s), X(3s), \dots$  für festes  $s > 0$  eine MKDZ zu  $a$  und der Matrix  $P(s)$  ist, wenn  $(X(t))_{t \geq 0}$  eine MKSZ zu  $a$  und  $(P(t))_{t>0}$  ist. Deutliche Unterschiede gibt es immer dort, wo man überabzählbar viele Zeitindizes gleichzeitig betrachtet, z.B. bei  $\sup_{s \in [0,1]} X(s)$ , wenn  $E = \mathbb{N}$  ist. Wir werden später darauf zurückkommen.

Ein weiterer Unterschied besteht offenbar darin, dass wir das einfache Objekt  $P$  im diskreten Fall durch ein kompliziertes – nämlich  $(P(t))_{t>0}$  – ersetzen müssen. Während man  $P$  sofort ansieht, ob es eine Übergangsmatrix ist, ist dies für (sub-)markovsche Halbgruppen viel schwieriger. In den Beispielen sahen wir bereits, dass die Beschreibung von MKSZ durch die  $Q$ -Matrix viel günstiger ist.  $Q$  beschreibt das Verhalten von  $P_{ij}(t)$  für infinitesimal kleine  $t > 0$ . Da man die komplette Halbgruppe aus der Kenntnis von  $(P_{ij}(t))_{0 < t \leq t_0}$  für alle  $i, j \in E$  und festes  $t_0 > 0$  rekonstruieren kann, ist es plausibel, dass dies auch aus der Kenntnis von  $\lim_{t \downarrow 0} t^{-1} P_{ij}(t) = q_{ij}$

möglich ist. Inwieweit dies stimmt, werden wir später sehen. Zunächst beweisen wir die Existenz der Grenzwerte  $q_{ij}$ . Dazu – und im folgenden fast immer – setzen wir voraus, dass  $(P(t))_{t>0}$  standard ist. Dies gilt nicht automatisch, aber in Anwendungsbeispielen praktisch immer. Wer an Aussagen über den Nichtstandardfall interessiert ist, dem sei das Buch von Chung ([Ch67]) empfohlen. Dort kann man auch solche Beweise nachlesen, die wir nicht bringen.

Die Voraussetzung “standard” (die übrigens bereits folgt, wenn man die Bedingung (iv) in Definition 3.2.1 nur für alle  $i = j \in E$  fordert) besagt, dass man nach einer kurzen Zeit mit großer Wahrscheinlichkeit im Startzustand liegt (sie besagt *nicht*, dass man innerhalb kurzer Zeiten den Startzustand nicht verlässt). Wir zeigen zunächst, dass im Standardfall  $P_{ij}(t)$  als Funktion von  $t$  gleichmäßig stetig (für feste  $i, j \in E$ ) ist. Ohne die Voraussetzung “standard” braucht  $t \mapsto P_{ij}(t)$  nicht einmal messbar zu sein (siehe [Ch67]).

Es ist klar, dass  $(P(t))_{t>0}$  standard auf  $E$  ist genau dann, wenn die in Proposition 3.2.3 definierte Fortsetzung auf  $E \cup \{\partial\}$  standard ist.

**Lemma 3.2.6.** *Sei  $(P(t))_{t \geq 0}$  standard. Dann gilt für  $i \in E$*

$$\limsup_{h \downarrow 0} \sup_{t \geq 0} \sum_{j \in E} |P_{ij}(t+h) - P_{ij}(t)| = 0. \quad (3.2.3)$$

*Insbesondere ist für  $i, j \in E$  die Abbildung  $t \mapsto P_{ij}(t)$  gleichmäßig stetig auf  $[0, \infty)$ .*

**Beweis.** Sei  $h > 0$ . Wir schätzen ab:

$$\begin{aligned} \sum_{j \in E} |P_{ij}(t+h) - P_{ij}(t)| &= \sum_{j \in E} \left| \sum_{k \in E} (P_{ik}(h)P_{kj}(t) - \delta_{ik}P_{kj}(t)) \right| \\ &\leq \sum_{j \in E} \sum_{k \in E} |P_{ik}(h) - \delta_{ik}| P_{kj}(t) = \sum_{k \in E} |P_{ik}(h) - \delta_{ik}| \sum_{j \in E} P_{kj}(t) \\ &\leq \sum_{k \in E} |P_{ik}(h) - \delta_{ik}| = 1 - P_{ii}(h) + \sum_{k \in E \setminus \{i\}} P_{ik}(h) \leq 2(1 - P_{ii}(h)). \end{aligned} \quad (3.2.4)$$

Nun folgt die Aussage, da der Term am Ende der Ungleichungskette nicht von  $t$  abhängt und für  $h \downarrow 0$  gegen Null konvergiert.  $\square$

**Satz 3.2.7.** *Sei  $(P(t))_{t \geq 0}$  standard. Dann existiert für alle  $i, j \in E$  der Grenzwert*

$$q_{ij} = \lim_{h \downarrow 0} \frac{P_{ij}(h) - \delta_{ij}}{h}. \quad (3.2.5)$$

*Im Fall  $i \neq j$  gilt  $0 \leq q_{ij} < \infty$ , im Fall  $i = j$  gilt  $0 \geq q_{ii} \geq -\infty$ . Ferner gilt für alle  $i \in E$ :*

$$\sum_{j \in E: j \neq i} q_{ij} \leq -q_{ii} =: q_i. \quad (3.2.6)$$

**Beweis.**<sup>2</sup> (Siehe z. B. [KT81] oder [GS75]).

1. *Teil:*  $i = j$ . Definiere

$$q'_i := \sup_{h > 0} \frac{1 - P_{ii}(h)}{h} \in [0, \infty].$$

<sup>2</sup>Der Beweis wird bei Prüfungen nicht abgefragt.

Wir werden zeigen, dass  $q'_i = -q_{ii} = \lim_{h \downarrow 0} h^{-1}(1 - P_{ii}(h))$  gilt. Seien  $c < q'_i$  und  $0 < t_0 < \infty$  so gewählt, dass

$$\frac{1 - P_{ii}(t_0)}{t_0} > c,$$

und seien  $0 < \tau < t_0$  und  $n \in \mathbb{N}$  so gewählt, dass

$$\tau \in \left[ \frac{t_0}{n+1}, \frac{t_0}{n} \right).$$

Dann gilt

$$\begin{aligned} c &< \frac{1}{t_0}(1 - P_{ii}(t_0)) \leq \frac{1}{t_0}(1 - (P_{ii}(\tau))^n P_{ii}(t_0 - n\tau)) \\ &\leq \frac{1 - (P_{ii}(\tau))^n}{t_0} + \frac{1 - P_{ii}(t_0 - n\tau)}{t_0} \leq \frac{n(1 - P_{ii}(\tau))}{n\tau} + \frac{1 - P_{ii}(t_0 - n\tau)}{t_0}, \end{aligned}$$

wobei wir beim zweiten Ungleichheitszeichen die Chapman-Kolmogorov-Gleichung, beim dritten die allgemeine Ungleichung  $(1 - ab) \leq 2 - a - b$ , falls  $a, b \in [0, 1]$  mit  $a = (P_{ii}(\tau))^n$  und  $b = P_{ii}(t_0 - n\tau)$  und beim vierten die Beziehung  $(1 - a^n) = (1 - a)(1 + a + a^2 + \dots + a^{n-1}) \leq n(1 - a)$  für  $a = P_{ii}(\tau)$  verwendet haben.

Bei festem  $t_0$  lassen wir nun  $\tau$  gegen Null und damit  $n \rightarrow \infty$  gehen, und somit konvergiert der letzte Summand wegen der "standard" Eigenschaft gegen Null. Es folgt

$$c < \liminf_{\tau \downarrow 0} \frac{1 - P_{ii}(\tau)}{\tau} \leq q'_i.$$

Da  $c < q'_i$  beliebig war, haben wir sogar

$$q'_i = \lim_{\tau \downarrow 0} \frac{1 - P_{ii}(\tau)}{\tau},$$

wie behauptet.

2. Teil:  $i \neq j$ . Sei  $\mathbb{X} = (X(t))_{t \geq 0}$  eine MKSZ mit submarkovscher (standard) Halbgruppe  $(P(t))_{t \geq 0}$  und Start in  $i$ . Definiere für  $i, j \in E$ ,  $n \in \mathbb{N}$  und  $h > 0$

$$\begin{aligned} {}_j P_{ii}^{(n)}(h) &= \mathbb{P}(X_{nh} = i, X_{rh} \neq j \text{ für alle } r = 1, \dots, n-1) \\ f_{ij}^{(n)}(h) &= \mathbb{P}(X_{nh} = j, X_{rh} \neq j \text{ für alle } r = 1, \dots, n-1). \end{aligned}$$

Sei  $\varepsilon \in (0, 1/3)$  vorgegeben. Da  $(P(t))_{t \geq 0}$  standard ist, existiert ein  $t_0 = t_0(\varepsilon) > 0$ , so dass für alle  $t \in [0, t_0]$  gilt:  $1 - P_{ii}(t) < \varepsilon$ ,  $1 - P_{jj}(t) < \varepsilon$  und  $P_{ij}(t) < \varepsilon$ . Seien  $n \in \mathbb{N}$  und  $h > 0$  gegeben mit  $nh \leq t_0$ , und sei  $u \in [nh, t_0]$ . Dann haben wir

$$\varepsilon > P_{ij}(u) \geq \sum_{r=1}^n f_{ij}^{(r)}(h) P_{jj}(u - rh) \geq (1 - \varepsilon) \sum_{r=1}^n f_{ij}^{(r)}(h), \quad \text{also gilt } \sum_{r=1}^n f_{ij}^{(r)}(h) \leq \frac{\varepsilon}{1 - \varepsilon}. \quad (3.2.7)$$

Außerdem haben wir

$$P_{ii}(rh) = {}_j P_{ii}^{(r)}(h) + \sum_{m=1}^{r-1} f_{ij}^{(m)}(h) P_{ji}((r-m)h), \quad r = 1, \dots, n. \quad (3.2.8)$$

Also folgt mit (3.2.7):

$${}_jP_{ii}^{(r)}(h) \geq P_{ii}(rh) - \sum_{m=1}^{r-1} f_{ij}^{(m)}(h) \geq 1 - \varepsilon - \frac{\varepsilon}{1 - \varepsilon} \geq \frac{1 - 3\varepsilon}{1 - \varepsilon}. \quad (3.2.9)$$

Daher haben wir

$$P_{ij}(u) \geq \sum_{r=0}^{n-1} {}_jP_{ii}^{(r)}(h) P_{ij}(h) P_{jj}(u - (r+1)h) \geq n(1 - 3\varepsilon) P_{ij}(h), \quad (3.2.10)$$

wobei wir  ${}_jP_{ii}^{(0)}(h) = 1$  setzten und bei der letzten Ungleichung (3.2.9) verwendeten.

Für festes  $u \in (0, t_0]$  wähle nun  $h > 0$  und  $n \in \mathbb{N}$  mit  $n \geq 2$ , so dass  $u \in [nh, (n+1)h]$ . Dann folgt  $n \geq (u-h)/h$  und aus (3.2.10):

$$\frac{P_{ij}(u)}{u-h} \geq (1 - 3\varepsilon) \frac{P_{ij}(h)}{h}. \quad (3.2.11)$$

Wir lassen für festes  $u > 0$  beidseitig  $h \downarrow 0$  gehen und erhalten

$$\infty > \frac{1}{1 - 3\varepsilon} \frac{P_{ij}(u)}{u} \geq \limsup_{h \downarrow 0} \frac{P_{ij}(h)}{h}. \quad (3.2.12)$$

Nun geht  $u \downarrow 0$ :

$$\frac{1}{1 - 3\varepsilon} \liminf_{u \downarrow 0} \frac{P_{ij}(u)}{u} \geq \limsup_{h \downarrow 0} \frac{P_{ij}(h)}{h}. \quad (3.2.13)$$

Mit  $\varepsilon \downarrow 0$  folgt, dass  $q_{ij} = \lim_{h \downarrow 0} P_{ij}(h)/h$  existiert und endlich ist.

3. Teil:

Sei  $M \subset E$  endlich. Dann gilt

$$\frac{1 - P_{ii}(h)}{h} \geq \sum_{j \in E \setminus \{i\}} \frac{P_{ij}(h)}{h} \geq \sum_{j \in M \setminus \{i\}} \frac{P_{ij}(h)}{h}. \quad (3.2.14)$$

Die linke Seite geht für  $h \downarrow 0$  gegen  $q_i = -q_{ii}$ , die rechte gegen  $\sum_{j \in M \setminus \{i\}} q_{ij}$ . Da  $M$  beliebig war, folgt  $-q_{ii} \geq \sum_{j \in E \setminus \{i\}} q_{ij}$ .  $\square$

**Bemerkung 3.2.8.** Der Fall  $-q_{ii} > \sum_{j \in E \setminus \{i\}} q_{ij}$  kann wirklich auftreten (siehe [Ch67, S. 275 ff]). Das am Anfang präsentierte Simulationsverfahren funktioniert in diesem Fall nicht, da man in einem solchen Fall mit positiver Wahrscheinlichkeit unendlich viele Sprünge in endlicher Zeit vollführt. Auch der Fall  $-q_{ii} = \infty$  ist möglich ([Ch67, S. 278ff]). Dies widerspricht nicht der Standardeigenschaft von  $(P(t))_{t \geq 0}$ , obwohl – wie wir gleich sehen werden – in diesem Fall der Zustand  $i$  sofort verlassen wird. Wir werden diese Phänomene in Beispiel 3.2.10 näher untersuchen.

**Beispiel 3.2.9.** Zur Vorbereitung auf das nächste Beispiel starten wir mit einem sehr einfachen Fall. Sei  $E = \{0, 1\}$ . Wir betrachten die MKSZ, die mit Rate  $\lambda > 0$  von 0 nach 1 und mit Rate  $\mu > 0$  von 1 nach 0 springt. Wir behaupten, dass diese die markovsche Halbgruppe

$$P(t) = I + B - Be^{-(\lambda+\mu)t}, \quad t > 0 \quad (3.2.15)$$

mit  $B = \frac{1}{\lambda + \mu} \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$  hat. Man prüft leicht nach, dass dies wirklich eine standard markovsche Halbgruppe ist und die zugehörige  $Q$ -Matrix gerade

$$Q = P'(0) = (\lambda + \mu)B = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

ist wie gewünscht (beim Nachweis der Halbgruppeneigenschaft ist dabei die Identität  $B = -B^2$  nützlich). Übrigens sieht man aus der Formel (3.2.15), dass

$$\lim_{t \rightarrow \infty} P(t) = I + B = \frac{1}{\lambda + \mu} \begin{pmatrix} \mu & \lambda \\ \mu & \lambda \end{pmatrix}$$

und damit wie im Fall einer irreduziblen und aperiodischen MKDZ die Übergangswahrscheinlichkeiten gegen einen von Startzustand unabhängigen Wert konvergieren.

**Beispiel 3.2.10.** Wir betrachten nun ein abzählbar unendliches Produkt von Ketten wie im vorherigen Beispiel. Die einzelnen Ketten seien mit  $i \in \mathbb{N}_0$  indiziert und unabhängig. Die zugehörigen Raten seien  $\lambda_i = 1$  und  $\mu_i > 0$ . Wir stellen uns vor, dass jede einzelne Kette eine Lampe beschreibt, die mit Rate 1 anspringt (den Wert 1 annimmt) und mit Rate  $\mu_i$  ausgeht (wenn sie an ist). Die Gesamtkonfiguration, die beschreibt, welche Lampen zur Zeit  $t$  eingeschaltet sind, ist also ein Element in  $\{0, 1\}^{\mathbb{N}_0}$ . Nun ist dieser Raum überabzählbar, aber wenn die  $\mu_i$  mit  $i \rightarrow \infty$  hinreichend schnell gegen unendlich gehen, dann besteht die Hoffnung, dass zu jedem festen Zeitpunkt fast sicher nur endlich viele Lampen an sind und wir mit dem abzählbaren Zustandsraum  $E$  aller Elemente in  $\{0, 1\}^{\mathbb{N}_0}$ , bei denen alle bis auf endlich viele Folgenglieder 0 sind, arbeiten können.

Wir sind in diesem Beispiel in der glücklichen Situation, dass wir die Übergangswahrscheinlichkeiten aufgrund des letzten Beispiels explizit als Produkt angeben können. Ein hinreichendes Kriterium dafür, dass zur Zeit  $t > 0$  bei Start in  $(i_0, i_1, \dots)$  fast sicher nur endlich viele Lampen an sind, ist, dass der Erwartungswert der Anzahl eingeschalteter Lampen endlich ist. Dieser lässt sich aber mit dem letzten Beispiel berechnen und zeigen, dass dazu die Konvergenz der Reihe  $\sum_i (1 + \mu_i)^{-1}$  notwendig und hinreichend ist (das werden wir in der Vorlesung zeigen). Man kann leicht zeigen, dass die markovsche Halbgruppe dann sogar standard ist, das heißt: kurz nach dem Start (sagen wir zur Zeit  $\varepsilon > 0$ ) werden mit großer Wahrscheinlichkeit dieselben Lichter an oder aus sein wie am Anfang. Dennoch wird es mit Wahrscheinlichkeit 1 sogar unendlich viele Lampen geben, die zur Zeit 0 aus waren und bis zur Zeit  $\varepsilon$  angesprungen und wieder ausgegangen sind und es wird sogar unendlich viele Lampen geben, die bis  $\varepsilon$  mindestens 100 mal an und wieder ausgingen. Es ist interessant in diesem Beispiel die  $Q$ -Matrix zu berechnen. Wir werden das (voraussichtlich) in der Vorlesung tun.

Wir wollen nun im Standardfall für  $i \in E$  die Verteilung des Zeitpunkts des ersten Sprunges,

$$\tau = \inf\{s > 0: X(s) \neq X(0)\} \tag{3.2.16}$$

bei Start in  $i$  bestimmen. Wie zu Beginn des Kapitels vereinbart, bezeichnet  $\mathbb{P}_i$  die Wahrscheinlichkeit bei Start in  $i$ .

Wir hatten uns anschaulich bereits überlegt, dass bei Start in  $i$  die Zeit  $\tau$   $\text{Exp}(q_i)$ -verteilt sein müsste. Dies ist allerdings ohne weitere Voraussetzungen nicht richtig. Man braucht eine Bedingung an die Trajektorien, die sichert, dass diese (auf Lebesguenullmengen) nicht zu wild



aussehen. Die Tatsache, dass  $(P(t))_{t \geq 0}$  standard ist, garantiert dies nicht! Allerdings kann man im Standardfall eine  $E \cup \{\partial\}$ -wertige MKSZ  $\mathbb{X} = (X(t))_{t \geq 0}$  so definieren, dass sie *separabel* im folgenden Sinn ist: Für jede abzählbare dichte Menge  $M \subset [0, \infty)$  existiert ein  $N \in \mathcal{F}$  mit  $\mathbb{P}(N) = 0$ , so dass für alle  $\omega \notin N$ ,  $t \geq 0$  und  $\varepsilon > 0$  die Zahl  $X(t, \omega)$  im Abschluss der Menge  $\{X(s, \omega) : s \in [t - \varepsilon, t + \varepsilon] \cap M\}$  liegt, wobei wir eine Teilmenge von  $E \cup \{\partial\}$  als abgeschlossen bezeichnen, wenn sie endlich ist oder  $\partial$  enthält (siehe [Ch67], S. 143 f. und S. 145 unten).

**Proposition 3.2.11.** *Sei  $(P(t))_{t \geq 0}$  standard und  $\mathbb{X}$  eine zugehörige separable MKSZ. Dann gilt für die in (3.2.16) definierte Zeit  $\tau$*

$$\mathbb{P}_i(\tau \geq t) = e^{-q_i t} \quad \text{für alle } t \geq 0, \quad (3.2.17)$$

(wobei wir die Konvention  $\infty \cdot 0 = 0$  benutzen).

**Beweis.** Für  $t = 0$  ist die Behauptung klar. Sei  $t > 0$  und zunächst  $q_i < \infty$ . Dann gilt

$$\begin{aligned} \mathbb{P}_i(\tau \geq t) &= \mathbb{P}_i(X(s) = i \text{ für alle } s \in [0, t)) \\ &= \mathbb{P}_i(X(k2^{-n}t) = i \text{ für alle } n \in \mathbb{N} \text{ und alle } k = 0, \dots, 2^n - 1) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}_i(X(k2^{-n}t) = i \text{ für alle } k = 0, \dots, 2^n - 1) \\ &= \lim_{n \rightarrow \infty} (P_{ii}(t2^{-n}))^{2^n - 1} = \lim_{n \rightarrow \infty} (1 - q_i t 2^{-n} + o(t 2^{-n}))^{2^n} = e^{-q_i t}, \end{aligned} \quad (3.2.18)$$

wobei wir beim zweiten Gleichheitszeichen die Separabilität, beim dritten die Stetigkeit von  $\mathbb{P}$  und beim vierten die Markoveigenschaft benutzt haben. Der Fall  $q_i = \infty$  ergibt sich leicht mit Hilfe eines Grenzübergangs.  $\square$

Wir würden nun gerne – um das Simulationsverfahren zu rechtfertigen – zeigen, dass bei Start in  $i$  die Zufallsgrößen  $\tau$  und  $X(\tau)$  unabhängig sind und  $\mathbb{P}(X(\tau) = j) = q_{ij}/q_i$  gilt, sofern  $q_i > 0$ . Unter der Voraussetzung “standard” ist dies selbst für separable MKSZ nicht unbedingt richtig. Bevor wir zeigen, in welchem Sinne und unter welchen Voraussetzungen die Aussage richtig ist, definieren wir, was eine konservative  $Q$ -Matrix ist.

**Definition 3.2.12.** Eine Abbildung  $Q: E \times E \rightarrow \mathbb{R}$  heißt eine *konservative  $Q$ -Matrix*, wenn gelten

- (i)  $q_{ij} \geq 0$  für alle  $i, j \in E$  mit  $i \neq j$ ,
- (ii)  $q_i := -q_{ii} = \sum_{j \in E \setminus \{i\}} q_{ij} < \infty$  für alle  $i \in E$ .

**Bemerkung 3.2.13.** Nicht jede Matrix  $Q = (q_{ij})_{i, j \in E}$ , die sich aus einer Standardhalbgruppe  $(P(t))_{t \geq 0}$  wie in Satz 3.2.7 ergibt, ist konservativ. Andererseits ist Konservativität eine notwendige Voraussetzung dafür, dass das beschriebene Simulationsverfahren funktioniert. Wir werden später sehen, dass es zu jeder konservativen  $Q$ -Matrix mindestens eine Standardhalbgruppe  $(P(t))_{t \geq 0}$  gibt mit der Eigenschaft, dass  $q_{ij}$  gleich den Grenzwerten in Satz 3.2.7 ist.

**Satz 3.2.14.** *Sei  $(P(t))_{t \geq 0}$  standard und  $(X(t))_{t \geq 0}$  eine zugehörige separable MKSZ mit Start in  $i \in E$ . Seien  $j \in E \setminus \{i\}$ ,  $q_i \in (0, \infty)$ ,  $q_j < \infty$  und  $\tau$  wie in (3.2.16). Dann gilt für alle  $u \geq 0$ :*

$$\mathbb{P}_i(\tau \geq u, \text{ und es existiert } s = s(\omega) > 0: X(t) = j \text{ für alle } t \in (\tau, \tau + s]) = e^{-q_i u} \frac{q_{ij}}{q_i}. \quad (3.2.19)$$

*Ist  $Q$  konservativ, so existiert  $\lim_{h \downarrow 0} X(\tau + h) = J(\omega)$  fast sicher, ist unabhängig von  $\tau$  und hat die Verteilung  $\mathbb{P}_i(J = j) = q_{ij}/q_i$ ,  $j \in E \setminus \{i\}$ .*

**Beweis.** Wir definieren die Ereignisse

$$B_\gamma = \{\omega: \tau \geq u, X(t) = j \text{ für alle } t \in (\tau, \tau + \gamma)\}, \quad \gamma > 0, \quad (3.2.20)$$

$$B = \{\omega: \tau \geq u, \text{ es existiert } s = s(\omega) > 0: X(t) = j \text{ für alle } t \in (\tau, \tau + s]\}. \quad (3.2.21)$$

Dann gilt  $B_\gamma \uparrow B$  für  $\gamma \downarrow 0$ .

Wir werden zeigen, dass

$$\mathbb{P}_i(B_\gamma) = e^{-q_i u} \frac{q_{ij}}{q_i} e^{-q_j \gamma} \quad (3.2.22)$$

gilt, woraus dann aufgrund der Stetigkeit von  $\mathbb{P}_i$  folgt, dass  $\mathbb{P}_i(B) = e^{-q_i u} q_{ij}/q_i$  gilt (beachte, dass  $q_j < \infty$ ).

Fixiere  $u \geq 0$ . Sei für  $n \in \mathbb{N}, m \in \mathbb{N}$  und  $\gamma := 2^{-m}$

$$B_{n,\gamma} := \{\omega: \text{ es existiert } s \geq u, \text{ so dass } X(k2^{-n}) = i \text{ für alle } k2^{-n} < s, \\ \text{ und } X(k2^{-n}) = j \text{ für alle } s \leq k2^{-n} < s + \gamma\}. \quad (3.2.23)$$

Dann existiert eine Menge  $B'_\gamma$  mit  $B_{n,\gamma} \downarrow B'_\gamma$  für  $n \rightarrow \infty$ , und wegen der Separabilität von  $\mathbb{X}$  gilt  $\mathbb{P}_i(B'_\gamma) = \mathbb{P}_i(B_\gamma)$ . Also genügt es zu zeigen, dass  $\lim_{n \rightarrow \infty} \mathbb{P}_i(B_{n,\gamma}) = e^{-q_i u} \frac{q_{ij}}{q_i} e^{-q_j \gamma}$  gilt. Für  $n \geq m$  folgt:

$$\mathbb{P}_i(B_{n,\gamma}) = \sum_{k=\lfloor u2^n \rfloor}^{\infty} (P_{ii}(2^{-n}))^k P_{ij}(2^{-n}) (P_{jj}(2^{-n}))^{2^{n-m}-1} \\ = P_{ij}(2^{-n}) P_{jj}(2^{-n})^{2^{n-m}-1} P_{ii}(2^{-n})^{\lfloor u2^n \rfloor} \frac{1}{1 - P_{ii}(2^{-n})}, \quad (3.2.24)$$

wobei  $\lfloor x \rfloor$  die größte ganze Zahl echt kleiner als  $x$  sei (wegen  $P_{ii}(2^{-n}) < 1$  für  $n$  hinreichend groß konvergiert die Reihe). Die vier Faktoren sind für  $n \rightarrow \infty$  in obiger Reihenfolge asymptotisch gleich

$$q_{ij}2^{-n}, \quad e^{-q_j \gamma}, \quad e^{-q_i u}, \quad \frac{1}{q_i 2^{-n}}.$$

Also ist die erste Behauptung, (3.2.22), bewiesen.

Ist  $Q$  konservativ, dann folgt die Existenz des Grenzwerts  $J(\omega)$  und die Formel für die Verteilung, indem man in (3.2.19)  $u = 0$  setzt und über alle  $j \in E \setminus \{i\}$  summiert. Die Unabhängigkeit ist klar, da

$$\mathbb{P}_i(J = j \mid \tau \geq u) = \frac{q_{ij}}{q_i}$$

nicht von  $u$  abhängt. □

### 3.3 Vorwärts- und Rückwärtsgleichungen

Mit Satz 3.2.14 ist das Simulationsverfahren leider immer noch nicht vollständig gerechtfertigt, auch nicht im konservativen Fall. Dazu müsste man wissen, dass z. B. auch der zweite Sprung gegeben den zweiten Zustand unabhängig von der Zeit ist, die man im ersten und zweiten Zustand verbracht hat. Diese Tatsache ist richtig und lässt sich entweder durch eine Erweiterung von Satz 3.2.14 zeigen (indem man die gemeinsame Verteilung der ersten  $k$  Sprünge und der Sprungzeiten berechnet, was möglich, aber mühsam ist) oder aus der starken Markoveigenschaft,

die wir aber weder formulieren noch beweisen wollen (siehe z. B. [Ch67]). Wir betonen, dass die (ähnlich wie im diskreten Fall definierte) starke Markoveigenschaft ohne Separabilität im allgemeinen nicht gilt und selbst im separablen Fall nicht gelten muss; z. B. dann nicht, wenn man als Stoppzeit  $\tau = \inf\{s \geq 0: X(s) = \partial\}$  wählt und ein Rücksprung nach  $E$  möglich ist. Die Rücksprungverteilung kann nämlich explizit von dem Verhalten von  $\mathbb{X}$  kurz vor  $\tau$  abhängen, ohne dass dies die Markoveigenschaft zerstört! Die starke Markoveigenschaft kann dadurch zerstört werden, dass die ‘‘Kompaktifizierung’’  $E \cup \{\partial\}$  von  $E$  zu grob ist. Die Theorie der *Ray-Knight-Kompaktifizierung* zeigt, wie man, abhängig von der Standardhalbgruppe  $(P(t))_{t \geq 0}$ , die Menge  $E$  so kompaktifiziert, dass eine MKSZ  $\mathbb{X}$  mit Werten in der Kompaktifizierung existiert mit den geforderten endlich-dimensionalen Verteilungen (zu  $(P(t))_{t \geq 0}$ ), rechtsstetige Pfade hat und die starke Markoveigenschaft erfüllt. Der interessierte Leser sei auf [Wi79] verwiesen.

Als nächstes stellen wir uns die Frage, ob und gegebenenfalls wie man aus einer konservativen  $Q$ -Matrix die Standardhalbgruppe  $(P(t))_{t \geq 0}$  berechnen kann.

**Satz 3.3.1** (Rückwärtsgleichung). *Sei  $(P(t))_{t \geq 0}$  standard mit konservativer  $Q$ -Matrix  $Q$ . Dann ist  $t \mapsto P_{ij}(t)$  auf  $[0, \infty)$  stetig differenzierbar für alle  $i, j \in E$ , und es gelten*

$$P'(t) = QP(t), \quad t \geq 0, \quad \text{und} \quad P(0) = I, \quad (3.3.1)$$

wobei  $I$  die Identität auf  $E$  ist (d. h.  $I_{ij} = \delta_{ij}$ ) und die Ableitung komponentenweise zu verstehen ist.

**Beweis.** Für  $h > 0$  und  $s \geq 0$  gilt:

$$\begin{aligned} \frac{P_{ij}(h+s) - P_{ij}(s)}{h} &= \frac{1}{h} \left[ \sum_{k \in E} P_{ik}(h)P_{kj}(s) - P_{ij}(s) \right] \\ &= \frac{1}{h} \left[ (P_{ii}(h) - 1)P_{ij}(s) + \sum_{k \in E \setminus \{i\}} P_{ik}(h)P_{kj}(s) \right]. \end{aligned} \quad (3.3.2)$$

Falls die eventuell unendliche Summe mit dem Grenzwert  $h \downarrow 0$  vertauscht werden darf, so erhalten wir aus (3.3.2) leicht:

$$\lim_{h \downarrow 0} \frac{P_{ij}(h+s) - P_{ij}(s)}{h} = q_{ii}P_{ij}(s) + \sum_{k \in E \setminus \{i\}} q_{ik}P_{kj}(s) = \sum_{k \in E} q_{ik}P_{kj}(s). \quad (3.3.3)$$

Wir rechtfertigen nun die Vertauschung. Seien  $\varepsilon > 0$  und  $J \subset E$  mit  $i \in J$  und  $|J| < \infty$ , so dass  $\sum_{k \in E \setminus J} q_{ik} < \varepsilon/2$  (hier benutzen wir, dass  $Q$  konservativ ist). Dann gilt

$$\begin{aligned} \left| \sum_{k \in E \setminus J} \left( \frac{P_{ik}(h)}{h} - q_{ik} \right) P_{kj}(s) \right| &\leq \left| \sum_{k \in E \setminus J} \frac{P_{ik}(h)}{h} \right| + \sum_{k \in E \setminus J} q_{ik} \\ &< \left| \frac{1 - P_{ii}(h)}{h} - \sum_{k \in J \setminus \{i\}} \frac{P_{ik}(h)}{h} \right| + \frac{\varepsilon}{2} \\ &\xrightarrow{h \downarrow 0} -q_{ii} - \sum_{k \in J \setminus \{i\}} q_{ik} + \frac{\varepsilon}{2} = \sum_{k \in E \setminus J} q_{ik} + \frac{\varepsilon}{2} < \varepsilon, \end{aligned} \quad (3.3.4)$$

wobei wir beim letzten Gleichheitszeichen die Konservativität von  $Q$  benutzen.

Damit ist die Konvergenz in (3.3.3) gerechtfertigt, und es folgt

$$P'(s) = QP(s), \quad (3.3.5)$$

wobei allerdings der Strich zunächst nur als rechtsseitige Ableitung zu verstehen ist, da  $h > 0$  war.

Aus Lemma 3.2.6 wissen wir, dass  $P_{ij}(\cdot)$  gleichmäßig stetig ist. Wegen der Konservativität von  $Q$  und Beschränktheit der  $P_{ij}(\cdot)$  durch 1 konvergiert  $\sum_{k \in E} q_{ik} P_{kj}(s)$  gleichmäßig für alle  $s \in [0, \infty)$ . Da gleichmäßige Grenzwerte stetiger Funktionen stetig sind, folgt, dass  $P'_{ij}(\cdot)$  stetig ist. Nun gilt aber allgemein, dass eine stetige Funktion mit existierender und stetiger rechtsseitiger Ableitung sogar stetig differenzierbar ist (da dies nicht ganz so trivial ist, wie man meinen könnte, zeigen wir dies in Lemma 3.3.2). Da  $P(0) = I$  im Standardfall immer gilt, folgt die Behauptung.  $\square$

**Lemma 3.3.2.** *Sei  $f: [0, t_0] \rightarrow \mathbb{R}$  stetig und auf  $[0, t_0)$  rechtsseitig differenzierbar mit stetiger Ableitung  $f^+$ . Dann ist  $f$  auf  $(0, t_0)$  stetig differenzierbar mit Ableitung  $f^+$ .*

**Beweis (nach M. Schäl).** Setze  $F(t) := f(0) + \int_0^t f^+(s) ds$  für  $t \in [0, t_0]$ . Da nach dem Hauptsatz der Differential- und Integralrechnung  $F$  stetig differenzierbar mit Ableitung  $f^+$  ist, genügt es zu zeigen, dass  $F = f$  ist.

Sei  $\varphi(t) := F(t) - f(t)$  für  $t \in [0, t_0]$ . Dann ist  $\varphi$  stetig,  $\varphi(0) = 0$  und  $\varphi^+(t) = 0$  für  $t \in [0, t_0)$ . Zu zeigen ist  $\varphi = 0$ . Wäre dem nicht so, dann gälte  $\max_{t \in [0, t_0]} \varphi(t) > 0$  oder  $\min_{t \in [0, t_0]} \varphi(t) < 0$ . Es genügt, den ersten Fall zu betrachten. Wegen der Stetigkeit von  $\varphi$  existiert ein  $t^* \in (0, t_0]$  mit  $\varphi(t^*) = \max_{t \in [0, t_0]} \varphi(t)$ . Sei  $\gamma(s) := \varphi(s) - \frac{s}{t^*} \varphi(t^*)$  für  $s \in [0, t^*]$ . Dann gelten

$$\gamma^+(s) = -\frac{\varphi(t^*)}{t^*} < 0 \quad \text{für } s \in [0, t^*) \quad \text{und} \quad \gamma(0) = 0 = \gamma(t^*). \quad (3.3.6)$$

Sei  $s^* \in [0, t^*)$  so gewählt, dass  $\gamma(s^*) = \min_{s \in [0, t^*]} \gamma(s)$ . Dann folgt

$$\gamma^+(s^*) = \lim_{h \downarrow 0} \frac{\gamma(s^* + h) - \gamma(s^*)}{h} \geq 0 \quad (3.3.7)$$

im Widerspruch zu (3.3.6).  $\square$

**Satz 3.3.3** (Vorwärtsgleichung). *Sei  $(P(t))_{t \geq 0}$  standard mit konservativer  $Q$ -Matrix  $Q$ . Weiter sei*

$$c := \sup_{i \in E} q_i < \infty. \quad (3.3.8)$$

Dann gelten

$$P'(t) = P(t)Q, \quad t \geq 0, \quad \text{und} \quad P(0) = I. \quad (3.3.9)$$

**Beweis.** Sei  $h > 0$ . Es gilt

$$\left| \frac{P_{kj}(h) - \delta_{kj}}{h} \right| \leq \frac{1 - P_{kk}(h)}{h} \leq q_k; \quad (3.3.10)$$

letzteres wurde im 1. Teil des Beweises von Satz 3.2.7 gezeigt. Somit gilt für festes  $t \geq 0$

$$\frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \sum_{k \in E} P_{ik}(t) \frac{P_{kj}(h) - \delta_{kj}}{h} \xrightarrow{h \downarrow 0} \sum_{k \in E} P_{ik}(t) q_{kj}, \quad (3.3.11)$$

denn zu jedem  $\varepsilon > 0$  existiert ein endliches (von  $t$  abhängiges)  $J \subset E$  mit  $j \in J$  und  $\sum_{k \in E \setminus J} P_{ik}(t) < \varepsilon/c$  und

$$\sum_{k \in E \setminus J} \left| P_{ik}(t) \left( \frac{P_{kj}(h) - \delta_{kj}}{h} - q_{kj} \right) \right| \leq c \frac{\varepsilon}{c} = \varepsilon. \quad (3.3.12)$$

Nach Satz 3.3.1 wissen wir, dass  $P_{ij}(\cdot)$  stetig differenzierbar ist, also folgt die Behauptung.  $\square$

**Beispiel 3.3.4.** Sei  $Q$  die (konservative)  $Q$ -Matrix des Poissonprozesses. Bislang wissen wir noch nicht, ob die zum Poissonprozess gehörige Halbgruppe wirklich standard ist, dennoch versuchen wir, die Rückwärts- und Vorwärtsgleichung zu  $Q$  zu lösen. Die Rückwärtsgleichung lautet ausgeschrieben:

$$P'_{ij}(t) = \sum_{k \in E} q_{ik} P_{kj}(t) = \lambda P_{i+1,j}(t) - \lambda P_{ij}(t), \quad t > 0, i, j \in \mathbb{N}_0, \quad (3.3.13)$$

$$P_{ij}(0) = \delta_{ij}, \quad i, j \in \mathbb{N}_0. \quad (3.3.14)$$

Wir werden später sehen, dass dieses System eine eindeutige Lösung hat, aber man sieht bereits jetzt, dass sich hier die Rückwärtsgleichung nicht besonders zur Berechnung von  $P(t)$  eignet (zur Berechnung von  $P_{0j}(t)$  muss man bereits  $P_{1j}(t)$  kennen usw.). Die Vorwärtsgleichung lautet ausgeschrieben:

$$P'_{ij}(t) = \sum_{k \in E} P_{ik}(t) q_{kj} = \begin{cases} \lambda P_{i,j-1}(t) - \lambda P_{ij}(t) & \text{falls } j \geq 1, \\ -\lambda P_{i0}(t), & \text{falls } j = 0, \end{cases} \quad (3.3.15)$$

$$P_{ij}(0) = \delta_{ij}, \quad i, j \in \mathbb{N}_0. \quad (3.3.16)$$

Hier kann man für festes  $i \in \mathbb{N}_0$  das System rekursiv für  $j = 0, 1, 2, \dots$  lösen. Für  $j = 0$  erhält man:

$$P_{i0}(t) = \begin{cases} 0 & \text{falls } i \geq 1, \\ e^{-\lambda t}, & \text{falls } i = 0. \end{cases} \quad (3.3.17)$$

Dies setzt man in die Gleichung für  $j = 1$  ein. Mit dieser Methode zeigt man leicht, dass

$$P_{ij}(t) = \begin{cases} 0 & \text{falls } i > j, \\ e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}, & \text{falls } i \leq j \end{cases} \quad (3.3.18)$$

gilt. Die Anzahl der Sprünge des Poissonprozesses in einem Zeitintervall der Länge  $t$  ist also Poisson-verteilt mit Parameter  $\lambda t$ . Man kann nun explizit nachrechnen, dass diese (einzige) Lösung der Vorwärtsgleichung wirklich eine markovsche Standardhalbgruppe ist. Dies wird sich allerdings gleich als unnötig herausstellen (siehe Satz 3.3.6).

Bislang wissen wir nicht, ob zu einer (konservativen)  $Q$ -Matrix immer eine Standardhalbgruppe gehört und wann diese eindeutig ist. Die Existenz wird im folgenden Satz sichergestellt. Dabei können wir die Konservativität von  $Q$  etwas abschwächen.

**Definition 3.3.5.**  $Q: E \times E \rightarrow \mathbb{R}$  heißt *schwach konservativ*, wenn gelten:

- (i)  $q_{ij} \geq 0$  für alle  $i, j \in E$  mit  $i \neq j$ ,

- (ii)  $q_{ii} > -\infty$  für alle  $i \in E$ ,
- (iii)  $\sum_{j \in E} q_{ij} \leq 0$  für alle  $i \in E$ .

Wieder definieren wir  $q_i := -q_{ii}$  für alle  $i \in E$ .

Unser Ziel besteht darin, zu einer gegebenen schwach konservativen Matrix  $Q$  eine Standardhalbgruppe  $(\bar{P}(t))_{t \geq 0}$  zu finden, die die Vorwärts- und Rückwärtsgleichung löst, für die also insbesondere (Rückwärtsgleichung für  $t = 0$ )  $\bar{P}'(0) = Q$  ist, d.h.  $Q$  ist die  $Q$ -Matrix von  $(\bar{P}(t))_{t \geq 0}$ . Wir wissen dann insbesondere, dass, wenn  $Q$  schwach konservativ ist und die VWG (oder RWG) eine eindeutige Lösung hat, diese automatisch eine Standardhalbgruppe sein muss (vgl. Bemerkung am Ende von Beispiel 3.3.4).

**Satz 3.3.6.** *Sei  $Q$  eine schwach konservative Matrix. Dann existiert eine Standardhalbgruppe  $(\bar{P}(t))_{t \geq 0}$ , die die VWG und RWG löst und für die gilt*

$$\bar{P}_{ij}(t) \leq Z_{ij}(t), \quad t \geq 0, i, j \in E, \quad (3.3.19)$$

für jede Standardhalbgruppe  $(Z(t))_{t \geq 0}$  mit  $Q$ -Matrix  $Q$ . Diese Standardhalbgruppe  $(\bar{P}(t))_{t \geq 0}$  heißt Minimallösung.

**Beweisansatz.** Wir zeigen nur die Konstruktion der Standardhalbgruppe  $(\bar{P}(t))_{t \geq 0}$  (die uns beim Beweis von Satz 3.4.4 nützlich sein wird), ohne aber zu beweisen, dass sie wirklich die geforderten Eigenschaften hat. Für den vollständigen Beweis siehe [Ch67, S. 251 ff].

Definiere für  $i, j \in E, t \geq 0$

$$P_{ij}^0(t) := \delta_{ij} e^{-q_i t} \quad (3.3.20)$$

und für  $n \in \mathbb{N}_0$  induktiv

$$P_{ij}^{n+1}(t) := \sum_{k \in E \setminus \{j\}} \int_0^t P_{ik}^n(s) q_{kj} e^{-q_j(t-s)} ds. \quad (3.3.21)$$

Induktiv zeigt man, dass  $P_{ij}^n(\cdot)$  stetig differenzierbar ist und

$$H_{ij}^N(t) := \sum_{n=0}^N P_{ij}^n(t) \leq 1 \quad (3.3.22)$$

ist. Schließlich definiert man für  $t \geq 0$

$$\bar{P}_{ij}(t) := \lim_{N \uparrow \infty} H_{ij}^N(t) \quad (3.3.23)$$

(der Grenzwert existiert und ist  $\leq 1$ ). Es bleibt zu zeigen, dass  $(\bar{P}(t))_{t \geq 0}$  standard ist und die VWG und RWG löst, sowie die Minimalitätseigenschaft. Der Beweis der Gültigkeit der VWG ist mit der obigen Rekursion relativ leicht, für den Nachweis der RWG zeigt man zunächst, dass die oben definierten  $P_{ij}^n$  auch der Rekursion

$$P_{ij}^{n+1}(t) = \sum_{k \in E \setminus \{i\}} \int_0^t e^{-q_i(t-s)} q_{ik} P_{kj}^n(s) ds \quad (3.3.24)$$

genügen. Die letzte Gleichung läßt sich anschaulich wie folgt interpretieren:  $P_{ij}^n(t)$  ist die Wahrscheinlichkeit, bei Start in  $i$  mit genau  $n$  Sprüngen zur Zeit  $t$  in  $j$  zu landen.  $\bar{P}_{ij}(t)$  ist dann die Wahrscheinlichkeit, bei Start in  $i$  mit endlich vielen Sprüngen zur Zeit  $t$  in  $j$  zu landen.  $\square$

**Bemerkung 3.3.7.** Die Minimallösung  $(\bar{P}(t))_{t \geq 0}$  ist die Halbgruppe derjenigen MKSZ  $\mathbb{X}$  zu  $Q$ , die nach der ersten *Explosion* – d. h. dem ersten Zeitpunkt zu dem  $\mathbb{X}$  entweder in den Zustand  $\partial$  springt, oder dem Infimum der Zeitpunkte, an denen  $\mathbb{X}$  unendlich viele Zustände besucht hat – für immer in  $\partial$  bleibt. Jede MKSZ  $\tilde{\mathbb{X}}$  zu einer anderen Standardlösung  $Z$  verhält sich bis zur Explosion genauso wie  $\mathbb{X}$ , springt dann aber von  $\partial$  aus auf irgendeine Weise zurück nach  $E$ , weswegen  $\bar{P}_{ij}(t) \leq Z_{ij}(t)$  gilt.

**Korollar 3.3.8.** Sei  $Q$  schwach konservativ. Wenn die Minimallösung  $(\bar{P}(t))_{t \geq 0}$  markovsch ist, dann ist sie die einzige Standardhalbgruppe mit  $Q$ -Matrix  $Q$ .

**Beweis.** Sei  $(Z(t))_{t \geq 0}$  auch eine Standardhalbgruppe zu  $Q$ , dann folgt aus Satz 3.3.6, dass  $Z_{ij}(t) \geq \bar{P}_{ij}(t)$  für alle  $i, j \in E$  und  $t \geq 0$ , also

$$1 \geq \sum_{j \in E} Z_{ij}(t) \geq \sum_{j \in E} \bar{P}_{ij}(t) = 1, \quad i \in E, t \geq 0, \quad (3.3.25)$$

also folgt  $Z_{ij}(t) = \bar{P}_{ij}(t)$  für alle  $i, j \in E$  und  $t \geq 0$ . □

**Proposition 3.3.9.** Sei  $Q$  konservativ und  $c := \sup_{i \in E} q_i < \infty$ . Dann ist die Minimallösung  $(\bar{P}(t))_{t \geq 0}$  markovsch.

**Beweis.** Unser Ziel besteht darin, zu zeigen, dass  $\frac{d}{dt} \sum_{j \in E} \bar{P}_{ij}(t) = 0$  für alle  $t \geq 0$  gilt, woraus wegen  $\sum_{j \in E} \bar{P}_{ij}(0) = 1$  folgt, dass  $(\bar{P}(t))_{t \geq 0}$  markovsch ist.

Sei  $J \subset E$  endlich. Da  $(\bar{P}(t))_{t \geq 0}$  die VWG erfüllt, gilt

$$\frac{d}{dt} \sum_{j \in J} \bar{P}_{ij}(t) = \sum_{j \in J} \bar{P}'_{ij}(t) = \sum_{j \in J} \sum_{k \in E \setminus \{j\}} \bar{P}_{ik}(t) q_{kj} - \sum_{j \in J} \bar{P}_{ij}(t) q_j. \quad (3.3.26)$$

Nun ist  $\sum_{j \in J} \sum_{k \in E \setminus \{j\}} \bar{P}_{ik}(\cdot) q_{kj}$  stetig, denn  $|J| < \infty$ , und  $\sum_{k \in E \setminus \{j\}} \bar{P}_{ik}(t) q_{kj} = \bar{P}'_{ij}(t) + \bar{P}_{ij}(t) q_j$  ist stetig in  $t$  nach Satz 3.3.1.

Weiter konvergiert mit  $J \uparrow E$

$$\sum_{j \in J} \sum_{k \in E \setminus \{j\}} \bar{P}_{ik}(\cdot) q_{kj} \xrightarrow{\text{monoton}} \sum_{j \in E} \sum_{k \in E \setminus \{j\}} \bar{P}_{ik}(\cdot) q_{kj} = \sum_{k \in E} \bar{P}_{ik}(\cdot) q_k, \quad (3.3.27)$$

da  $Q$  konservativ ist, und

$$\sum_{j \in J} \bar{P}_{ij}(t) q_j \xrightarrow{\text{monoton}} \sum_{j \in E} \bar{P}_{ij}(t) q_j. \quad (3.3.28)$$

Wir zeigen, dass  $t \mapsto \sum_{j \in E} \bar{P}_{ij}(t) q_j$  stetig ist:

$$\left| \sum_{j \in E} (\bar{P}_{ij}(t+h) - \bar{P}_{ij}(t)) q_j \right| \leq c \sum_{j \in E} |\bar{P}_{ij}(t+h) - \bar{P}_{ij}(t)| \xrightarrow{h \downarrow 0} 0 \quad (3.3.29)$$

gleichmäßig für alle  $t \in [0, \infty)$  nach Lemma 3.2.6. Nun besagt der *Satz von Dini*, dass, wenn eine Folge stetiger Funktionen auf einem kompakten Intervall monoton gegen eine stetige Funktion punktweise konvergiert, die Konvergenz sogar gleichmäßig ist (Übungsaufgabe: man zeige dies!),

vgl. [He86, S. 578]). Somit konvergiert die rechte Seite von (3.3.26) gleichmäßig auf kompakten Intervallen gegen

$$\sum_{k \in E} \bar{P}_{ik}(t)q_k - \sum_{j \in E} \bar{P}_{ij}(t)q_j = 0. \quad (3.3.30)$$

Aus dem aus Analysisvorlesungen bekannten Satz, dass eine Folge stetig differenzierbarer Funktionen  $f_n: [0, t_0] \rightarrow \mathbb{R}$  mit  $f_n(0) \rightarrow a \in \mathbb{R}$ , deren Ableitungen gleichmäßig konvergieren, gleichmäßig auf  $[0, t_0]$  gegen eine Funktion  $f$  konvergiert, die stetig differenzierbar ist und deren Ableitung gleich dem Grenzwert der Ableitungen der  $f_n$  ist, folgt

$$\frac{d}{dt} \sum_{j \in E} \bar{P}_{ij}(t) = 0. \quad (3.3.31)$$

□

**Bemerkung 3.3.10.** Aus den bisherigen Resultaten folgt: Ist  $Q$  schwach konservativ, und hat die VWG eine eindeutige Lösung  $(P(t))_{t \geq 0}$  und erfüllt diese  $\sum_{j \in E} P_{ij}(t) = 1$  für alle  $i \in E$ , dann ist  $(P(t))_{t \geq 0}$  standard und die einzige Standardhalbgruppe mit  $Q$ -Matrix  $Q$ .

### 3.4 Langzeitverhalten und invariante Maße

Wir studieren nun die Existenz von Grenzwerten von  $P_{ij}(t)$  für  $t \rightarrow \infty$ . Interessanterweise ist dies weniger kompliziert als im diskreten Fall, da das Problem der Periodizität bei MKSZ nicht auftaucht.

**Satz 3.4.1.** Sei  $(P(t))_{t \geq 0}$  standard. Dann existiert für alle  $i, j \in E$

$$\pi_{ij} := \lim_{t \rightarrow \infty} P_{ij}(t), \quad (3.4.1)$$

und es gilt

$$\sum_{j \in E} \pi_{ij} \leq 1. \quad (3.4.2)$$

**Beweis.** Setze zunächst  $(P(t))_{t \geq 0}$  wie in Proposition 3.2.3 zu einer Standard-Markovhalbgruppe auf  $E \cup \{\partial\}$  fort, falls  $(P(t))_{t \geq 0}$  nicht markovsch ist. Also können wir oBdA annehmen, dass  $(P(t))_{t \geq 0}$  standard und markovsch ist. Für  $j \in E$  existiert (da  $(P(t))_{t \geq 0}$  standard ist) ein  $h_0 > 0$ , so dass  $P_{jj}(h) > 0$  für alle  $0 \leq h \leq h_0$ . Mit der Chapman-Kolmogorov-Gleichung folgt somit für  $t > 0$ , indem man  $t = nh$  mit  $n \in \mathbb{N}$  und  $h \leq h_0$  setzt:

$$P_{jj}(t) \geq (P_{jj}(h))^n > 0. \quad (3.4.3)$$

Für beliebiges  $h > 0$  betrachte nun die MKDZ  $(X(0), X(h), X(2h), \dots)$  mit Übergangsmatrix  $P(h)$ . Diese ist wegen  $P_{jj}(h) > 0$  für alle  $j \in E$  aperiodisch, also existiert nach Satz 2.6.18 der Grenzwert

$$\pi_{ij}(h) := \lim_{n \rightarrow \infty} P_{ij}(nh). \quad (3.4.4)$$

Sei  $\varepsilon > 0$  vorgegeben und  $i, j \in E$  fest. Da  $P_{ij}(\cdot)$  nach Lemma 3.2.6 gleichmäßig stetig ist, existiert ein  $h$ , so dass  $|P_{ij}(t+s) - P_{ij}(t)| < \varepsilon/4$  für alle  $t \geq 0$  und alle  $s \leq h$ . Wähle nun  $n_0 = n_0(h)$ , so dass

$$|P_{ij}(nh) - \pi_{ij}(h)| < \frac{\varepsilon}{4}, \quad n \geq n_0. \quad (3.4.5)$$



Dann gilt für  $t, t' \geq n_0 h$  mit  $kh \leq t \leq (k+1)h$  und  $mh \leq t' \leq (m+1)h$

$$\begin{aligned} |P_{ij}(t) - P_{ij}(t')| &\leq |P_{ij}(t) - P_{ij}(kh)| + |P_{ij}(kh) - \pi_{ij}(h)| \\ &\quad + |\pi_{ij}(h) - P_{ij}(mh)| + |P_{ij}(mh) - P_{ij}(t')| < 4\frac{\varepsilon}{4} = \varepsilon. \end{aligned} \quad (3.4.6)$$

Also existiert  $\lim_{t \rightarrow \infty} P_{ij}(t)$ , und es ist  $\pi_{ij} := \lim_{t \rightarrow \infty} P_{ij}(t) = \lim_{n \rightarrow \infty} P_{ij}(nh) = \pi_{ij}(h)$ . Insbesondere hängt  $\pi_{ij}(h)$  gar nicht von  $h$  ab.

Sei nun  $J \subset E$  endlich. Dann gilt

$$\sum_{j \in J} \pi_{ij} = \sum_{j \in J} \lim_{t \rightarrow \infty} P_{ij}(t) = \lim_{t \rightarrow \infty} \sum_{j \in J} P_{ij}(t) \leq 1, \quad (3.4.7)$$

also folgt  $\sum_{j \in E} \pi_{ij} \leq 1$ . □

Wie im diskreten Fall wollen wir die Beziehung zwischen den Grenzwerten  $\pi_{ij}$  und invarianten Maßen näher studieren.

**Definition 3.4.2.** Sei  $(P(t))_{t>0}$  submarkovsch und  $\mathbb{X}$  eine zugehörige MKSZ. Eine Abbildung  $\pi: E \rightarrow [0, \infty]$  heißt ein *invariantes Maß* für  $(P(t))_{t>0}$  (oder für  $\mathbb{X}$ ), wenn  $\pi P(t) = \pi$  für alle  $t > 0$ .

Gilt zusätzlich  $\sum_{i \in E} \pi_i = 1$ , dann heißt  $\pi$  *invariantes Wahrscheinlichkeitsmaß* (oder *invariante Verteilung*) von  $(P(t))_{t>0}$ .

**Proposition 3.4.3.** Ist  $(P(t))_{t \geq 0}$  standard und  $i \in E$ , dann ist  $(\pi_{ij})_{j \in E}$  (wie in Satz 3.4.1 definiert) ein invariantes Maß.

**Beweis.** Setze wie in Proposition 3.2.3  $(P(t))_{t \geq 0}$  zu einer Markovhalbgruppe auf  $E \cup \{\partial\}$  fort. Dann gilt für  $s, t \geq 0$  und  $k \in E \cup \{\partial\}$

$$\sum_{j \in E \cup \{\partial\}} P_{ij}(s) P_{jk}(t) = P_{ik}(t+s). \quad (3.4.8)$$

Lässt man  $s \rightarrow \infty$  gehen, so folgt (mit der üblichen ‘‘Abschneidetechnik’’)

$$\sum_{j \in E \cup \{\partial\}} \pi_{ij} P_{jk}(t) \leq \pi_{ik}. \quad (3.4.9)$$

Summiert man über alle  $k \in E \cup \{\partial\}$ , so folgt  $\sum_{j \in E \cup \{\partial\}} \pi_{ij} \leq \sum_{k \in E \cup \{\partial\}} \pi_{ik}$ , also – da die Summe endlich ist – die Gleichheit in (3.4.9) für alle  $k \in E \cup \{\partial\}$ . Für  $k \in E$  ist  $P_{\partial k}(t) = 0$  für alle  $t \geq 0$ , also ist  $(\pi_{ij})_{j \in E}$  ein invariantes Maß. □

**Satz 3.4.4.** Sei  $Q$  konservativ und die Minimallösung  $(P(t))_{t \geq 0}$  markovsch. Dann ist  $\pi: E \rightarrow \mathbb{R}$  ein invariantes Wahrscheinlichkeitsmaß von  $(P(t))_{t \geq 0}$  genau dann, wenn es

$$(i) \pi Q = 0, \quad (ii) \pi_i \geq 0, \quad i \in E, \quad (iii) \sum_{i \in E} \pi_i = 1 \quad (3.4.10)$$

löst.

**Beweis.** (nach [Mi63]):

1. *Schritt:* Sei  $\pi$  ein invariantes Wahrscheinlichkeitsmaß, also  $\pi_i \geq 0$  für alle  $i \in E$ ,  $\sum_{i \in E} \pi_i = 1$  und  $\pi P(t) = \pi$  für alle  $t \geq 0$ . Dann gilt

$$\pi_j(1 - P_{jj}(t)) = \sum_{i \in E \setminus \{j\}} \pi_i P_{ij}(t). \quad (3.4.11)$$

Dividiert man beidseitig durch  $t > 0$  und läßt  $t \downarrow 0$  gehen, so folgt

$$\infty > \pi_j q_j \geq \sum_{i \in E \setminus \{j\}} \pi_i q_{ij} \geq 0 \quad (3.4.12)$$

(um das mittlere “ $\geq$ ” zu erhalten, summiere man zunächst über endliche Mengen).

Sei  $J \subset E$  endlich. Dann folgt mit der VWG (die ja von der Minimallösung erfüllt wird)

$$\frac{d}{dt} \sum_{i \in J} \pi_i P_{ij}(t) = \sum_{i \in J} \pi_i \frac{d}{dt} P_{ij}(t) = -q_j \sum_{i \in J} \pi_i P_{ij}(t) + \sum_{i \in J} \pi_i \sum_{k \in E \setminus \{j\}} P_{ik}(t) q_{kj}. \quad (3.4.13)$$

Wie im Beweis von Satz 3.3.9 folgt die Stetigkeit der letzten Summe. Die beiden Terme auf der rechten Seite von (3.4.13) konvergieren mit  $J \uparrow E$  jeweils monoton gegen  $-q_j \sum_{i \in E} \pi_i P_{ij}(t) = -q_j \pi_j$  bzw.

$$\sum_{i \in E} \pi_i \sum_{k \in E \setminus \{j\}} P_{ik}(t) q_{kj} = \sum_{k \in E \setminus \{j\}} q_{kj} \sum_{i \in E} \pi_i P_{ik}(t) = \sum_{k \in E \setminus \{j\}} q_{kj} \pi_k,$$

da  $\pi$  invariant ist.

Der Grenzwert ist jeweils konstant und endlich nach (3.4.12), insbesondere also stetig. Wie im Beweis von Satz 3.3.9 folgt mit dem Satz von Dini und dem Satz über die Vertauschbarkeit von Grenzwert und Ableitung

$$0 = \frac{d}{dt} \pi_j = \frac{d}{dt} \sum_{i \in E} \pi_i P_{ij}(t) = -q_j \pi_j + \sum_{k \in E \setminus \{j\}} q_{kj} \pi_k = \sum_{k \in E} \pi_k q_{kj}. \quad (3.4.14)$$

Also gilt (i) (die Aussagen (ii) und (iii) sind ohnehin klar).

2. *Schritt:*

$\pi$  erfülle (i), (ii) und (iii). Definiere  $P_{ij}^n(t)$  und  $H_{ij}^N(t)$  für  $n, N \in \mathbb{N}_0$ , und  $i, j \in E$  und  $t \geq 0$  wie im Beweis von Satz 3.3.6. Wir zeigen per Induktion nach  $N \in \mathbb{N}_0$ , dass gilt:

$$\sum_{i \in E} \pi_i H_{ij}^N(t) \leq \pi_j \quad \text{für alle } j \in E, t \geq 0. \quad (3.4.15)$$

Für  $N = 0$  gilt

$$\sum_{i \in E} \pi_i H_{ij}^0(t) = \sum_{i \in E} \pi_i \delta_{ij} e^{-q_j t} = \pi_j e^{-q_j t} \leq \pi_j. \quad (3.4.16)$$

Gelte also (3.4.15) für ein  $N \in \mathbb{N}_0$ . Dann folgt

$$\begin{aligned}
 \sum_{i \in E} \pi_i H_{ij}^{N+1}(t) &= \pi_j e^{-q_j t} + \sum_{k \in E \setminus \{j\}} \int_0^t \sum_{i \in E} \pi_i H_{ik}^N(s) q_{kj} e^{-q_j(t-s)} ds \\
 &\leq \pi_j e^{-q_j t} + \sum_{k \in E \setminus \{j\}} \int_0^t \pi_k q_{kj} e^{-q_j(t-s)} ds \\
 &= \pi_j e^{-q_j t} + \pi_j q_j \int_0^t e^{-q_j(t-s)} ds = \pi_j,
 \end{aligned} \tag{3.4.17}$$

wobei in “ $\leq$ ” die Induktionsvoraussetzung einging und beim vorletzten Gleichheitszeichen die Eigenschaft (i).

Da  $P_{ij}(t) = \bar{P}_{ij}(t) = \lim_{N \uparrow \infty} H_{ij}^N(t)$ , folgt aus (3.4.15)

$$\sum_{i \in E} \pi_i P_{ij}(t) \leq \pi_j \quad \text{für alle } j \in E, t \geq 0. \tag{3.4.18}$$

Summiert man über alle  $j \in E$ , so sieht man, dass in (3.4.18) in Wirklichkeit Gleichheit gilt, d. h.  $\pi$  ist ein invariantes Wahrscheinlichkeitsmaß.  $\square$

**Bemerkung 3.4.5.** Satz 3.4.4 zeigt, dass man unter den angegebenen Voraussetzungen alle invarianten Wahrscheinlichkeitsmaße durch Lösen von  $\pi Q = 0$  mit den Nebenbedingungen (ii), (iii) in 3.4.10 erhält. Dies ist für die Berechnung von  $\pi$  sehr bedeutsam, da die Matrizen  $P(t)$  für  $t \geq 0$  im Gegensatz zu  $Q$  meist nicht explizit gegeben sind. Man beachte aber die Voraussetzungen von Satz 3.4.4! Miller gibt in der zitierten Arbeit ein Beispiel, bei dem für ein konservatives  $Q$  (i) - (iii) eine eindeutige Lösung hat, ohne dass  $\pi$  ein invariantes Wahrscheinlichkeitsmaß ist (die Minimallösung ist in seinem Beispiel nicht markovsch).

**Definition 3.4.6** (Irreduzibilität). Eine Standardhalbgruppe  $(P(t))_{t \geq 0}$  mit schwach konservativer  $Q$ -Matrix  $Q$  (oder  $Q$  selbst) heißt *irreduzibel*, wenn für alle  $i, j \in E$  mit  $i \neq j$  ein  $n \in \mathbb{N}_0$  und  $i = i_0, i_1, \dots, i_n = j$  aus  $E$  existieren mit  $q_{i_m, i_{m+1}} > 0$  für alle  $m = 0, \dots, n-1$ .

**Korollar 3.4.7.** Sei  $Q$  konservativ und irreduzibel und die Minimallösung markovsch. Wenn (3.4.10) in Satz 3.4.4 eine Lösung  $\pi$  hat, dann ist diese eindeutig, und es gilt  $\lim_{t \rightarrow \infty} P_{ij}(t) = \pi_j$  für alle  $i, j \in E$ .

**Beweis.** Nach Satz 3.4.4 ist die Lösung  $\pi$  invariant unter  $(P(t))_{t \geq 0}$ , also ist  $\pi$  auch invariantes Wahrscheinlichkeitsmaß der MKDZ  $(X(0), X(1), X(2), \dots)$  mit Übergangsmatrix  $P(1)$ . Diese MKDZ ist irreduzibel: Im Beweis von Satz 3.4.1 sahen wir, dass  $P_{ii}(t) > 0$  für alle  $t \geq 0$ . Für  $i \neq j$  seien  $i_0, \dots, i_n$  wie bei der Definition 3.4.6 gegeben. Es genügt zu zeigen, dass  $i \rightarrow i_1$  für die MKDZ  $(X(0), X(1), \dots)$  gilt. Wegen  $P_{i, i_1}(t) = q_{i, i_1} t + o(t)$  und  $q_{i, i_1} > 0$  existiert ein  $h_0 > 0$ , so dass  $P_{i, i_1}(t) > 0$  für alle  $t \in (0, h_0]$  gilt. Weiter gilt für  $t > h_0$ , dass  $P_{i, i_1}(t) \geq P_{ii}(t - h_0) P_{i, i_1}(h_0) > 0$ , womit insbesondere die Irreduzibilität von  $P(1)$  gezeigt ist. Wegen  $P_{ii}(1) > 0$  ist die Aperiodizität klar. Also folgt nach Satz 2.6.15 für  $j \in E$ , dass  $\pi_j = \lim_{n \rightarrow \infty} P_{ij}(n)$ , und wegen Satz 3.4.1 gilt sogar  $\pi_j = \lim_{t \rightarrow \infty} P_{ij}(t)$ . Dies zeigt auch die Eindeutigkeit der Lösung  $\pi$  von (i) - (iii) in Satz 3.4.4.  $\square$

**Korollar 3.4.8.** Sei  $Q$  konservativ und die Minimallösung markovsch. Wenn (3.4.10) in Satz 3.4.4 keine Lösung hat, dann folgt

$$\lim_{t \rightarrow \infty} P_{ij}(t) = 0 \quad \text{für alle } i, j \in E. \tag{3.4.19}$$

**Beweis.** Nach Satz 3.4.1 existiert  $\pi_{ik} = \lim_{t \rightarrow \infty} P_{ik}(t)$  und es gilt  $\sum_{k \in E} \pi_{ik} \leq 1$ . Angenommen, es existieren  $i, j \in E$  mit  $\lim_{t \rightarrow \infty} P_{ij}(t) = \pi_{ij} > 0$ , dann definiere

$$\bar{\pi}_{ik} := \frac{\pi_{ik}}{\sum_{r \in E} \pi_{ir}}. \quad (3.4.20)$$

Nach Proposition 3.4.3 ist  $(\bar{\pi}_{ik})_{k \in E}$  ein invariantes Wahrscheinlichkeitsmaß und erfüllt (3.4.10) im Widerspruch zur Voraussetzung. Also gilt  $\pi_{ij} = 0$  für alle  $i, j \in E$ .  $\square$

**Beispiel 3.4.9.** Für Geburts- und Todesprozesse lassen sich die invarianten Wahrscheinlichkeitsverteilungen (wenn sie existieren) explizit berechnen. Wir nehmen an, dass die Geburts- und Sterberaten alle strikt positiv sind. Definiert man  $b_0 = 1$  und

$$b_j = \frac{\lambda_0 \cdots \lambda_{j-1}}{\mu_1 \cdots \mu_j}, \quad j \in \mathbb{N}, \quad (3.4.21)$$

so kann man zeigen, dass genau im Fall

$$\sum_{i=0}^{\infty} \left( (\lambda_i b_i)^{-1} \sum_{j=0}^i b_j \right) = \infty \quad (3.4.22)$$

der Prozess nicht explodiert (d. h. die Minimallösung markovsch ist). Ein Beweis der Formel (mit anderer Notation) findet sich in [As03, S.73] (wobei in der Formulierung der dortigen Proposition 2.2 offensichtlich “nonexplosive” durch “explosive” ersetzt werden muss). Eine Richtung ist übrigens recht einfach zu sehen: man zeigt, dass der Summand Nummer  $i$  in der Reihe genau der Erwartungswert der Zeit ist, die es dauert um bei Start  $i$  den Zustand  $i + 1$  zu erreichen. Dann ist die unendliche Reihe genau der Erwartungswert der Explosionszeit bei Start in 0. Ist die Reihe daher konvergent, so muss auch die Explosionszeit fast sicher endlich sein. Wir setzen die Gültigkeit von (3.4.22) im Folgenden voraus. Solche expliziten notwendige und hinreichende Nichtexplosionskriterien sind übrigens für allgemeine MKSZ nicht bekannt (siehe aber [As03, Proposition II.3.3] für ein nicht sehr explizites notwendiges und hinreichendes Kriterium).

Der Geburts- und Todesprozess ist wegen der Voraussetzung der Positivität der  $\lambda_i$  und  $\mu_i$  offenbar irreduzibel. Anstatt das Gleichungssystem in Satz 3.4.4 direkt zu lösen, führt auch die folgende Überlegung zu einer Lösung:

Wenn ein invariantes Wahrscheinlichkeitsmaß  $\pi$  existiert, dann muss (da der Prozess im Gleichgewicht ist), für jedes  $i \in \mathbb{N}_0$  genausoviel “Masse” pro Zeit von  $i$  nach  $i + 1$  fließen wie umgekehrt, d. h. es muss gelten:

$$\pi_i \lambda_i = \pi_{i+1} \mu_{i+1} \quad \text{für alle } i \in \mathbb{N}_0. \quad (3.4.23)$$

Man zeigt leicht, dass (3.4.23) äquivalent zu (i) in Satz 3.4.4 ist. Die Aussage in (3.4.23) erhält man, indem man die ersten  $i + 1$  Gleichungen von  $\pi Q = 0$  addiert. Aus (3.4.23) folgt für  $k \in \mathbb{N}$

$$\pi_k = \pi_{k-1} \frac{\lambda_{k-1}}{\mu_k} = \pi_{k-2} \frac{\lambda_{k-2} \lambda_{k-1}}{\mu_{k-1} \mu_k} = \cdots = \pi_0 b_k. \quad (3.4.24)$$

Wenn  $\sum_{k=0}^{\infty} \pi_k = 1$  ist, so muss  $1 = \pi_0 \sum_{k=0}^{\infty} b_k$  gelten. Wäre  $\sum_{k=0}^{\infty} b_k = \infty$ , so müsste  $\pi_0 = 0$  sein und nach (3.4.24)  $\pi_k = 0$  für alle  $k \in \mathbb{N}$  gelten, was  $\sum_{k=0}^{\infty} \pi_k = 1$  widerspricht. Ist also

$\sum_{k=0}^{\infty} b_k = \infty$ , dann existiert kein invariantes Wahrscheinlichkeitsmaß. Ist dagegen  $\sum_{k=0}^{\infty} b_k < \infty$ , so definiert

$$\pi_j = \frac{1}{\sum_{k=0}^{\infty} b_k} b_j, \quad j \in \mathbb{N}_0, \quad (3.4.25)$$

eine Lösung von (3.4.23) und damit ein invariantes Wahrscheinlichkeitsmaß. Also ist  $\sum_{k=0}^{\infty} b_k < \infty$  eine notwendige und hinreichende Bedingung für die Existenz eines invarianten Wahrscheinlichkeitsmaßes eines irreduziblen Geburts- und Todesprozesses. Aus Korollar 3.4.7 folgt dann sogar  $\pi_j = \lim_{t \rightarrow \infty} P_{ij}(t)$  für alle  $i, j \in \mathbb{N}_0$ .

### 3.5 Beispiele: Warteschlangen

#### 3.5.1 Warteschlangen

Wir benutzen das eben gewonnene Resultat, um zu entscheiden, ob die Zahl der Kunden in einer M/M/c-Schlange eine invariante Verteilung hat und um gegebenenfalls die invariante Wahrscheinlichkeitsverteilung zu berechnen.

**Beispiel 3.5.1** (M/M/c-Warteschlange). Mit den Bezeichnungen aus Beispiel 3.1.2 gilt:

$$\sum_{j=0}^{\infty} b_j = \sum_{j=0}^{c-1} \frac{\lambda^j}{j! \mu^j} + \sum_{j=c}^{\infty} \frac{\lambda^j}{c! c^{j-c} \mu^j}. \quad (3.5.1)$$

Die erste Summe ist immer endlich, die zweite ist endlich genau dann, wenn  $\lambda < c\mu$  ist. Definiert man die *Verkehrsdichte*  $\rho$  als  $\lambda/(c\mu)$ , so gilt also wiederum (vgl. Satz 2.7.3), dass eine invariante Verteilung genau dann existiert, wenn  $\rho < 1$  ist.

Ist  $\rho < 1$ , so folgt

$$\sum_{j=0}^{\infty} b_j = \sum_{j=0}^{c-1} \frac{\lambda^j}{j! \mu^j} + \frac{c^c \rho^c}{c!(1-\rho)}, \quad (3.5.2)$$

woraus sich  $\pi_j = b_j (\sum_{k=0}^{\infty} b_k)^{-1}$  explizit berechnen lässt:

$$\pi_j = \pi_0 \left(\frac{\lambda}{\mu}\right)^j \times \begin{cases} \frac{1}{j!} & \text{falls } j \leq c, \\ \frac{1}{c! c^{j-c}} & \text{falls } j \geq c. \end{cases} \quad (3.5.3)$$

Im Spezialfall  $c = 1$  erhält man  $\pi_j = (1 - \rho)\rho^j$  für alle  $j = 0, 1, 2, \dots$ , also eine geometrische Verteilung.

Wir berechnen noch weitere interessante Größen für die M/M/c-Schlange im Gleichgewicht (im Fall  $\rho < 1$ ). Die Wahrscheinlichkeit, dass alle Bediener beschäftigt sind, ist

$$p_{\text{voll}} = \sum_{j=c}^{\infty} \pi_j = \pi_0 \frac{c^c}{c!} \sum_{j=c}^{\infty} \rho^j = \pi_0 \frac{c^c \rho^c}{c!} \frac{1}{1-\rho}.$$

Die erwartete Anzahl der im Betrieb befindlichen Bediener ist

$$S := \sum_{j=0}^{c-1} j\pi_j + c \sum_{j=c}^{\infty} \pi_j = \sum_{j=0}^{c-1} j\pi_j + c p_{\text{voll}} = \lambda/\mu (= c\rho),$$

wobei die vorletzte Identität (die letzte ausserhalb der Klammern) durch Einsetzen von  $p_{\text{voll}}$  und ein paar Umformungen folgt. Man kann die Formel  $S = \lambda/\mu$  mit etwas Heuristik auch ohne Berechnungen einsehen (das diskutieren wir in der Vorlesung, bitte nicht verpassen!).

Schließlich interessiert noch die erwartete Anzahl Kunden im System, also

$$L = \sum_{j=1}^{\infty} j\pi_j = \sum_{j=c}^{\infty} j\pi_j + S - cp_{\text{voll}} = c\rho + \pi_0\rho \frac{(c\rho)^c}{(1-\rho)^2 c!},$$

wobei die letzte Identität durch einsetzen und umformen folgt. Man beachte, dass in der letzten Formel der erste Summand gerade  $S$  ist und daher der zweite Summand die mittlere Anzahl wartender Kunden (ohne diejenigen, die gerade bedient werden) angibt.

Bei der M/M/c-Schlange haben wir bislang angenommen, dass es eine gemeinsame Warteschlange gibt und Kunden erst dann einem Bediener zugeordnet werden, wenn ihre Bedienungszeit beginnt. Alternativ kann man auch annehmen, dass diese Zuordnung zu einem Bediener direkt bei Ankunft erfolgt (wie zum Beispiel an Supermarktkassen). Eine Möglichkeit ist, diese Zuordnung rein zufällig vorzunehmen – unabhängig davon, wie stark die einzelnen Bediener belastet sind (andere Fälle werden wir in der Vorlesung diskutieren). Anschaulich ist ziemlich klar, dass bei gegebenen Parametern eine zufällige Zuordnung ungünstiger sein wird (dh zu längeren Wartezeiten führt) als eine gemeinsame Schlange. Wir wollen sehen ob das stimmt. Bei zufälliger Zuordnung ist das System sehr einfach zu analysieren, denn bei jedem einzelnen Bediener kommen die Kunden mit  $\text{Exp}(\lambda/c)$ -verteilten Zwischenankunftszeiten an. Also hat man an jeder Station eine M/M/1-Schlange. Die Anzahl der Kunden an dieser Schlange ist eine positiv rekurrente MKSZ genau dann, wenn  $\lambda/c < \mu$ , was auch die Stabilitätsbedingung für die zugehörige M/M/c-Schlange ist. Bei der mittleren Zahl von beschäftigten Bedienern gibt es keinen Unterschied zwischen beiden Modellen (jeweils  $c\rho$ ), bei der erwarteten Anzahl der wartenden Kunden (ohne die, die gerade bedient werden) aber schon. Bei der zufälligen Zuordnung ist dieser Erwartungswert (im gesamten System)  $c\rho^2/(1-\rho)$  (man nehme obige Formel für den zweiten Summanden von  $L$  mit  $c = 1$  und multipliziere mit  $c$ ), Betrachtet man die erwartete Anzahl wartender Kunden pro Bediener, dann ist diese Zahl bei festem  $\rho < 1$  im Modell mit zufälliger Zuordnung eine positive Konstante in  $c$ , beim Modell mit einer gemeinsamen Schlange dagegen konvergiert nicht nur dieser Erwartungswert mit  $c \rightarrow \infty$  gegen Null sondern sogar die erwartete Gesamtzahl wartender Kunden (man zeige dies! Dazu muss man untersuchen, wie schnell  $\pi_0$  als Funktion von  $c$  fällt). Wir werden in der Vorlesung noch weitere Kundenzuordnungsprotokolle analysieren.

### 3.5.2 Warteschlangennetze

In den letzten 20 Jahren sind Warteschlangennetze vor allem mit Anwendung auf Rechnersysteme, aber auch auf Produktionsabläufe untersucht worden. Man modelliert solche Netze dadurch, dass man jeden der Bediener (o.ä.) als Eckpunkt eines (endlichen) Graphen auffasst. Man verbindet  $i$  mit  $j$  durch einen gerichteten Pfeil und beschriftet ihn mit  $\pi_{ij}$ , wenn  $\pi_{ij} > 0$  die Wahrscheinlichkeit ist, dass ein Kunde, nachdem er bei  $i$  bedient wurde, sich bei  $j$  anstellt. Zusätzlich kann man eine Ecke des Graphen einführen, die die “Außenwelt” darstellt. Ein Kunde, der das ganze System verlässt, geht also in jene Ecke. Außerdem legt man fest, mit welcher Verteilung Kunden von außen in den einzelnen Ecken eintreffen, und beschriftet die Ecken mit der zugehörigen Bedienungszeitverteilung. Nimmt man an, dass alle Bedienungszeiten, Sprünge und Zwischenankunftszeiten unabhängig sind, und legt man eine geeignete Startbedingung fest,

so ist damit die Dynamik des Prozesses festgelegt. Interessant sind die gemeinsame Verteilung der Warteschlangenlängen in allen Ecken (außer der “Außenwelt”), insbesondere für  $t \rightarrow \infty$ , aber auch die Verweildauerverteilung eines Kunden im System. Wenn alle Bedienungs- und Zwischenankunftsverteilungen endliche Mischungen von  $E_k$ -Verteilungen sind, so kann man mit der in Beispiel 3.1.3 - 3.1.5 vorgestellten Methode den Vektor der Zahl der Kunden in jeder Ecke zu einer MKSZ “aufblähen”. Dies wird auch häufig gemacht, um damit invariante Wahrscheinlichkeitsmaße (numerisch) für obigen Vektor zu berechnen.

Ein Spezialfall, für den die Grenzverteilungen eine besonders einfache Gestalt haben, sind *Jackson networks* (vgl. [HS82, S. 456 ff]). Eine gute Referenz für allgemeine Warteschlangennetze ist die Monographie [Bra08] von M. Bramson.

### Jackson-Netzwerke.

Wir machen die folgenden Annahmen:

1. Es gibt  $N \in \mathbb{N}$  Knoten – durchnummeriert von 1 bis  $N$ .
2. Im Knoten  $i$  befinden sich  $c_i$  Bediener. Die Bedienungszeiten seien  $\text{Exp}(\mu_i)$ -verteilt.
3. Kunden, die von außen (und nicht von einem anderen Knoten) bei Knoten  $i$  eintreffen, haben unabhängige  $\text{Exp}(\lambda_i)$ -verteilte Zwischenankunftszeiten mit  $\lambda_i \geq 0$ , wobei  $\lambda_i = 0$  bedeutet, dass am Knoten  $i$  keine Kunden von außen eintreffen.
4. Gegeben ist eine substochastische  $N \times N$ -Matrix  $P$ , wobei  $p_{ij}$  die Wahrscheinlichkeit sei, dass ein Kunde nach der Bedienung im Knoten  $i$  sich (sofort!) am Knoten  $j$  anstellt.  $p_{i0} := 1 - \sum_{j=1}^N p_{ij}$  sei die Wahrscheinlichkeit, dass der Kunde nach Bedienung am Knoten  $i$  das System verlässt.
5. Es gelte überall FIFO (“first in first out”).

Notwendig und hinreichend für die Existenz einer Gleichgewichtsverteilung ist offenbar, dass alle  $N$  Knoten so schnell arbeiten, dass langfristig genausoviel hinein- wie herausfließt. Wir leiten zunächst heuristisch eine Verkehrsgleichung her, die wir dann (mathematisch exakt) analysieren, womit wir (im folgenden Satz) ein notwendiges und hinreichendes Kriterium für die Existenz einer invarianten Verteilung – sowie eine Formel dafür – erhalten.

Wenn  $\alpha_i \geq 0$  die Rate ist, mit der Kunden im stationären Zustand den Knoten  $i$  verlassen, dann gilt anschaulich:

$$\alpha_i = \lambda_i + \sum_{j=1}^N \alpha_j p_{ji}, \quad i = 1, \dots, N, \quad (3.5.4)$$

denn was abfließt ( $\alpha_i$ ) muss auch reinfließen (rechte Seite). Wir nehmen nun zusätzlich an, dass gilt

$$P^n \xrightarrow{n \rightarrow \infty} 0 \quad \text{komponentenweise,} \quad (3.5.5)$$

was besagt, dass Kunden von jedem Knoten aus irgendwann das System verlassen – sicher keine allzu starke und eine leicht überprüfbare Voraussetzung an  $P$ .

**Behauptung:** Unter obigen Annahmen hat (3.5.4) genau eine Lösung  $\alpha$ , und zwar

$$\alpha^T = \lambda^T (I - P)^{-1} = \lambda^T \sum_{k=0}^{\infty} P^k. \quad (3.5.6)$$

**Beweis.** (3.5.4) kann man in der Form  $\alpha^T = \lambda^T + \alpha^T P$ , also  $\alpha^T(I - P) = \lambda^T$  schreiben. Nun gilt

$$I - P^n = (I - P)(I + P + P^2 + \dots + P^{n-1}). \quad (3.5.7)$$

Wegen unserer Voraussetzung (3.5.5) ist  $I - P^n$  für hinreichend große  $n$  invertierbar und daher  $\det(I - P^n) \neq 0$ . Also ist nach (3.5.7) auch  $\det(I - P) \neq 0$ , d. h.,  $I - P$  ist invertierbar, woraus das erste Gleichheitszeichen der Behauptung folgt. Weiter folgt aus (3.5.7) nach Grenzübergang  $n \rightarrow \infty$  die Gleichung  $(I - P)^{-1} = \sum_{k=0}^{\infty} P^k$ .  $\square$

Wir machen nun noch die weitere Annahme

$$\alpha_i = \sum_{j=1}^N \lambda_j \left( \sum_{k=0}^{\infty} P^k \right)_{ji} > 0 \quad \text{für alle } i, \quad (3.5.8)$$

was zum Beispiel aus der stärkeren Forderung  $\lambda_j > 0$  für alle  $j$  folgt. Nun wählen wir – naheliegenderweise – als Zustandsraum des Netzwerks  $E = \mathbb{N}_0^N$ , wobei  $X(t) = (n_1, \dots, n_N)$  bedeuten soll, dass sich zur Zeit  $t$  genau  $n_i \in \mathbb{N}_0$  Kunden am Knoten  $i$  befinden für alle  $i = 1, \dots, N$ . Offenbar ist  $(X(t))_{t \geq 0}$  eine MKSZ, deren  $Q$ -Matrix wir im Beweis des folgenden Satzes explizit angeben werden.

**Satz 3.5.2** (Jackson). *Unter den obigen Annahmen hat die MKSZ  $(X(t))_{t \geq 0}$  genau dann ein invariantes Wahrscheinlichkeitsmaß  $\pi$ , wenn für alle  $i = 1, \dots, N$  gilt:  $\alpha_i < c_i \mu_i$ .*

*In diesem Fall ist  $\pi$  eindeutig, und es gilt*

$$\pi_{n_1, \dots, n_N} = \Psi_1(n_1) \cdots \Psi_N(n_N), \quad (3.5.9)$$

wobei  $\Psi_i(\cdot)$  das invariante Wahrscheinlichkeitsmaß einer  $M/M/c_i$ -Schlange mit Ankunftsrate  $\alpha_i$  und Bedienungsrate  $\mu_i$  ist, also

$$\Psi_i(n) = \Psi_i(0) \left( \frac{\alpha_i}{\mu_i} \right)^n \times \begin{cases} \frac{1}{n!}, & \text{falls } n \leq c_i, \\ \frac{1}{c_i!} \frac{1}{c_i^{n-c_i}}, & \text{falls } n \geq c_i. \end{cases} \quad (3.5.10)$$

**Bemerkung 3.5.3.** Man beachte, dass trotz der Abhängigkeiten, die zwischen den Knoten bestehen, die Anzahl der Kunden an den einzelnen Knoten im stationären Zustand zu einem festen Zeitpunkt unabhängige Zufallsgrößen sind!

**Beweis von Satz 3.5.2.** Offenbar hat die  $Q$ -Matrix folgende Gestalt (mit  $\epsilon_i$  bezeichnen wir das Element aus  $\mathbb{N}_0^N$  mit einer Eins an der  $i$ -ten Stelle und Nullen sonst). Für alle  $\mathbf{n} \in E = \mathbb{N}_0^N$  gilt

$$\begin{aligned} q_{\mathbf{n}, \mathbf{n} + \epsilon_i} &= \lambda_i, \\ q_{\mathbf{n}, \mathbf{n} - \epsilon_i} &= (n_i \wedge c_i) \mu_i p_{i0}, \\ q_{\mathbf{n}, \mathbf{n} - \epsilon_i + \epsilon_j} &= (n_i \wedge c_i) \mu_i p_{ij}, \end{aligned} \quad (3.5.11)$$

alle anderen  $q_{\mathbf{m}, \mathbf{n}}$  mit  $\mathbf{m} \neq \mathbf{n}$  sind Null. Weiter ist  $Q$  irreduzibel und die Minimallösung ist Markovsch (da die  $q_{\mathbf{n}}$  beschränkt sind).

Wir setzen nun voraus, dass  $\alpha_i < c_i \mu_i$  für alle  $i$  (den Beweis im umgekehrten Fall ersparen wir uns). Definiere  $\pi$  durch (3.5.9). Nach Satz 3.4.4 ist zu zeigen, dass  $\pi Q = 0$  (die Eindeutigkeit



von  $\pi$  folgt dann aus Korollar 3.4.7). Wir zeigen sogar, dass für jedes  $\mathbf{n} = (n_1, \dots, n_N)$  gilt

$$\sum_{j=1}^N \pi_{\mathbf{n}+\mathbf{e}_j} q_{\mathbf{n}+\mathbf{e}_j, \mathbf{n}} = \pi_{\mathbf{n}} \sum_{j=1}^N q_{\mathbf{n}, \mathbf{n}+\mathbf{e}_j}, \quad (3.5.12)$$

$$\sum_{i=1}^N \pi_{\mathbf{n}-\mathbf{e}_i} q_{\mathbf{n}-\mathbf{e}_i, \mathbf{n}} + \sum_{\substack{i,j=1 \\ i \neq j}}^N \pi_{\mathbf{n}-\mathbf{e}_i+\mathbf{e}_j} q_{\mathbf{n}-\mathbf{e}_i+\mathbf{e}_j, \mathbf{n}} = \pi_{\mathbf{n}} \sum_{i=1}^N q_{\mathbf{n}, \mathbf{n}-\mathbf{e}_i} + \pi_{\mathbf{n}} \sum_{\substack{i,j=1 \\ i \neq j}}^N q_{\mathbf{n}, \mathbf{n}-\mathbf{e}_i+\mathbf{e}_j}, \quad (3.5.13)$$

woraus sofort  $\pi Q = 0$  folgt.

Wir zeigen nur (3.5.12), da (3.5.13) analog bewiesen wird. Die linke Seite von (3.5.12) ist

$$\begin{aligned} & \sum_{j=1}^N \left( \prod_{\substack{i=1 \\ i \neq j}}^N \Psi_i(n_i) \right) \Psi_j(n_j + 1) ((n_j + 1) \wedge c_j) \mu_j p_{j0} \\ &= \prod_{i=1}^N \Psi_i(n_i) \sum_{j=1}^N \frac{\Psi_j(n_j + 1)}{\Psi_j(n_j)} ((n_j + 1) \wedge c_j) \mu_j p_{j0} \\ &= \prod_{i=1}^N \Psi_i(n_i) \sum_{j=1}^N \alpha_j p_{j0} = \prod_{i=1}^N \Psi_i(n_i) \sum_{j=1}^N \alpha_j \left( 1 - \sum_{k=1}^N p_{jk} \right) \\ &= \prod_{i=1}^N \Psi_i(n_i) \sum_{j=1}^N \lambda_j, \end{aligned} \quad (3.5.14)$$

was gleich der rechten Seite von (3.5.12) ist, wobei wir beim letzten Gleichheitszeichen die Verkehrsgleichung (3.5.4) und beim drittletzten die explizite Formel für  $\Psi_j(n_j + 1)/\Psi_j(n_j)$  benutzen.  $\square$

### Underload instability.

Zu Beginn der neunziger Jahre des letzten Jahrhunderts wurde ein bemerkenswertes Phänomen bei bestimmten Netzwerken beobachtet (zuerst durch Simulation und später mit Beweis), welches in der angelsächsischen Literatur als *underload instability* bezeichnet wird. Grob gesagt bedeutet *underload*, dass jeder Knoten in dem Netzwerk schnell genug arbeitet um die anfallende Arbeit zu bewältigen und man daher erwarten sollte, dass in diesem Fall das System stabil ist in dem Sinne, dass die Verteilung des Systemzustands gegen ein Gleichgewicht konvergiert. Interessanterweise ist das aber nicht immer so. Als ein Beispiel präsentieren wir ohne Beweis das *Rybko-Stolyar Netzwerk* (Details und andere Beispiele findet man in Kapitel 3 von [Bra08]): es gibt zwei Bediener (mit Nummern 1 und 2) und zwei Arten von Kunden: die einen, die wir als Kunden vom Typ 1 bezeichnen wollen, stellen sich zuerst bei Bediener 1 an und dann, sobald sie bedient wurden, bei Bediener 2. Sobald sie dort bedient wurden verlassen sie das System. Analoges gilt für Kunden vom Typ 2 (die zuerst von Bediener 2 und dann von 1 bedient werden). Wir nehmen an, dass die Zwischenankunftszeiten beider Kundenströme unabhängig  $\text{Exp}(1)$ -verteilt sind und die Bedienungszeiten unabhängig  $\text{Exp}(\mu)$ -verteilt sind. Ist  $\mu > 2$ , so hat man offenbar *underload*, denn jeder Bediener muss im Schnitt pro Zeiteinheit zwei Kunden bedienen, deren Bedienungszeit im Mittel zusammen aber nur  $2/\mu < 1$  ist. Nimmt man weiter an, dass bei beiden Bedienern diejenigen Kunden, die von dem anderen Knoten kommen, Priorität gegenüber den Kunden haben, für die dies nicht gilt, dann existiert ein  $\mu_0 > 2$ , so dass für alle  $\mu \in (2, \mu_0]$  das System instabil ist in dem Sinn, dass die Zahl der Kunden im System mit  $t \rightarrow \infty$  fast sicher gegen unendlich geht. Wir werden in der Vorlesung sehen, wie dieser Effekt zustande kommt.



# Kapitel 4

## Gaußsche Prozesse und Zeitreihenanalyse

### 4.1 Einleitung

In diesem Kapitel behandeln wir die wichtige Klasse der Gaußprozesse und einige Fragestellungen aus der Zeitreihenanalyse – wenn auch sehr oberflächlich.

**Definition 4.1.1.** Ein reellwertiger stochastischer Prozess  $X(t), t \in I$  heißt *Gaußprozess*, wenn alle (endlichen) Linearkombinationen

$$Y_{(\alpha_1, \dots, \alpha_k)}^{(t_1, \dots, t_k)} = \sum_{i=1}^k \alpha_i X(t_i)$$

$k \in \mathbb{N}, t_1, \dots, t_k \in I, \alpha_1, \dots, \alpha_k \in \mathbb{R}$ , normalverteilt sind. Im Fall  $|I| < \infty$  heißt  $X(t), t \in I$  auch *Gaußvektor*.

**Bemerkung 4.1.2.**

- Wir bezeichnen eine Zufallsgröße auch dann als normalverteilt, wenn sie fast sicher konstant ist!
- Da  $(X_1, \dots, X_n)$  genau dann (gemeinsam) Gaußverteilt (oder normalverteilt) ist, wenn alle Linearkombinationen (eindimensional) normalverteilt sind, ist  $X(t), t \in I$  ein Gaußprozess genau dann, wenn alle endlich-dimensionalen Verteilungen Normalverteilungen sind.

**Definition 4.1.3.** Sei  $X(t), t \in I$  ein stochastischer Prozess mit  $\mathbb{E}X^2(t) < \infty$  für alle  $t \in I$ . Sei  $\mu(t) := \mathbb{E}X(t)$ . Dann heißt die Funktion

$$r : I \times I \rightarrow \mathbb{R}$$

definiert durch

$$r(s, t) := \mathbb{E}((X(s) - \mu(s))(X(t) - \mu(t))), s, t \in I$$

*Kovarianzfunktion* von  $X(t), t \in I$ .

**Satz 4.1.4.** Sei  $r$  die Kovarianzfunktion des Prozesses  $X(t), t \in I$ . Dann ist  $r$  positiv semidefinit, d.h.

$$\sum_{i,j=1}^k z_i r(t_i, t_j) \overline{z_j}$$

ist reell und nicht negativ für alle  $k \in \mathbb{N}, z = (z_1, \dots, z_k) \in \mathbb{C}^k, t_1, \dots, t_k \in I$ .

**Beweis.**

$$\begin{aligned} \sum_{i,j=1}^k z_i r(t_i, t_j) \overline{z_j} &= \sum_{i,j=1}^k z_i \mathbb{E} \left( (X(t_i) - \mu(t_i)) \overline{(X(t_j) - \mu(t_j))} \right) \overline{z_j} \\ &= \mathbb{E} \left( \left( \sum_{i=1}^k z_i (X(t_i) - \mu(t_i)) \right) \overline{\left( \sum_{j=1}^k z_j (X(t_j) - \mu(t_j)) \right)} \right) \\ &= \mathbb{E} \left| \sum_{i=1}^k z_i (X(t_i) - \mu(t_i)) \right|^2 \geq 0 \end{aligned}$$

□

**Bemerkung 4.1.5.** Den Querstrich über  $X(t_j) - \mu(t_j)$  im zweiten Ausdruck des Beweises kann man natürlich auch weglassen, da  $X(t_j) - \mu(t_j)$  reell ist. Man sieht aber, dass auch im Falle  $\mathbb{C}$ -wertiger Prozesse  $r$  positiv semidefinit ist, wenn man  $r(t_i, t_j)$  als  $\mathbb{E}((X(t_i) - \mu(t_i)) \overline{(X(t_j) - \mu(t_j))})$  definiert.

**Bemerkung 4.1.6.** Wir werden später sehen, dass auch umgekehrt jede reelle positiv semidefinite Funktion Kovarianzfunktion eines (reellwertigen) stochastischen Prozesses ist (sogar eines Gaußprozesses).

**Satz 4.1.7.** a) Ein Gaußprozess hat endliche zweite Momente, d.h.  $\mathbb{E}X^2(t) < \infty$  für alle  $t \in I$ .

b) Wenn  $X(t), t \in I$  ein Gaußprozess ist,  $t_1, t_2, \dots, t_k \in I$ ,

$$\mu(t) := \mathbb{E}X(t), t \in I, X - \mu := (X(t_1) - \mu(t_1), \dots, X(t_k) - \mu(t_k))^T$$

und

$$\Gamma := \text{cov}(X(t_1), \dots, X(t_k)) := \mathbb{E}((X - \mu)(X - \mu)^T)$$

invertierbar ist, dann hat die Verteilung von  $(X(t_1), \dots, X(t_k))$  eine Dichte, die gegeben ist durch

$$f_{t_1, \dots, t_k}(x) = (2\pi)^{-k/2} (\det \Gamma)^{-1/2} \exp \left( -\frac{1}{2} (x - \mu)^T \Gamma^{-1} (x - \mu) \right),$$

wobei  $x = (x_1, \dots, x_k)^T, \mu = (\mu(t_1), \dots, \mu(t_k))^T$ .

c) Die Verteilung eines Gaußprozesses ist durch seine Erwartungwertfunktion  $\mathbb{E}X(t) = \mu(t)$  und Kovarianzfunktion eindeutig festgelegt.

**Beweis.** a) Nach Definition 4.1.1 ist  $X(t) = Y_{(1)}^{(t)}$  eindimensional normalverteilt und hat daher ein zweites (und sogar beliebiges  $n$ -tes für  $n \in \mathbb{N}$ ) Moment.

b) und c) sind aus der Vorlesung WT I bekannt. Bei c) beachte man, dass die endlich-dimensionalen Verteilungen die Verteilung festlegen. □

Der folgende Satz zeigt, wie man mehrdimensional-normalverteilte Zufallsvariablen aus unabhängig  $\mathcal{N}(0, 1)$ -verteilten erhält.

**Satz 4.1.8.** Sei  $m \in \mathbb{N}$ ,  $\mu \in \mathbb{R}^m$  und  $\Sigma \in \mathbb{R}^{m \times m}$  positiv semidefinit.

Sei  $\text{rg}(\Sigma) = k$  und  $n \geq \max(k, 1)$ . Dann existiert ein  $A \in \mathbb{R}^{m \times n}$ , so dass  $\Sigma = AA^T$ . Sind  $Y_1, \dots, Y_n$  unabhängig  $\mathcal{N}(0, 1)$ -verteilte ZVn, dann ist für  $Y := (Y_1, \dots, Y_n)^T$

$$X := AY + \mu$$

$\mathcal{N}(\mu, \Sigma)$ -verteilt.

**Beweis.** Die erste Behauptung ist ein bekanntes Resultat aus der linearen Algebra.

(i)  $X$  ist ein Gaußvektor, da alle Linearkombinationen der Komponenten von  $X$  gleichzeitig Linearkombinationen der Komponenten von  $Y$  plus einer Konstanten und damit normalverteilt sind.

(ii)  $\mathbb{E}X = \mathbb{E}(AY + \mu) = A\mathbb{E}Y + \mu = \mu$ .

(iii)  $\text{cov } X = \mathbb{E}((X - \mu)(X - \mu)^T) = \mathbb{E}((AY)(AY)^T) = \mathbb{E}(AYY^T A^T)$   
 $= A\mathbb{E}(YY^T)A^T = AA^T = \Sigma$ .

□

**Bemerkung 4.1.9.** Satz 4.1.8 lässt sich zur Simulation von Gaußvektoren anwenden. Will man  $X \sim \mathcal{N}(\mu, \Sigma)$  simulieren, so finde man für  $n \geq \text{rg}(\Sigma) \vee 1$  ein  $A$  mit  $\Sigma = AA^T$ . Man kann z.B.  $n = m$  wählen und  $A$  mit dem Choleskyverfahren berechnen. Die unabhängigen  $\mathcal{N}(0, 1)$ -verteilten  $Y_1, Y_2, \dots$  simuliert man aus unabhängig, auf  $[0, 1]$  gleichverteilten Zufallsvariablen  $U_1, U_2, \dots$  am einfachsten paarweise durch

$$Y_1 := (-2 \log U_1)^{1/2} \cos(2\pi U_2)$$

$$Y_2 := (-2 \log U_1)^{1/2} \sin(2\pi U_2)$$

usw. (vgl. letzte Bemerkung in Kapitel 1).

**Satz 4.1.10.** Sei  $I$  eine nichtleere Menge,  $\mu : I \rightarrow \mathbb{R}$  beliebig und  $r : I \times I \rightarrow \mathbb{R}$  positiv semidefinit. Dann existiert (bzgl. Verteilung) genau ein Gaußprozess mit Erwartungwertfunktion  $\mu$  und Kovarianzfunktion  $r$ .

**Beweis.** Definiere die Verteilungen

$$P_{t_1, \dots, t_k} := \mathcal{N} \left( \begin{pmatrix} \mu(t_1) \\ \vdots \\ \mu(t_k) \end{pmatrix}, r(t_i, t_j)_{i,j=1, \dots, k} \right)$$

für  $k \in \mathbb{N}$ ,  $t_1, \dots, t_k \in I$ . Man kann leicht zeigen, dass die Familie dieser Maße konsistent im Sinne von Definition 1.4.2 ist. Daher existiert nach Satz 1.4.3 eine eindeutige Verteilung  $\mathbb{P}$  auf  $(\mathbb{R}^I, \mathcal{B}^I)$  mit obigen endlich dimensionalen Verteilungen.  $\mathbb{P}$  ist offensichtlich die eindeutige Verteilung eines Gaußprozesses mit den vorgeschriebenen Erwartungswerten und Kovarianzen.  $\square$

**Satz 4.1.11.** Sei  $X(t), t \in I$  ein Gaußprozess. Dann sind äquivalent:

- (i)  $X(t), t \in I$  sind unabhängig (d.h.  $\{X(t_1), \dots, X(t_k)\}$  sind unabhängig für alle  $k \in \mathbb{N}, t_1, t_2, \dots, t_k \in I$  paarweise verschieden).
- (ii)  $X(t), t \in I$  sind paarweise orthogonal, d.h.  $\mathbb{E}((X(t) - \mu(t))(X(u) - \mu(u))) = 0$  für alle  $t, u \in I, t \neq u$ .

**Beweis.**

(i)  $\Rightarrow$  (ii): Gilt allgemein für Prozesse mit endlichen zweiten Momenten.

(ii)  $\Rightarrow$  (i): Seien  $t_1, \dots, t_k \in I$  paarweise verschieden. Wenn  $X(t_1), \dots, X(t_k)$  paarweise unkorreliert sind, gilt

$$\begin{pmatrix} X(t_1) \\ \vdots \\ X(t_k) \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \mu(t_1) \\ \vdots \\ \mu(t_k) \end{pmatrix}, \begin{pmatrix} \sigma^2(t_1) & & 0 \\ & \ddots & \\ 0 & & \sigma^2(t_k) \end{pmatrix} \right),$$

wobei  $\sigma^2(t_i) = \text{var}(X(t_i))$ . Nach Satz 4.1.8 gilt, dass für unabhängige  $\mathcal{N}(0, 1)$ -verteilte  $Y_1, \dots, Y_k$

$$\begin{pmatrix} \tilde{X}(t_1) \\ \vdots \\ \tilde{X}(t_k) \end{pmatrix} = \begin{pmatrix} \mu(t_1) \\ \vdots \\ \mu(t_k) \end{pmatrix} + \begin{pmatrix} \sigma(t_1) & & 0 \\ & \ddots & \\ 0 & & \sigma(t_k) \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix}$$

dieselbe Verteilung wie  $(X(t_1), \dots, X(t_k))^T$  hat, also  $(\tilde{X}(t_1), \dots, \tilde{X}(t_k))$  paarweise unkorreliert sind. Da aber  $\tilde{X}(t_i) = \mu(t_i) + \sigma(t_i)Y_i$  für  $i = 1, \dots, k$  als Funktionen unabhängiger Zufallsvariablen ebenfalls unabhängig sind, gilt dasselbe auch für  $X(t_1), \dots, X(t_k)$ .  $\square$

## 4.2 Vorhersage stochastischer Prozesse

Wir studieren nun das für Anwendungen wichtige Vorhersageproblem für stochastische Prozesse:

**Problem:** Gegeben sei ein reellwertiger stochastischer Prozess  $X(t), t \in I$  auf  $(\Omega, \mathcal{F}, \mathbb{P})$  (mit bekannter Verteilung  $\hat{P}$ ). Wir nehmen an, dass für  $\tilde{I} \subset I$  die Realisierung  $X(t), t \in \tilde{I}$  beobachtet wurde. Sei  $t_0 \in I \setminus \tilde{I}$ . Man finde eine gute (optimale) Vorhersage für  $X(t_0)$  aufgrund der Beobachtungen und der bekannten Verteilung.

Offenbar besteht die Aufgabe darin, eine (von  $\hat{P}$  abhängige) Funktion  $f : \mathbb{R}^{\tilde{I}} \rightarrow \mathbb{R}$  zu finden, so dass

$$\hat{X}(t_0) = f(X(t), t \in \tilde{I})$$

eine gute Schätzung für  $X(t_0)$  darstellt. Da sowohl  $X(t_0)$  also auch  $\hat{X}(t_0)$  Zufallsvariablen sind, bietet es sich an, z.B. eine Metrik  $d$  auf  $\mathbb{R}$  zu definieren und  $\hat{X}(t_0)$  (oder  $f$ ) als „optimal“ zu bezeichnen, wenn

$$\mathbb{E} d(X(t_0), \hat{X}(t_0))$$

minimal ist, wobei das Minimum über alle (messbaren) Funktionen  $f : \mathbb{R}^{\tilde{I}} \rightarrow \mathbb{R}$  gebildet wird. Aus Gründen der mathematischen Einfachheit wählt man meist das Quadrat des euklidischen Abstands  $d(x, y) = (x - y)^2$ , was zwar keine Metrik ist, aber zumindest symmetrisch ist und  $d(x, y) = 0$  genau dann, wenn  $x = y$  erfüllt. Diese Wahl von  $d$  ist aber keineswegs zwingend, die Berechnung des optimalen  $f$  wird aber durch andere Wahlen von  $d$  unter Umständen beträchtlich aufwendiger. Die Verwendung von  $d(x, y) = (x - y)^2$  setzt natürlich voraus, dass  $X(t), t \in I$  endliche zweite Momente hat (sonst wäre typischerweise  $\mathbb{E}d(X(t_0), \hat{X}(t_0))$  unendlich für jedes  $f$ ).

Zur Lösung des Problems im Fall  $d(x, y) = (x - y)^2$  definieren wir den Raum  $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$  als den Hilbertraum aller über  $(\Omega, \mathcal{F}, \mathbb{P})$  definierten (reellwertigen) quadratintegrierbaren Funktionen (Zufallsvariablen) mit dem Skalarprodukt

$$\langle Y, Z \rangle = \mathbb{E}(YZ).$$

Sei  $K = L^2_{\tilde{I}}(\Omega, \mathcal{F}, \mathbb{P})$  der von den  $X(t), t \in \tilde{I}$  erzeugte abgeschlossene Teilraum von  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ . Er enthält zum Beispiel  $|X(t)X(u)|^{1/2}$ , falls  $t, u \in \tilde{I}$ . Alle Elemente von  $L^2_{\tilde{I}}(\Omega, \mathcal{F}, \mathbb{P})$  sind Funktionen der (beobachteten) Zufallsvariablen  $X(t), t \in \tilde{I}$ . Die beste Approximation (im obigen Sinn) einer Zufallsvariablen  $X(t_0)$ , die in  $H$ , aber im allgemeinen nicht in  $L^2_{\tilde{I}}(\Omega, \mathcal{F}, \mathbb{P})$  liegt, ist gerade die orthogonale Projektion  $\hat{X}(t_0)$  von  $X(t_0)$  auf  $L^2_{\tilde{I}}(\Omega, \mathcal{F}, \mathbb{P})$ . Aus der Vorlesung WTII ist bekannt, dass

$$\hat{X}(t_0) = \mathbb{E}(X(t_0)|X(t), t \in \tilde{I}).$$

Es gilt also:

$$\mathbb{E}(X(t_0) - \hat{X}(t_0))^2 = \min_{Y \in L^2_{\tilde{I}}(\Omega, \mathcal{F}, \mathbb{P})} \mathbb{E}(X(t_0) - Y)^2. \quad (4.2.1)$$

Im Fall  $\tilde{I} = \emptyset$  enthält  $L^2_{\tilde{I}}(\Omega, \mathcal{F}, \mathbb{P})$  übrigens nur die Konstanten, ist also eindimensional. In diesem Fall ist  $\hat{X}(t_0) = \mathbb{E}X(t_0)$ , also die bedingte Erwartung gleich dem (unbedingten) Erwartungswert.

Unser Problem ist bislang aber nur theoretisch gelöst. Die entscheidende Frage ist, wie man die orthogonale Projektion  $\hat{X}(t_0)$  als Funktion  $f$  von  $X(t), t \in \tilde{I}$  explizit ausdrückt. Das Problem ist deswegen oft sehr schwer zu lösen, weil auch im Fall  $|\tilde{I}| < \infty$  der Raum  $K = L^2_{\tilde{I}}(\Omega, \mathcal{F}, \mathbb{P})$  typischerweise unendlich-dimensional ist – selbst im Fall  $|\tilde{I}| = 1$ .

Um wenigstens im Fall  $|\tilde{I}| < \infty$  auf ein endlichdimensionales Approximationsproblem zu kommen, vereinfacht man oft die Fragestellung dadurch, dass man, anstatt in (4.2.1) das Minimum über *alle*  $Y \in L^2_{\tilde{I}}(\Omega, \mathcal{F}, \mathbb{P})$  zu bilden, nur über die  $Y$  minimiert, die sich als *Linearkombination* der  $X(t), t \in \tilde{I}$  und der Konstanten schreiben lassen. Damit die zugelassenen  $Y$  einen abgeschlossenen Teilraum von  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  bilden, vervollständigt man noch bezüglich der Norm von  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  (im Fall  $|\tilde{I}| = \infty$ ).

Wir bezeichnen diese Vervollständigung mit  $L^2_{\tilde{I},lin}(\Omega, \mathcal{F}, \mathbb{P})$ . Ist  $|\tilde{I}| = k < \infty$ , dann ist dieser Raum maximal  $k+1$ -dimensional. Im Fall linearer Abhängigkeiten zwischen den  $X(t), t \in \tilde{I}$  und den Konstanten ist die Dimension kleiner als  $k+1$ , sonst gleich  $k+1$ . Das optimale  $Y$  in dieser Klasse bezeichnen wir mit  $\tilde{X}(t_0)$ .  $\tilde{X}(t_0)$  ist also die orthogonale Projektion von  $X(t_0)$  auf den Raum  $L^2_{\tilde{I},lin}(\Omega, \mathcal{F}, \mathbb{P})$ . Da  $L^2_{\tilde{I},lin}(\Omega, \mathcal{F}, \mathbb{P}) \subseteq L^2_{\tilde{I}}(\Omega, \mathcal{F}, \mathbb{P})$ , ist  $\mathbb{E}(X(t_0) - \tilde{X}(t_0))^2$  im allgemeinen größer als  $\mathbb{E}(X(t_0) - \hat{X}(t_0))^2$ . Also ist  $\tilde{X}(t_0)$  ein schlechterer Prädiktor für  $X(t_0)$  als  $\hat{X}(t_0)$ , ist dafür aber meist leichter berechenbar.

Wir zeigen nun, wie man  $\tilde{X}(t_0)$  im Fall  $|\tilde{I}| < \infty$  als Lösung eines linearen Gleichungssystems berechnen kann.

**Satz 4.2.1.** *Sei  $X(t), t \in I$  ein reellwertiger stochastischer Prozess,  $\tilde{I} = \{t_1, \dots, t_m\} \subseteq I$  und  $t_0 \in I \setminus \tilde{I}$ . Weiter sei  $\mathbb{E}X^2(t) < \infty$  für  $t \in \tilde{I} \cup \{t_0\}$  und*

$$\begin{aligned} \mu(t) &:= \mathbb{E}X(t) \\ R(t, u) &:= \mathbb{E}X(t)X(u), \quad t, u \in \tilde{I} \cup \{t_0\}. \end{aligned}$$

*Sei  $\tilde{X}(t_0)$  die orthogonale Projektion von  $X(t_0)$  auf den von den Linearkombinationen von  $X(t), t \in \tilde{I}$  und den Konstanten erzeugten Teilraum  $L^2_{\tilde{I},lin}(\Omega, \mathcal{F}, \mathbb{P})$  von  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ . Dann gilt*

$$\tilde{X}(t_0) = \sum_{i=1}^m a_i X(t_i) + a,$$

wobei  $a, a_1, \dots, a_m$  irgendeine Lösung von

$$\begin{cases} \sum_{j=1}^m a_j R(t_i, t_j) + a\mu(t_i) = R(t_0, t_i), \quad i = 1, \dots, m \\ \sum_{j=1}^m a_j \mu(t_j) + a = \mu(t_0) \end{cases} \quad (4.2.2)$$

ist.

**Beweis.** Die orthogonale Projektion  $\tilde{X}(t_0)$  von  $X(t_0)$  auf den Spann von  $X(t_1), \dots, X(t_m), \mathbb{1}$  ( $\mathbb{1} =$  Konstante 1) ist charakterisiert durch

$$\begin{aligned} \mathbb{E}((X(t_0) - \tilde{X}(t_0))X(t_i)) &= 0, \quad i = 1, \dots, m \\ \mathbb{E}(X(t_0) - \tilde{X}(t_0))\mathbb{1} &= 0 \end{aligned} \quad (4.2.3)$$

und die Bedingung, dass  $\tilde{X}(t_0)$  in diesem Spann liegt, also eine Darstellung

$$\tilde{X}(t_0) = \sum_{i=1}^m a_i X(t_i) + a$$



hat. Setzt man letztere Formel in (4.2.3) ein, so ist das dadurch entstehende lineare Gleichungssystem identisch mit (4.2.2). □

**Bemerkungen.** Man sieht aus (4.2.2), dass man zur Berechnung der optimalen linearen Vorhersage nur  $\mu$  und  $r$  braucht. Dies hat eine große praktische Bedeutung, da man bei unbekannter Verteilung nur  $\mu$  und  $r$  und nicht die komplette Verteilung schätzen muss, um die optimale lineare Vorhersage zu berechnen. Auch der sogenannte *Vorhersagefehler*  $\mathbb{E}(X(t_0) - \tilde{X}(t_0))^2$  hängt nur von  $\mu$  und  $r$  (oder  $\mu$  und  $R$ ) ab.

### 4.3 Rekursive lineare Vorhersage

Der Aufwand zur Lösung von (4.2.2) mit Standardmethoden ist von der Größenordnung  $m^3$ . Kommt eine neue Beobachtung hinzu, so muss man ein neues Gleichungssystem mit  $m + 1$  Unbekannten lösen. Es stellt sich die Frage, ob man sich dabei Arbeit sparen kann indem man (Zwischen)ergebnisse früherer Berechnungen mitverwendet. Konkret betrachten wir den Fall, dass sukzessive  $X_1, X_2, X_3, \dots$  beobachtet werden und man jeweils die beste lineare Vorhersage der nächsten Zufallsvariable berechnen will. Wir nehmen der Einfachheit halber an, dass  $\mathbb{E}X_i = 0$  für alle  $i \in \mathbb{N}$  gilt und setzen wieder  $r(i, j) = R(i, j) = \text{cov}(X_i, X_j) = \mathbb{E}(X_i X_j)$  für  $i, j \in \mathbb{N}$ . Weiter sei

$$H_n = \text{sp}\{X_1, \dots, X_n\}$$

der (lineare) Spann von  $X_1$  bis  $X_n$  und  $\tilde{X}_{n+1}$  die orthogonale Projektion (= beste lineare Vorhersage) von  $X_{n+1}$  auf  $H_n$ . Weiter sei  $v_n = \|X_{n+1} - \tilde{X}_{n+1}\|^2 = \mathbb{E}((X_{n+1} - \tilde{X}_{n+1})^2)$  der Vorhersagefehler bei der optimalen linearen Vorhersage von  $X_{n+1}$  aufgrund von  $X_1$  bis  $X_n$ .

Offenbar gilt:

$$H_n = \text{sp}\{X_n - \tilde{X}_n, \dots, X_1 - \tilde{X}_1\}$$

(man beachte, dass  $\tilde{X}_1 = 0$  ist). Da  $\tilde{X}_{n+1} \in H_n$  für jedes  $n \in \mathbb{N}$  gilt, existiert eine Darstellung

$$\tilde{X}_{n+1} = \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \tilde{X}_{n+1-j}),$$

die sogenannte *Innovationsdarstellung*. Man beachte, dass die *Innovationen*  $X_n - \tilde{X}_n$  für  $n \in \mathbb{N}$  orthogonal sind (wobei es möglich ist, dass  $X_n - \tilde{X}_n = 0$  für gewisse  $n \in \mathbb{N}$  ist). Bei der Innovationsdarstellung wird die (neue) Vorhersage im  $n$ -ten Schritt durch die früheren Beobachtungen  $X_{n+1-j}$  und die früheren Vorhersagen  $\tilde{X}_{n+1-j}$  ausgedrückt. Die Aufgabe besteht nun darin, die zunächst unbekanntesten Koeffizienten  $\theta_{nj}$  zu berechnen und zwar möglichst so, dass sie ohne großen Aufwand aus den  $\theta_{kj}$  für  $1 \leq k < n$  berechenbar sind. Der folgende Satz (der sich auch in [BD87], S. 165 findet) zeigt, wie dies funktioniert:

**Satz 4.3.1.** *Sei  $X_1, X_2, \dots$  eine Folge von quadratintegrierbaren Zufallsvariablen mit  $\mathbb{E}X_i = 0, i \in \mathbb{N}$ . Weiter sei für alle  $n$  die Matrix  $(r(i, j))_{i,j=1,\dots,n}$  invertierbar. Mit den obigen Bezeichnungen gilt dann:*

$$\tilde{X}_{n+1} = \begin{cases} 0 & n = 0 \\ \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \tilde{X}_{n+1-j}) & n \geq 1 \end{cases} \quad (4.3.1)$$

und

$$\begin{cases} v_0 = r(1, 1) \\ \theta_{n,n-k} = v_k^{-1}(r(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j) & k = 0, \dots, n-1 \\ v_n = r(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j & n \in \mathbb{N}. \end{cases} \quad (4.3.2)$$

**Bemerkung 4.3.2.** (4.3.2) erlaubt eine rekursive Berechnung der  $\theta_{nj}$  und  $v_j$  in der Reihenfolge  $v_0; \theta_{11}, v_1; \theta_{22}, \theta_{21}, v_2; \theta_{33}, \theta_{32}, \theta_{31}, v_3$  usw. Der Zusatzaufwand zur Berechnung von  $\theta_{nm}$  bis  $\theta_{n1}$  ist von der Größenordnung  $n^2$  anstatt  $n^3$  bei einem nichtrekursiven Verfahren. Für große  $n$  ist der Speicheraufwand allerdings nicht unerheblich, da man sich alle früheren  $\theta_{nj}$  und  $v_j$  aufheben muss.

**Beweis von Satz 4.3.1.** Bildet man in (4.3.1) beidseitig das Skalarprodukt mit  $X_{k+1} - \tilde{X}_{k+1}$  für  $0 \leq k < n$ , so erhält man (wegen der Orthogonalität der Innovationen)

$$\langle \tilde{X}_{n+1}, X_{k+1} - \tilde{X}_{k+1} \rangle = \theta_{n,n-k} v_k$$

und somit (unter erneuter Verwendung der Orthogonalität)

$$\theta_{n,n-k} = v_k^{-1} \langle X_{n+1}, X_{k+1} - \tilde{X}_{k+1} \rangle.$$

Die Tatsache, dass  $v_k \neq 0$  ist, folgt aus der Invertierbarkeit der Kovarianzmatrizen. Nun setzen wir (4.3.1) mit  $k$  statt  $n$  ein und erhalten:

$$\theta_{n,n-k} = v_k^{-1} (r(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \langle X_{n+1}, X_{j+1} - \tilde{X}_{j+1} \rangle).$$

Ersetzt man  $\langle X_{n+1}, X_{j+1} - \tilde{X}_{j+1} \rangle$  durch  $\theta_{n,n-j} v_j$  (siehe oben), so erhält man, wie behauptet,

$$\theta_{n,n-k} = v_k^{-1} (r(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j).$$

Weiter gilt mit dem (verallgemeinerten) Satz von Pythagoras:

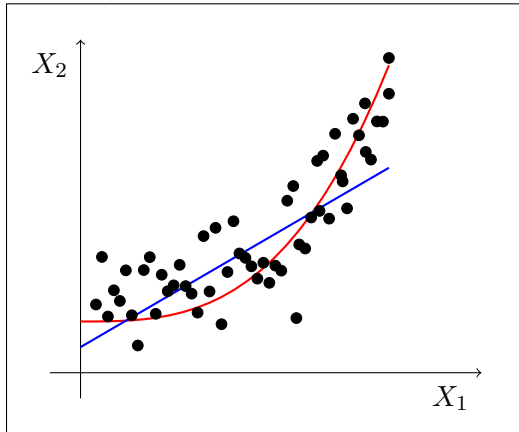
$$v_n = \|X_{n+1} - \tilde{X}_{n+1}\|^2 = \|X_{n+1}\|^2 - \|\tilde{X}_{n+1}\|^2 = r(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j.$$

□

Wir präsentieren (nach der folgenden Bemerkung) zwei Beispiele, die zeigen, dass optimale lineare Vorhersagen gelegentlich deutlich schlechter sind als optimale (nichtlineare) Vorhersagen.

**Bemerkung 4.3.3.** Satz 4.3.1 bleibt im Prinzip richtig, wenn man die Invertierbarkeit von  $r$  nicht voraussetzt. Dann ändert sich natürlich die rekursive Formel für  $\theta_{n,n-k}$  im Fall  $v_k = 0$  (Aufgabe: man finde die richtige Formel!).

**Beispiel 4.3.4.** Wir wählen  $I = \{1, 2\}$  und  $X_1 =$  Größe und  $X_2 =$  Gewicht einer zufällig gewählten Person. Sicher sind  $X_1$  und  $X_2$  *nicht* unabhängig. Ein “Scatterplot” vieler Messungen könnte (etwas übertrieben) so aussehen:



Die beste lineare Vorhersage des Gewichts gegeben die Größe ist die eingezeichnete Regressionsgerade (jedenfalls im Grenzfall unendlich vieler Beobachtungen). Man sieht, dass vor allem in den Bereichen extremer Größen, aber auch im Zentralbereich die Vorhersage schlecht ist. Eine geeignete kubische statt einer linearen Funktion führt dagegen zu wesentlich besseren Vorhersagen.

**Beispiel 4.3.5.** Sei  $I = \{1, 2\}$ ,  $X_1 \sim \mathcal{N}(0, 1)$  und  $X_2 = X_1^2$ . Offenbar ist  $\mathbb{E}(X_2|X_1) = X_1^2 = X_2$ , d.h.  $X_2$  lässt sich aus  $X_1$  perfekt vorhersagen. Wir berechnen nun die beste lineare Vorhersage  $\tilde{X}_2 = a_1 X_1 + a$ . Das Gleichungssystem (4.2.2) in Satz 4.2.1 lautet:

$$\begin{aligned} a_1 R(1, 1) + a \mu(1) &= R(1, 2) \\ a_1 \mu(1) + a &= \mu(2). \end{aligned}$$

Hier ist  $R(1, 1) = \mathbb{E}X_1^2 = 1$ ,  $R(1, 2) = \mathbb{E}(X_1 X_2) = \mathbb{E} X_1^3 = 0$ ,  $\mu(1) = 0$  und  $\mu(2) = \mathbb{E}X_1^2 = 1$ . Die Lösung des Systems lautet daher

$$\begin{aligned} a_1 &= 0 \\ a &= 1. \end{aligned}$$

Also ist die beste lineare Schätzung von  $X_1^2$  aufgrund von  $X_1$  die Konstante 1. Diese Vorhersage ist also viel schlechter als die beste nichtlineare Vorhersage  $X_1^2$ . Letztere erfordert in diesem Fall sogar geringeren Rechenaufwand.

Während im allgemeinen die beste lineare Vorhersage schlechter als die beste (nichtlineare) Vorhersage ist, sind im Fall von *Gaußprozessen* beide gleich, wie der folgende Satz zeigt.

**Satz 4.3.6.** Für Gaußprozesse  $X(t), t \in I$  ist die beste Vorhersage linear.

**Beweis.** Die Aussage gilt für beliebige  $\tilde{I} \subseteq I$ , wir zeigen sie aber nur für  $|\tilde{I}| < \infty$ . Sei  $\tilde{I} = \{t_1, \dots, t_m\}, t_0 \in I \setminus \tilde{I}$  und

$$\tilde{X}(t_0) = \sum_{i=1}^m a_i X(t_i) + a$$

die beste lineare Vorhersage von  $X(t_0)$  gegeben  $X(t_1), \dots, X(t_m)$ . Also gilt

$$(X(t_0) - \sum_{i=1}^m a_i X(t_i) - a) \perp \mathbb{1}, X(t_1), \dots, X(t_m).$$

Nun orthogonalisieren wir  $\mathbb{1}, X(t_1), \dots, X(t_m)$ : das Resultat sei  $\mathbb{1}, X_1, \dots, X_r$ . Nun ist  $(X(t_0) - \sum_{i=1}^m a_i X(t_i) - a, \mathbb{1}, X_1, \dots, X_r)$  ein Gaußvektor mit paarweise unkorrelierten Komponenten. Nach Satz 4.1.11 sind die Komponenten sogar unabhängig. Insbesondere sind die beiden Zufallsvariablen

$$X(t_0) - \sum_{i=1}^m a_i X(t_i) - a \text{ und } f(\mathbb{1}, X_1, \dots, X_r) = g(\mathbb{1}, X(t_1), \dots, X(t_m))$$

unabhängig für jedes messbare  $f$  (oder  $g$ ) und damit gilt

$$\mathbb{E}((X(t_0) - \sum_{i=1}^m a_i X(t_i) - a)g(\mathbb{1}, X(t_1), \dots, X(t_m))) = 0,$$

für jedes messbare  $g: \mathbb{R}^{m+1} \rightarrow \mathbb{R}$  mit  $\mathbb{E}g^2(\mathbb{1}, X(t_1), \dots, X(t_m)) < \infty$ .

Daraus folgt

$$\hat{X}(t_0) = \mathbb{E}(X(t_0)|\mathbb{1}, X(t_1), \dots, X(t_m)) = \check{X}(t_0).$$

□

**Bemerkung 4.3.7.** Man kann Satz 4.3.6 wie folgt interpretieren:

Unter allen Prozessen mit vorgegebener Erwartungswertfunktion und vorgegebener Kovarianzfunktion ist der Gaußprozess am schlechtesten (nichtlinear) vorhersagbar. Die Güte der linearen Vorhersage ist nämlich für all diese Prozesse identisch. Für die meisten dieser Prozesse ist aber die beste nichtlineare Vorhersage strikt besser als die beste lineare. Zu den Ausnahmen gehört aber nach dem letzten Satz der (einzige) Gaußprozess in der Klasse.

**Bemerkung 4.3.8.** Zwischen der besten linearen Vorhersage und der bedingten Erwartung (= beste (nichtlineare) Vorhersage) gibt es die besten polynomialen Vorhersagen mit vorgegebenem Grad. So ist im Fall  $|\tilde{I}| < \infty$  die beste quadratische Vorhersage für  $X(t_0)$  von der Form

$$\check{X}(t_0) = a + \sum_{i=1}^m a_i X(t_i) + \sum_{i,j=1}^m a_{ij} X(t_i) X(t_j). \quad (4.3.3)$$

Die Koeffizienten bestimmen sich wieder aus einem *linearen* Gleichungssystem, wenn man in

$$\begin{aligned} \mathbb{E}((X(t_0) - \check{X}(t_0))X(t_i)) &= 0 & i = 1, \dots, m \\ \mathbb{E}(X(t_0) - \check{X}(t_0)) &= 0 \\ \mathbb{E}((X(t_0) - \check{X}(t_0))X(t_i)X(t_j)) &= 0 & i, j = 1, \dots, m \end{aligned}$$

Formel (4.3.3) einsetzt. Damit dieser Ansatz funktioniert, müssen alle Momente bis zur vierten Ordnung endlich und bekannt sein. Die bessere Qualität der quadratischen gegenüber der linearen Vorhersage wird also dadurch erkauft, dass man mehr Informationen über den Prozess benötigt, was vor allem dann problematisch sein kann, wenn alle diese Parameter aus Realdaten *geschätzt* werden müssen.

## 4.4 Filterprobleme, der Kalmanfilter

Spezialfälle der Betrachtungen im letzten Abschnitt sind neben Vorhersageproblemen (im engeren Sinne) auch *Interpolations-* und *Filterprobleme*. Ein Beispiel für ein Interpolationsproblem ist: bei einem reellen Prozess  $X_t, t \in \mathbb{N}_0$  liegen Messwerte für  $t = 0, 1, 2, 4$  vor, der für  $t = 3$  fehle. Er lässt sich optimal (im Sinne minimalen quadratischen Fehlers) interpolieren durch  $\hat{X}_3 = \mathbb{E}(X_3|X_0, X_1, X_2, X_4)$  oder optimal linear interpolieren durch  $\tilde{X}_3$  wie zuvor (“linear” heißt hier natürlich *nicht*, dass  $\tilde{X}_3$  notwendig der Mittelwert von  $X_2$  und  $X_4$  ist).

Bei *Filterproblemen* hat man es immer mit einem höherdimensionalen Prozess zu tun, bei dem gewisse Komponenten den interessierenden Zustand eines Systems beschreiben und andere Komponenten typischerweise verrauschte Beobachtungen (von Funktionen) der Zustandskomponenten. Der Zustand kann z.B. die Position und Geschwindigkeit einer Rakete sein. Das Filterproblem besteht darin, bei gegebenem Modell und Beobachtung nur einiger Komponenten bis zur Zeit  $n$  die anderen Komponenten zur Zeit  $n$  (oder  $n + m$  für ein  $m \in \mathbb{Z}$ ) optimal zu schätzen. Dies kann man wie vorher für jede zu schätzende Komponente getrennt durch orthogonale Projektion lösen. Besonders berühmt ist der Spezialfall des *Kalmanfilters*, den wir etwas näher untersuchen wollen. Wir folgen dabei der Darstellung von Jazwinski: *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

Sei  $X_0, X_1, X_2, \dots$  ein  $\mathbb{R}^d$ -wertiger Prozess auf  $(\Omega, \mathcal{F}, \mathbb{P})$ , der durch folgende Rekursion gegeben ist:

$$X_{n+1} = F_n X_n + V_{n+1}, \quad n \in \mathbb{N}_0,$$

wobei  $F_n$  für jedes  $n \in \mathbb{N}_0$  eine vorgegebene (deterministische)  $d \times d$ -Matrix ist. Der  $\mathbb{R}^k$ -wertige Beobachtungsprozess  $Y_0, Y_1, \dots$  sei gegeben durch

$$Y_n = G_n X_n + W_n, \quad n \in \mathbb{N}_0$$

mit bekannten  $k \times d$ -Matrizen  $G_n$ . Wir nehmen an, dass alle Komponenten von  $X_0, V_{n+1}$  und  $W_n, n \in \mathbb{N}_0$  ein endliches zweites Moment und Erwartungswert Null haben und je zwei Komponenten verschiedener Elemente von  $\{X_0, V_{n+1}, W_n, n \in \mathbb{N}_0\}$  unkorreliert sind. Sei  $Q_n = \text{cov}(V_n)$  und  $R_n = \text{cov}(W_n)$ .

Sei  $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$ ,  $H_n$  der Teilraum der Linearkombinationen der Komponenten von  $Y_0, \dots, Y_n, n \in \mathbb{N}_0$  und  $H_{-1} := \{0\}$ .  $\pi_n$  bezeichne die orthogonale Projektion von  $H$  auf  $H_n$ . Weiter sei  $X_{-1} := 0, V_0 := X_0, F_{-1} := 0$  (womit  $X_0 = F_{-1}X_{-1} + V_0$ ) und  $\hat{X}_n^m := \pi_m(X_n)$  (komponentenweise) für  $m, n + 1 \in \mathbb{N}_0$ . Unser Ziel besteht darin, für festes  $m \in \mathbb{N}_0$  eine Rekursionsformel für  $\hat{X}_{n+m}^m$  zu finden. Eine solche Formel geben wir im folgenden Satz an. Dazu brauchen wir noch die Matrizen

$$P_n^m := \text{cov}(X_n - \hat{X}_n^m), \quad m, n + 1 \in \mathbb{N}_0,$$

die die Vorhersagefehler beschreiben.

**Satz 4.4.1.** *Sei  $R_n$  invertierbar für jedes  $n \in \mathbb{N}_0$ , dann gilt für  $n \in \mathbb{N}_0$ :*

- (i)  $\hat{X}_{n+1}^n = F_n \hat{X}_n^n,$
- (ii)  $P_{n+1}^n = F_n P_n^n F_n^T + Q_{n+1},$

$$(iii) \hat{X}_n^n = \hat{X}_n^{n-1} + K_n(Y_n - G_n \hat{X}_n^{n-1}),$$

$$(iv) P_n^n = P_n^{n-1} - K_n G_n P_n^{n-1},$$

wobei

$$(v) K_n := P_n^{n-1} G_n^T (G_n P_n^{n-1} G_n^T + R_n)^{-1}$$

der "Kalmanzuwachs" ist. Weiter gelten die Anfangsbedingungen

$$(vi) \hat{X}_0^{-1} = 0, P_0^{-1} = \text{cov}(X_0) =: Q_0.$$

**Bemerkung 4.4.2.** Die Formeln erlauben – startend mit  $\hat{X}_0^{-1}$  und  $P_0^{-1}$  – eine rekursive Berechnung von  $\hat{X}_n^n, \hat{X}_{n+1}^n, P_n^n$  und  $P_{n+1}^n$ , wobei die  $P$ s deterministisch sind. Man sieht, dass die Rekursion für die  $P$ s nicht von den Beobachtungen abhängt und sich daher "off-line" im Voraus berechnen lässt, während in die  $\hat{X}$  (wie zu erwarten) die Beobachtungen (linear) einfließen. Die Berechnung der  $P$ s ist auch notwendig, wenn man an ihnen kein Interesse hat, da sie via  $K_n$  in die Rekursion der  $\hat{X}$  eingehen.

Ist man nur an  $\hat{X}_{n+1}^n$  und nicht an  $\hat{X}_n^n$  interessiert, so kann man die dritte in die erste Gleichung einsetzen und entsprechend, wenn man nur an  $\hat{X}_n^n$  interessiert ist, die erste in die dritte (mit verschobenem  $n$ ) einsetzen.

Interessiert man sich (auch) für  $\hat{X}_{n+m}^n$  für  $m \geq 2$ , so kann man die Formel

$$\hat{X}_{n+m}^n = \prod_{k=n+m-1}^n F_k \hat{X}_n^n$$

verwenden, denn

$$\begin{aligned} \hat{X}_{n+m}^n &= \pi_n(X_{n+m}) = \pi_n(F_{n+m-1} X_{n+m-1} + V_{n+m}) \\ &= F_{n+m-1} \pi_n(X_{n+m-1}) = \dots = \prod_{k=n+m-1}^n F_k \pi_n(X_n) \end{aligned}$$

per Induktion, wobei wir  $\pi_n(V_{n+k}) = 0$  für  $k \in \mathbb{N}$  benutzen.

#### Beweis von Satz 4.4.1.

(i) haben wir gerade eben gezeigt (setze  $m = 1$ ). Auch (vi) ist offensichtlich.

(ii)

$$\begin{aligned} P_{n+1}^n &= \mathbb{E}((X_{n+1} - \hat{X}_{n+1}^n)(X_{n+1} - \hat{X}_{n+1}^n)^T) \\ &= \mathbb{E}((F_n X_n + V_{n+1} - F_n \hat{X}_n^n)(F_n X_n + V_{n+1} - F_n \hat{X}_n^n)^T) \\ &= \mathbb{E}(F_n (X_n - \hat{X}_n^n)(X_n - \hat{X}_n^n)^T F_n^T) + Q_{n+1} \\ &= F_n P_n^n F_n^T + Q_{n+1} \quad \text{für } n \in \mathbb{N}_0, \end{aligned}$$

wobei wir (i) benutzen, sowie die Unkorreliertheit der Komponenten von  $V_{n+1}$  mit den Komponenten von  $X_n$  und  $\hat{X}_n^n$ .

(iii) Wir berechnen zunächst  $\pi_{n-1}(Y_n)$  für  $n \in \mathbb{N}_0$ :

$$\pi_{n-1}(Y_n) = \pi_{n-1}(G_n X_n + W_n) = G_n \pi_{n-1}(X_n) + \pi_{n-1}(W_n) = G_n \hat{X}_n^{n-1} + 0.$$

Nach Definition von  $H_n$  erzeugen die Komponenten von  $Y_n - \pi_{n-1}(Y_n)$  das orthogonale Komplement von  $H_{n-1}$  in  $H_n$ , somit existiert eine  $d \times k$ -Matrix  $K_n$  mit

$$\pi_n(X_n) = \pi_{n-1}(X_n) + K_n(Y_n - \pi_{n-1}(Y_n)),$$

also

$$\hat{X}_n^n = \hat{X}_n^{n-1} + K_n(Y_n - G_n \hat{X}_n^{n-1}). \quad (4.4.1)$$

(v) Zu zeigen ist, dass  $K_n$  die in (v) angegebene Gestalt hat. Für  $n \in \mathbb{N}_0$  gilt

$$Y_n = G_n X_n + W_n = G_n F_{n-1} X_{n-1} + G_n V_n + W_n,$$

sowie mit (iii), (i)

$$\begin{aligned} X_n - \hat{X}_n^n &= F_{n-1} X_{n-1} + V_n - \hat{X}_n^{n-1} - K_n(Y_n - G_n \hat{X}_n^{n-1}) \\ &= F_{n-1}(X_{n-1} - \hat{X}_{n-1}^{n-1}) + V_n - K_n G_n F_{n-1}(X_{n-1} - \hat{X}_{n-1}^{n-1}) \\ &\quad - K_n(G_n V_n + W_n). \end{aligned}$$

Nun gilt  $X_n - \hat{X}_n^n \perp Y_n$  (für je 2 Komponenten), also – unter Benutzung von

$$Y_n = G_n F_{n-1}(X_{n-1} - \hat{X}_{n-1}^{n-1} + \hat{X}_{n-1}^{n-1}) + G_n V_n + W_n :$$

$$\begin{aligned} 0 &= \mathbb{E} \left[ (X_n - \hat{X}_n^n) Y_n^T \right] = F_{n-1} P_{n-1}^{n-1} F_{n-1}^T G_n^T + Q_n G_n^T \\ &\quad - K_n(G_n F_{n-1} P_{n-1}^{n-1} F_{n-1}^T G_n^T + G_n Q_n G_n^T + R_n), \end{aligned}$$

woraus mit (ii) die Formel (v) folgt (da  $R_n$  und damit  $G_n P_{n-1}^{n-1} G_n^T + R_n$  invertierbar ist).

(iv) Aus (4.4.1) folgt (mit  $I = d$ -dim. Einheitsmatrix)

$$\begin{aligned} P_n^n &= \text{cov}(X_n - \hat{X}_n^{n-1} - K_n G_n (X_n - \hat{X}_n^{n-1}) - K_n W_n) \\ &= (I - K_n G_n) P_{n-1}^{n-1} (I - K_n G_n)^T + K_n R_n K_n^T \\ &= (I - K_n G_n) P_{n-1}^{n-1} + (K_n G_n - I) P_{n-1}^{n-1} G_n^T K_n^T + K_n R_n K_n^T \\ &= (I - K_n G_n) P_{n-1}^{n-1} + K_n (G_n P_{n-1}^{n-1} G_n^T + R_n) K_n^T - P_{n-1}^{n-1} G_n^T K_n^T \\ &= (I - K_n G_n) P_{n-1}^{n-1} \end{aligned}$$

nach (v).

□

## 4.5 Schwach stationäre Prozesse

Wir beschäftigen uns nun mit der Frage, wie man Datenreihen vorhersagt, für die a-priori keine Information über die Verteilung des Prozesses, als deren Realisierung sie interpretiert werden, vorliegt. Man wird in diesem Fall versuchen, entweder die komplette Verteilung aus den Daten zu schätzen und dann die jeweils für die besten (nichtlineare) Vorhersage benötigten bedingten Erwartungen zu berechnen (wobei die Frage nach einem Berechnungsverfahren zunächst offen bleibt), oder lediglich die Mittelwerte und Kovarianzen zu schätzen und dann nach Satz 4.2.1 oder 4.3.1 die besten *linearen* Vorhersagen explizit berechnen. Natürlich könnte man analog auch gemäß Bemerkung 4.3.8 die besten quadratischen Vorhersagen usw. berechnen.

Wir behandeln im weiteren nur den Fall der optimalen *linearen* Vorhersage. Wir stellen uns vor, dass  $Y_1, Y_2, \dots$  zeitlich äquidistante Messungen sind. Liegen die Beobachtungen  $y_1, \dots, y_n$  vor, so sieht man aus Satz 4.3.1, dass man zur Vorhersage von  $X_{n+1} := Y_{n+1} - \mathbb{E}Y_{n+1}$  den Mittelwert  $\mu(n+1)$  und die Kovarianzen  $r(n+1, k+1)$  benötigt. Da bislang keine Beobachtung von  $Y_{n+1}$  vorliegt, hat man zunächst keinerlei Grundlage für eine Schätzung dieser Größen. Man behilft sich dadurch, dass man darauf vertraut, dass  $Y_{n+1}$  sich im Vergleich zu den bisherigen Daten  $y_1, \dots, y_n$  nicht völlig anders verhält. Man wird also versuchen, "Muster" oder "Strukturen" in den Beobachtungen  $y_1, \dots, y_n$  zu erkennen und auf  $y_{n+1}$  zu extrapolieren.

Hierfür gibt es prinzipiell sehr viele Möglichkeiten. Die folgende hat sich in vielen Fällen aber als besonders geeignet erwiesen. Sie beruht darauf, dass viele Datenreihen *Trends* und/oder *saisonale Komponenten* aufweisen und ansonsten – d.h. nach Trend- und Saisonbereinigung – "stationär aussehen". Damit ist ein konkreter Weg bereits angedeutet: Man schätzt zunächst Trend und saisonale Komponenten, zieht diese ab und schätzt dann für den als stationär angenommenen Prozess den (konstanten) Mittelwert und die Kovarianzen, die dann nur noch von der Zeitdifferenz abhängen. Man geht also von der Annahme aus, dass nach Trend- und Saisonbereinigung die statistische Kopplung von  $Y_{n+1}$  an  $Y_n, \dots, Y_{n-k}$  dieselbe ist wie die von  $Y_n$  an  $Y_{n-1}, \dots, Y_{n-k-1}$  und  $Y_{n-1}$  an  $Y_{n-2}, \dots, Y_{n-k-2}$  usw.

Wir werden nun diese etwas vagen Äußerungen präzisieren, indem wir zunächst sehr allgemein definieren, was ein *stationärer* und *schwach stationärer* Prozess ist und dann andeuten, wie man Trend- und Saisonbereinigung durchführen kann und wie man Mittelwert und Kovarianzen eines schwach stationären Prozesses schätzen kann.

**Definition 4.5.1.** Sei  $(I, +)$  eine Halbgruppe und  $X(t), t \in I$  ein reellwertiger stochastischer Prozess. Dieser heißt (*strikt*) *stationär*, wenn

$$\mathbb{P}(\{\omega : (X(t_1), \dots, X(t_k)) \in B\}) = \mathbb{P}(\{\omega : (X(t_1 + h), \dots, X(t_k + h)) \in B\})$$

für alle  $k \in \mathbb{N}, B \in \mathcal{B}^k$  und  $t_1, \dots, t_k, h \in I$  gilt.

**Bemerkung 4.5.2.** Nach Satz 1.4.3 bleiben nicht nur die endlich-dimensionalen Verteilungen unter einem "Zeitshift" invariant, sondern sogar die Verteilung selbst.

**Definition 4.5.3.** Sei  $(I, +)$  eine Halbgruppe und  $X(t), t \in I$  ein stochastischer Prozess. Dieser heißt *schwach stationär* (oder *stationär von zweiter Ordnung*), wenn  $EX^2(t) < \infty$  für alle  $t \in I$



und

$$\begin{aligned}\mathbb{E}X(t_1) &= \mathbb{E}X(t_1 + h) \\ \mathbb{E}(X(t_1)X(t_2)) &= \mathbb{E}(X(t_1 + h)X(t_2 + h))\end{aligned}$$

für alle  $t_1, t_2, h \in I$  gilt.

**Bemerkung 4.5.4.** Fordert man  $\mathbb{E}|X(t)|^m < \infty$  für alle  $t \in I$  und  $\mathbb{E}(X(t_1)X(t_2) \dots X(t_k)) = \mathbb{E}(X(t_1 + h) \dots X(t_k + h))$  für alle  $k \leq m \in \mathbb{N}$  und  $t_1, \dots, t_k, h \in I$ , dann nennt man den Prozess stationär von  $m$ -ter Ordnung. Wir werden solche Prozesse aber nicht speziell betrachten.

**Bemerkung 4.5.5.** Wenn ein Prozess endliche zweite Momente hat, dann folgt im Falle seiner (strikten) Stationarität die schwache Stationarität. Die Umkehrung gilt dagegen nicht (siehe Beispiel 4.5.7). Wie wir im folgenden Satz sehen, ist die Umkehrung allerdings unter der Zusatzvoraussetzung richtig, dass der Prozess Gaußsch ist.

**Satz 4.5.6.** Ein schwach stationärer Gaußprozess ist stationär.

**Beweis.** Sei  $X(t), t \in (I, +)$  Gaußsch und schwach stationär. Dann gilt für  $\mu(t) := \mathbb{E}X(t), r(t, u) = \text{cov}(X(t), X(u))$  nach Definition 4.5.3

$$\mu(t) = \mu(t + h), \quad r(t, u) = r(t + h, u + h), \quad t, u, h \in I. \quad (4.5.1)$$

Die Verteilungen von  $(X(t_1) \dots X(t_k))$  und  $(X(t_1 + h) \dots X(t_k + h))$  sind dann  $\mathcal{N}((\mu(t_1), \dots, \mu(t_k))^T, r(t_i, t_j)_{i,j=1, \dots, k})$  bzw.  $\mathcal{N}((\mu(t_1 + h), \dots, \mu(t_k + h))^T, r(t_i + h, t_j + h)_{i,j=1, \dots, k})$  und stimmen daher nach (4.5.1) überein, d.h.  $X(t), t \in I$  ist stationär.  $\square$

**Beispiel 4.5.7.** Das folgende Beispiel zeigt, dass ein schwach stationärer Prozess nicht immer stationär ist. Sei  $I = \mathbb{Z}$  und  $X_i, i \in \mathbb{Z}$  unabhängig mit  $X_i \sim \mathcal{N}(0, 1)$  für  $i \neq 0$  und  $\mathbb{P}(\{X_0 = -1\}) = \mathbb{P}(\{X_0 = 1\}) = 1/2$ . Dann gilt  $\mu(i) = 0, r(i, j) = \delta_{ij}$  für  $i, j \in \mathbb{Z}$ , d.h. der Prozess ist schwach stationär. Andererseits gilt aber z.B.  $0 = \mathbb{P}(\{X_0 > 2\}) \neq \mathbb{P}(\{X_1 > 2\}) > 0$ , d.h. nicht einmal die eindimensionalen Verteilungen sind invariant unter Zeitshifts.

Wir diskutieren zuerst, wie man für eine als schwach stationär angenommene Datenreihe  $x_1, \dots, x_n$  den (konstanten) Erwartungswert und die Kovarianzfunktion schätzt. Danach gehen wir kurz auf die Frage der Trend- und Saisonbereinigung von Datenreihen ein.

## 4.6 Schätzung von Erwartungswert und Kovarianzfunktion

Ein naheliegender und in sehr vielen Fällen sinnvoller Schätzer für den Erwartungswert  $\mu$  eines schwach stationären Prozesses  $X_i, i \in \mathbb{N}$  ist der (empirische) Mittelwert der Zeitreihe

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

**Satz 4.6.1. a)**  $\hat{\mu}_n, n \in \mathbb{N}$  ist eine Folge von erwartungstreuen Schätzern für  $\mu = \mathbb{E}X_i, i \in \mathbb{N}$ , d.h.  $\mathbb{E}\hat{\mu}_n = \mu$  für alle  $n \in \mathbb{N}$

b) Wenn  $\lim_{|k| \rightarrow \infty} |r(k)| = 0$  für  $r(k) := \text{cov}(X_1, X_{k+1})$  gilt, dann ist die Folge  $\hat{\mu}_n, n \in \mathbb{N}$  eine  $L^2$ -konsistente Folge von Schätzern für  $\mu$ , d.h.  $\mathbb{E}((\mu - \hat{\mu}_n)^2) \xrightarrow{n \rightarrow \infty} 0$ .

**Bemerkung 4.6.2.** Die Bedingung für  $r$  in b) lässt sich nicht ersatzlos streichen. So ist zum Beispiel für  $X_1 = X_2 = \dots \sim \mathcal{N}(\mu, 1)$   $\hat{\mu}_n = X_1$  für alle  $n \in \mathbb{N}$  und daher *nicht*  $L^2$ -konsistent (aber erwartungstreu). Man beachte, dass  $L^2$ -Konsistenz (schwache) Konsistenz impliziert, d.h.

$$\mathbb{P}(|\mu - \hat{\mu}_n| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0.$$

**Beweis.**

$$\text{a) } \mathbb{E}\hat{\mu}_n = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \mu.$$

$$\begin{aligned} \text{b) } \mathbb{E}((\hat{\mu}_n - \mu)^2) &= \mathbb{E}\left(\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)\right)^2\right) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}((X_i - \mu)(X_j - \mu)) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n r(i-j) = \frac{1}{n^2} \sum_{k=-(n-1)}^{n-1} (n - |k|)r(k) \\ &= \frac{1}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right)r(k) \\ &\leq \frac{1}{n} \sum_{k=-(n-1)}^{n-1} |r(k)| \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

□

Als Schätzer für die Kovarianzen  $r(k) = \text{cov}(X_1, X_{k+1}), k \in \mathbb{N}_0$  sind üblich:

$$\hat{r}_n^*(k) := \begin{cases} \frac{1}{n-k} \sum_{i=1}^{n-k} (x_i - \hat{\mu}_n)(x_{i+k} - \hat{\mu}_n) & k = 0, \dots, n-1 \\ 0 & \text{sonst.} \end{cases}$$

$$\hat{r}_n(k) := \begin{cases} \frac{1}{n} \sum_{i=1}^{n-k} (x_i - \hat{\mu}_n)(x_{i+k} - \hat{\mu}_n) & k = 0, \dots, n-1 \\ 0 & \text{sonst.} \end{cases}$$

Beide unterscheiden sich nur durch den Vorfaktor  $\frac{1}{n-k}$  bzw.  $\frac{1}{n}$ . Der erste Schätzer erscheint zunächst vernünftiger, da man durch die Anzahl der Summanden dividiert. Dennoch wird meist der zweite verwendet. Ein Grund dafür ist der, dass  $\hat{r}_n(k)$  als Funktion von  $k$  immer eine positiv semidefinite Funktion ist (das werden wir in Satz 4.7.9 sehen),  $\hat{r}_n^*(k)$  dagegen im allgemeinen nicht (Beispiel:  $n = 3, x_1 = 1, x_2 = 0, x_3 = -1$ ). Darüberhinaus hat  $\hat{r}_n$  in vielen Fällen asymptotisch günstigere Eigenschaften.

Für Eigenschaften wie (asymptotische) Erwartungstreue, Konsistenz, asymptotische Normalität usw. der Schätzungen  $\hat{r}_n^*$  und  $\hat{r}_n$  verweisen wir auf die Literatur ([Pr81], [BD87]).

Neben der eben erwähnten Methode sind zwei weitere Zugänge zur Schätzung der Kovarianzfunktion eines Prozesses üblich – eine *parametrische* im *Zeitbereich* und eine *nichtparametrische* im *Frequenzbereich*.

Bei dem *parametrischen* Zugang im *Zeitbereich* versucht man, aus einer durch eine endliche Zahl von Parametern beschriebenen Klasse von stationären Prozessen einen Prozess auszuwählen, der am besten zu den beobachteten Daten passt. Besonders beliebt ist die Klasse der (stationären) ARMA-Prozesse. Dies sind stationäre Prozesse (auf einem geeigneten Wahrscheinlichkeitsraum), die eine Darstellung der Form

$$X_n + \sum_{i=1}^p a_i X_{n-i} = \sum_{j=0}^q b_j \varepsilon_{n-j} \quad (4.6.1)$$

haben mit  $p, q \in \mathbb{N}_0, a_1, \dots, a_p, b_0, \dots, b_q \in \mathbb{R}$  und  $\varepsilon_n, n \in \mathbb{Z}$  *weißes Rauschen*, d.h.  $\mathbb{E}\varepsilon_n = 0, n \in \mathbb{Z}$  und  $\mathbb{E}\varepsilon_m \varepsilon_n = \delta_{mn}$ . Gibt es eine Darstellung mit  $p = 0$ , so heißt der Prozess MA (“Moving Average”)-Prozess, gibt es eine solche mit  $q = 0$ , dann heißt der Prozess AR (*autoregressiver*)-Prozess. Macht man eine spezielle Verteilungsannahme für die  $\varepsilon_n, n \in \mathbb{Z}$ , z.B. uiv  $\mathcal{N}(0, 1)$ , dann kann man die Parameter mit der Maximum-Likelihood-Methode (ML) schätzen. Es stellt sich dabei als wenig sinnvoll heraus, auch  $p$  und  $q$  so zu schätzen. Vielmehr wird man zunächst  $p$  und  $q$  mit anderen Methoden (möglichst klein) schätzen und dann die verbliebenen  $a_1, \dots, a_p, b_0, \dots, b_q$  nach ML schätzen. In vielen Fällen ist diese Berechnung mühsam und man verwendet daher Näherungsverfahren. Bei der Schätzung der  $a_i$  und  $b_j$  ist übrigens darauf zu achten, dass es auch wirklich einen *stationären* Prozess  $X_n, n \in \mathbb{Z}$  gibt mit einer Darstellung der Form (4.6.1). Für Details verweisen wir auf die Literatur ([Pr81], [BD87]). Man beachte die Ähnlichkeit von (4.6.1) zu den beim Kalmanfilter betrachteten Zustandsprozessen.

Der *nichtparametrische* Zugang über den *Frequenzbereich* benutzt die Tatsache, dass die positiv-semidefinite Kovarianzfunktion  $r_{n,n \in \mathbb{Z}}$  eines schwach stationären Prozesses, definiert als  $r_n = \mathbb{E}((X_0 - \mu)(X_n - \mu))$ , eine Darstellung der Form

$$r_n = \int_{-1/2}^{1/2} e^{2\pi i \lambda n} dm(\lambda), \quad n \in \mathbb{Z} \quad (4.6.2)$$

hat, wobei  $m$  ein endliches, symmetrisches Maß auf  $[-1/2, 1/2]$  ist, d.h.  $m([-1/2, 1/2]) < \infty$  und  $m(A) = m(-A)$ . Es gilt auch die Umkehrung, dass jedes solche Maß mit (4.6.2) eine positiv-semidefinite Funktion erzeugt.  $m$  nennt man *Spektralmaß* des Prozesses. Es ist eindeutig durch die Funktion  $r_n, n \in \mathbb{Z}$  festgelegt (Satz von Herglotz, in allgemeiner Form: Satz von Bochner). Anstatt  $r$  direkt zu schätzen, kann man auch  $m$  schätzen und erhält dann mittels (4.6.2) eine Schätzung für  $r_n$  für alle  $n \in \mathbb{Z}$ . Vor allem in dem wichtigen Spezialfall, dass  $m$  eine Dichte hat (die “Spektraldichte”), gibt es umfangreiche Literatur zur Frage des Schätzens von  $m$  (oder der Dichte von  $m$ ). Unter numerischen Gesichtspunkten erweist sich der Zugang über das Spektralmaß als sinnvoll, da man für die direkte Berechnung von  $\hat{r}_n(k)$  für alle  $k$   $O(n^2)$  Multiplikationen ausführen muss, beim Zugang über den Frequenzbereich dagegen nur  $O(n \log n)$ , da man mit der schnellen Fouriertransformation (FFT) arbeiten kann. Wenn  $n$  größer

als 1000 ist, merkt man den Unterschied sehr deutlich. Im nächsten Abschnitt gehen wir darauf etwas genauer ein.

## 4.7 Der nichtparametrische Ansatz; Spektraldichtenschätzung

**Satz 4.7.1.** *Sei  $m$  ein endliches symmetrisches Maß auf  $[-1/2, 1/2]$  (symmetrisch heißt:  $m(A) = m(-A)$  für jede Borelmenge  $A \subseteq [-1/2, 1/2]$ ).*

*Dann ist die Funktion  $r : \mathbb{Z} \rightarrow \mathbb{R}$  definiert durch*

$$r(n) := \int_{-1/2}^{1/2} e^{2\pi i n \lambda} dm(\lambda), \quad n \in \mathbb{Z}$$

*positiv semidefinit d.h.  $\sum_{j,k=1}^n z_j r(t_j - t_k) \overline{z_k}$  ist reell und  $\geq 0$  für alle  $n \in \mathbb{N}$ ,  $t_1, \dots, t_n \in \mathbb{Z}$  und  $(z_1, \dots, z_n) \in \mathbb{C}^n$ .*

**Beweis.**  $r(n)$  ist reellwertig, da wegen der Symmetrie von  $m$  gilt:

$$r(n) = \int_{-1/2}^{1/2} \cos(2\pi n \lambda) dm(\lambda) + i \int_{-1/2}^{1/2} \sin(2\pi n \lambda) dm(\lambda) = \int_{-1/2}^{1/2} \cos(2\pi n \lambda) dm(\lambda).$$

Sei  $n \in \mathbb{N}$ ,  $t_1, \dots, t_n \in \mathbb{Z}$  und  $(z_1, \dots, z_n)^T \in \mathbb{C}^n$ . Dann gilt

$$\begin{aligned} \sum_{j,k=1}^n z_j r(t_j - t_k) \overline{z_k} &= \sum_{j,k=1}^n z_j \int_{-1/2}^{1/2} e^{2\pi i (t_j - t_k) \lambda} dm(\lambda) \overline{z_k} \\ &= \int_{-1/2}^{1/2} \left| \sum_{j=1}^n z_j e^{2\pi i t_j \lambda} \right|^2 dm(\lambda) \geq 0. \end{aligned}$$

□

**Satz 4.7.2.** (Herglotz): *Zu jeder positiv semidefiniten Funktion  $r : \mathbb{Z} \rightarrow \mathbb{R}$  existiert genau ein symmetrisches endliches Maß  $m$  auf  $[-1/2, 1/2]$  mit*

$$r(n) = \int_{-1/2}^{1/2} e^{2\pi i n \lambda} dm(\lambda), \quad n \in \mathbb{Z}.$$

**Beweis.** Wir beweisen die Aussage nur unter der Zusatzbedingung  $\sum_{k \in \mathbb{Z}} |r(k)| < \infty$ . Wir definieren in diesem Fall

$$f(\lambda) := \sum_{k \in \mathbb{Z}} r(k) e^{-2\pi i k \lambda}, \quad -1/2 \leq \lambda \leq 1/2.$$

Die Reihe konvergiert und ihr Wert ist reell. Weiter gilt  $f(\lambda) = f(-\lambda)$ . Da  $r$  positiv semidefinit ist, folgt

$$0 \leq \frac{1}{N} \sum_{k,l=1}^N e^{2\pi i k \lambda} r(k-l) e^{-2\pi i l \lambda} = \sum_{k=-N+1}^{N-1} \left(1 - \frac{|k|}{N}\right) r(k) e^{-2\pi i k \lambda} \xrightarrow{N \rightarrow \infty} f(\lambda), \text{ d.h. } f(\lambda) \geq 0.$$

Die Funktion  $\lambda \mapsto f(\lambda)$  ist stetig als gleichmäßiger Grenzwert stetiger Funktionen und es gilt

$$\int_{-1/2}^{1/2} f(\lambda) e^{2\pi i n \lambda} d\lambda = \sum_{k \in \mathbb{Z}} r(k) \int_{-1/2}^{1/2} e^{-2\pi i k \lambda} e^{2\pi i n \lambda} d\lambda = r(n), \quad n \in \mathbb{N}.$$

Setzt man  $m(A) := \int_A f(\lambda) d\lambda$  für jede Borelmenge  $A \subseteq [-\frac{1}{2}, \frac{1}{2}]$ , dann folgt  $r(n) = \int_{-1/2}^{1/2} e^{2\pi i n \lambda} dm(\lambda)$ .

Die Eindeutigkeit zeigen wir nicht. □

**Definition 4.7.3.** Ist  $r : \mathbb{Z} \rightarrow \mathbb{R}$  die Kovarianzfunktion eines schwach stationären Prozesses, dann heißt das nach Satz 4.7.2 existierende Maß  $m$  *Spektralmaß* des Prozesses (oder Spektralmaß von  $r$ ) und  $f(\lambda) := \frac{dm}{d\lambda}(\lambda)$  die *Spektraldichte* (sofern sie existiert; sie existiert und ist stetig falls  $\sum |r(k)| < \infty$ ).

Wir zeigen nun, wie man die Spektraldichte schätzen kann. Gegeben seien Beobachtungen  $X_0, \dots, X_{N-1}$  eines schwach stationären Prozesses mit stetiger Spektraldichte  $f$ . Wir wollen  $f$  schätzen.

$$Z_N(\lambda) := \sum_{k=0}^{N-1} X_k e^{-2\pi i k \lambda}, \quad \lambda \in \mathbb{R}$$

heißt *diskrete Fouriertransformation* (DFT) von  $X_0, \dots, X_{N-1}$ .

**Proposition 4.7.4.** a)  $Z_N$  ist periodisch mit Periode 1.

b)  $Z_N(-\lambda) = \overline{Z_N(\lambda)}$ .

c)  $X_k = \frac{1}{N} \sum_{l=0}^{N-1} Z_N\left(\frac{l}{N}\right) e^{2\pi i \frac{kl}{N}} = \int_{-1/2}^{1/2} e^{2\pi i k \lambda} Z_N(\lambda) d\lambda, \quad k = 0, \dots, N-1$

**Beweis.** a) und b) sind klar

c)  $\frac{1}{N} \sum_{l=0}^{N-1} Z_N\left(\frac{l}{N}\right) e^{2\pi i \frac{kl}{N}} = \frac{1}{N} \sum_{l=0}^{N-1} \left( \sum_{m=0}^{N-1} X_m e^{-2\pi i \frac{ml}{N}} \right) e^{2\pi i \frac{kl}{N}} = \sum_{m=0}^{N-1} X_m \frac{1}{N} \sum_{l=0}^{N-1} e^{2\pi i l \frac{k-m}{N}} = X_k.$

Weiter gilt:

$$\begin{aligned} \int_{-1/2}^{1/2} e^{2\pi i k \lambda} Z_N(\lambda) d\lambda &= \int_{-1/2}^{1/2} \sum_{m=0}^{N-1} X_m e^{-2\pi i m \lambda} e^{2\pi i k \lambda} d\lambda \\ &= \sum_{m=0}^{N-1} X_m \int_{-1/2}^{1/2} e^{2\pi i (k-m)\lambda} d\lambda = X_k. \end{aligned}$$

□

**Bemerkung 4.7.5.** Proposition 4.7.4 c) zeigt, dass die Funktion  $Z_N(\lambda)$  durch die Werte an den Stellen  $\lambda = \frac{l}{N}, l = 0, \dots, N-1$  bereits festgelegt ist. Oft bezeichnet man auch die Zahlen  $Z_N(\frac{l}{N}), l = 0, \dots, N-1$  als diskrete Fouriertransformierte. Mit der sogenannten FFT (fast Fourier transform) lässt sich  $Z_N(\frac{l}{N}), l = 0, \dots, N-1$  in  $O(N \log N)$  Schritten "schnell" berechnen.

**Definition 4.7.6.** Die Funktion  $I_N(\lambda) := \frac{1}{N} |Z_N(\lambda)|^2, \lambda \in \mathbb{R}$  heißt *Periodogramm* der Daten  $X_0, \dots, X_{N-1}$ . Weiter definieren wir  $\tilde{I}_N(\lambda) := \frac{1}{N} |\tilde{Z}_N(\lambda)|^2, \lambda \in \mathbb{R}$ , wobei  $\tilde{Z}_N$  die DFT der zentrierten Daten  $X_l - \frac{1}{N} \sum_{j=0}^{N-1} X_j, l = 0, \dots, N-1$  ist.

**Proposition 4.7.7.** •  $I_N(\lambda) = I_N(-\lambda); I_N$  hat Periode 1.

•  $\tilde{I}_N(\lambda) = \tilde{I}_N(-\lambda); \tilde{I}_N$  hat Periode 1;  $\tilde{I}_N(0) = 0$ .

**Beweis.** Klar. □

**Satz 4.7.8.**  $\tilde{I}_N(\lambda) = \sum_{k \in \mathbb{Z}} \hat{r}_N(k) e^{-2\pi i k \lambda}, \lambda \in \mathbb{R}$ .

**Beweis.**

$$\begin{aligned} \tilde{I}_N(\lambda) &= \frac{1}{N} \left| \sum_{k=0}^{N-1} (X_k - \hat{\mu}_N) e^{-2\pi i k \lambda} \right|^2 \\ &= \frac{1}{N} \sum_{m=0}^{N-1} \sum_{l=0}^{N-1} (X_m - \hat{\mu}_N)(X_l - \hat{\mu}_N) e^{2\pi i (l-m)\lambda} \\ &= \frac{1}{N} \sum_{k=-N+1}^{N-1} \sum_{l=0}^{N-|k|-1} (X_l - \hat{\mu}_N)(X_{l+|k|} - \hat{\mu}_N) e^{-2\pi i k \lambda} \\ &= \sum_{k=-N+1}^{N-1} \hat{r}_N(k) e^{-2\pi i k \lambda} = \sum_{k \in \mathbb{Z}} \hat{r}_N(k) e^{-2\pi i k \lambda}. \end{aligned}$$

□

**Satz 4.7.9.** *Es gilt*

$$\hat{r}_N(k) = \int_{-1/2}^{1/2} \tilde{I}_N(\lambda) e^{2\pi i k \lambda} d\lambda, k \in \mathbb{Z}.$$

*Insbesondere ist  $\hat{r}_N(k), k \in \mathbb{Z}$  positiv semidefinit (nach Satz 4.7.1).*

**Beweis.** Es gilt

$$\begin{aligned} \int_{-1/2}^{1/2} \tilde{I}_N(\lambda) e^{2\pi i k \lambda} d\lambda &= \int_{-1/2}^{1/2} \sum_{m \in \mathbb{Z}} \hat{r}_N(m) e^{-2\pi i m \lambda} e^{2\pi i k \lambda} d\lambda \\ &= \sum_{m \in \mathbb{Z}} \hat{r}_N(m) \delta_{k,m} = \hat{r}_N(k). \end{aligned}$$

□

Da  $\hat{r}_N(k)$  ein sinnvoller Schätzer für  $r(k)$  zu sein scheint, sollte man erwarten, dass  $\tilde{I}_N(\lambda)$  ein sinnvoller Schätzer für die Spektraldichte  $f(\lambda)$  ist (denn der Zusammenhang zwischen  $r$  und  $f$  ist derselbe wie der zwischen  $\hat{r}_N$  und  $\tilde{I}_N$ ). Zwar kann man relativ leicht unter schwachen Annahmen zeigen, dass  $\tilde{I}_N(\lambda)$  ein asymptotisch erwartungstreuer Schätzer für  $f(\lambda)$  ist, der allerdings in der Regel *nicht* konsistent ist (nicht einmal schwach konsistent). Dies hängt damit zusammen, dass  $\lambda \mapsto \tilde{I}_N(\lambda)$  für große  $N$  sehr kräftig oszilliert. Man behilft sich dadurch, dass man die Funktion  $\lambda \mapsto \tilde{I}_N(\lambda)$  mit einem ‘‘Fenster’’  $w_N$  (‘‘window’’) glättet. Dabei sei  $w_N$  für jedes  $N \in \mathbb{N}$  eine Funktion von  $\mathbb{R}$  nach  $\mathbb{R}$  mit

(i)  $w_N$  ist stetig

(ii)  $w_N$  hat Periode 1

(iii)  $w_N$  ist gerade (d.h.  $w_N(\lambda) = w_N(-\lambda)$ )

(iv)  $\int_{-1/2}^{1/2} w_N(\lambda) d\lambda = 1$

(v) Für jedes  $\varepsilon > 0, \varepsilon < 1/2$  gilt  $\lim_{N \rightarrow \infty} \int_{\varepsilon}^{1/2} |w_N(\lambda)| d\lambda = 0$ .

Insbesondere konvergiert  $w_N$  eingeschränkt auf  $[-1/2, 1/2]$  schwach gegen ein Punktmaß mit Masse 1 in der Null.

Unter geeigneten Voraussetzungen an  $w_N$  und den zugrunde liegenden Prozess konvergiert dann

$$\tilde{f}_N(\lambda) := \int_{-1/2}^{1/2} w_N(\lambda - \mu) \tilde{I}_N(\mu) d\mu$$

gleichmäßig für  $N \rightarrow \infty$  gegen die Spektraldichte  $f(\lambda)$  (fast sicher oder in  $L^2$  oder in Wahrscheinlichkeit).

Zu jedem Fenster  $w_N$  korrespondiert eine Schätzung der Kovarianzfunktion  $r$ :

$$\begin{aligned}\tilde{r}_N(k) &:= \int_{-1/2}^{1/2} \tilde{f}_N(\lambda) e^{2\pi i k \lambda} d\lambda = \int_{-1/2}^{1/2} \tilde{I}_N(\mu) e^{2\pi i k \mu} \int_{-1/2}^{1/2} w_N(\lambda - \mu) e^{2\pi i k(\lambda - \mu)} d\lambda d\mu \\ &= \hat{r}_N(k) \int_{-1/2}^{1/2} w_N(\lambda) e^{2\pi i k \lambda} d\lambda, \quad k \in \mathbb{Z}.\end{aligned}$$

Gilt  $w_N(\lambda) \geq 0$  für alle  $\lambda$ , dann ist das letzte Integral höchstens 1 und die Funktion  $\tilde{r}_N(\cdot)$  ist positiv semidefinit.

### Zusammenfassung des nichtparametrischen Zugangs

Gegeben seien reelle Zahlen  $x_0, \dots, x_{N-1}$ , die als Realisierung eines schwach stationären Prozesses mit stetiger Spektraldichte gedeutet werden. Gesucht (zu berechnen) sind  $\hat{r}_N(k)$ ,  $k \in \{0, \dots, N-1\}$  oder  $\tilde{r}_N(k)$ ,  $k \in \{0, \dots, N-1\}$  oder  $\tilde{f}_N(\lambda)$ ,  $-1/2 \leq \lambda \leq 1/2$ .

1. Ziehe von den Daten den Mittelwert  $\hat{\mu}_N$  ab  $\rightsquigarrow y_0, \dots, y_{N-1}$
2. Füge an die Daten  $N-1$  Nullen hinten an ("zero padding")
3. Berechne  $\tilde{Z}_{2N-1}(\frac{l}{2N-1}) (= \tilde{Z}_N(\frac{l}{2N-1}))$ ,  $l = 0, \dots, 2N-1$  via FFT
4. Berechne  $\tilde{I}_N(\frac{l}{2N-1}) = \frac{1}{N} |\tilde{Z}_{2N-1}(\frac{l}{2N-1})|^2$
5. Berechne  $\hat{r}_N(k)$ ,  $-N+1 \leq k \leq N-1$  durch FFT aus  $\tilde{I}_N(\frac{l}{2N-1})$ ,  $l = 0, \dots, 2N-2$
6. Berechne (ggf)  $\tilde{r}_N(k)$ ,  $-N+1 \leq k \leq N-1$  wie oben
7. Berechne (ggf)  $\tilde{f}_N(\frac{l}{2N-1})$ ,  $l = 0, \dots, N-1$  mit FFT aus  $\tilde{r}(k)$ ,  $k = -N+1, \dots, N-1$

**Bemerkung 4.7.10.** Wenn  $N$  groß ist, dann ist obige Berechnung selbst dann sinnvoll, wenn man nur an der Funktion  $\hat{r}_N(k)$ ,  $k \in \mathbb{Z}$  interessiert ist, da man dann nur  $O(N \log N)$  Operationen benötigt statt  $O(N^2)$  bei Berechnung der  $\hat{r}_N(k)$ ,  $k \in \mathbb{Z}$  nach der Definition.

## 4.8 Datentransformation, Trendbereinigung, Saisonbereinigung

Der erste Schritt bei der Analyse einer Zeitreihe  $x_1, \dots, x_n$  mit Werten im Intervall  $I \subseteq \mathbb{R}$  ist in der Regel der, dass man eine streng monoton wachsende Funktion  $f: I \rightarrow \mathbb{R}$  findet, so dass die Reihe  $y_i = f(x_i)$ ,  $i = 1, \dots, n$  sich als Teil einer Realisierung eines Prozesses der Form

$$Y_i = m_i + s_i + Z_i, \quad i \in \mathbb{Z} \tag{4.8.1}$$

auffassen lässt, wobei  $m_i$  den *Trend* darstellt,  $s_i$  periodisch ist (oder jedenfalls  $s_i = \tilde{s}(i)$  für eine periodische Funktion  $\tilde{s}: \mathbb{R} \rightarrow \mathbb{R}$  gilt) und  $Z_i$ ,  $i \in \mathbb{Z}$  schwach stationär ist. Wir wollen nicht genau definieren, was ein *Trend* ist. Die Vorstellung ist aber die, dass  $m_i$  eine Funktion einfacher Gestalt ist, die sich nicht zu schnell ändert (jedenfalls nicht "oszilliert" - was auch immer dies genau bedeutet).  $m_i$  könnte zum Beispiel eine Exponentialfunktion oder ein Polynom niedrigen



Grades sein.

In manchen Fällen braucht man nicht zu transformieren, d.h.  $f$  kann als die Identität gewählt werden. Eine Transformation empfiehlt sich aber immer dann, wenn die Stärke der lokalen Fluktuation vom Niveau der  $x$ -Werte abhängt - zum Beispiel monoton wachsend. In diesem Fall sollte  $f$  so gewählt werden, dass  $f'(x)$  ungefähr umgekehrt proportional zur Fluktuationsintensität beim Niveau  $x$  ist. Wachsen die Fluktuationen (im Mittel) zum Beispiel proportional zu  $x$  (dies ist bei Preisen oder Aktienkursen plausibel), dann ist  $f(x) = \log x$  eine angemessene Transformation, die zudem den Wertebereich von  $I = ]0, \infty[$  auf  $\mathbb{R}$  transformiert.

Man kann sich fragen, warum man überhaupt nach der Elimination des Trends noch den saisonalen Anteil abtrennt, denn die Summe eines stationären Prozesses und einer periodischen Funktion (mit gleichverteilt zufälliger vom Prozess unabhängiger Phase) ist ebenfalls stationär. Der Grund ist der, dass man gerne mit solchen (schwach) stationären Prozessen arbeitet, deren Kovarianzfunktionen (möglichst schnell) gegen Null konvergieren (wir sahen die Bedeutung dieser Eigenschaft in Satz 4.6.1 und periodische stationäre Prozesse haben periodische Kovarianzfunktionen (die deswegen *nicht* konvergieren)).

Nach diesen allgemeinen Bemerkungen gehen wir nun auf die Frage der Trend- und Saisonschätzung bzw. -elimination etwas genauer ein. Dazu nehmen wir an, dass erstens die Periode  $d$  der Funktion  $s_{i,i \in \mathbb{Z}}$  ganzzahlig und bekannt sei (zu der Frage, wie man  $d$  aus der Zeitreihe schätzen kann, verweisen wir auf die genannte Literatur) und dass (oBdA)  $\mathbb{E}Z_i = 0, i \in \mathbb{Z}$  und  $\sum_{i=1}^d s_i = 0$  gelte. Die letzten Voraussetzungen sind deswegen keine Einschränkungen, weil man anderenfalls die additiven Konstanten immer dem Trend  $m_i, i \in \mathbb{Z}$  zuschlagen kann.

Für die Trendeinschätzung hat man die Wahl zwischen parametrischen und nichtparametrischen Ansätzen. Bei parametrischen Ansätzen versucht man für eine durch endlich viele Parameter beschriebene Funktionsklasse die Parameter so zu bestimmen, dass die beobachtete (evtl. transformierte) Datenreihe  $y_1, \dots, y_n$  gut "passt" (z.B. im Sinne kleinster Quadrate). Bei nichtparametrischen Schätzungen stellt man zunächst keine Voraussetzungen an die Klasse der Schätzungen  $\hat{m}_i, i \in \mathbb{Z}$ . Gängige nichtparametrische Schätzungen sind *Moving-average-Filter* der Form

$$\hat{m}_i = \sum_{j \in \mathbb{Z}} \alpha_j y_{i-j}, \quad (4.8.2)$$

wobei  $\sum_{j \in \mathbb{Z}} \alpha_j = 1$  gilt. Typischerweise wird man  $\alpha_j = \alpha_{-j}$  und  $\alpha_0 \geq \alpha_1 \geq \dots$  und nicht zu viele  $\alpha_j \neq 0$  wählen. Probleme treten an den Rändern des Beobachtungsintervalls auf, da dort (4.8.2) wegen nichtvorliegender Daten nicht ausgewertet werden kann. Bei gleichzeitigem Vorliegen einer saisonalen Komponente sollte man die  $\alpha_j$  so wählen, dass  $\hat{m}_i$  die Periodizität möglichst nicht spürt z.B. dadurch, dass man in (4.8.2) über ein ganzzahliges Vielfaches der Periode die  $y_i$  mittelt.

Eine andere übliche Methode ist das Differenzieren der Datenreihe, womit sich polynomiale

Trends vollständig beseitigen lassen. Gilt nämlich

$$Y_i = \sum_{j=0}^k \gamma_j i^j + s_i + Z_i, \quad i \in \mathbb{Z}$$

und bildet man  $k$  mal die Differenzen jeweils aufeinanderfolgender Werte, so erhält man

$$(\nabla^k Y)_i = k! \gamma_k + (\nabla^k s)_i + (\nabla^k Z)_i,$$

einen schwach stationären Prozess mit saisonaler Komponente ohne Trend (bis auf die Konstante  $k! \gamma_k$ , die man - so sie stört - durch eine weitere Differenzbildung auch noch eliminieren kann). Man beachte, dass die Saisonkomponente von  $\nabla^k Y$  durch  $\nabla^k s$  gegeben ist.

Oft verwendet man das Differenzieren nur dazu, den Grad des Trendpolynoms zu schätzen und schätzt dann die Koeffizienten des Polynoms auf andere Weise (z.B. mit der Methode der kleinsten Quadrate).

Die Saisonkomponente  $s_i$  schätzt man immer, *nachdem* man den Trend eliminiert hat. Wir nehmen daher an, dass  $Y_i$  von der Form  $Y_i = s_i + Z_i$  ist ( $m_i$  also Null ist). Eine übliche Schätzung  $\hat{s}_i$  für  $s_i$  ist

$$\hat{s}_i = \frac{1}{\lfloor \frac{n-i}{d} \rfloor + 1} \sum_{k=0}^{\lfloor \frac{n-i}{d} \rfloor} Y_{i+kd}, \quad i = 1, \dots, d, \quad (4.8.3)$$

wobei  $\lfloor x \rfloor$  die größte ganze Zahl  $\leq x$  ist. Für  $i > d$  setzt man  $\hat{s}_i$  periodisch mit Periode  $d$  fort.

Hat man den Trend geschätzt (parametrisch oder nichtparametrisch) und anschließend die Saisonkomponente (z.B. nach (4.8.3)), so wird man für die trend- und saisonbereinigte Zeitreihe  $z_i$  die Kovarianzfunktion (z.B. so wie früher vorgeschlagen) schätzen und mit Satz 4.2.1 oder 4.3.1 vorhersagen. Eine Vorhersage für den nächsten Wert  $y_{n+1}$  der ursprünglichen Daten  $y_1, \dots, y_n$  erhält man, indem man zur Vorhersage  $\hat{z}_{n+1}$  den Trend  $\hat{m}_{n+1}$  und die Saisonkomponente  $\hat{s}_{n+1}$  addiert.

Wir fassen noch einmal zusammen, wie man beobachtete Zeitreihen vorhersagen kann.

## 4.9 Zusammenfassung

- Man transformiere die Daten so, dass sie als Realisierung eines schwach stationären Prozesses plus Trend plus saisonaler Komponente interpretierbar sind.
- Man schätze und eliminiere den Trend.
- Man schätze und eliminiere den saisonalen Anteil.
- Für den verbliebenen schwach stationären Anteil schätze man Erwartungswert und die Kovarianzfunktion (vgl. Satz 4.6.1, die Formel für  $\hat{r}_n(k)$  im Anschluss daran, sowie Abschnitt 4.7).

- Mit diesen Größen gehe man in Satz 4.2.1 oder 4.3.1 und berechne die nächste Vorhersage.
- Man transformiere diese zurück auf die ursprüngliche Zeitreihe.

Wir betonen, dass dies keineswegs die beste oder gar die einzige Methode der Vorhersage ist. Die Literatur zur Zeitreihenanalyse ist voll mit Verbesserungsvorschlägen für die einzelnen Punkte. Wir erwähnten schon den (vermeidbaren) numerischen Aufwand bei der direkten Berechnung von  $\hat{r}_n$ .

Dieses Kapitel sollte lediglich einige Fragestellungen in der Zeitreihenanalyse und erste Lösungsansätze nahebringen. Auf neuere Zugänge, wie “Bootstrap”methoden konnte aus Zeitgründen überhaupt nicht eingegangen werden. Es ist zu erwarten, dass mit solchen (und ähnlichen) Methoden ganz neue Vorhersagemethoden in der Zukunft entwickelt werden, die sich völlig von dem Konzept der besten linearen Vorhersage lösen und gleichzeitig algorithmisch einfacher – allerdings rechenzeitintensiver – sein werden.



# Bibliographie

- [Ai04] M. AIGNER und G. ZIEGLER, *Das BUCH der Beweise*, 2. Auflage, Springer, Berlin, 2004.
- [As03] S. ASMUSSEN, *Applied Probability and Queues*, Second edition, Springer, Berlin, 2003.
- [Ba68] H. BAUER, *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, de Gruyter, Berlin, 1968.
- [Bra08] M. BRAMSON, *Stability of Queueing Networks*, Springer, Berlin, 2008.
- [Br68] L. BREIMAN, *Probability*, Addison-Wesley, Reading, 1968.
- [BD87] P. BROCKWELL und R. DAVIS, *Time Series: Theory and Methods*, Springer, Berlin, 1987.
- [Ch67] K. L. CHUNG, *Markov Chains with Stationary Transition Probabilities*, Springer, Berlin, 1967.
- [Ci75] E. CINLAR, *Introduction to Stochastic Processes*, Prentice-Hall, Englewood, 1975.
- [Fe68] W. FELLER, *An Introduction to Probability Theory and its Applications*, Vol. I, Wiley, New York, 1968.
- [GS75] I. I. GIKHMAN und A. V. SKOROHOD, *The Theory of Stochastic Processes II*, Springer, Berlin, 1975.
- [Ha02] O. HÄGGSTRÖM, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, Cambridge, 2002.
- [He86] H. HEUSER, *Analysis 1*, Teubner, Stuttgart, 1986.
- [HS82] D. HEYMAN und M. SOBEL, *Stochastic Models in Operations Research, Vol. I*, McGraw Hill, New York, 1982.
- [KT75] S. KARLIN und H. M. TAYLOR, *A First Course in Stochastic Processes*, Academic Press, New York, 1975.
- [KT81] S. KARLIN und H. M. TAYLOR, *A Second Course in Stochastic Processes*, Academic Press, New York, 1975.
- [Le09] D. LEVIN, Y. PERES und E. WILMER, *Markov Chains and Mixing Times*, AMS, Providence, 2009.
- [Mi63] R. MILLER, Stationary equations in continuous time Markov chains, *Trans. Amer. Math. Soc.* **109**, 35-44, 1963.
- [No97] J. NORRIS, *Markov Chains*, Cambridge Univ. Press, Cambridge, 1997.

- [Pi90] B. PITTEL, On a Daley-Kendall model of random rumours, *Journal of Applied Probability* **27**, 1990.
- [Pr81] M. PRIESTLEY, *Spectral Analysis and Time Series*, Wiley, New York, 1981.
- [Sh84] A.N. SHIRYAYEV, *Probability*, Springer, Berlin, 1984.
- [Ti94] H. TIJMS, *Stochastic Models*, Wiley, Chichester, 1994.
- [Wi79] D. WILLIAMS, *Diffusions, Markov Processes and Martingales*, Wiley, 1979.