STOCHASTISCHE PROZESSE

I: Markovketten in diskreter und stetiger Zeit

Wolfgang König

Vorlesungsskript

Universität Leipzig

Sommersemester 2005

Inhaltsverzeichnis

Vorwort

Dies ist das Vorlesungsskript einer zweistündigen einführenden Vorlesung über stochastische Prozesse. Es werden zunächst die grundlegenden Begriffe motiviert und entwickelt, wie endlich dimensionale Verteilungen und die Verteilung eines stochastischen Prozesses. Der Rest dieses Skriptes ist den Markovketten in diskreter und in stetiger Zeit auf endlichen oder höchstens abzählbaren Zustandsräumen gewidmet. Wir führen die grundlegenden Begriffe und wichtigsten Beispiele ein und diskutieren ausführlich die Rekurrenz- und Transienzeigenschaften, Klasseneinteilung der Zustände, invariante Maße, Konvergenz in das Gleichgewicht, Sub-Markov'sche Halbgruppen, Q-Matrizen, Vorwärts- und Rückwärtsgleichungen und ihre Beziehungen untereinander. Fast alle Aussagen werden vollständig bewiesen.

Der Skriptteil über Markovketten in diskreter Zeit ist im Wesentlichen meinem Skript *Stochastik I* (Universität zu Köln, Sommersemester 2003) entnommen, der Teil über kontinuierliche Zeit folgt einem Skript *Stochastische Modelle* (TU Berlin) von Michael Scheutzow, dem ich hiermit für die freundliche Überlassung danke.

Leipzig, im März 2005

In der vorliegenden Überarbeitung wurden etliche Punkte korrigiert, klarer erläutert, gestrafft und geordnet.

Leipzig, im Mai 2006

Kapitel 1

Grundlagen

In diesem Kapitel wollen wir den Begriff des stochastischen Prozesses anschaulich und formal einführen.

1.1 Einleitung

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),
- Ausbreitung von Epidemien oder von Genen,
- Populationsentwicklung,
- Bewegung eines Elektrons in einem magnetischen Feld,
- Simulation einer Wahrscheinlichkeitsverteilung,
- Stochastisches Optimieren,
- Kartenmischen,
- Aktienkurse,
- Kapital eines Versicherungsunternehmens,
- 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 in einem geeigneten Raum von (Zeit-)funktionen, also eine zufällige Funktion.

Zu einem stochastischen Prozess gehört immer ein *Zustandsraum* E und eine *Parametermenge* I. Der Zustandsraum ist der Wertebereich oder eine Obermenge davon, die Parametermenge ist der Definitionsbereich der zufälligen Funktion, meist ein Zeitintervall. 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 die Spezialfälle, wo E endlich oder abzählbar unendlich ist oder $E = \mathbb{R}$, $\mathbb{R}^n$ oder ein Funktionenraum und I diskret (also endlich oder z. B. gleich $\mathbb{N}$, $\mathbb{N}_0$ oder $\mathbb{Z}$) oder kontinuierlich ist (also etwa $\mathbb{R}$, $[0,\infty)$, $\mathbb{R}^n$ oder die Kugeloberfläche).

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.

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 der Wahrscheinlichkeiten gewisser Ereignisse oder qualitativen Aussagen wie Langzeitverhalten, Rückkehreigenschaften, Existenz von Gleichgewichtszuständen. In diesem Skript werden wir diese Fragen und noch mehr für zwei der fundamentalen Typen von stochastischen Prozessen allgemein untersuchen und an Hand von Beispielen illustrieren: Markovketten in diskreter und in stetiger Zeit. Doch zuvor legen wir noch die allgemeinen mathematischen Fundamente der Theorie der stochastischen Prozesse.

1.2 Stochastische Prozesse und ihre Verteilung

Wir wollen nun formal den Begriff eines stochastischen Prozesses definieren. Zunächst geben wir die Definition für reelle Werte.

Definition 1.2.1 (Stochastischer Prozess). *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.*

Meist ist I eine Teilmenge von $\mathbb{R}$ und wird als Zeit interpretiert.

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

Eine ausführlichere und allgemeine Definition eines stochastischen Prozesses lautet wie folgt:

Definition 1.2.2 (Stochastischer Prozess). *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 $X_\cdot(\omega)\colon I \to E$* 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 $X\colon (E_1, \mathcal{E}_1) \to (E_2, \mathcal{E}_2)$ zwischen zwei Messräumen *messbar* heisst, wenn alle Urbilder unter X von Elementen in $\mathcal{E}_2$ in $\mathcal{E}_1$ liegen, also

$$B \in \mathcal{E}_2 \qquad \Longrightarrow \qquad X^{-1}(B) = \{\omega \in E_1 \colon X(\omega) \in B\} \in \mathcal{E}_1. \tag{1.2.1}$$

Bekanntermaßen ist für Zufallsvariablen der Begriff der *Verteilung* sehr wichtig. Sehr oft sind die Größen, die einen an einer Zufallsvariablen interessieren, Funktionen der Verteilung der Zufallsvariablen, zum Beispiel Erwartungswert, Varianz oder 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 $|I| = 1$ und $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$, wobei $\mathcal{B}$ die *Borel-σ-Algebra* über $\mathbb{R}$ ist, nämlich die von den Intervallen erzeugte. Eine reelle Zufallsvariable X ist eine messbare Abbildung

$$X\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}), \tag{1.2.2}$$

und die Verteilung $\mathbb{P} \circ X^{-1}$ von X ist das Bildmaß auf $(\mathbb{R}, \mathcal{B})$ von $\mathbb{P}$ unter X d. h.

$$\mathbb{P} \circ X^{-1}(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega \in \Omega \colon X(\omega) \in B\}), \qquad B \in \mathcal{B}. \tag{1.2.3}$$

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 Situation, wo X ein stochastischer Prozess ist, hat man statt der Wertemenge $\mathbb{R}$ eine Menge von Funktionen $I \to E$, also eine Teilmenge von E^I, d. h. zufällige Funktionen statt zufälliger Zahlen. Um wie im reellwertigen Fall wie oben eine Verteilung $\mathbb{P} \circ X^{-1}$ definieren zu können, braucht man auf E^I eine σ-Algebra. Die am meisten verwendete und üblichste ist die Produkt-σ-Algebra $\mathcal{E}^{\otimes I}$:

Definition 1.2.3 (Produkt-σ-Algebra). *$\mathcal{E}^{\otimes I}$ ist die kleinste σ-Algebra über E^I, die alle endlich-dimensionalen Rechtecke enthält, d. h. alle Mengen der Form*

$$\{f \in E^I \colon f(i_1) \in E_1, \dots, f(i_k) \in E_k\}, \qquad \text{mit } k \in \mathbb{N}, i_1, \dots, i_k \in I, E_1, \dots, E_k \in \mathcal{E}. \tag{1.2.4}$$

Dann stellt sich auf elementare Weise (Übungsaufgabe) heraus, dass ein stochastischer Prozess nichts weiter als eine $(E^I, \mathcal{E}^{\otimes I})$-wertige Zufallsgröße ist:

Lemma 1.2.4. *Sei $(X_t)_{t \in I}$ ein stochastischer Prozess mit Zustandsraum $(E, \mathcal{E})$. Definiere eine Abbildung $X\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (E^I, \mathcal{E}^{\otimes I})$ durch*

$$(X(\omega))(t) := X_t(\omega), \qquad \omega \in \Omega, t \in I. \tag{1.2.5}$$

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

Wir schreiben auch $X = (X_t)_{t \in I}$ für die in (1.2.5) formal definierte Zufallsgröße, also den stochastischen Prozess. Also können wir die seine Verteilung ganz analog zur Verteilung einer Zufallsvariablen definieren.

Definition 1.2.5 (Verteilung eines stochastischen Prozesses). *Die* Verteilung 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* (1.2.5) *definierten Abbildung X; in Formeln:*

$$\mathbb{P} \circ X^{-1}(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega \colon X(\omega) \in B\}), \qquad B \in \mathcal{E}^{\otimes I}. \tag{1.2.6}$$

Aus der Verteilung 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. 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 einem gegebenen Wahrscheinlichkeitsmaß $\widehat{P}$ auf $(E^I, \mathcal{E}^{\otimes I})$ 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.

1.3 Endlich-dimensionale Verteilungen

Die Verteilung eines stochastischen Prozesses ist oft ein recht kompliziertes Objekt und ist meist nicht – oder nur mit großer Mühe – explizit angebbar. 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, was viel leichter und expliziter ist. 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 der σ-Algebra gegeben sind.

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

Definition 1.3.1 (Endlich-dimensionale Verteilungen). *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.*

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

Eine viel schwierigere Frage ist, unter welchen Umständen eine gegebene Familie von Wahrscheinlichkeitsmaßen $\widehat{P}_J$ auf $(E^J, \mathcal{E}^{\otimes J})$ mit $J \sqsubset I$ gerade die Familie der endlich dimensionalen Verteilungen einer Verteilung $\widehat{P}$ sind. Dies wird in dem wichtigen Satz 1.3.3 geklärt werden. Es stellt sich heraus, dass die $\widehat{P}_J$ gewisse Konsistenzbedingungen erfüllen müssen:

Definition 1.3.2 (Konsistenz). *Sei I eine nichtleere Indexmenge, $(E, \mathcal{E})$ ein Messraum, und für jedes $J \sqsubset I$ sei ein Wahrscheinlichkeitsmaß $\widehat{P}_J$ auf $(E^J, \mathcal{E}^{\otimes J})$ vorgegeben. Die Familie $(\widehat{P}_J)_{J\sqsubset I}$ heißt* konsistent*, wenn für alle $J_1 \sqsubset I$ und $J_2 \sqsubset I$ mit $J_1 \subset J_2$ gilt, dass $\widehat{P}_{J_1}$ gleich der Projektion von $\widehat{P}_{J_2}$ auf J_1 ist.*

Ein Beispiel für eine konsistente Folge ist etwa die Folge der Gleichverteilungen aller Nächstnachbarschaftspfade $(X_0, \dots, X_n) \in (\mathbb{Z}^d)^{n+1}$, die in $X_0 = 0$ starten (dies ist nichts Anderes als die sogenannte *gewöhnliche Irrfahrt*), und ein Beispiel für eine nicht-konsistente Folge ist die Folge der Gleichverteilungen auf der Menge aller solchen Pfade mit der Zusatzbedingung, dass er keinen Punkt mehr als einmal besucht, also dass $X_i \neq X_j$ für $i, j \in \{0, 1, \dots, n\}$ mit $i \neq j$ ist (dies nennt man die *selbstvermeidende Irrfahrt*).

Es stellt sich in Satz 1.3.3 heraus, dass die geeignete Struktur auf dem Zustandsraum E für unsere Frage nach der eindeutigen Charakterisierung eines stochastischen Prozesses durch endlich dimensionale Verteilungen die eines polnischen Raumes 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 σ-Algebra der Borelmengen, d. h. der kleinsten σ-Algebra, die alle offenen Mengen enthält. Polnische Räume haben sich als sehr natürliche Zustandsräume von stochastischen Prozessen erwiesen. Die meisten interessanten Zustandsräume $(E, \mathcal{E})$ *sind* polnisch, z. B. $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{C}^n$, $\mathbb{R}^{\mathbb{N}}$, $C([0,1], \mathbb{R}^n)$.

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

Satz 1.3.3 (Existenzsatz von Kolmogorov). *Wenn E ein polnischer Raum ist, I eine Indexmenge und $(\widehat{P}_J)_{J \sqsubset I}$ eine konsistente Familie von Wahrscheinlichkeitsmaßen, dann existiert genau ein Wahrscheinlichkeitsmaß $\widehat{P}$ auf $(E^I, \mathcal{E}^{\otimes I})$, so dass die $\widehat{P}_J$ die endlich dimensionalen Verteilungen von $\widehat{P}$ sind. Weiter existiert ein Wahrscheinlichkeitsraum und darauf ein stochastischer Prozess $(X_t)_{t \in I}$ mit Zustandsraum $(E, \mathcal{E})$ und Verteilung $\widehat{P}$.*

Der Beweis ist zu aufwendig, um ihn hier zu präsentieren. Der interessierte Leser sei auf [Ba91] verwiesen. Die letzte Aussage ist allerdings recht leicht zu sehen. Man kann als Wahrscheinlichkeitsraum immer $(E^I, \mathcal{E}^{\otimes I}, \widehat{P})$ wählen und $X_i(f) := f(i)$ für $i \in I$ und $f \in E^I$ wählen. Dann ist die in Lemma 1.2.4 definierte Abbildung X gerade die Identität und somit die Verteilung von X gleich $\widehat{P}$.

Eine Konsequenz von Satz 1.3.3 ist, dass, um einen Prozess $(X_n)_{n \in \mathbb{N}_0}$ oder $(X_t)_{t \in [0,\infty)}$ 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 bzw. von $(X_{t_0}, \dots, X_{t_n})$ für jede Wahl von $0 \leq t_0 < t_1 < \dots < t_n$ unter der Voraussetzung, dass die Familie der so definierten Verteilungen konsistent ist.

Kapitel 2

Markovketten in diskreter Zeit

In diesem Kapitel behandeln wir einen der wichtigsten stochastischen Prozesse, die Markovketten auf einem diskreten Raum in diskreter Zeit. Wir werden also immer die Parametermenge $I = \mathbb{N}_0$ wählen, und der Zustandsraum E ist eine beliebige diskrete Menge, also endlich oder höchstens abzählbar unendlich. Es ist üblich, den Zustandsraum in diesem Fall nicht E, sondern I zu nennen, und das tun wir hier auch. Im gesamten Kapitel sei I eine nichtleere, endliche oder höchstens abzählbar unendliche Menge, der Zustandsraum.

Man stelle sich ein Teilchen vor, das sich durch eine höchstens abzählbare Menge I zufällig bewegt und zu den Zeitpunkten $1, 2, 3, \dots$ jeweils zu einem (eventuell anderen) Punkt springt. Die besondere Eigenschaft der Sprungentscheidungen, die die Markoveigenschaft ausmacht, ist die Tatsache, dass diese Entscheidung nur von der aktuellen Position des Teilchens abhängt, aber (gegeben, man kennt diese Position) nicht von dem Verlauf des ganzen bisherigen Pfades, den das Teilchen zurück gelegt hat.

Markovketten spielen eine wichtige Rolle bei der Modellierung vieler zeitlich sich entwickelnder Prozesse: bei Mischungsvorgängen, bei Sortieralgorithmen und anderen stochastischen Algorithmen, bei der Modellierung physikalischer Prozesse oder von Finanzmärkten und Vielem mehr.

2.1 Definition und einfache Eigenschaften

Wir beginnen mit der Definition der Markoveigenschaft: Selbst wenn der gesamte bisherige Verlauf bekannt ist, hat nur die aktuelle Position einen Einfluss auf die Sprungentscheidung.

Definition 2.1.1 (Markoveigenschaft). *Wir sagen, eine (endliche oder unendliche) Folge $X_0, X_1, X_2, \dots$ von I-wertigen Zufallsgrößen* besitzt die Markoveigenschaft, *wenn für jedes $n \in \mathbb{N}$ und alle $i_0, i_1, \dots, i_{n+1} \in I$ gilt:*

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n), \qquad (2.1.1)$$

sofern alle auftretenden bedingten Wahrscheinlichkeiten wohldefiniert sind.

Das legt nahe, dass eine Markovkette im Wesentlichen durch die bedingten Wahrscheinlichkeiten $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n)$ festgelegt wird. Ihre Kollektion ist also ein ganz wesentliches Objekt:

Definition 2.1.2 (stochastische Matrix). *Eine Matrix $P = (p_{i,j})_{i,j\in I}$ heißt* stochastisch*, falls $p_{i,j} \in [0,1]$ für alle $i,j \in I$ gilt und $\sum_{j\in I} p_{i,j} = 1$ für jedes $i \in I$ gilt.*

Wir werden im Folgenden immer still schweigend davon ausgehen, dass die Koeffizienten einer stochastischen Matrix P mit $p_{i,j}$ bezeichnet sind.

Die Sprungwahrscheinlichkeiten einer Folge von Zufallsgrößen, die die Markoveigenschaft besitzt, sind also durch stochastische Matrizen gegeben, die *a priori* noch von dem Zeitpunkt des Sprunges abhängen dürfen. Wir werden im Folgenden nur solche Folgen betrachten, deren Sprungwahrscheinlichkeiten nicht von diesem Zeitpunkt abhängen:

Definition 2.1.3 (Markovkette). *Sei P eine stochastische Matrix. Eine (endliche oder unendliche) Folge $X_0, X_1, X_2, \dots$ von I-wertigen Zufallsgrößen heißt eine (zeitlich homogene)* Markovkette mit Übergangsmatrix *P, falls für alle $n \in \mathbb{N}$ und alle $i_0, i_1, \dots, i_{n+1} \in I$ mit $\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$ gilt:*

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p_{i_n,i_{n+1}}. \tag{2.1.2}$$

Die Einträge $p_{i,j}$ von P heißen die Übergangswahrscheinlichkeiten*, und die* Startverteilung *ν der Kette ist definiert durch $\nu(i) = \mathbb{P}(X_0 = i)$ für $i \in I$.*

Eine Startverteilung ist also eine Wahrscheinlichkeitsverteilung, oder kurz eine Verteilung, auf I, genau wie jede Zeile einer stochastischen Matrix. Wir schreiben auch oft $\mathbb{P}_\nu$ an Stelle von $\mathbb{P}$, um die Startverteilung zu betonen. Im Fall, dass ν in einem $i \in I$ konzentriert ist (also $\nu(i) = 1$), schreiben wir $\mathbb{P}_i$. Die Elemente von I nennt man auch oft *Zustände*, die Menge I selber den *Zustandsraum*.

Beispiel 2.1.4 (Irrfahrten). Es seien I eine kommutative abzählbare additive Gruppe (etwa $\mathbb{Z}^d$) und $(Y_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch verteilter I-wertiger Zufallsgrößen. Wir betrachten die durch $X_0 = 0$ und $X_n = \sum_{k=1}^n Y_k$ definierte Partialsummenfolge $(X_n)_{n\in\mathbb{N}_0}$. Als Übungsaufgabe beweist man, dass $(X_n)_{n\in\mathbb{N}_0}$ eine Markovkette ist, deren Übergangsmatrix die Koeffizienten $p_{i,j} = \mathbb{P}(Y_1 = j - i)$ besitzt, die also nur von der Differenz der Indizes abhängt. Man nennt $(X_n)_{n\in\mathbb{N}_0}$ eine *Irrfahrt* auf I (wobei allerdings der englische Begriff *random walk* sicherlich sinnvoller ist). ◇

Es folgen Charakterisierungen von Markovketten. Notationell ist es angenehm, für $s < t$ den Pfad $(X_s, X_{s+1}, \dots, X_t)$ mit $X_{[s,t]}$ abzukürzen, ebenso schreiben wir für (nicht zufällige) Vektoren $i_{[s,t]}$ statt $(i_s, i_{s+1}, \dots, i_t) \in I^{t-s+1}$.

Satz 2.1.5. *Es seien $(X_n)_{n\in\mathbb{N}_0}$ eine Folge von I-wertigen Zufallsgrößen, ν eine Verteilung auf I und P eine stochastische Matrix.*

(a) $(X_n)_{n\in\mathbb{N}_0}$ ist genau dann eine Markovkette mit Übergangsmatrix P und Startverteilung ν, wenn für alle $n \in \mathbb{N}_0$ und alle $i_0, i_1, \dots, i_n \in I$ gilt

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \nu(i_0) p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n}. \tag{2.1.3}$$

(b) Falls $(X_n)_{n\in\mathbb{N}_0}$ eine solche Markovkette ist, so gilt für alle $n < m$, $i_n \in I$ und alle $A \subset I^n$ mit $\mathbb{P}(X_{[0,n-1]} \in A, X_n = i_n) > 0$ und für alle $B \subset I^{m-n}$:

$$\mathbb{P}(X_{[n+1,m]} \in B \mid X_{[0,n-1]} \in A, X_n = i_n) = \mathbb{P}(X_{[n+1,m]} \in B \mid X_n = i_n). \tag{2.1.4}$$

Beweis. (a) Der Beweis von (2.1.3) wird leicht mit einer Vollständigen Induktion über n geführt, und eine Folge $(X_n)_{n\in\mathbb{N}_0}$, die (2.1.3) für alle $n \in \mathbb{N}_0$ und alle $i_0, i_1, \dots, i_n \in I$ erfüllt, wird mit Hilfe der Definition der bedingten Wahrscheinlichkeit leicht als eine Markovkette identifiziert.

(b) Mit Hilfe der Definition der bedingten Wahrscheinlichkeit und Teil (a) errechnet man

$$\begin{aligned}\mathbb{P}(X_{[n+1,m]} \in B \mid X_{[0,n-1]} \in A, X_n = i_n) &= \frac{\mathbb{P}(X_{[n+1,m]} \in B, X_{[0,n-1]} \in A, X_n = i_n)}{\mathbb{P}(X_{[0,n-1]} \in A, X_n = i_n)} \\ &= \frac{\sum_{i_{[n+1,m]}\in B} \sum_{i_{[0,n-1]}\in A} \nu(i_0) p_{i_0,i_1} \cdots p_{i_{m-1},i_m}}{\sum_{i_{[0,n-1]}\in A} \nu(i_0) p_{i_0,i_1} \cdots p_{i_{n-1},i_n}} \\ &= \sum_{i_{[n+1,m]}\in B} p_{i_n,i_{n+1}} p_{i_{n+1},i_{n+2}} \cdots p_{i_{m-1},i_m}.\end{aligned}$$

Der Ausdruck auf der rechten Seite hängt nicht von A ab. Wir können insbesondere $A = I^n$ setzen und erhalten

$$\mathbb{P}(X_{[n+1,m]} \in B \mid X_n = i_n) = \sum_{i_{[n+1,m]}\in B} p_{i_n,i_{n+1}} p_{i_{n+1},i_{n+2}} \cdots p_{i_{m-1},i_m}.$$

Wenn wir dies in der obigen Rechnung wieder einsetzen, ist der Beweis von (2.1.4) beendet. □

Die Aussage in (b) kann man auch einprägsam wie folgt formulieren:

Korollar 2.1.6 (Unabhängigkeit von Zukunft und Vergangenheit bei gegebener Gegenwart). *Falls $(X_n)_{n\in\mathbb{N}_0}$ eine Markovkette ist, so gilt für alle $n < m$, $i_n \in I$ mit $\mathbb{P}(X_n = i_n) > 0$ und alle $A \subset I^n$ und $B \subset I^{m-n}$:*

$$\mathbb{P}(X_{[0,n-1]} \in A, X_{[n+1,m]} \in B \mid X_n = i_n) = \mathbb{P}(X_{[0,n-1]} \in A \mid X_n = i_n)\mathbb{P}(X_{[n+1,m]} \in B \mid X_n = i_n). \tag{2.1.5}$$

Beweis. Im Fall $\mathbb{P}(X_{[0,n-1]} \in A, X_n = i_n) > 0$ ergibt sich (2.1.5) direkt aus (2.1.4) nach Multiplikation mit $\mathbb{P}(X_{[0,n-1]} \in A \mid X_n = i_n)$. Ansonsten steht Null auf beiden Seiten von (2.1.5), und die Aussage gilt trivialerweise. □

Man überlege sich an einem Beispiel, dass die Aussage von Korollar 2.1.6 im Allgemeinen falsch wird, wenn man das Ereignis $\{X_n = i_n\}$ ersetzt durch $\{X_n \in C\}$ für beliebige Teilmengen C von I.

Wir schneiden kurz die Frage der Existenz und Konstruktion von Markovketten an.

Bemerkung 2.1.7 (Konstruktion von Markovketten). Eine endlich lange Markovkette $(X_0, X_1, \dots, X_n)$ im Sinne von Definition 2.1.3 kann leicht auf einem diskreten Wahrscheinlichkeitsraum konstruiert werden. Wir setzen $\Omega_n = I^{n+1}$ und definieren die Einzelwahrscheinlichkeiten als $p((i_0, \dots, i_n)) = \nu(i_0) p_{i_0,i_1} \dots p_{i_{n-1},i_n}$, wobei ν eine Verteilung auf I ist und P eine stochastische Matrix. Dann bilden die Projektionsabbildungen $X_i \colon I^{n+1} \to I$ eine Markovkette mit Startverteilung ν und Übergangsmatrix P.

Man beachte, dass dieser Wahrscheinlichkeitsraum von der Länge der Kette abhängt und dass man genauer $\mathbb{P}_n$ statt $\mathbb{P}$ für das induzierte Wahrscheinlichkeitsmaß schreiben müsste. Allerdings überzeugt man sich leicht davon, dass die ersten $n+1$ Zufallsgrößen einer Markovkette der Länge $m > n$ selber eine Markovkette der Länge n bilden. Mit anderen Worten, man kann leicht zeigen, dass die so konstruierte Folge von Wahrscheinlichkeitsmaßen konsistent ist, siehe Definition 1.3.2. Nach dem Existenzsatz 1.3.3 von Kolmogorov existiert eine unendliche lange Markovkette, deren n-schrittiger Beginn gerade die oben konstruierten endlich langen Ketten sind.

Bei der Untersuchung einer Markovkette in endlicher Zeit reicht es also für viele Zwecke aus, eine genügend späte Zeit m zu fixieren und den für dieses m konstruierten Wahrscheinlichkeitsraum zu Grunde zu legen. Wir werden daher im Folgenden aus Bequemlichkeit immer von einer unendlich langen Markovkette ausgehen und den benutzten Wahrscheinlichkeitsraum nicht mit einer endlichen Länge indizieren. $\diamond$

Der Zusammenhang zwischen dem Markovschen Mechanismus und der Matrixmultiplikation ist sehr eng, wie wir uns kurz klar machen wollen.

Bemerkung 2.1.8 (Potenzen stochastischer Matrizen). Stochastische Matrizen P und Q kann man ohne Probleme im Sinne der Matrixmultiplikation mit einander multiplizieren, und das Produkt ist ebenfalls eine stochastische Matrix. Die Koeffizienten der n-ten Potenz P^n von P bezeichnen wir mit $P^n = (p_{i,j}^{(n)})_{i,j\in I}$. Es ist $P^0 = (\delta_{i,j})_{i,j\in I}$ die Einheitsmatrix, wobei $\delta_{i,j}$ das Kroneckersymbol bezeichnet. Wenn man die Gleichung $P^n P^m = P^{n+m}$ ausschreibt, erhält man die sogenannten *Chapman-Kolmogorov-Gleichungen*:

$$p_{i,j}^{(n+m)} = \sum_{k\in I} p_{i,k}^{(n)} p_{k,j}^{(m)}, \qquad i,j \in I. \tag{2.1.6}$$

Insbesondere haben wir

$$p_{i,j}^{(n+m)} \geq p_{i,k}^{(n)} p_{k,j}^{(m)}, \qquad i,j,k \in I, n,m \in \mathbb{N}_0. \tag{2.1.7}$$

Auf Grund des folgenden Lemmas 2.1.9 nennt man die Koeffizienten $p_{i,j}^{(n)}$ von P^n auch die *n-stufigen Übergangswahrscheinlichkeiten*. $\diamond$

Wir stellen uns eine Verteilung ν als *Zeilenvektoren* vor (wie auch z. B. die Zeilen einer stochastischen Matrix), so dass das Matrixprodukt νP wohldefiniert ist, also $(\nu P)_j = \sum_{i\in I} \nu(i) p_{i,j}$.

Lemma 2.1.9. *Es sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markovkette mit Startverteilung ν und Übergangsmatrix P. Dann gilt $\mathbb{P}_\nu(X_n = j) = (\nu P^n)_j$ für alle $n \in \mathbb{N}$ und alle $j \in I$. Insbesondere gilt $\mathbb{P}_i(X_n = j) = p_{i,j}^{(n)}$ für alle $i,j \in I$.*

Beweis. Wir summieren die Gleichung (2.1.3) (mit $j = i_n$) über alle $i_0, \dots, i_n \in I$ und beachten die Regeln der Matrixmultiplikation. $\square$

Die Verteilung einer Markovkette zum Zeitpunkt n ist also nichts Anderes als die n-te Potenz der Übergangsmatrix, multipliziert von links mit der Startverteilung. Wir halten noch fest, wie sich dies auf zeitliche Verschiebung auswirkt:

Korollar 2.1.10. *Es sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markovkette mit Übergangsmatrix P. Dann gilt für alle $n,m \in \mathbb{N}_0$ und alle $i,j \in I$ mit $\mathbb{P}(X_m = i) > 0$:*

$$\mathbb{P}(X_{m+n} = j \mid X_m = i) = p_{i,j}^{(n)}.$$

2.2 Beispiele

Ein sehr allgemeines Prinzip, nach dem man Markovketten konstruieren kann (siehe Übungsaufgabe), ist das folgende. Mit einer Folge von unabhängigen identisch verteilten Zufallsgrößen $Y_1, Y_2, Y_3, \dots$ mit Werten in einer beliebigen abzählbaren Menge J und mit einer Funktion $f\colon I \times J \to I$ setzt man rekursiv $X_{n+1} = f(X_n, Y_{n+1})$ für $n \in \mathbb{N}_0$. Die Kette wird also rekursiv fortgesetzt, indem man den aktuellen Wert der Kette mit einem festgelegten Mechanismus einem unabhängigen Zufall unterwirft.

Beispiel 2.2.1 (unabhängige identisch verteilte Folgen). Wenn $(X_n)_{n\in\mathbb{N}_0}$ eine Folge unabhängiger und identisch verteilter I-wertiger Zufallsgrößen ist, so ist sie auch eine Markovkette. Die Übergangsmatrix ist gegeben durch $p_{i,j} = q_j$ für alle $i, j \in I$, wobei q die Verteilung von X_0 ist. Natürlich ist q auch die Startverteilung. Andersherum ist eine Markovkette, deren Übergangsmatrix identische Zeilen besitzt (d. h. deren $p_{i,j}$ nicht von i abhängen), eine Folge unabhängiger und identisch verteilter Zufallsgrößen. $\diamond$

Beispiel 2.2.2 (Irrfahrten auf Gruppen). Es sei $I = G$ eine abzählbare Gruppe, die wir multiplikativ schreiben wollen und nicht unbedingt als kommutativ voraus setzen wollen. Ferner sei μ eine beliebige Verteilung auf G. Wir definieren $p_{g,h} = \mu(g^{-1}h)$ für alle $g, h \in G$. Wegen der Gruppeneigenschaft ist für jedes g die Abbildung $h \mapsto g^{-1}h$ bijektiv auf G, und es gilt

$$\sum_{h\in G} p_{g,h} = \sum_{h\in G} \mu(g^{-1}h) = \sum_{h'\in G} \mu(h') = 1,$$

also ist $P = (p_{g,h})_{g,h\in G}$ eine stochastische Matrix. Die zugehörige Markovkette heißt die *μ-Irrfahrt* auf G. Dieses Beispiel ist die multiplikative Variante der Irrfahrt von Beispiel 2.1.4 (bis auf die dort voraus gesetzte Kommutativität). Mit Hilfe einer Folge von unabhängigen, nach μ verteilten Zufallsgrößen $Y_1, Y_2, Y_3, \dots$ erhält man eine μ-Irrfahrt, indem man rekursiv setzt: $X_0 = 1$ und $X_{n+1} = X_n Y_n$, denn dann gilt für alle $n \in \mathbb{N}_0$ und alle $g, h \in G$ mit $\mathbb{P}(X_n = g) > 0$:

$$\mathbb{P}(X_{n+1} = h \mid X_n = g) = \mathbb{P}(gY_n = h) = \mu(g^{-1}h) = p_{g,h}.$$

Die Wahl von $I = G$ als die Menge der Permutationen einer endlichen Menge führt zum Beispiel auf ein Modell für die Mischung eines Kartenstapels; man beachte, dass diese Gruppe nicht kommutativ ist. $\diamond$

Beispiel 2.2.3 (eindimensionale Irrfahrt). Der folgende Spezialfall von Beispiel 2.1.4 wird die *eindimensionale Irrfahrt* genannt. Setze $I = \mathbb{Z}$, und Y_n nehme die Werte 1 und -1 mit Wahrscheinlichkeiten p und $q = 1 - p$ an, wobei $p \in [0, 1]$ ein Parameter sei. Dann beschreibt die Markovkette $(X_n)_{n\in\mathbb{N}_0}$ den Weg eines Teilchens durch die diskrete Achse mit unabhängigen Sprüngen, wobei es zu jedem Zeitpunkt mit Wahrscheinlichkeit p um eine Einheit nach rechts springt und sonst nach links. Die Übergangsmatrix besitzt die Einträge p auf der rechten Nebendiagonalen und $1 - p$ auf der linken, ansonsten besteht sie aus Nullen. Im Fall $p = \frac{1}{2}$ wird die Irrfahrt *symmetrisch* genannt. $\diamond$

Beispiel 2.2.4 (Irrfahrt auf $\mathbb{Z}^d$). Die d-dimensionale Variante von Beispiel 2.2.3 (die ebenfalls ein Spezialfall von Beispiel 2.1.4 ist) ist auf $I = \mathbb{Z}^d$ gegeben, indem die Übergangsmatrix durch $p_{i,j} = \frac{1}{2d}$ für $|i - j| = 1$ fest gelegt wird. (Hier ist $|\cdot|$ die ℓ^1-Norm auf $\mathbb{Z}^d$.) Die zugehörige Markovkette beschreibt einen Nächstnachbarschaftspfad durch $\mathbb{Z}^d$, wobei jeder Nachbar mit gleicher Wahrscheinlichkeit ausgewählt wird, unabhängig von allen anderen Sprungentscheidungen. $\diamond$

Beispiel 2.2.5 (Irrfahrten auf $\{0,\dots,N\}$). Ähnlich wie in Beispiel 2.2.3 soll ein Sprung innerhalb von $I = \{0,\dots,N\}$ mit Wahrscheinlichkeit p zum rechten Nachbarn und mit Wahrscheinlichkeit $1-p$ zum linken ausgeführt werden. Für die Sprungentscheidungen an den Rändern 0 und N müssen wir allerdings gesonderte Vereinbarungen treffen, und es gibt dafür mehrere Möglichkeiten. Ein Randpunkt, sagen wir 0, heißt *absorbierend*, falls $p_{0,0} = 1$ gilt, falls also das springende Teilchen nie mehr von der 0 sich entfernen kann. Im Fall $p_{0,1} = 1$, wenn also das Teilchen sofort wieder unweigerlich zurück springen muss, heißt der Randpunkt 0 *reflektierend*. $\diamond$

Beispiel 2.2.6 (Polyas Urnenschema). In einer Urne liegen gewisse (endliche) Anzahlen roter und schwarzer Kugeln. Zu jedem Zeitpunkt wird eine Kugel zufällig gezogen und zusammen mit einer neuen Kugel der selben Farbe in die Urne zurück gelegt. Dann bildet das Paar der Anzahlen der roten und der schwarzen Kugeln zu den Zeitpunkten $0,1,2,\dots$ eine Markovkette auf $\mathbb{N}_0^2$. Die Übergangswahrscheinlichkeiten sind gegeben durch $p_{(r,s),(r+1,s)} = \frac{r}{r+s}$ und $p_{(r,s),(r,s+1)} = \frac{s}{r+s}$; alle anderen sind Null. $\diamond$

Beispiel 2.2.7 (Ehrenfests Urnenmodell). Insgesamt N Kugeln liegen in zwei Urnen. Zu jedem Zeitpunkt $1,2,\dots$ wählen wir eine der Kugeln mit gleicher Wahrscheinlichkeit und lassen sie die Urne wechseln. Dann ist die Anzahl der Kugeln in der linken Urne zum Zeitpunkt n eine Markovkette auf $I = \{0,\dots,N\}$ im Zeitparameter n. Die Übergangsmatrix P ist gegeben durch $p_{k,k-1} = \frac{k}{N}$ und $p_{k,k+1} = 1 - \frac{k}{N}$, und alle anderen Übergangswahrscheinlichkeiten sind Null. Siehe Beispiel 2.7.4 für interessante Eigenschaften dieses Modells. $\diamond$

Beispiel 2.2.8 (Bernoulli-Laplace-Diffusionsmodell). In zwei Behältern A und B befinden sich insgesamt w weiße und s schwarze Kugeln, wobei s Kugeln in A liegen, und es sei $w \le s$. Zu den diskreten Zeitpunkten $n = 1,2,3,\dots$ wird jeweils in A und in B eine Kugel zufällig ausgewählt und in den jeweils anderen Behälter gelegt. Dann ist die Anzahl der weißen Kugeln in A eine Markovkette auf $\{0,1,\dots,w\}$ im Zeitparameter n. Siehe Beispiel 2.5.9 für interessante Eigenschaften dieses Modells. $\diamond$

Beispiel 2.2.9 (Irrfahrten-Maxima). Falls $(S_n)_{n\in\mathbb{N}_0}$ eine Irrfahrt auf $\mathbb{Z}$ wie in Beispiel 2.1.4 ist, dann bildet die Folge der Maxima $M_n = \max\{S_0,S_1,\dots,S_n\}$ im Allgemeinen keine Markovkette, aber die Folge $(M_n,S_n)_{n\in\mathbb{N}_0}$ der Paare (Übungsaufgabe). $\diamond$

Beispiel 2.2.10 (Verzweigungsprozesse). Wir nehmen an, dass zur Zeit $n = 0$ ein Individuum existiert, das mit Wahrscheinlichkeit $p_k \in [0,1]$ genau $k \in \mathbb{N}_0$ Nachkommen produziert. Jeder der Nachkommen produziert wieder unabhängig voneinander Nachkommen mit derselben Verteilung. 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 können diesen Prozess auch wie folgt konstruieren. Sei $\xi_i^{(n)}$ die Anzahl der Nachkommen des i-ten Individuums der n-ten Generation, dann gilt

$$S_{n+1} = \sum_{i=1}^{S_n} \xi_i^{(n)}, \qquad n \in \mathbb{N}.$$

Wir können annehmen, dass die Familie $(\xi_i^{(n)})_{n\in\mathbb{N}_0, i\in\mathbb{N}}$ u. i. v. ist. Damit ist der Galton-Watson-Prozess von der Form, die wir am Beginn dieses Abschnitts erwähnten. $\diamond$

2.3 Klasseneigenschaften, Rekurrenz, Transienz

In diesem Abschnitt sei P eine stochastische Matrix auf I. Wir wollen die Frage untersuchen, ob die zugehörige Markovkette gegebene Punkte in I mit Sicherheit besucht oder nicht. Es wird sich in diesem Abschnitt heraus stellen, dass der Zustandsraum zerlegt werden kann in rekurrente und transiente Klassen. Eine in einer rekurrenten Klasse gestartete Markovkette verlässt die Klasse nie und besucht jeden Punkt dieser Klasse mit Sicherheit.

Wir müssen zunächst die Frage untersuchen, ob ein gegebener Punkt überhaupt mit positiver Wahrscheinlichkeit jemals erreicht werden kann. Diese Unterscheidung induziert auf natürliche Weise eine Einteilung von I in Klassen.

Definition 2.3.1 (erreichbar). *Ein Punkt $j \in I$ heißt von einem Punkt $i \in I$ aus* erreichbar*, falls ein $n \in \mathbb{N}_0$ existiert mit $p_{i,j}^{(n)} > 0$. Wir schreiben dann $i \rightsquigarrow j$. Falls $i \rightsquigarrow j$ und $j \rightsquigarrow i$, so schreiben wir $i \leftrightsquigarrow j$.*

Bemerkung 2.3.2 ($\leftrightsquigarrow$ als Äquivalenzrelation). Die Relation $\rightsquigarrow$ auf I ist reflexiv und transitiv, denn wegen $p_{i,i}^{(0)} = 1$ gilt $i \rightsquigarrow i$ für jedes $i \in I$, und falls $i \rightsquigarrow j$ und $j \rightsquigarrow k$ gelten, so folgt aus (2.1.7) leicht, dass auch $i \rightsquigarrow k$ gilt. Also ist $\leftrightsquigarrow$ eine Äquivalenzrelation auf I und teilt I folglich in Klassen ein. Für zwei Klassen $A, B \subset I$ schreiben wir $A \rightsquigarrow B$, falls es $i \in A$ und $j \in B$ gibt mit $i \rightsquigarrow j$, und wir sagen dann, dass B von A aus erreichbar ist. (Offensichtlich ist diese Sprechweise wohldefiniert, d. h. die Definition unabhängig von den gewählten Repräsentanten.) $\diamond$

Definition 2.3.3 (abgeschlossen, irreduzibel). *(a) Eine Teilmenge J von I heißt* abgeschlossen*, wenn $J \not\rightsquigarrow I \setminus J$ gilt, d. h. wenn keine zwei Elemente $j \in J$ und $i \in I \setminus J$ existieren mit $j \rightsquigarrow i$.*

(b) P heißt irreduzibel*, falls I aus einer einzigen Klasse besteht, d. h. wenn je zwei Elemente aus I äquivalent sind.*

Man sieht leicht, dass die Einschränkung $P_J = (p_{i,j})_{i,j \in J}$ von P auf eine abgeschlossene Klasse J von I ihrerseits eine stochastische Matrix auf J ist. Falls P irreduzibel ist, so existieren keine abgeschlossenen echten Teilmengen von I.

Beispiel 2.3.4. (a) Die symmetrische Irrfahrt auf $\mathbb{Z}^d$ ist irreduzibel.

(b) In Polyas Urnenschema (siehe Beispiel 2.2.6) sind keine zwei Elemente aus $I = \mathbb{N}_0^2$ äquivalent. Für jede $r_0, s_0 \in \mathbb{N}$ ist die Menge $\{(r, s) \in I : r \geq r_0, s \geq s_0\}$ abgeschlossen.

(c) Die Irrfahrt auf $\{0, \dots, n\}$ (siehe Beispiel 2.2.5) mit absorbierenden Rändern besitzt die Äquivalenzklassen $\{0\}$, $\{1, \dots, n-1\}$ und $\{n\}$. Die Mengen $\{0\}$ und $\{n\}$ sind abgeschlossen, und es gelten $\{1, \dots, n-1\} \rightsquigarrow \{0\}$ und $\{1, \dots, n-1\} \rightsquigarrow \{n\}$.

$\diamond$

Es sei $(X_n)_{n \in \mathbb{N}_0}$ die zur stochastischen Matrix gehörige Markovkette. Wir führen nun die *Ersteintrittszeit* von $i \in I$ ein:

$$T_i = \inf\{n \in \mathbb{N} : X_n = i\}. \tag{2.3.1}$$

Falls die Kette den Punkt i gar nicht besucht, dann setzen wir $T_i = \infty$. Man beachte, dass die Definition von T_i die gesamte unendlich lange Kette $(X_n)_{n \in \mathbb{N}_0}$ erfordert, deren Konstruktion wir noch nicht durchgeführt haben (siehe Bemerkung 2.1.7). Allerdings hängt das Ereignis $\{T_i = n\}$ nur von der Zeit bis n ab, kann also mit unseren Mitteln korrekt behandelt werden. Wir definieren

$$f_{i,j}^{(n)} = \mathbb{P}_i(T_j = n) = \mathbb{P}_i(X_1 \neq j, X_2 \neq j, \dots, X_{n-1} \neq j, X_n = j), \qquad i, j \in I, n \in \mathbb{N}. \tag{2.3.2}$$

In Worten: $f_{i,j}^{(n)}$ ist die Wahrscheinlichkeit, dass eine in i gestartete Kette zum Zeitpunkt n den Punkt j zum ersten Mal trifft. Die Summe

$$f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{(n)} \tag{2.3.3}$$

liegt in $[0, 1]$, da die Ereignisse $\{T_j = 1\}, \{T_j = 2\}, \dots$ disjunkt sind. Man kann (und sollte) die unendliche Summe $f_{i,j}$ als $\mathbb{P}_i(T_j < \infty)$ interpretieren, d. h. als die Wahrscheinlichkeit, dass eine in i gestartete Kette jemals den Punkt j besucht.

Satz 2.3.5 (Erneuerungsgleichung). *Es gilt*

$$p_{i,i}^{(n)} = \sum_{k=1}^{n} f_{i,i}^{(k)} p_{i,i}^{(n-k)}, \qquad n \in \mathbb{N}, i \in I. \tag{2.3.4}$$

Beweis. Gemäß Lemma 2.1.9 gilt $p_{i,i}^{(n)} = \mathbb{P}_i(X_n = i)$. Wir spalten das Ereignis $\{X_n = i\}$ nach dem ersten Zeitpunkt, an dem die Kette i erreicht, auf und erhalten

$$p_{i,i}^{(n)} = \sum_{k=1}^{n} \mathbb{P}_i(T_i = k, X_n = i) = \sum_{k=1}^{n} \mathbb{P}_i(X_n = i \mid X_1 \neq i, \dots, X_{k-1} \neq i, X_k = i) f_{i,i}^{(k)}.$$

Nun wenden wir (2.1.4) und Korollar 2.1.10 an und erhalten

$$p_{i,i}^{(n)} = \sum_{k=1}^{n} \mathbb{P}_i(X_n = i \mid X_k = i) f_{i,i}^{(k)} = \sum_{k=1}^{n} f_{i,i}^{(k)} p_{i,i}^{(n-k)}.$$

□

Nun kommen wir zu weiteren zentralen Begriffen der Theorie der Markovketten.

Definition 2.3.6 (Rekurrenz, Transienz). *Ein Zustand $i \in I$ heißt* rekurrent, *falls $f_{i,i} = 1$, und* transient *sonst.*

Interpretationsgemäß heißt also i rekurrent, falls die in i gestartete Kette mit Sicherheit wieder irgendwann einmal zu i zurück kehrt. Wir können diese Eigenschaft in Termen der Potenzen von P charakterisieren:

Satz 2.3.7. *Ein Zustand $i \in I$ ist genau dann rekurrent, wenn $\sum_{n \in \mathbb{N}_0} p_{i,i}^{(n)} = \infty$ gilt.*

Beweis. Für jedes $s \in (0, 1)$ erhalten wir aus der Erneuerungsgleichung (siehe Satz 2.3.5)

$$\begin{aligned}\sum_{n \in \mathbb{N}_0} p_{i,i}^{(n)} s^n &= 1 + \sum_{n \in \mathbb{N}} s^n p_{i,i}^{(n)} = 1 + \sum_{n \in \mathbb{N}} s^n \sum_{k=1}^{n} f_{i,i}^{(k)} p_{i,i}^{(n-k)} = 1 + \sum_{k \in \mathbb{N}} f_{i,i}^{(k)} s^k \sum_{n=k}^{\infty} p_{i,i}^{(n-k)} s^{n-k} \\ &= 1 + \sum_{k \in \mathbb{N}} f_{i,i}^{(k)} s^k \sum_{n \in \mathbb{N}_0} p_{i,i}^{(n)} s^n.\end{aligned}$$

Also haben wir für die beiden Funktionen

$$\pi(s) = \sum_{n\in\mathbb{N}_0} p_{i,i}^{(n)} s^n \qquad \text{und} \qquad \phi(s) = \sum_{k\in\mathbb{N}} f_{i,i}^{(k)} s^k$$

die Beziehung $\pi(s) = 1+\pi(s)\phi(s)$ hergeleitet. Im Fall $1 = f_{i,i} = \phi(1)$ machen wir den Grenzübergang $s \uparrow 1$ und erhalten $\pi(1) = 1 + \pi(1)$, was natürlich impliziert, dass $\sum_{n\in\mathbb{N}_0} p_{i,i}^{(n)} = \pi(1) = \infty$. Falls $f_{i,i} < 1$, so formen wir zunächst um zu $\pi(s) = \frac{1}{1-\phi(s)}$ und lassen dann $s \uparrow 1$, was auf $\sum_{n\in\mathbb{N}_0} p_{i,i}^{(n)} = \frac{1}{1-f_{i,i}} < \infty$ führt und den Beweis beendet. □

Die in Satz 2.3.7 auftretende Reihe kann interpretiert werden:

$$\sum_{n\in\mathbb{N}_0} p_{i,i}^{(n)} = \sum_{n\in\mathbb{N}_0} \mathbb{E}_i(\mathbb{1}_{\{X_n=i\}}) = \mathbb{E}_i\Big(\sum_{n\in\mathbb{N}_0} \mathbb{1}_{\{X_n=i\}}\Big),$$

also ist sie gleich der erwarteten Anzahl der Besuche in i für die unendlich lange in i gestartete Markovkette. Der Zustand i ist also nach Satz 2.3.7 genau dann transient, wenn diese Kette ihren Startpunkt erwartungsgemäß nur endlich oft besucht.

Rekurrenz und Transienz sind Klasseneigenschaften:

Satz 2.3.8. *Es seien $i, j \in I$ mit $i \leftrightsquigarrow j$. Dann ist i genau dann rekurrent, wenn j es ist.*

Beweis. Übungsaufgabe. □

Also können und werden wir in Zukunft auch von rekurrenten bzw. transienten Klassen sprechen und bei Irreduzibilität von rekurrenten bzw. transienten Markovketten. Das Kriterium aus Satz 2.3.7 kann man benutzen, um beispielsweise die Rekurrenz der d-dimensionalen Irrfahrt aus Beispiel 2.2.4 zu entscheiden:

Satz 2.3.9. *Die d-dimensionale symmetrische Irrfahrt auf $\mathbb{Z}^d$ ist rekurrent für $d \in \{1, 2\}$ und transient für $d \geq 3$.*

Beweisskizze. Wegen der Verschiebungsinvarianz genügt es, die Divergenz bzw. Konvergenz der Reihe $\sum_{n\in\mathbb{N}} p_{0,0}^{(n)}$ zu untersuchen. Wir sind in der glücklichen Lage, den Wert von $p_{0,0}^{(n)}$ ausrechnen zu können. Natürlich ist immer $p_{0,0}^{(2n-1)} = 0$ für alle $n \in \mathbb{N}$.

In $d = 1$ ist leicht zu sehen, dass $p_{0,0}^{(2n)} = \binom{2n}{n} 2^{-2n}$ für alle $n \in \mathbb{N}$, und Stirlings Formel zeigt, dass $p_{0,0}^{(2n)} \sim (\pi n)^{-1/2}$ für $n \to \infty$. Also ist $\sum_{n\in\mathbb{N}} p_{0,0}^{(n)} = \infty$ für $d = 1$.

In $d = 2$ schreiben wir $X_n = (X_n^{(1)}, X_n^{(2)})$ für die beiden Komponenten und nutzen die Beobachtung aus, dass die beiden Folgen $(X_n^{(1)} - X_n^{(2)})_{n\in\mathbb{N}_0}$ und $(X_n^{(1)} + X_n^{(2)})_{n\in\mathbb{N}_0}$ zwei unabhängige eindimensionale symmetrische Irrfahrten auf $\mathbb{Z}$ sind. Folglich ist $p_{0,0}^{(2n)} = \big(\binom{2n}{n} 2^{-2n}\big)^2$, und wie oben sieht man, dass dies sich wie $\frac{1}{\pi n}$ verhält, also ebenfalls nicht summierbar ist.

Den Fall $d \geq 3$ kann man auf den Grenzfall $d = 3$ zurück führen. Im Fall $d = 3$ stellt man den Wert von $p_{0,0}^{(2n)}$ mit Hilfe einer Doppelsumme dar (nach einer Aufspaltung, wieviele Schritte jeweils in die drei Dimensionsrichtungen geschehen) und schätzt diese mit einiger Arbeit geeignet ab, bis man sieht, dass $p_{0,0}^{(2n)}$ von der Ordnung $n^{-3/2}$ ist. □

Im Folgenden zeigen wir insbesondere, dass die Markovkette, wenn sie in einer rekurrenten Klasse gestartet wird, *jeden* Zustand in dieser Klasse mit Sicherheit besucht, siehe Lemma 2.3.13. Zunächst ziehen wir eine Folgerung aus der Erneuerungsgleichung.

Lemma 2.3.10. *Für alle $i, j \in I$ gilt*

$$\sum_{n\in\mathbb{N}} p_{i,j}^{(n)} = f_{i,j} \sum_{n\in\mathbb{N}} p_{j,j}^{(n)}.$$

Beweis. Genau wie im Beweis der Erneuerungsgleichung (Satz 2.3.5) leitet man die Beziehung

$$p_{i,j}^{(n)} = \sum_{k=1}^{n} f_{i,j}^{(k)} p_{j,j}^{(n-k)}$$

für alle $n \in \mathbb{N}$ her. Nun summiert man diese Beziehung über $n \in \mathbb{N}$, vertauscht auf der rechten Seite die beiden Summationen und verschiebt die eine davon (ähnlich wie im Beweis von Satz 2.3.7). □

Korollar 2.3.11. *Für alle $i, j \in I$ gilt:*

$$j \text{ transient} \quad \Longrightarrow \quad \sum_{n\in\mathbb{N}_0} p_{i,j}^{(n)} < \infty.$$

Aus dem folgenden Ergebnis folgt insbesondere, dass Klassen, die nicht abgeschlossen sind, zwangsläufig transient sind. (Abgeschlossene Klassen können hingegen sowohl rekurrent als auch transient sein.)

Lemma 2.3.12. *Es seien $i, j \in I$ mit $i \rightsquigarrow j$. Falls i rekurrent ist, so gilt auch $j \rightsquigarrow i$, und j ist dann ebenfalls rekurrent.*

Beweis. Wir dürfen $i \neq j$ annehmen. Wir führen einen Widerspruchsbeweis und nehmen an, dass i nicht von j aus erreichbar ist, d. h. dass $p_{j,i}^{(n)} = 0$ für alle $n \in \mathbb{N}_0$ ist.

Sei $N \in \mathbb{N}$ die kleinste Zahl n mit $p_{i,j}^{(n)} > 0$. Für alle $n \in \mathbb{N}$ gilt dann $\mathbb{P}_i(X_N = j, X_n = i) = 0$, denn für $n > N$ gilt ja $\mathbb{P}_i(X_N = j, X_n = i) = p_{i,j}^{(N)} p_{j,i}^{(n)} = 0$, und für $n < N$ gilt $\mathbb{P}_i(X_N = j, X_n = i) = p_{i,i}^{(n)} p_{i,j}^{(N-n)} = 0$, denn N ist ja das *kleinste* n mit $p_{i,j}^{(n)} > 0$. Daher haben wir

$$\mathbb{P}_i(T_i \leq M, X_N = j) = \sum_{n=1}^{M} \mathbb{P}_i(T_i = n, X_N = j) \leq \sum_{n=1}^{M} \mathbb{P}_i(X_n = i, X_N = j) = 0.$$

Also folgt

$$\sum_{n=1}^{M} f_{i,i}^{(n)} = \mathbb{P}_i(T_i \leq M) = \mathbb{P}_i(T_i \leq M, X_N \neq j) \leq \mathbb{P}_i(X_N \neq j) = 1 - \mathbb{P}_i(X_N = j) = 1 - p_{i,j}^{(N)}.$$

Nun lassen wir $M \uparrow \infty$ und beachten, dass die rechte Seite der Abschätzung nicht von M abhängt. Mit der Rekurrenz von i folgt der Widerspruch $1 = f_{i,i} = \sum_{n\in\mathbb{N}} f_{i,i}^{(n)} \leq 1 - p_{i,j}^{(N)} < 1$. □

Insbesondere sind rekurrente Klassen abgeschlossen. (Transiente Klassen müssen nicht abgeschlossen sein.) Insbesondere ist die Einschränkung von P auf eine rekurrente Klasse (siehe die Bemerkung nach Definition 2.3.3) wieder eine stochastische Matrix, die dann natürlich auch irreduzibel ist. Daher lassen sich also die einzelnen rekurrenten Klassen getrennt diskutieren.

Lemma 2.3.13. *Wenn i und j in der selben rekurrenten Klasse liegen, so gilt $f_{i,j} = f_{j,i} = 1$.*

Beweis. Wir dürfen $i \neq j$ annehmen. Sei N das kleinste $n \in \mathbb{N}$ mit $p_{j,i}^{(n)} > 0$. Für $M > N$ gilt

$$\begin{aligned}\mathbb{P}_j(T_j \leq M, X_N = i) &= \sum_{n=1}^{M} \mathbb{P}_j(T_j = n, X_N = i)\\ &= \sum_{n=1}^{N-1} f_{j,j}^{(n)} p_{j,i}^{(N-n)} + \sum_{n=N+1}^{M} \mathbb{P}_j(T_j \geq N, X_N = i) f_{i,j}^{(n-N)}.\end{aligned}$$

Nach Definition ist aber $p_{j,i}^{(N-n)} = 0$ für jedes $n \in \{1, \dots, N-1\}$, also ist die erste Summe auf der rechten Seite gleich Null. In der zweiten schätzen wir ab: $\mathbb{P}_j(T_j \geq N, X_N = i) \leq p_{j,i}^{(N)}$, also erhalten wir aus der obigen Rechnung

$$\mathbb{P}_j(T_j \leq M, X_N = i) \leq \sum_{n=N+1}^{M} p_{j,i}^{(N)} f_{i,j}^{(n-N)} \leq p_{j,i}^{(N)} f_{i,j}.$$

Nun lassen wir $M \uparrow \infty$ und beachten, dass $\lim_{M\to\infty} \mathbb{P}_j(T_j \leq M) = \lim_{M\to\infty} \sum_{k=1}^{M} f_{j,j}^{(k)} = f_{j,j} = 1$, also folgt

$$p_{j,i}^{(N)} = \mathbb{P}_j(X_N = i) = \lim_{M\to\infty} \mathbb{P}_j(T_j \leq M, X_N = i) \leq p_{j,i}^{(N)} f_{i,j}.$$

Wegen $f_{i,j} \leq 1$ und $p_{j,i}^{(N)} > 0$ ergibt sich $f_{i,j} = 1$, und dies beendet den Beweis. □

Ein sehr handliches und simples hinreichendes Kriterium für Rekurrenz ist das Folgende.

Satz 2.3.14. *Wenn I endlich ist, so ist jede irreduzible Kette rekurrent.*

Beweis. Wegen $\sum_{j\in I} p_{i,j}^{(n)} = 1$ für jedes $n \in \mathbb{N}$ und jedes $i \in I$ folgt, dass zu jedem i ein j existiert mit $\sum_{n\in\mathbb{N}} p_{i,j}^{(n)} = \infty$, denn es gilt ja $\sum_{j\in I}(\sum_{n\in\mathbb{N}} p_{i,j}^{(n)}) = \infty$, und irgendeiner der endlich vielen Summanden $\sum_{n\in\mathbb{N}} p_{i,j}^{(n)}$ muss gleich ∞ sein. Aus Lemma 2.3.10 folgt $\sum_{n\in\mathbb{N}} p_{j,j}^{(n)} = \infty$, und es folgt die Rekurrenz von j. □

2.4 Stoppzeiten und die starke Markov-Eigenschaft

Stoppzeiten sind zufällige Zeiten, die nicht in die Zukunft blicken können. Im Zusammenhang mit Markovketten sind Stoppzeiten die Eintrittszeitpunkte gewisser Ereignisse, deren Eintreten durch die bisherige 'Geschichte' der Kette beurteilt werden kann.

Im Folgenden sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markovkette auf I mit Übergangsmatrix P. In diesem Abschnitt empfiehlt es sich anzunehmen, dass alle Zufallsgrößen X_n mit $n \in \mathbb{N}$ auf einem gemeinsamen Wahrscheinlichkeitsraum Ω definiert sind, und das werden wir tun.

Definition 2.4.1 (Filtrierung, Stoppzeit). *(a) Für jedes $n \in \mathbb{N}_0$ sei $\mathcal{F}_n$ die Menge der Ereignisse von der Form $\{\omega \in \Omega \colon X_{[0,n]}(\omega) \in B\}$ mit $B \subset I^{n+1}$, d. h. die Menge der Ereignisse, die mit Hilfe der Zufallsgrößen $X_0, \dots, X_n$ beschrieben werden können. Die Familie $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ nennt man die zu $(X_n)_{n\in\mathbb{N}_0}$ gehörige* Filtrierung.

(b) *Eine Abbildung* $T\colon \Omega \to \mathbb{N}_0 \cup \{\infty\}$ *heißt eine* Stoppzeit, *wenn für jedes* $n \in \mathbb{N}_0$ *das Ereignis* $\{T = n\}$ *in* $\mathcal{F}_n$ *liegt.*

Eine Filtrierung ist immer aufsteigend, denn es gilt ja $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ für jedes n. Das Ereignis $\{T = \infty\}$ kann man interpretieren als das Ereignis, dass die Stoppzeit T nie eintritt. Ein Beispiel von Stoppzeiten sind die Ersteintrittszeiten $T_i = \inf\{n \in \mathbb{N}\colon X_n = i\}$ in einen Punkt i für eine Markovkette $(X_n)_{n\in\mathbb{N}_0}$. Etwas allgemeiner definiert man die Ersteintrittszeit in eine nichtleere Menge $A \subset I$ als

$$T_A = \inf\{n \in \mathbb{N}\colon X_n \in A\}.$$

Falls man den Zeitpunkt 0 einbeziehen möchte, benutzt man die Stoppzeit

$$S_A = \inf\{n \in \mathbb{N}_0\colon X_n \in A\}.$$

Falls $i \in A$, so gilt $\mathbb{P}_i(S_A = 0) = 1$. Ferner gilt für $i \notin A$ die Beziehung $\mathbb{P}_i(T_A = S_A) = 1$.

Wir beginnen mit einer technischen Vorbereitung.

Lemma 2.4.2. *Für alle* $i \in I$, $n \in \mathbb{N}_0$ *und alle nichtleeren* $A \subset I$ *gilt*

$$\mathbb{P}_i(T_A \leq n+1) = \sum_{j\in I} p_{i,j}\mathbb{P}_j(S_A \leq n).$$

Beweis. Übungsaufgabe. □

Sei $A \subset I$ nicht leer. Die Funktion $h_A\colon I \to [0,1]$, definiert durch

$$h_A(i) = \mathbb{P}_i(S_A < \infty),$$

spielt in der Theorie der Markovketten eine wichtige Rolle. Insbesondere ist sie die minimale Lösung eines gewissen Systems von linearen Gleichungen:

Satz 2.4.3. *Sei* $A \subset I$ *nicht leer. Dann ist* h_A *die kleinste Funktion* $I \to [0,\infty)$, *die das Gleichungssystem*

$$h_A(i) = \sum_{j\in I} p_{i,j} h_A(j), \qquad i \in I \setminus A,$$

löst mit Randbedingung $h_A(i) = 1$ *für alle* $i \in A$.

Beweis. Dass h_A die Randbedingung $h_A(i) = 1$ für alle $i \in A$ erfüllt, ist klar. Sei nun $i \in I \setminus A$, und wir wollen die Gleichung $h_A(i) = \sum_{j\in I} p_{i,j} h_A(j)$ beweisen. Zunächst benutzen wir die Markoveigenschaft zum Zeitpunkt 1, um zu sehen, dass für alle $N \in \mathbb{N}$ gilt:

$$\begin{aligned}
\mathbb{P}_i(1 \leq S_A \leq N) &= \sum_{j\in I} \mathbb{P}_i(X_1 = j, 1 \leq S_A \leq N) \\
&= \sum_{j\in I} \mathbb{P}_i(X_1 = j)\mathbb{P}_i(1 \leq S_A \leq N \mid X_1 = j) \\
&= \sum_{j\in I} p_{i,j}\mathbb{P}_j(0 \leq S_A \leq N-1).
\end{aligned}$$

Nun lassen wir auf beiden Seiten $N \to \infty$ gehen und beachten, dass $h_A(i) = \mathbb{P}_i(S_A < \infty) = \lim_{N\to\infty} \mathbb{P}_i(1 \leq S_A \leq N)$. Da die Summanden der rechten Seite, $\mathbb{P}_j(0 \leq S_A \leq N-1)$, monoton gegen $h_A(j)$ konvergieren, dürfen wir die Summation mit dem Grenzwert vertauschen, und es folgt die behauptete Gleichung.

Als Letztes zeigen wir die Minimalität von h_A. Sei also $g\colon I \to [0,\infty)$ eine Lösung von $g(i) = \sum_{j\in I} p_{i,j} g(j)$ für $i \in I \setminus A$ mit Randbedingung $g(i) = 1$ für $i \in A$. Wir zeigen mit Induktion nach $n \in \mathbb{N}_0$, dass

$$g(i) \geq \mathbb{P}_i(S_A \leq n) \qquad \text{für alle } i \in I. \tag{2.4.1}$$

(Daraus folgt durch Grenzübergang $n \to \infty$, dass $g(i) \geq \lim_{n\to\infty} \mathbb{P}_i(S_A \leq n) = \mathbb{P}_i(S_A < \infty) = h_A(i)$, was zu zeigen war.)

Für $n = 0$ ist (2.4.1) offensichtlich, ebenso für alle $i \in A$. Sei (2.4.1) für ein n vorausgesetzt, und sei $i \in I \setminus A$. Dann ist

$$g(i) = \sum_{j\in I} p_{i,j} g(j) \geq \sum_{j\in I} p_{i,j} \mathbb{P}_j(S_A \leq n) = \mathbb{P}_i(T_A \leq n+1)$$

nach Lemma 2.4.2. Da $i \notin A$ ist, gilt aber $\mathbb{P}_i(T_A \leq n+1) = \mathbb{P}_i(S_A \leq n+1)$. Dies zeigt (2.4.1) für $n+1$ und beendet den Beweis. □

Wir kommen nun zur *starken Markov-Eigenschaft*. Diese besagt im Wesentlichen, dass die gewöhnliche Markov-Eigenschaft, die in Satz 2.1.5 formuliert wurde, auch an Stoppzeiten gilt, d. h. dass die ab einer Stoppzeit T betrachtete Restkette $(X_T, X_{T+1}, \dots)$, gegeben das Ereignis $\{X_T = i\}$ für ein $i \in I$, wieder eine Markovkette mit der selben Übergangsmatrix und Start in i ist und dass sie ferner unabhängig von der Kette vor dem Zeitpunkt T ist, also unabhängig von $(X_0, X_1, \dots, X_T)$. Die Menge aller Ereignisse, die der Kette $(X_0, X_1, \dots, X_T)$ zustoßen können, ist ein wichtiges System von Ereignissen:

Definition 2.4.4 (Prä-T-Ereignis). *Es sei $T\colon \Omega \to \mathbb{N}_0 \cup \{\infty\}$ eine Stoppzeit. Wir nennen ein Ereignis $A \subset \Omega$ ein* Prä-T-Ereignis, *falls das Ereignis $A \cap \{T = n\}$ für jedes $n \in \mathbb{N}_0$ in $\mathcal{F}_n$ liegt. Die Menge aller Prä-T-Ereignisse bezeichnen wir mit $\mathcal{F}_T$.*

Wir definieren die I-wertige Zufallsgröße X_T als $X_{T(\omega)}(\omega)$, allerdings nur auf der Menge $\{\omega \in \Omega\colon T(\omega) < \infty\}$. Falls $T(\omega) = \infty$, so ist also X_T nicht definiert. Daraus folgt, dass das Ereignis $\{X_T = i\}$ insbesondere impliziert, dass $T < \infty$.

Nun folgt eine exakte Formulierung der starken Markov-Eigenschaft.

Satz 2.4.5 (Starke Markov-Eigenschaft). *Es sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markovkette, T eine Stoppzeit, $A \in \mathcal{F}_T$ und $B \subset I^m$ für ein $m \in \mathbb{N}$. Falls $\mathbb{P}(X_T = i, A) > 0$, so gilt*

$$\mathbb{P}(X_{[T+1,T+m]} \in B \mid X_T = i, A) = \mathbb{P}_i(X_{[1,m]} \in B). \tag{2.4.2}$$

Beweis. Wir multiplizieren mit $\mathbb{P}(X_T = i, A)$ und brauchen nur die Gleichung

$$\mathbb{P}(X_{[T+1,T+m]} \in B, X_T = i, A) = \mathbb{P}_i(X_{[1,m]} \in B)\mathbb{P}(X_T = i, A)$$

zu zeigen. Dies tun wir durch Aufspaltung nach allen (endlichen) Werten, die T annehmen kann: Die linke Seite ist gleich

$$\sum_{n\in\mathbb{N}_0} \mathbb{P}(X_{[n+1,n+m]} \in B, X_n = i, A, T = n).$$

Da $A \in \mathcal{F}_T$, liegt das Ereignis $A \cap \{T = n\}$ in $\mathcal{F}_n$. Also können wir die gewöhnliche Markov-Eigenschaft anwenden (siehe etwa Korollar 2.1.6) und erhalten

$$\mathbb{P}(X_{[n+1,n+m]} \in B, X_n = i, A, T = n) = \mathbb{P}(X_n = i, A, T = n)\mathbb{P}_i(X_{[1,m]} \in B).$$

Aufsummation über $n \in \mathbb{N}_0$ beendet den Beweis. □

Damit können wir die Intuition hinter Satz 2.3.7 (siehe die Bemerkung danach) verschärfen: Die Anzahl der Besuche zu einem gegebenen *rekurrenten* Punkt hat nicht nur unendlichen Erwartungswert, sondern ist sogar konstant gleich ∞:

Korollar 2.4.6. *Es sei $i \in I$ mit $p_i = \mathbb{P}_i(T_i = \infty) < 1$. Es sei $V_i = \sum_{n\in\mathbb{N}_0} \mathbb{1}_{\{X_n=i\}}$ die Anzahl der Besuche in i. Dann gilt $\mathbb{P}_i(V_i > k) = (1 - p_i)^k$ für jedes $k \in \mathbb{N}$, d. h., V_i ist unter $\mathbb{P}_i$ geometrisch verteilt, falls i transient ist, und konstant gleich ∞, falls i rekurrent ist.*

Beweis. Wir definieren die sukzessiven Zeitpunkte der Besuche in i durch $T_i^{(0)} = 0$ und $T_i^{(k+1)} = \inf\{n > T_i^{(k)} : X_n = i\}$ für $k \in \mathbb{N}_0$. Das Ereignis $\{V_i > k\}$ ist also gleich dem Ereignis $\{T_i^{(k)} < \infty\}$, falls die Kette in i startet. Ferner gilt offensichtlich $\mathbb{P}_i(X_{T_i^{(k)}} = i \mid T_i^{(k)} < \infty) = 1$. Also gilt für alle $k, l \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{P}_i(V_i > k + l \mid V_i > l) &= \mathbb{P}_i(T_i^{(k+l)} < \infty \mid T_i^{(l)} < \infty)\\
&= \lim_{n\to\infty} \mathbb{P}_i(T_i^{(k+l)} \leq T_i^{(l)} + n \mid X_{T_i^{(l)}} = i, T_i^{(l)} < \infty)\\
&= \lim_{n\to\infty} \mathbb{P}_i(T_i^{(k)} \leq n) = \mathbb{P}_i(T_i^{(k)} < \infty)\\
&= \mathbb{P}_i(V_i > k).
\end{aligned}$$

Im dritten Schritt benutzten wir die starke Markov-Eigenschaft (2.4.2) mit $A = \{T_i^{(l)} < \infty\}$ und $B = \{T_i^{(k)} \leq n\}$.

Also ist V_i auf $\mathbb{N}_0$ geometrisch verteilt, und zwar mit Parameter $\mathbb{P}_i(V_i = 1) = 1 - \mathbb{P}_i(V_i > 1) = 1 - \mathbb{P}_i(T_i < \infty) = \mathbb{P}_i(T_i = \infty) = p_i$. Wie man leicht sieht, gilt dies auch im Fall $p_i = 0$, wobei dann $V_i = \infty$ mit $\mathbb{P}_i$-Wahrscheinlichkeit Eins. □

Beispiel 2.4.7. Es sei $(X_n)_{n\in\mathbb{N}_0}$ die Irrfahrt auf $\mathbb{Z}$ wie in Beispiel 2.2.3. Wir wollen berechnen, wie lange die Irrfahrt im Durchschnitt benötigt, um einen gegebenen Punkt zu erreichen. Genauer: Wir wollen $\mathbb{P}_0(T_1 < \infty)$ und $\mathbb{E}_0(T_1)$ berechnen.

Zunächst sieht man mit Hilfe der starken Markov-Eigenschaft und der räumlichen Verschiebungsinvarianz der Irrfahrt ein, dass $\mathbb{P}_0(T_2 < \infty) = \mathbb{P}_0(T_1 < \infty)\mathbb{P}_1(T_2 < \infty) = \mathbb{P}_0(T_1 < \infty)^2$ gilt (der formale Beweis ist eine Übungsaufgabe). Ferner benutzt man die (gewöhnliche) Markov-Eigenschaft zum Zeitpunkt Eins, um die Gleichung $\mathbb{P}_0(T_1 < \infty) = p + (1 - p)\mathbb{P}_{-1}(T_1 < \infty)$ aufzustellen. Wieder mit Hilfe der Verschiebungsinvarianz und der ersten Beobachtung erhält man die quadratische Gleichung $\mathbb{P}_0(T_1 < \infty) = p + (1 - p)\mathbb{P}_0(T_1 < \infty)^2$. Diese hat die beiden Lösungen

$$\mathbb{P}_0(T_1 < \infty) = \frac{1 \pm \sqrt{1 - 4p(1 - p)}}{2(1 - p)} = \frac{1 \pm |2p - 1|}{2(1 - p)} = \begin{cases} 1 \\ \text{oder} \\ \frac{p}{1-p}. \end{cases}$$

Im Fall $p \geq \frac{1}{2}$ ergibt sich sofort, dass $\mathbb{P}_0(T_1 < \infty) = 1$, denn $\frac{p}{1-p} > 1$ für $p > \frac{1}{2}$. Im Fall $p < \frac{1}{2}$

gilt $\mathbb{P}_0(T_1 < \infty) = \frac{p}{1-p} < 1$, aber unsere Mittel reichen derzeit noch nicht aus, dies zu beweisen.[1] Für $p \geq \frac{1}{2}$ erreicht die Irrfahrt also den Punkt 1 von 0 aus mit Sicherheit, und man zeigt mit der starken Markov-Eigenschaft leicht, dass dann auch jedes andere $i \in \mathbb{N}$ mit Sicherheit erreicht wird.

Mit der (gewöhnlichen) Markov-Eigenschaft leitet man im Fall $p \geq \frac{1}{2}$ auch leicht die Beziehung $\mathbb{E}_0(T_1) = p + (1-p)[1 + \mathbb{E}_{-1}(T_1)]$ her, und die starke impliziert, dass $\mathbb{E}_{-1}(T_1) = \mathbb{E}_0(T_2) = 2\mathbb{E}_0(T_1)$. Also gilt $\mathbb{E}_0(T_1) = \frac{1}{1-2(1-p)}$, was im symmetrischen Fall gleich ∞ ist. ◇

2.5 Gleichgewichtsverteilungen

Im Allgemeinen hängt die Verteilung einer Markovkette zu einem gegebenen Zeitpunkt auf sehr komplizierte Weise ab von der Startverteilung (und natürlich von dem Zeitpunkt). Oft gibt es aber für eine gegebene Übergangsmatrix spezielle Verteilungen, so dass die mit dieser Verteilung gestartete Kette zu jedem Zeitpunkt dieselbe Verteilung besitzt, nämlich diese Startverteilung. Man sagt, die Kette sei dann *im Gleichgewicht* bzw. sie sei *stationär*.

Wie bisher sei I eine höchstens abzählbare Menge und P eine stochastische Matrix auf I. Wir wollen jede Abbildung $\nu\colon I \to [0,\infty)$ ein *Maß* nennen und erinnern uns, dass ein Maß ν eine *Verteilung* heißt, falls $\sum_{i\in I} \nu(i) = 1$ gilt. Außerdem erinnern wir uns, dass νP das Matrixprodukt $(\nu P)(j) = \sum_{i\in I} \nu(i)p_{i,j}$ bezeichnet.

Definition 2.5.1 (invariantes Maß, Gleichgewichtsverteilung). *Ein Maß ν heißt* invariant *für P, falls $\nu P = \nu$ gilt, d. h. falls ν ein (nichtnegativer) Links-Eigenvektor von P zum Eigenwert 1 ist. Falls ν eine Verteilung und invariant ist, nennt man ν auch eine* Gleichgewichtsverteilung *für P.*

Bemerkung 2.5.2. (a) Wenn I endlich ist, kann jedes invariante Maß zu einer Gleichgewichtsverteilung normiert werden.

(b) Die Frage, ob P ein invariantes Maß besitzt, ist *a priori* nicht leicht zu beantworten. Man muss im Prinzip das Gleichungssystem $\nu P = \nu$ lösen und prüfen, ob es eine nichtnegative Lösung ν gibt.

(c) Offensichtlich ist ein für P invariantes Maß auch für jede Potenz von P invariant. Falls P irreduzibel ist und $\nu \neq 0$ ein invariantes Maß ist, so ist $\nu(i) > 0$ für jedes $i \in I$. Denn wegen $\nu \neq 0$ existiert ein $i_0 \in I$ mit $\nu(i_0) > 0$, und dann folgt für jedes $i \in I$ (mit einem $n \in \mathbb{N}$, so dass $p^{(n)}_{i_0,i} > 0$): $\nu(i) \geq \nu(i_0)p^{(n)}_{i_0,i} > 0$.

(d) Wenn ν eine Gleichgewichtsverteilung ist, dann gilt für jedes $n \in \mathbb{N}_0$ und jedes $j \in I$:

$$\mathbb{P}_\nu(X_n = j) = \sum_{i\in I} \nu(i)p^{(n)}_{i,j} = \nu(j),$$

d. h. die mit Verteilung ν gestartete Kette hat zu jedem Zeitpunkt die selbe Verteilung ν. ◇

[1] Dies folgt etwa aus dem Starken Gesetz der Großen Zahlen, zusammen mit dem Hinweis auf $\mathbb{E}_0(S_1) < 0$, d. h. es gilt insbesondere $\lim_{n\to\infty} S_n = -\infty$ mit $\mathbb{P}_0$-Wahrscheinlichkeit Eins.

Sehr oft kann das invariante Maß nicht sinnvoll explizit angegeben werden, und es gibt wenige Beispiele, in denen das invariante Maß eine der bekannten Verteilungen ist. Für die symmetrische Irrfahrt auf $\mathbb{Z}$ (siehe Beispiel 2.2.3) ist die Gleichverteilung auf $\mathbb{Z}$ invariant, aber offensichtlich kann man sie nicht zu einer Gleichgewichtsverteilung normieren. Für die Irrfahrt auf $\{0,\dots,N\}$ (siehe Beispiel 2.2.5) mit absorbierenden Rändern 0 und N sind die beiden Verteilungen invariant, die auf $\{0\}$ bzw. auf $\{N\}$ konzentriert sind. Für weitere Beispiele siehe Lemma 2.5.8 und Beispiele 2.5.9 und 2.7.4.

Für den Rest dieses Abschnittes setzen wir voraus, dass P irreduzibel ist, also dass I aus einer einzigen Äquivalenzklasse besteht. Zunächst zeigen wir, dass immer mindestens ein invariantes Maß existiert, und zwar, für beliebiges $k \in I$, das Maß γ_k, gegeben als

$$\gamma_k(i) = \mathbb{E}_k\Big(\sum_{n=1}^{T_k} 1\!\!1_{\{X_n=i\}}\Big).$$

In Worten: $\gamma_k(i)$ ist die erwartete Zahl der Besuche in i für eine in k gestartete Markovkette, die betrachtet wird bis zur ersten Rückkehr zum Startpunkt k.

Satz 2.5.3. *Sei P irreduzibel, und sei $k \in I$. Dann gelten:*

(a) γ_k ist ein invariantes Maß.

(b) Für jedes $i \in I$ gilt $0 < \gamma_k(i) < \infty$.

(c) γ_k ist das einzige invariante Maß mit Wert 1 in k.

Beweis. (a) Wir schreiben zunächst $\gamma_k(i)$ wie folgt:

$$\begin{aligned}\gamma_k(i) &= \mathbb{E}_k\Big(\sum_{n\in\mathbb{N}} 1\!\!1_{\{X_n=i,n\le T_k\}}\Big) = \sum_{n\in\mathbb{N}} \mathbb{P}_k(X_n=i, n\le T_k)\\ &= \sum_{n\in\mathbb{N}}\sum_{j\in I} \mathbb{P}_k(X_n=i, X_{n-1}=j, n\le T_k).\end{aligned}$$

Man beachte, dass das Ereignis $\{n \le T_k\} = \{T_k \le n-1\}^{\mathrm{c}}$ in $\mathcal{F}_{n-1}$ liegt. Also können wir die Markov-Eigenschaft zum Zeitpunkt $n-1$ anwenden:

$$\begin{aligned}\mathbb{P}_k(X_n=i, X_{n-1}=j, n\le T_k) &= \mathbb{P}_k(X_{n-1}=j, n\le T_k)\mathbb{P}_j(X_1=i)\\ &= \mathbb{P}_k(X_{n-1}=j, n-1\le T_k-1)p_{j,i}.\end{aligned}$$

Dies setzen wir oben ein und erhalten nach einer Indexverschiebung

$$\begin{aligned}\gamma_k(i) &= \sum_{j\in I} p_{j,i}\sum_{n\in\mathbb{N}_0}\mathbb{P}_k(X_n=j, n\le T_k-1) = \sum_{j\in I} p_{j,i}\mathbb{E}_k\Big(\sum_{n=0}^{T_k-1} 1\!\!1_{\{X_n=j\}}\Big)\\ &= \sum_{j\in I} p_{j,i}\mathbb{E}_k\Big(\sum_{n=1}^{T_k} 1\!\!1_{\{X_n=j\}}\Big) = \sum_{j\in I}\gamma_k(j)p_{j,i}.\end{aligned}$$

(b) Auf Grund von (a) ist γ_k auch invariant für P^n für jedes $n \in \mathbb{N}$, also folgt insbesondere $1 = \gamma_k(k) \ge \gamma_k(j)p_{j,k}^{(n)}$ für jedes $j \in I$. Wegen der Irreduzibilität gibt es für jedes j ein n mit $p_{j,k}^{(n)} > 0$, und somit folgt $\gamma_k(j) < \infty$ für jedes j. Andererseits folgt auch $\gamma_k(j) \ge \gamma_k(k)p_{k,j}^{(n)} = p_{k,j}^{(n)}$, was positiv ist für geeignetes n.

(c) Es sei λ ein invariantes Maß mit $\lambda(k) = 1$. Dann haben wir $\lambda(j) = \sum_{i\in I\setminus\{k\}} \lambda(i)p_{i,j} + p_{k,j}$ für jedes j. Nun ersetzen wir auf der rechten Seite $\lambda(i)$ ebenfalls mit Hilfe der Invarianz und erhalten

$$\begin{aligned}
\lambda(j) &= \sum_{i\in I\setminus\{k\}} \Big(\sum_{i_1\in I\setminus\{k\}} \lambda(i_1)p_{i_1,i} + p_{k,i}\Big) p_{i,j} + p_{k,j}\\
&= \sum_{i,i_1\in I\setminus\{k\}} \lambda(i_1)p_{i_1,i}p_{i,j} + \sum_{i\in I\setminus\{k\}} p_{k,i}p_{i,j} + p_{k,j}\\
&= \sum_{i,i_1\in I\setminus\{k\}} \lambda(i_1)p_{i_1,i}p_{i,j} + \mathbb{P}_k(T_k \geq 2, X_2 = j) + \mathbb{P}_k(T_k \geq 1, X_1 = j).
\end{aligned}$$

In dieser Weise fahren wir iterativ fort und erhalten für jedes $n \in \mathbb{N}$

$$\begin{aligned}
\lambda(j) &= \sum_{i_0,i_1,\dots,i_n\in I\setminus\{k\}} \lambda(i_n)\Big(\prod_{r=1}^{n} p_{i_r,i_{r-1}}\Big)p_{i_0,j} + \sum_{r=1}^{n+1} \mathbb{P}_k(T_k \geq r, X_r = j)\\
&\geq \sum_{r=1}^{n+1} \mathbb{P}_k(T_k \geq r, X_r = j)\\
&= \mathbb{E}_k\Big(\sum_{r=1}^{\min\{T_k,n+1\}} 1\!\!1_{\{X_r=j\}}\Big).
\end{aligned}$$

Nun lassen wir $n \to \infty$ und erhalten auf der rechten Seite $\gamma_k(j)$ als Grenzwert. Also haben wir gezeigt, dass $\lambda(j) \geq \gamma_k(j)$ für alle $j \in I$. Daher ist $\lambda - \gamma_k$ ebenfalls ein invariantes Maß. Da dieses Maß allerdings eine Nullstelle besitzt (nämlich in k), ist es nach Bemerkung 2.5.2(b) identisch gleich Null. Dies beendet den Beweis. □

Zusammen mit Satz 2.3.14 wissen wir nun, dass jede *endliche* irreduzible stochastische Matrix immer eine Gleichgewichtsverteilung besitzt.

Der folgende Satz 2.5.5 klärt die Frage nach der Existenz von Gleichgewichtsverteilungen und ist einer der wichtigsten Sätze über Markovketten. Die Antwort ist eng verknüpft mit der *erwarteten Rückkehrzeit* zu einem Startpunkt $i \in I$:

$$\mu_i = \mathbb{E}_i(T_i) = \sum_{n\in\mathbb{N}} n\mathbb{P}_i(T_i = n) + \infty\mathbb{P}_i(T_i = \infty) = \sum_{n\in\mathbb{N}} n f_{i,i}^{(n)} + \infty\mathbb{P}_i(T_i = \infty) \in [0, \infty].$$

Wir benutzen die Konvention $\infty 0 = 0$. Falls $\mu_i < \infty$ (d. h. falls die Irrfahrt nach endlicher erwarteter Zeit zu ihrem Startpunkt i zurück kehrt), dann ist natürlich $\mathbb{P}_i(T_i < \infty) = 1$, die Irrfahrt also rekurrent. Mit Hilfe von μ_i können wir also rekurrente Zustände feiner unterteilen:

Definition 2.5.4 (positive und Nullrekurrenz). *Ein Zustand $i \in I$ heißt* positiv rekurrent*, falls $\mu_i < \infty$, und* nullrekurrent*, falls i rekurrent, aber nicht positiv rekurrent ist.*

Der folgende Satz zeigt insbesondere, dass die Existenz von Gleichgewichtsverteilungen äquivalent zur positiven Rekurrenz ist:

Satz 2.5.5. *Sei P irreduzibel, dann sind die folgenden drei Aussagen äquivalent.*

(i) Es existiert eine Gleichgewichtsverteilung.

(ii) Es gibt einen positiv rekurrenten Zustand in I.

(iii) Alle Zustände in I sind positiv rekurrent.

Sind diese Bedingungen erfüllt, so ist die Gleichgewichtsverteilung π eindeutig bestimmt und durch $\pi(i) = \frac{1}{\mu_i}$ gegeben.

Beweis. Die Richtung $(iii) \Longrightarrow (ii)$ ist trivial. Wir zeigen nun $(ii) \Longrightarrow (i)$. Sei $k \in I$ mit $\mu_k < \infty$, dann ist wegen Irreduzibilität die Kette rekurrent. Nach Satz 2.5.3 ist γ_k ein invariantes Maß. Man sieht, dass

$$\sum_{j\in I} \gamma_k(j) = \sum_{j\in I} \mathbb{E}_k\Big(\sum_{n=1}^{T_k} \mathbb{1}_{\{X_n=j\}}\Big) = \mathbb{E}_k\Big(\sum_{n=1}^{T_k}\sum_{j\in I} \mathbb{1}_{\{X_n=j\}}\Big) = \mathbb{E}_k(T_k) = \mu_k < \infty. \tag{2.5.1}$$

Also kann man γ_k zu einer Gleichgewichtsverteilung normieren, d. h. (i) folgt.

Nun beweisen wir $(i) \Longrightarrow (iii)$. Sei π eine Gleichgewichtsverteilung, und sei $k \in I$. Dann ist $\gamma = \pi/\pi(k)$ ein invariantes Maß mit $\gamma(k) = 1$. Nach Satz 2.5.3 gilt $\gamma = \gamma_k$. Wie in (2.5.1) sieht man, dass

$$\mu_k = \sum_{j\in I} \gamma_k(j) = \sum_{j\in I} \gamma(j) = \frac{1}{\pi(k)} \sum_{j\in I} \pi(j) = \frac{1}{\pi(k)} < \infty.$$

Also ist (iii) gezeigt.

Die Zusatzaussage ergibt sich aus den obigen Beweisen. □

Bemerkung 2.5.6 (Klassifikation). Also ergibt sich eine Trichotomie für irreduzible Markovketten, d. h. eine irreduzible Markovkette gehört immer zu genau einem der folgenden drei Fälle:

- Die Kette ist *transient*, für alle $i, j \in I$ gilt $\mathbb{P}_i(T_j < \infty) < 1$, und es gibt keine invariante Verteilung.
- Die Kette ist *nullrekurrent*, für alle $i, j \in I$ gilt $\mathbb{P}_i(T_j < \infty) = 1$ und $\mathbb{E}_i(T_j) = \infty$, und es gibt keine invariante Verteilung.
- Die Kette ist *positiv rekurrent*, für alle $i, j \in I$ gilt $\mathbb{E}_i(T_j) < \infty$, und es gibt eine invariante Verteilung.

Auf Grund von Satz 2.3.14 kann bei endlichem Zustandsraum I nur der positiv rekurrente Fall auftreten. ◇

Beispiel 2.5.7. Die eindimensionale Irrfahrt (siehe Beispiel 2.2.3) ist nullrekurrent. Denn sie besitzt ja das konstante Maß als ein invariantes Maß, und dies lässt sich nicht normieren. ◇

In den meisten Fällen, selbst bei endlichem I, kann man die Gleichgewichtsverteilung nicht sinnvoll explizit angeben, die Formel wäre zu kompliziert. Ein schönes Gegenbeispiel ist die μ-Irrfahrt auf einer Gruppe, siehe Beispiel 2.2.2:

Lemma 2.5.8. *Es seien G eine endliche Gruppe und μ eine Verteilung auf G. Dann ist die Gleichverteilung auf G eine Gleichgewichtsverteilung der μ-Irrfahrt. Falls die Irrfahrt irreduzibel ist, so ist sie auch die einzige Gleichgewichtsverteilung.*

Beweis. Für jedes $h \in G$ ist die Abbildung $g \mapsto g^{-1}h$ bijektiv auf G, und es gilt

$$\sum_{g\in G} p_{g,h} = \sum_{g\in G} \mu(g^{-1}h) = \sum_{g'\in G} \mu(g') = 1.$$

□

Beispiel 2.5.9 (Bernoulli-Laplace-Diffusionsmodell). Noch ein Beispiel, in dem die invariante Verteilung explizit angegeben werden kann, ist das Diffusionsmodell aus Beispiel 2.2.8. Hier ist die hypergeometrische Verteilung $\nu(i) = \binom{s}{i}\binom{w}{w-i}/\binom{s+w}{w}$ invariant auf $\{0, 1, \dots, w\}$. ◇

2.6 Konvergenz gegen die Gleichgewichtsverteilung

Das Hauptergebnis dieses Abschnittes (und eine der wichtigsten Eigenschaften von Markovketten) ist die Tatsache, dass jede irreduzible aperiodische und positiv rekurrente Markovkette bei divergierender Laufzeit gegen ihre Gleichgewichtsverteilung konvergiert. Hierbei nennen wir eine Markovkette bzw. ihre Übergangsmatrix P *aperiodisch*, wenn für jedes $i \in I$ die sogenannte *Periode von i*,

$$d_i = \mathrm{ggT}\{n \in \mathbb{N} \colon p_{i,i}^{(n)} > 0\},$$

gleich Eins ist. Wir wollen uns im Folgenden auf aperiodische Markovketten beschränken, d. h. wir werden voraussetzen, dass $d_i = 1$ für jedes $i \in I$ gilt. Unser Hauptergebnis ist:

Satz 2.6.1. *Es sei P irreduzibel, aperiodisch und positiv rekurrent mit Gleichgewichtsverteilung π. Dann gilt für alle $i, j \in I$:*

$$\lim_{n\to\infty} p_{i,j}^{(n)} = \pi(j).$$

Als Übungsaufgabe zeigt man leicht, dass auch für jede Startverteilung ν und jedes $j \in I$ gilt: $\lim_{n\to\infty} \mathbb{P}_\nu(X_n = j) = \pi(j)$. Falls die Markovkette nicht aperiodisch ist, also wenn sie etwa die Periode $d > 1$ hat[2], dann zeigt man leicht, dass die Matrix P^d auf gewissen Teilmengen aperiodisch ist, und kann Satz 2.6.1 auf geeignete Teilmatrizen von P^d anwenden.

Als Vorbereitung des Beweises von Satz 2.6.1 charakterisieren wir die Aperiodizität:

Lemma 2.6.2. *P ist genau dann irreduzibel und aperiodisch, wenn für jede $i, j \in I$ ein $n_0 \in \mathbb{N}$ existiert, so dass $p_{i,j}^{(n)} > 0$ für alle $n \geq n_0$ gilt.*

Beweis. Dieser Beweis ist länglich und benutzt ausschließlich zahlentheoretische Argumente und wird daher weggelassen. Siehe etwa [Se81]. □

Beweis von Satz 2.6.1. Wie immer sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markovkette mit Übergangsmatrix P und Start in i unter $\mathbb{P}_i$. Wir wollen also zeigen, dass $\mathbb{P}_i(X_n = j)$ für $n \to \infty$ gegen $\pi(j)$ konvergiert.

[2] Man kann leicht zeigen, dass die Periode eine Klasseneigenschaft ist, d. h. dass die Abbildung $i \mapsto d_i$ konstant auf den Klassen ist.

Wir benutzen ein sogenanntes *Kopplungsargument*, das die Untersuchung der betrachteten Markovkette $(X_n)_{n\in\mathbb{N}_0}$ koppelt mit einer weiteren, unabhängigen Markovkette $(Y_n)_{n\in\mathbb{N}_0}$ mit der selben Übergangsmatrix, wobei allerdings $(Y_n)_{n\in\mathbb{N}_0}$ mit der Gleichgewichtsverteilung π gestartet wird. Daraus folgt, dass die Verteilung von Y_n zu jedem festen Zeitpunkt n exakt gleich π ist, siehe Bemerkung 2.5.2(d). Unsere Hauptaufgabe besteht nun darin zu beweisen, dass die erste Kette irgendwann einmal die zweite trifft, d. h. dass sich beide Ketten jemals im selben Zustand befinden. Mit anderen Worten, wir werden zeigen, dass die *Treffzeit*

$$T = \inf\{n \in \mathbb{N}_0 \colon X_n = Y_n\}$$

mit Sicherheit endlich ist. Dann definieren wir die gekoppelte Kette $(Z_n)_{n\in\mathbb{N}_0}$ durch

$$Z_n = \begin{cases} X_n & \text{für } n \leq T, \\ Y_n & \text{für } n > T, \end{cases}$$

und zeigen, dass $(Z_n)_{n\in\mathbb{N}_0}$ ebenfalls eine Markovkette mit Übergangsmatrix P ist. Danach ist der Beweis schnell beendet.

Wir kommen zu den Details. Als Erstes sieht man, dass die Paar-Markovkette $(X_n, Y_n)_{n\in\mathbb{N}_0}$ die Übergangsmatrix $\widehat{P}$ auf $I \times I$ mit $\widehat{p}_{(i,j),(k,l)} = p_{i,k}p_{j,l}$ für alle $i,j,k,l \in I$ besitzt. Für alle $n \in \mathbb{N}$ gilt auch $\widehat{p}^{(n)}_{(i,j),(k,l)} = p^{(n)}_{i,k}p^{(n)}_{j,l}$, wovon man sich leicht überzeugt. Die Aperiodizität von $\widehat{P}$ zeigt man leicht als Übungsaufgabe mit Hilfe von Lemma 2.6.2 und (2.1.7). Da P die Gleichgewichtsverteilung π besitzt, hat $\widehat{P}$ die Gleichgewichtsverteilung $\widehat{\pi}$, die gegeben ist durch $\widehat{\pi}(i,j) = \pi(i)\pi(j)$ für $i,j \in I$. Also ist $\widehat{P}$ insbesondere positiv rekurrent (siehe Satz 2.5.5).

Ab jetzt sei $X_0 = i$ fest. Die Markovkette $(X_n, Y_n)_{n\in\mathbb{N}_0}$ auf $I \times I$ besitzt die Startverteilung $\widehat{\nu}(k,l) = \delta_i(k)\pi(l)$, wobei $\delta_i(k) = \delta_{i,k}$ das Kroneckersymbol ist.

Nun argumentieren wir, dass die Treffzeit T mit Wahrscheinlichkeit Eins endlich ist. Für einen beliebigen Zustand $b \in I$ sei $T_{(b,b)}$ der erste Zeitpunkt, an dem die Kette $(X_n, Y_n)_{n\in\mathbb{N}_0}$ den Punkt (b,b) trifft. Offensichtlich ist $T \leq T_{(b,b)}$. Wegen der Rekurrenz von $\widehat{P}$ ist $T_{(b,b)}$ endlich, also auch T.

Nun beweisen wir, dass die oben definierte Folge $(Z_n)_{n\in\mathbb{N}_0}$ eine Markovkette mit Übergangsmatrix P ist. Dazu reicht es nach Lemma 2.1.5(a) zu zeigen, dass für alle $n \in \mathbb{N}$ und alle $i_0, \dots, i_n \in I$ gilt:

$$\mathbb{P}_{\widehat{\nu}}(Z_{[0,n]} = i_{[0,n]}) = \delta_i(i_0) \prod_{r=1}^{n} p_{i_{r-1},i_r}. \tag{2.6.1}$$

Dies zeigen wir durch Aufspaltung nach allen Werten, die T annehmen kann:

$$\begin{aligned}
\mathbb{P}_{\widehat{\nu}}(Z_{[0,n]} = i_{[0,n]}) &= \sum_{r=0}^{n} \mathbb{P}_{\widehat{\nu}}(Z_{[0,n]} = i_{[0,n]}, T = r) + \mathbb{P}_{\widehat{\nu}}(Z_{[0,n]} = i_{[0,n]}, T > n) \\
&= \sum_{r=0}^{n} \mathbb{P}_{\widehat{\nu}}(X_{[0,r]} = i_{[0,r]}, Y_{[r+1,n]} = i_{[r+1,n]}, Y_0 \neq i_0, \dots, Y_{r-1} \neq i_{r-1}, Y_r = i_r) \\
&\quad + \mathbb{P}_{\widehat{\nu}}(X_{[0,n]} = i_{[0,n]}, Y_0 \neq i_0, \dots, Y_n \neq i_n).
\end{aligned}$$

Nun nutzt man die Unabhängigkeit der Ketten $(X_n)_{n\in\mathbb{N}_0}$ und $(Y_n)_{n\in\mathbb{N}_0}$ sowie die Markov-

Eigenschaft der Kette $(Y_n)_{n\in\mathbb{N}_0}$ zum Zeitpunkt r bzw. n aus, um zu erhalten:

$$\begin{aligned}
&\mathbb{P}_{\widehat{\nu}}(X_{[0,r]} = i_{[0,r]}, Y_{[r+1,n]} = i_{[r+1,n]}, Y_0 \neq i_0, \dots, Y_{r-1} \neq i_{r-1}, Y_r = i_r) \\
&\quad = \mathbb{P}_i(X_{[0,r]} = i_{[0,r]})\mathbb{P}_\pi(Y_{[r+1,n]} = i_{[r+1,n]} \mid Y_0 \neq i_0, \dots, Y_{r-1} \neq i_{r-1}, Y_r = i_r) \\
&\qquad\qquad \times \mathbb{P}_\pi(Y_0 \neq i_0, \dots, Y_{r-1} \neq i_{r-1}, Y_r = i_r) \\
&\quad = \delta_i(i_0)\Big(\prod_{j=1}^n p_{i_{j-1},i_j}\Big)\mathbb{P}_\pi(Y_0 \neq i_0, \dots, Y_{r-1} \neq i_{r-1}, Y_r = i_r)
\end{aligned}$$

und

$$\mathbb{P}_{\widehat{\nu}}(X_{[0,n]} = i_{[0,n]}, Y_0 \neq i_0, \dots, Y_n \neq i_n) = \delta_i(i_0)\Big(\prod_{j=1}^n p_{i_{j-1},i_j}\Big)\mathbb{P}_\pi(Y_0 \neq i_0, \dots, Y_n \neq i_n).$$

Nun kombiniere man all diese Gleichungen und beachte, dass

$$\sum_{r=0}^n \mathbb{P}_\pi(Y_0 \neq i_0, \dots, Y_{r-1} \neq i_{r-1}, Y_r = i_r) + \mathbb{P}_\pi(Y_0 \neq i_0, \dots, Y_n \neq i_n) = 1,$$

um bei (2.6.1) anzukommen. Also ist $(Z_n)_{n\in\mathbb{N}_0}$ eine Markovkette mit Übergangsmatrix P.

Nun ist der Beweis von Satz 2.6.1 leicht zu beenden. Wir schreiben $p_{i,j}^{(n)}$ mit Hilfe von $(Z_n)_{n\in\mathbb{N}_0}$ als

$$p_{i,j}^{(n)} = \mathbb{P}_{\widehat{\nu}}(Z_n = j) = \mathbb{P}_{\widehat{\nu}}(Z_n = j, T \leq n) + \mathbb{P}_{\widehat{\nu}}(Z_n = j, T > n)$$

und analog $\pi(j)$ als

$$\pi(j) = \mathbb{P}_\pi(Y_n = j) = \mathbb{P}_{\widehat{\nu}}(Z_n = j, T \leq n) + \mathbb{P}_{\widehat{\nu}}(Y_n = j, T > n).$$

Somit folgt

$$\big|p_{i,j}^{(n)} - \pi(j)\big| \leq 2\mathbb{P}_{\widehat{\nu}}(T > n),$$

und dies konvergiert für $n \to \infty$ gegen Null, da ja $\mathbb{P}_{\widehat{\nu}}(T < \infty) = 1$. □

2.7 Reversible Markovketten

Einen wichtigen Spezialfall von Gleichgewichtsverteilungen stellen Verteilungen dar, die die sogenannte *detailed balance condition* erfüllen:

Definition 2.7.1 (reversibel). *Ein Maß ν auf I heißt* reversibel *bezüglich der stochastischen Matrix P, falls für alle $i,j \in I$ gilt: $\nu(i)p_{i,j} = \nu(j)p_{j,i}$. In diesem Fall nennen wir auch P reversibel.*

Bemerkung 2.7.2. (a) Es ist offensichtlich, dass jedes reversible Maß invariant ist. Im positiv rekurrenten Fall ist also die Gleichgewichtsverteilung der einzige Kandidat für eine reversible Verteilung. Allerdings sind bei weitem nicht alle Gleichgewichtsverteilungen reversibel. Die Reversibilität ist eine sehr spezielle Eigenschaft.

(b) Der Begriff der Reversibilität kommt von der folgenden Eigenschaft her: Wenn π eine reversible Verteilung ist, so ist die Verteilung der Markovkette unter $\mathbb{P}_\pi$ invariant unter Zeitumkehr, d. h. für alle $n \in \mathbb{N}$ und alle $i_0, i_1, \dots, i_n \in I$ gilt

$$\mathbb{P}_\pi(X_0 = i_0, \dots, X_n = i_n) = \mathbb{P}_\pi(X_0 = i_n, \dots, X_n = i_0).$$

Diese Eigenschaft rechnet man leicht als Übungsaufgabe nach.

◇

Es folgen zwei Beispiele, in denen die reversible Gleichgewichtsverteilung existiert und explizit angegeben werden kann.

Beispiel 2.7.3. Sei G eine endliche Gruppe und μ eine Verteilung auf G mit $\mu(g) = \mu(g^{-1})$ für jedes $g \in G$. Dann ist die Gleichverteilung auf G reversibel für die μ-Irrfahrt auf G (siehe Beispiele 2.2.2 und 2.5.8). Dies sieht man leicht, indem man nachrechnet, dass die Übergangsmatrix P symmetrisch ist. ◇

Beispiel 2.7.4 (Ehrenfests Urnenmodell). Wir betrachten das Ehrenfest-Modell von Beispiel 2.2.7. Als Übungsaufgabe rechnet man nach, dass die Binomialverteilung $\pi(k) = \binom{N}{k} 2^{-N}$ reversibel ist.

Eine andere Betrachtungsweise macht es möglich, die reversible Verteilung durch Deduktion zu finden. Wir erweitern das Modell, indem wir den Aufenthaltsort jeder einzelnen Kugel registrieren: '0' für die linke Urne, '1' für die rechte. Dann ist der Vektor der N Aufenthaltsorte der N Kugeln ein Element des neuen Zustandsraumes $\{0,1\}^N$. Wir können diesen Zustandsraum als eine Gruppe auffassen, wobei die Verknüpfung die Addition modulo 2 ist. Die Markovkette der N Labels gehorcht der Regel, dass zu den Zeitpunkten $1, 2, \dots$ ein zufällig gewähltes der N Indizes geflippt wird. Dies ist eine μ-Irrfahrt im Sinne von Beispiel 2.2.2 mit $\mu(x) = \frac{1}{N}$, falls $x = (x_1, \dots, x_N)$ genau eine von Null verschiedene Komponente hat. Nach Beispiel 2.7.3 ist die Gleichverteilung reversibel für diese Irrfahrt. Projiziert man diese Verteilung wieder zurück auf die Anzahl der Kugeln in der linken Urne, erhält man gerade die Binomialverteilung.

Da die Kehrwerte der Verteilung $\pi(k)$ gerade die erwarteten Rückkehrzeiten zu k sind, erhält man die interessante Folgerung, dass das System im Mittel viel später in seinen Anfangszustand zurückkehrt, wenn es in einem extremen Zustand startet (d. h. alle Kugeln in einer Urne), als wenn es mit etwa ausgeglichenen Kugelbeständen beginnt. ◇

Beispiel 2.7.5 (MCMC). In vielen Anwendungen des Konvergenzsatzes 2.6.1 dreht man den Spieß folgendermaßen um: Auf einer endlichen Menge I hat man eine gewisse Verteilung π gegeben und möchte eine Stichprobe mit dieser Verteilung ziehen, d. h. man möchte eine Zufallsvariable mit der Verteilung π generieren. Diese Aufgabe ist zwar endlich, aber in den meisten Fällen höchst nichttrivial, weil I oft gigantisch groß ist. Ein beliebter Ansatz ist die sogenannte *Markov Chain Monte Carlo* Methode: Man wählt eine Markovkette, deren invariante Verteilung π ist, und man lässt diese Kette mit einem beliebigen Startwert 'genügend lange' laufen, bis man auf Grund von Satz 2.6.1 meint, dass die Kette 'nahe genug' an π heran gekommen ist. Dabei wählt man meist eine Übergangsmatrix mit vielen Nullen, so dass π sogar reversibel ist, etwa die Matrix $P = (p_{i,j})_{i,j \in I}$ mit

$$p_{i,j} = \begin{cases} q_{i,j} \min\Big\{1, \frac{\pi(j) q_{j,i}}{\pi(i) q_{i,j}}\Big\}, & \text{falls } i \neq j, \\ 1 - \sum_{k \in I \setminus \{i\}} p_{i,k}, & \text{falls } i = j, \end{cases}$$

wobei $Q = (q_{i,j})_{i,j\in I}$ eine stochastische ‘Referenzmatrix’ auf I ist (mit vielen Nullen). Dieser Ansatz wird der *Metropolis-Algorithmus* genannt. Man beachte, dass nur die Quotienten $\pi(j)/\pi(i)$ bekannt sein müssen, die bei geeigneter Wahl von Q meist nicht extrem klein oder extrem groß sind und deren Berechnung dann auch nicht aufwendig ist.

Der Vorteil einer solchen MCMC-Methode ist, dass sie einfach zu definieren und zu implementieren ist. Der Nachteil ist, dass man oft keine guten Kriterien hat, wann man die Iterationen abbrechen sollte. In jedem Fall werden Abschätzungen für die Konvergenzgeschwindigkeit benötigt, und es ist oft nicht leicht, effektive Abschätzungen zu erhalten.

Viele Beispiele für zu simulierende Verteilungen kommen aus der statistischen Physik, wo auf Gitterkonfigurationsräumen der Form $I = \{0,1\}^{\Lambda_n}$, wobei $\Lambda_n = \mathbb{Z}^d \cap [-n,n]^d$ eine große Box ist, gewisse Verteilungen der Form $\pi(i) = \frac{1}{Z_{n,\beta}} e^{-\beta H_n(i)}$ gegeben sind, wobei $Z_{n,\beta}$ die Normierung ist, $\beta > 0$ ein Stärkenparameter und $H_n\colon I \to \mathbb{R}$ eine *Hamiltonsche Abbildung*. Dies ist ein Maß, das Konfigurationen i mit kleinem Wert von $H_n(i)$ bevorzugt, und dieser Effekt ist desto stärker, je größer β ist. Ein Beispiel ist die Anzahl der 0-1-Übergänge $H_n(i) = \sum_{x,y\in\Lambda_n\colon |x-y|=1} |i_x - i_y|$, das ein Stück Eisen modelliert (Ladung in Nachbarpunkten soll möglichst oft gleich gerichtet sein), oder die Anzahl der Nachbarpaare, in denen ein Molekül sitzt, $H_n = \#\{(x,y) \in \Lambda_n^2\colon |x-y| = 1, i_x = i_y = 1\}$. $\diamond$

Kapitel 3

Markovketten in stetiger Zeit

Markovketten in stetiger Zeit sind Prozesse $(X(t))_{t\in[0,\infty)}$ mit höchstens abzählbarem Zustandsraum I, die im Gegensatz zu Markovketten mit diskreter Zeit, die wir im vorigen Kapitel studiert haben, mit $t \in [0,\infty)$ statt $n \in \mathbb{N}_0$ indiziert sind und die sich ebenfalls der Markoveigenschaft erfreuen, d. h. für jedes $n \in \mathbb{N}$ und jede $0 \le t_1 \le t_2 \le \dots \le t_n \le t_{n+1}$ sowie jede $i_1, i_2, \dots, i_n, i_{n+1} \in I$ gilt

$$\mathbb{P}\big(X(t_{n+1}) = i_{n+1} \,\big|\, X(t_1) = i_1, \dots, X(t_n) = i_n\big) = \mathbb{P}\big(X(t_{n+1}) = i_{n+1} \,\big|\, X(t_n) = i_n\big). \qquad (3.0.1)$$

Wir werden im Folgenden kurz *MKSZ* schreiben für Markovketten in stetiger Zeit und *MKDZ* für Markovketten in diskreter Zeit wie im vorigen Kapitel. Wenn also $(X(t))_{t\in[0,\infty)}$ eine MKSZ ist, so ist für jedes $t_0 > 0$ die Kette $(X(nt_0))_{n\in\mathbb{N}_0}$ eine MKDZ mit Übergangsmatrix $(\mathbb{P}(X(t_0)) = j \mid X(0) = i))_{i,j\in I}$. Daher haben MKSZ und MKDZ viele Dinge mit einander gemein, aber wegen der Überabzählbarkeit der Indexmenge $[0,\infty)$ gibt es zumindest immer dort bedeutsame Unterschiede, wo potentiell überabzählbar viele Zeitpunkte involviert sind. Außerdem benötigt man *a priori* sehr viele stochastische Matrizen, um eine MKSZ zu beschreiben, und es ist klar, dass man nach Methoden suchen muss, eine MKSZ auf viel praktikablere Weise zu beschreiben. Dies führt dazu, dass die Theorie der MKSZ recht unterschiedlich von der der MKDZ ist.

3.1 Intuitive Konstruktion

In diesem Abschnitt skizzieren wir die Konstruktion einer MKSZ in einem wichtigen Spezialfall auf natürliche Weise, um eine intuitive Vorstellung von MKSZ zu erlangen. Wir werden hier recht oberflächlich bleiben; präzise Definitionen, Annahmen, Aussagen und Beweise folgen im Abschnitt 3.3.

Zunächst sieht man, dass eine MKSZ $(X(t))_{t\in[0,\infty)}$ zu gewissen *Sprungzeiten* $0 < T_1 < T_2 < \dots$ Sprünge vom aktuellen Zustand zu einem anderen Zustand ausführt, wobei wir "Sprünge" zum aktuellen Zustand ignorieren. Es sei also $T_0 = 0$ und $T_n = \inf\{t > T_{n-1} \colon X(t) \ne X(T_{n-1})\}$ die n-te Sprungzeit. Der Einfachheit halber setzen wir hier voraus, dass $T_n > T_{n-1}$ für alle $n \in \mathbb{N}$ mit Wahrscheinlichkeit Eins gilt. Außerdem setzen wir voraus, dass die Pfade $t \mapsto X(t)$ mit Wahrscheinlichkeit Eins rechtsstetig sind und überall linksseitige Grenzwerte besitzen. Dies garantiert, dass $X(T_n) = X(T_n+) \ne X(T_n-)$ ist, wobei wir $X(T_n+) = \lim_{t\downarrow 0} X(T_n + t)$ und $X(T_n-) = \lim_{t\uparrow 0} X(T_n + t)$ setzen.

Wir bezeichnen mit $\tau_n = T_n - T_{n-1}$ die Wartezeit von $(n-1)$-ten bis zum n-ten Sprung. Dann folgt aus einer etwas sorglosen Anwendung der Markoveigenschaft in (3.0.1), dass gilt:

$$\begin{aligned}\mathbb{P}(\tau_1 > s+t \mid \tau_1 > t, X(0)=i) &= \mathbb{P}\big(X(r)=i\,\forall r\in[0,t+s] \bigm| X(r)=i\,\forall r\in[0,t]\big)\\ &= \mathbb{P}\big(X(r)=i\,\forall r\in[t,t+s] \bigm| X(t)=i\big)\\ &= \mathbb{P}(\tau_1 > s \mid X(0)=i), \qquad i\in I, s,t>0.\end{aligned} \tag{3.1.1}$$

Also sollten die Wartezeiten zwischen den Sprüngen gedächtnislos sein. Als eine beliebte Übungsaufgabe zeigt man, dass dann diese Wartezeiten exponentialverteilt sein müssen. Außerdem sollte die Wahrscheinlichkeit, dann in einen bestimmten anderen Zustand zu springen, unabhängig von allem Vorherigen sein.

Dies legt die folgende Konstruktion einer MKSZ $X = (X(t))_{t\in[0,\infty)}$ nahe. Wir nehmen an, dass es für jedes $i \in I$ einen Vorrat an zum Parameter $c_i > 0$ exponentiell verteilten Zufallsgrößen $\xi_i^{(n)}$, $n \in \mathbb{N}$, gibt. Die Variable $\xi_i^{(n)}$ soll die Wartezeit sein, die die Kette beim n-ten Besuch in i verweilt, bevor sie wieder von i weg springt. Die Kette verbringt also bei jedem Besuch durchschnittlich $1/c_i$ Zeiteinheiten in i. Die Gesamtheit aller dieser Wartezeiten sei unabhängig. Ferner gebe es eine von all diesen Wartezeiten unabhängige Markovkette $Y = (Y_n)_{n\in\mathbb{N}_0}$ in diskreter Zeit. Wir setzen voraus, dass $Y_n \neq Y_{n-1}$ für jedes $n \in \mathbb{N}$. Wir starten mit $X_0 = Y_0$. Nach Ablauf der Wartezeit $\xi_{Y_0}^{(1)}$ springt die Kette X in den Zustand Y_1. Nun beginnt die Wartezeit $\xi_{Y_1}^{(1)}$ zu laufen, und nach ihrem Ablauf springt X in den Zustand Y_2. Wenn $Y_2 = Y_0$, so beginnt die Wartezeit $\xi_{Y_0}^{(2)}$ zu laufen, sonst die Wartezeit $\xi_{Y_2}^{(1)}$. Nach ihrem Ablauf springt X in den Zustand Y_3 und so weiter. Somit ist der Übergangsmechanismus der MKSZ X offenbar vollständig beschrieben. Zusammen mit einer Startverteilung ν auf I ist damit ein stochastischer Prozess definiert. Die MKDZ $(Y_n)_{n\in\mathbb{N}_0}$ heißt die in $(X(t))_{t\in[0,\infty)}$ eingebettete Kette. Es gilt dann nach Konstruktion $Y_n = X(T_n)$ für jedes $n \in \mathbb{N}_0$.

Die Tatsache, dass diese Konstruktion wirklich eine MKSZ definiert, also dass insbesondere die Markoveigenschaft in (3.0.1) erfüllt ist, kann man durch eine nicht wirklich schwierige, aber sehr umständliche und lästige Rechnung verifizieren, siehe etwa [Ge02].

Mit $P = (p_{i,j})_{i,j\in I}$ bezeichnen wir die Übergangsmatrix der MKDZ Y. Nach Voraussetzung gilt $p_{i,i} = 0$ für jedes $i \in I$. Man definiert

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

und bezeichnet $Q = (q_{i,j})_{i,j\in I}$ als die *Q-Matrix* der MKSZ. Die Zeilensummen von Q sind alle gleich Null. 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. Wir erwähnen (siehe Satz 3.3.7), dass

$$q_{i,j} = \lim_{t\downarrow 0} \frac{\mathbb{P}_i(X(t)=j)}{t}, \qquad i,j\in I, i\neq j, \tag{3.1.3}$$

gilt. Die $q_{i,j}$ heißen daher auch *Übergangsraten.*

Im Allgemeinen kann man den Fall $c_i = 0$ für ein $i \in I$ erlauben. In diesem Fall ist i absorbierend, und von i aus werden keinerlei Sprünge mehr ausgeführt. Dagegen ignorieren wir zunächst den möglichen Fall $c_i = \infty$, in welchem man sofort springt. Es kann allerdings passieren, dass der Prozess "steckenbleibt" in dem Sinne, dass der *Explosionszeitpunkt*

$$\zeta = \sum_{n=1}^{\infty} \tau_n = \lim_{n\to\infty} T_n \tag{3.1.4}$$

mit positiver Wahrscheinlichkeit endlich ist. Hier stellt sich die Frage, ob man den Prozess auch noch nach dem Zeitpunkt ζ definieren kann oder will; die bisherige Konstruktion reicht nur bis ζ. Hier ist ein Kriterium für die Endlichkeit des Explosionszeitpunktes:

Lemma 3.1.1. *Sei $i \in I$ fest. Dann gilt genau dann $\mathbb{P}_i(\zeta < \infty) = 1$, wenn $\mathbb{P}_i(\sum_{n\in\mathbb{N}} 1/c_{Y_n} < \infty) = 1$ gilt. Falls $\mathbb{P}_i(\sum_{n\in\mathbb{N}} 1/c_{Y_n} = \infty) = 1$, so gilt $\mathbb{P}_i(\zeta = \infty) = 1$.*

Beweis. Es sei $\widehat{P}_i$ die Verteilung der MKDZ $Y = (Y_n)_{n\in\mathbb{N}_0}$ bei Start in i, dann ergibt der Satz von der Totalen Wahrscheinlichkeit:

$$\mathbb{P}_i\Big(\sum_{n\in\mathbb{N}} \frac{1}{c_{Y_n}} < \infty\Big) = \int_{I^{\mathbb{N}_0}} \mathbb{P}_i\Big(\sum_{n\in\mathbb{N}} \frac{1}{c_{Y_n}} < \infty \,\Big|\, Y = y\Big)\, \widehat{P}_i(\mathrm{d}y).$$

Die Aussage des Lemmas lautet also:

$$\mathbb{P}_i(\zeta < \infty) = 1 \qquad \Longleftrightarrow \qquad \sum_{n\in\mathbb{N}} \frac{1}{c_{y_n}} < \infty \quad \text{für } \widehat{P}_i\text{-fast alle } y \in I^{\mathbb{N}_0}.$$

Auf dem Ereignis $\{Y = y\}$ gilt:

$$\mathbb{E}_i[\zeta] = \sum_{n=1}^{\infty} \mathbb{E}_i[\tau_n] = \sum_{n=1}^{\infty} \frac{1}{c_{y_n}},$$

also ist '$\Longleftarrow$' klar. Um die umgekehrte Richtung zu zeigen, betrachte man, wiederum auf dem Ereignis $\{Y = y\}$,

$$\mathbb{E}_i[\mathrm{e}^{-\zeta}] = \prod_{n=1}^{\infty} \mathbb{E}_i[\mathrm{e}^{-\tau_n}] = \prod_{n=1}^{\infty} \frac{c_{y_n}}{1 + c_{y_n}} = \exp\Big\{\sum_{n=1}^{\infty} \log\Big(1 - \frac{1}{1 + c_{y_n}}\Big)\Big\}.$$

Eine elementare Übungsaufgabe zeigt, dass die rechte Seite gleich Null ist, wenn die Reihe der $1/c_{y_n}$ divergiert. Dies zeigt die umgekehrte Richtung und beendet den Beweis des Lemmas. □

Wir erläutern nun eine *alternative Konstruktion* einer MKSZ $(X(t))_{t\geq 0}$ mit Hilfe von Poissonprozessen. Zunächst rufen wir in Erinnerung, was das ist:

Bemerkung 3.1.2 (Poissonprozess). Ein *Poisson'scher Punktprozess* oder kurz ein *Poissonprozess* mit Parameter $\lambda \in (0, \infty)$ ist ein $\mathbb{N}_0$-wertiger stochastischer Prozess $(N_t)_{t\in[0,\infty)}$ mit Start in $N_0 = 0$, der jeweils nach unabhängigen, zu λ exponentialverteilten Zeiten einen Sprung um Eins nach oben macht. Wenn $0 < T_1 < T_2 < \dots$ die Sprungzeiten sind, so ist also $N_t = \max\{n \in \mathbb{N} \colon T_n \leq t\}$. Eine alternative Beschreibung ist, dass $N_t = N_{(0,t]}$ gilt, wobei $(N_{(a,b]})_{0\leq a<b<\infty}$ eine Familie von Zufallsvariablen ist mit der Eigenschaft, dass für jedes $n \in \mathbb{N}$ und jede $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_n < b_n < \infty$ die Größen $N_{(a_1,b_1]}, \dots, N_{(a_n,b_n]}$ unabhängige, zu den Parametern $\lambda(b_1 - a_1), \dots, \lambda(b_n - a_n)$ Poisson-verteilte Zufallsgrößen sind. Insbesondere ist $\mathbb{P}(N_t - N_s = n) = \mathrm{e}^{-\lambda(t-s)}[\lambda(t - s)]^n/n!$ für jedes $n \in \mathbb{N}_0$ und $0 \leq s < t$.

Eine wichtige Eigenschaft von Poissonprozessen ist, dass eine Überlagerung (oder Vereinigung) zweier unabhängiger Poissonprozesse wieder einer ist, und dass sich die Parameter addieren. Genauer gesagt, wenn $0 < T_1 < T_2 < \dots$ bzw. $0 < S_1 < S_2 < \dots$ die Sprungzeiten der beiden Prozesse mit Parametern λ bzw. μ sind und $\{U_n : n \in \mathbb{N}\} = \{T_n : n \in \mathbb{N}\} \cup \{S_n : n \in \mathbb{N}\}$ mit $0 < U_1 < U_2 < \dots$ die Vereinigung aller Sprungzeiten ist, so definiert $N_t = \max\{n \in \mathbb{N} : U_n \leq t\}$ einen Poissonprozess mit Parameter $\lambda + \mu$. Es ist leicht zu sehen, dass $U_1 = \min\{T_1, S_1\}$ eine zum Parameter $\lambda + \mu$ exponentialverteilte Zufallsgröße ist, denn für jedes $t > 0$ gilt

$$\mathbb{P}(U_1 > t) = \mathbb{P}(T_1 > t, S_1 > t) = \mathbb{P}(T_1 > t)\mathbb{P}(S_1 > t) = \mathrm{e}^{-\lambda t}\mathrm{e}^{-\mu t} = \mathrm{e}^{-(\lambda+\mu)t}.$$

◇

Mit Hilfe eines Poissonprozesses $(N_t)_{t\in[0,\infty)}$ und einer MKDZ $(Y(n))_{n\in\mathbb{N}_0}$ kann man leicht eine spezielle MKSZ $(X(t))_{t\in[0,\infty)}$ konstruieren, indem man $X(t) = Y(N_t)$ setzt. Dann benutzt man den Poissonprozess als eine zufällige Uhr. Dann ist $(Y(n))_{n\in\mathbb{N}_0}$ die in $(X(t))_{t\in[0,\infty)}$ eingebettete MKDZ. (Wir haben übrigens wiederum vorausgesetzt, dass $Y_n \neq Y_{n-1}$ für jedes $n \in \mathbb{N}$ gilt.) Die Wartezeiten dieser MKSZ haben alle den selben Parameter λ.

Die angekündigte alternative Konstruktion einer beliebigen MKSZ auf einem Zustandsraum I (den wir hier zur Vereinfachung als endlich voraussetzen wollen) geht nun wie folgt. Für jedes Paar (i, j) von verschiedenen Zuständen sei ein Poissonprozess $(N_{i,j}(t))_{t\in[0,\infty)}$ mit Parameter $\lambda_{i,j} \in (0, \infty)$ gegeben. Die Familie dieser Poissonprozesse wird als unabhängig vorausgesetzt. Zu einem Zeitpunkt, in dem die MKSZ in den Zustand i springt, beginnen alle zufälligen Uhren $N_{i,j}$ mit $j \in I \setminus \{i\}$ zu ticken. Sobald die erste dieser Uhren klingelt, springt die Markovkette in den zugehörigen Zustand. Hier wiederholt sich die Sache mit diesem Zustand an Stelle von i. Zuvor verwendete Uhren werden nicht mehr benutzt, d. h. wenn der Prozess einen Zustand i zum l-tem Male besucht, laufen die Poissonprozesse $N_{i,j}$ ab ihrer l-ten Sprungzeit.

Die Idee bei dieser Konstruktion ist, dass wir wieder einen Poissonprozess erhalten, wenn wir die Gesamtheit der Poissonprozesse $N_{i,j}$ mit $j \in I \setminus \{i\}$ zu einem zusammenfassen (siehe Bemerkung 3.1.2), und der Parameter dieses Prozesses ist $\lambda^{(i)} = \sum_{j\neq i} \lambda_{i,j}$. Insbesondere ist das erste Klingeln dieser Uhren exponentialverteilt mit diesem Parameter, und ferner ist die Wahrscheinlichkeit, dass es die j-te Uhr ist, gerade gleich $\lambda_{i,j}/\lambda^{(i)}$. Die stochastische Matrix $P = (p_{i,j})_{i,j\in I} = (\lambda_{i,j}/\lambda^{(i)})_{i,j\in I}$ hat Nullen auf der Diagonalen. Die Q-Matrix hat also die Darstellung $q_{i,j} = \lambda^{(i)} p_{i,j} = \lambda_{i,j}$ für $i \neq j$, und es gilt $c_i = \lambda^{(i)}$ in obiger Notation. Ein großer Vorteil dieser Konstruktionsvariante ist, dass die Übergangsmatrix P nicht explizit spezifiziert werden muss, sondern dass sie sich durch einen relativen Vergleich der Raten ergibt.

3.2 Beispiele

Beispiel 3.2.1 (Geburts- und Todesprozesse). Geburts- und Todesprozesse nennt man solche MKSZ, deren Zustandsraum $I = \mathbb{N}_0$ ist und für die die Q-Matrix die Form

$$q_{i,j} = \begin{cases} \lambda_i, & \text{falls } j = i + 1, \\ \mu_i, & \text{falls } j = i - 1, i \geq 1, \\ 0 & \text{sonst, wenn } i \neq j, \end{cases} \tag{3.2.1}$$

besitzt. Dies ist ein Prozess auf $\mathbb{N}_0$, der mit Rate λ_i um 1 größer wird und mit Rate μ_i um 1 kleiner, wenn er gerade in i ist. Die λ_i werden als *Geburtsraten*, die μ_i als *Sterberaten* bezeichnet.

Sind alle $\mu_i = 0$ (bzw. alle $\lambda_i = 0$), dann heisst die MKSZ (reiner) Geburtsprozess (bzw. (reiner) Todesprozess). Der Poissonprozess ist ein reiner Geburtsprozess mit Geburtsrate λ.

Geburts- und Todesprozesse sind Nächstnachbarschaftsirrfahrten auf $\mathbb{N}_0$ in stetiger Zeit und werden zum Beispiel als (sehr einfache) Modelle zur Beschreibung der zeitlichen Evolution der Größe einer Population verwendet. ◇

Beispiel 3.2.2 (Einfache Irrfahrt auf $\mathbb{Z}^d$). Das Standardbeispiel einer Irrfahrt auf $\mathbb{Z}^d$ ist die sogenannte *einfache Irrfahrt*, die im Ursprung startet und jeweils nach unabhängigen, zum Parameter Eins exponentialverteilten Wartezeiten mit gleicher Wahrscheinlichkeit zum nächsten Nachbrn springt. Ihre Q-Matrix ist gegeben als $q_{i,j} = \frac{1}{2d}$ für alle benachbarten Paare $i, j \in I$ und $q_{i,i} = -1$. ◇

Beispiel 3.2.3 (M/M/c-Warteschlange). Wir betrachten eine Warteschlange mit exponentiell verteilten Zwischenankunftszeiten der Kunden (dafür steht das erste 'M'; es erinnert an 'Markov'), exponentiell verteilten Bedienzeiten (dafür steht das zweite 'M') und $c \in \mathbb{N}$ Bedienern (dafür steht das 'c'). Wir setzen voraus, dass die Gesamtheit aller Zwischenankunftszeiten und Bedienzeiten unabhängig sind und dass die Zwischenankunftszeiten und die Bedienzeiten jeweils die gleiche Verteilung haben. Die Schlange funktioniert nach dem Prinzip *FIFO*, also 'first in – first out'. Wenn andere Verteilungen für die Zwischenankunftszeiten bzw. die Bedienzeiten gewählt werden als eine exponentielle, spezifiziert man dies im ersten bzw. zweiten Argument. Allerdings verliert man dabei im Allgemeinen die Markoveigenschaft.

Wir betrachten die Anzahl $X(t)$ der Kunden in der Warteschlange zum Zeitpunkt $t \in [0, \infty)$. Da wir die Exponentialverteilung zugrunde legen, ist $(X(t))_{t\in[0,\infty)}$ eine MKSZ mit Zustandsraum $\mathbb{N}_0$. Seien $\lambda > 0$ bzw. $\mu > 0$ die Parameter der Exponentialverteilungen der Zwischenankunfts- bzw. Bedienzeiten (für jeden der c Bediener). Es ist nicht mit positiver Wahrscheinlichkeit möglich, dass zu einem Zeitpunkt zwei Kunden gleichzeitig eintreffen oder abgefertigt werden. Also kann die MKSZ $(X(t))_{t\in[0,\infty)}$ Sprünge immer nur zu benachbarten Zuständen ausführen, ist also ein Geburts- und Todesprozess. Die Geburts- und Sterberaten sind gegeben durch

$$\lambda_i = \lambda \quad \text{für alle } i \in \mathbb{N}_0, \qquad \mu_i = \begin{cases} i\mu & \text{falls } 0 \leq i \leq c, \\ c\mu & \text{falls } i \geq c. \end{cases} \tag{3.2.2}$$

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

Beispiel 3.2.4 (Galton-Watson-Prozess). Wir betrachten einen Verzweigungsprozess in stetiger Zeit, also das kontinuierliche Gegenstück zu dem Prozess in Beispiel 2.2.10. Jeweils nach unabhängigen, zum Parameter Eins exponentialverteilten Wartezeiten stirbt ein Partikel und wird durch eine zufällige Anzahl von Partikeln ersetzt. Die Wahrscheinlichkeit, dass es genau j Nachkommen hat, sei $p_j \in [0, 1]$, wobei $\sum_{j\in\mathbb{N}_0} p_j = 1$ sei. Die Wahl der Anzahl der Nachkommen ist individuell pro Partikel und unabhängig. Es sei $X(t)$ die Anzahl der Partikel zum Zeitpunkt $t \in [0, \infty)$. Wir setzen voraus, dass $p_1 = 0$ ist, um zu vermeiden, dass die Ersetzung eines Partikels durch genau eines als ein Sprung der MKSZ angesehen wird. Dann ist $(X(t))_{t\in[0,\infty)}$ eine MKSZ mit Q-Matrix

$$q_{i,j} = \begin{cases} -i, & \text{wenn } i = j, \\ ip_{j-i+1}, & \text{wenn } j \geq i + 1 \text{ oder } j = i - 1, \\ 0 & \text{sonst.} \end{cases}$$

Insbesondere ist 0 ein absorbierender Zustand der MKSZ.

Wir diskutieren kurz die Endlichkeit des Explosionszeitpunkts ζ. Wir setzen dabei voraus, dass die erwartete Nachkommenschaft, $m = \sum_{j\in\mathbb{N}_0} jp_j$, endlich ist. Die eingebettete MKDZ kann rekursiv beschrieben werden durch

$$Y_n = Y_{n-1} + (W_n - 1)1\!\!1_{\{Y_{n-1}>0\}}, \qquad n \in \mathbb{N},$$

wobei $(W_n)_{n\in\mathbb{N}}$ eine u. i. v. Folge von Zufallsgrößen mit Verteilung $(p_j)_{j\in\mathbb{N}_0}$ ist. Dann ist

$$0 \le Y_n = Y_0 + \sum_{k=1}^n (Y_k - Y_{k-1}) = Y_0 + \sum_{k=1}^n (W_k - 1)1\!\!1_{\{Y_{k-1}>0\}} \le Y_0 + 1\!\!1_{\{Y_{n-1}>0\}} \sum_{k=1}^n (W_k - 1).$$

Nach dem Gesetz der Großen Zahlen haben wir also $\limsup_{n\to\infty} Y_n/n \le \max\{0, m-1\}$. Daher ist für jedes $i \in I$

$$\mathbb{P}_i\Big(\sum_{n\in\mathbb{N}} \frac{1}{c_{Y_n}} = \infty\Big) = \mathbb{P}_i\Big(\sum_{n\in\mathbb{N}} \frac{1}{Y_n} = \infty\Big) = \mathbb{P}_i\Big(\sum_{n\in\mathbb{N}} \frac{1}{n}\Big(\frac{Y_n}{n}\Big)^{-1} = \infty\Big) = 1.$$

Nach Lemma 3.1.1 ist der Explosionszeitpunkt ζ unendlich, $\mathbb{P}_i$-fast sicher für jedes $i \in I$. ◇

Beispiel 3.2.5 (Stochastische Vielteilchensysteme). Wenn man ein stochastisches System vieler Teilchen im Raum modellieren möchte, dann bietet sich der Zustandsraum $I = \Lambda^S$ an, wobei $\Lambda \subset \mathbb{Z}^d$ eine große Box ist (etwa $\Lambda = [-N, N]^d \cap \mathbb{Z}^d$) und S eine endliche Menge von Eigenschaften, die ein Teilchen haben kann. Wir betrachten dann eine Konfiguration von genau einem Teilchen in jedem Punkt der Box Λ, das jeweils eine Eigenschaft $s \in S$ besitzt, etwa einen Spin, eine Farbe, eine Meinung, ein Geschlecht oder andere Merkmale, je nach Anwendung. Jedes der Teilchen wechselt individuell seine Eigenschaft zufällig zu gewissen zufälligen Zeiten. Um die Markoveigenschaft des gesamten Teilchensystems zu erhalten, nehmen wir natürlich an, dass die Wartezeiten exponentiell verteilt sind. Der Eigenschaftswechsel kann abhängig sein von den Eigenschaften der anderen Teilchen, und ein interessanter Fall besteht darin, dass dieser Wechsel nur von den Eigenschaften der Nachbarn abhängt.

Eine andere Möglichkeit ist, eine beliebig große Anzahl von Teilchen in der Box Λ anzunehmen und nur zu registrieren, wie viele Teilchen mit einer gegebenen Eigenschaft in einem gegebenen Gitterpunkt sitzen. (Hierfür ist ein größerer Zustandsraum als Λ^S notwendig.) Zusätzlich kann man eine zufällige individuelle Migration jedes einzelnen Teilchens zulassen. Dies wäre ein räumlich dynamischer Verzweigungsprozess.

Die Konstruktion eines solchen Vielteilchensystems kann im Prinzip ohne große Probleme wie in Abschnitt 3.1 durchgeführt werden. Allerdings ist man hauptsächlich interessiert an räumlich *unendlichen* Systemen mit eventuell unendlich vielen Teilchen. Hier muss man zusätzliche Bedingungen an die Übergangsraten stellen, und die Konstruktion wird technisch aufwendig. ◇

Beispiel 3.2.6 (M/E_2/1-Warteschlange). Eine M/E_k/1-Warteschlange besitzt einen Schalter, exponentiell verteilte Zwischenankunftszeiten der Kunden, und Bedienzeiten, die die sogenannte *Erlang*-Verteilung E_k besitzen, also die Verteilung einer Summe von k unabhängigen exponentialverteilten Zufallsgrößen mit dem selben Parameter, sagen wir $\mu \in (0, \infty)$. Ansonsten gelten die in Beispiel 3.2.3 beschriebenen Voraussetzungen.

Die Zahl der Kunden in einer M/E_2/1-Warteschlange ist keine MKSZ, da die E_2-Verteilung nicht gedächtnislos ist. Wegen der besonderen Eigenschaft der Erlang-Verteilung lässt sich aber eine MKSZ mit größerem Zustandsraum so definieren, dass die Zahl der Kunden in einer M/E_2/1-Warteschlange eine Funktion jener MKSZ ist.

Der "Trick" besteht darin, dass man die E_2-verteilte Bedienzeit künstlich in zwei Phasen zerlegt, die jeweils exponentiell mit Parameter 2μ 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

$$E = \{0\} \cup \big(\mathbb{N} \times \{1, 2\}\big), \tag{3.2.3}$$

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 Bedienphase $i \in \{1, 2\}$ befindet. Übergänge von Phase 1 zur Phase 2 finden immer mit Rate 2μ statt. Mit derselben Rate 2μ endet die Bedienzeit 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 λ. ◇

Beispiel 3.2.7 (M/G/1-Warteschlange). Sei $X(t)$ die Anzahl der Kunden im System zur Zeit $t \geq 0$ bei einer M/G/1-Warteschlange, siehe auch Beispiel 3.2.3. Hierbei steht 'M' dafür, dass die Zwischenankunftszeiten der Kunden exponentiell verteilt sind, aber das 'G' steht dafür, dass die Verteilung der Bedienzeiten nicht näher spezifiziert wird ('general'). Der stochastische Prozess $(X(t))_{t\geq 0}$ besitzt also in der Regel nicht die Markoveigenschaft. Wenn aber $0 < S_1 < S_2 < \dots$ die (zufälligen) Zeitpunkte sind, an denen die Bedienzeit eines Kunden endet, dann definiert $Y_n := X(S_n+)$ (wobei $S_0 = 0$) eine MKDZ $(Y_n)_{n\in\mathbb{N}_0}$ auf $I = \mathbb{N}_0$, da die Zwischenankunftszeiten der Kunden gedächtnislos sind. Mit anderen Worten, $(Y_n)_n$ ist eine in $(X(t))_{t\geq 0}$ eingebettete Markovkette.

Wir werden nun die Übergangswahrscheinlichkeiten $p_{i,j}$ dieser MKDZ berechnen und sie auf Rekurrenz untersuchen. Sei G die Verteilung der Bedienzeit. Wir nehmen an, dass $G((0, \infty)) = 1$ ist (d. h. die Bedienzeiten positiv sind mit Wahrscheinlichkeit Eins) und der Erwartungswert ν der Bedienzeit endlich ist. Den Parameter der exponentialverteilten Zwischenankunftszeiten nennen wir $\lambda \in (0, \infty)$. Wir berechnen nun $p_{i,j} = \mathbb{P}(Y_1 = j \mid Y_0 = i)$. Es ist klar, dass $p_{i,j} = 0$, falls $j < i - 1$, denn es kann in einer Bedienzeit nur eine Person bedient werden (aber beliebig viele neu in der Schlange erscheinen). Sei zunächst $i \geq 1$ und K die Zahl der Kunden, die während der Bedienzeit B des ersten Kunden eintreffen. Dann gilt nach der Formel von der Totalen Wahrscheinlichkeit

$$\begin{aligned} p_{i,j} &= \mathbb{P}(Y_1 = j \mid Y_0 = i) = \mathbb{P}(K = j - i + 1 \mid Y_0 = i) \\ &= \int_0^\infty \mathbb{P}_i(K = j - i + 1 \mid B = x)\, G(\mathrm{d}x). \end{aligned} \tag{3.2.4}$$

Da genau ein Kunde während der Bedienzeit die Schlange verlässt, ist die bedingte Verteilung von $K + 1$ (der Zahl von neu erscheinenden Kunden) gegeben $\{B = x\}$ gleich der Poisson-Verteilung mit Parameter λx, siehe Bemerkung 3.1.2. Somit gilt für $i \in \mathbb{N}$ und $j - i + 1 \geq 0$:

$$p_{i,j} = c_{j-i+1}, \qquad \text{wobei } c_r = \mathbb{P}(K = r) = \int_0^\infty \frac{(\lambda x)^r}{r!} \mathrm{e}^{-\lambda x}\, G(\mathrm{d}x), \qquad r \in \mathbb{N}_0. \tag{3.2.5}$$

Weiter gilt $p_{0,j} = \mathbb{P}(K = j) = p_{1,j}$ für jedes $j \in \mathbb{N}_0$, denn wenn am Beginn niemand im System ist, beginnt die erste Bedienzeit erst beim Eintreffen des ersten Kunden. Also folgt

$$P = (p_{ij})_{i,j\in\mathbb{N}_0} = \begin{pmatrix} c_0 & c_1 & c_2 & c_3 & c_4 & \dots \\ c_0 & c_1 & c_2 & c_3 & c_4 & \dots \\ 0 & c_0 & c_1 & c_2 & c_3 & \dots \\ 0 & 0 & c_0 & c_1 & c_2 & \dots \\ 0 & 0 & 0 & c_0 & c_1 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{3.2.6}$$

Offenbar sind alle c_r positiv. Daher ist die Markovkette irreduzibel und aperiodisch.

Wir berechnen den Erwartungswert von K:

$$\mathbb{E}K = \sum_{i=0}^\infty ic_i = \sum_{i=0}^\infty i\int_0^\infty \frac{(\lambda x)^i}{i!}\mathrm{e}^{-\lambda x}\,G(\mathrm{d}x) = \int_0^\infty \lambda x\,G(\mathrm{d}x) = \lambda\nu =: \rho. \tag{3.2.7}$$

Die Zahl ρ heißt *Verkehrsdichte* der Warteschlange. Große Werte von ρ entstehen durch große λ, d. h. durch kurze mittlere Kundenabstände $1/\lambda$ oder durch langsame Bedienung d. h. große ν. Es ist daher zu erwarten, dass große ρ 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 Bedienzeit ν gleich der erwarteten Zwischenankunftszeit λ^{-1} ist.

Um dies alles zu zeigen, lösen wir das lineare Gleichungssystem $\pi = \pi P$. Zur Abkürzung definieren wir

$$\overline{c}_i = \sum_{j=i+1}^\infty c_j = \mathbb{P}(K \geq i+1), \qquad i \in \mathbb{N}_0, \tag{3.2.8}$$

dann gilt auch

$$\sum_{i=0}^\infty \overline{c}_i = \sum_{i=0}^\infty \mathbb{P}(K \geq i+1) = \mathbb{E}K = \rho. \tag{3.2.9}$$

Schreibt man die einzelnen Gleichungen des Systems $\pi = \pi P$ untereinander und addiert die ersten k für alle $k \in \mathbb{N}$, dann erhält man das äquivalente Gleichungssystem

$$\begin{aligned} \pi_1 c_0 &= \pi_0\overline{c}_0 \\ \pi_2 c_0 &= \pi_0\overline{c}_1 + \pi_1\overline{c}_1 \\ \pi_3 c_0 &= \pi_0\overline{c}_2 + \pi_1\overline{c}_2 + \pi_2\overline{c}_1 \\ \vdots \quad & \quad \vdots \end{aligned} \tag{3.2.10}$$

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

$$(1-\pi_0)c_0 = \pi_0\rho + \sum_{i=1}^\infty \pi_i \sum_{j=1}^\infty \overline{c}_j = \pi_0\rho + (1-\pi_0)(\rho - \overline{c}_0). \tag{3.2.11}$$

Auflösen nach π_0 ergibt $\pi_0 = c_0 - \rho + \overline{c}_0 = 1 - \rho$. Durch sukzessives Einsetzen in (3.2.10) erhält man rekursiv alle π_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 π_i aus (3.2.10) rekursiv, dann sind offenbar alle π_i positiv. Da (3.2.10) 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 (3.2.10), dann erhält man

$$\begin{aligned} s_k c_0 &= (1-\rho)\sum_{j=0}^{k-1} \overline{c}_j + \sum_{i=1}^{k-1} \pi_i \sum_{j=1}^{k-i} \overline{c}_j \\ &\le (1-\rho)\rho + s_{k-1}(\rho - \overline{c}_0) \\ &= (1-\rho)\rho + s_{k-1}(c_0 - (1-\rho)). \end{aligned} \tag{3.2.12}$$

Hieraus folgt

$$0 < \pi_k = s_k - s_{k-1} \le \frac{1-\rho}{c_0}(\rho - s_{k-1}), \tag{3.2.13}$$

woraus $s_{k-1} \le \rho$ für alle k und somit die Endlichkeit von $\sum_{i=0}^{\infty} \pi_i$ folgt. Falls $\sum_{i=0}^{\infty} \pi_i \ne 1$ ist, normiert man die π_i so, dass die Summe 1 ist, und sieht wie vorher, dass das normierte π_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.

Für weitere Diskussionen hierzu verweisen wir auf [HS82, S. 251] und [KT81, S. 502 ff]. $\diamond$

3.3 Markovketten und ihre Q-Matrix

Nun zurück zur Theorie. Gegeben sei als Zustandsraum eine abzählbare (nichtleere) Menge I. In diesem Abschnitt führen wir das Konzept einer MKSZ ein und viele damit verbundene Objekte wie Familien von Übergangsmatrizen und die schon erwähnte Q-Matrix.

Zunächst führen wir das Analogon der Übergangsmatrix P einer MKDZ einführen. Während im zeitlich diskreten Fall durch eine Verteilung ν und eine stochastische Matrix P die Verteilung der MKDZ mit Startverteilung ν und Übergangsmatrix P festgelegt ist, ist nicht klar, ob dies im stetigen Fall mit ν und der Matrix $P(1) := (\mathbb{P}_i(X(1) = j))_{i,j\in I}$ 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 gleich eine ganze geeignete Familie $(P(t))_{t\ge 0}$ von Übergangsmatrizen. Um flexibler zu sein, müssen die Zeilensummen nicht unbedingt gleich Eins sein.

Definition 3.3.1 ((Sub-)Markov'sche Halbgruppe). *Sei I eine abzählbare, nichtleere Menge. Eine Familie $(P(t))_{t>0}$ von Matrizen $(P_{i,j}(t))_{i,j\in I}$ heißt* sub-Markov'sche Halbgruppe *auf I, wenn für alle $i, j \in I$ und alle $t, s > 0$ gilt:*

(i) $P_{i,j}(t) \ge 0$,

(ii) $\sum_{k\in I} P_{i,k}(t) \le 1$,

(iii) $P_{i,j}(t+s) = \sum_{k\in I} P_{i,k}(t) P_{k,j}(s)$ (Chapman-Kolmogorov-Gleichung)

Die Familie $(P(t))_{t>0}$ heißt Markov'sch*, wenn zusätzlich in (ii) das Gleichheitszeichen für alle $i \in I$ 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 $i, j \in I$.*

Eine Markov'sche Halbgruppe besteht also aus stochastischen Matrizen. Wir werden später sehen, dass die Bedingung (iv) schon gilt, wenn sie nur für alle $i = j$ gefordert wird.

Nun bringen wir die Definition einer MKSZ. Wir wollen die Möglichkeit einer Explosion in endlicher Zeit (siehe Abschnitt 3.1) nicht *a priori* ausschließen. Im Falle einer Explosion gibt es vielfältige Möglichkeiten, den Prozess als MKSZ weiterlaufen 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 verharrt. Dies erreicht man dadurch, dass man einen zusätzlichen Zustand ∂ (*Grab*, *Sarg*) zu I hinzufügt und vereinbart, dass sich der Prozess nach der Explosion für immer in ∂ befindet. Dies war der Grund, dass wir in der Definition einer sub-Markov'schen Halbgruppe zugelassen haben, dass die Zeilensummen kleiner als Eins sind.

Definition 3.3.2 (Markovkette in stetiger Zeit). *Sei $(P(t))_{t>0}$ eine sub-Markov'sche Halbgruppe auf I. Weiter seien $\partial \notin I$ und ν eine Wahrscheinlichkeitsverteilung auf $I \cup \{\partial\}$. Dann heißt ein $I \cup \{\partial\}$-wertiger stochastischer Prozess $(X(t))_{t\geq 0}$ eine* Markovkette in stetiger Zeit (MKSZ) *mit Startverteilung ν und Halbgruppe $(P(t))_{t>0}$, wenn für alle $n \in \mathbb{N}$, alle $0 \leq t_1 < t_2 < \cdots < t_n < t$ und alle $i_1, \dots, i_n, i \in I$ gelten:*

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

(ii) $\mathbb{P}\big(X(0) = i\big) = \nu_i$ *für alle* $i \in I \cup \{\partial\}$,

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

Bemerkung 3.3.3. Eine sub-Markov'sche Halbgruppe auf I lässt sich zu einer Markov'schen auf $I \cup \{\partial\}$ fortsetzen durch die Definition

$$P_{i,\partial}(t) = \begin{cases} 1, & \text{falls } i = \partial \text{ und } t > 0, \\ 1 - \sum_{j\in I} P_{i,j}(t), & \text{falls } i \in I \text{ und } t > 0. \end{cases} \tag{3.3.1}$$

In diesem Fall gilt die Bedingung (i) aus Definition 3.3.2 automatisch auch für $i_1, \dots, i_n, i \in I \cup \{\partial\}$ (Übungsaufgabe). Die Fortsetzung ist die einzig mögliche genau dann, wenn die Halbgruppe nicht Markov'sch auf I ist (Übungsaufgabe). Es ist klar, dass $(P(t))_{t>0}$ genau dann standard auf I ist, wenn die Fortsetzung auf $I \cup \{\partial\}$ standard ist.

◇

Satz 3.3.4. *Sei ν eine Verteilung auf $I \cup \{\partial\}$, und sei $(P(t))_{t>0}$ sub-Markov'sch. Dann existiert eine MKSZ $\mathbb{X} = (X(t))_{t\geq 0}$ zu ν und $(P(t))_{t>0}$ im Sinne von Definition 3.3.2. Für diese MKSZ gilt für alle $n \in \mathbb{N}_0$, alle $0 = t_0 < t_1 < \cdots < t_n$ und alle $i_0, \dots, i_n \in I$:*

$$\mathbb{P}\big(X(t_0) = i_0, X(t_1) = i_1, \dots, X(t_n) = i_n\big) = \nu_{i_0} P_{i_0,i_1}(t_1 - t_0) \cdots P_{i_{n-1},i_n}(t_n - t_{n-1}). \tag{3.3.2}$$

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

Beweis. Dies folgt mit Hilfe des Satzes 1.3.3 von Kolmogorov. Wir verzichten auf eine genauere Begründung. □

Bemerkung 3.3.5. Wie im Fall einer MKDZ kann man (3.3.2) auch als Definition einer MKSZ zu ν 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 $(X(ns))_{n\in\mathbb{N}_0}$ für festes $s>0$ eine MKDZ mit Startverteilung ν und Übergangsmatrix $P(s)$ ist, wenn $(X(t))_{t\geq 0}$ eine MKSZ zu ν und $(P(t))_{t>0}$ ist. Deutliche Unterschiede gibt es immer dort, wo man überabzählbar viele Zeitindizes gleichzeitig betrachtet, z. B. bei der Betrachtung der Zufallsvariable $\sup_{s\in[0,1]} X(s)$, etwa wenn $I=\mathbb{N}$ ist.

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-)Markov'sche Halbgruppen viel schwieriger. In Abschnitt 3.1 sahen wir bereits, dass die Beschreibung von MKSZ durch die Q-Matrix viel günstiger ist, siehe (3.1.3). Q beschreibt das Verhalten von $P_{ij}(t)$ für infinitesimal kleine $t>0$. Da man die komplette Halbgruppe aus der Kenntnis von $(P_{i,j}(t))_{0<t\leq t_0}$ für alle $i,j\in I$ und festes $t_0>0$ rekonstruieren kann, ist es plausibel, dass dies auch aus der Kenntnis von $\lim_{t\downarrow 0} t^{-1}P_{i,j}(t)=q_{i,j}$ möglich ist. Inwieweit dies stimmt, werden wir später sehen. Zunächst beweisen wir die Existenz der Grenzwerte $q_{i,j}$. 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. Für den Nichtstandardfall konsultiere man [Ch67].

Die Voraussetzung "standard" besagt, dass die Kette nach einer kurzen Zeit mit großer Wahrscheinlichkeit im Startzustand ist. Man beachte jedoch, dass auch für kleine $t>0$ die Zahl $P_{i,i}(t)$ *nicht* gleich der Wahrscheinlichkeit ist, bis zum Zeitpunkt t nicht gesprungen zu sein. Wir zeigen zunächst, dass im Standardfall für jede $i,j\in I$ die Abbildung $t\mapsto P_{i,j}(t)$ gleichmäßig stetig ist. Ohne die Voraussetzung "standard" braucht $t\mapsto P_{i,j}(t)$ nicht einmal messbar zu sein (siehe [Ch67]).

Lemma 3.3.6. *Sei $(P(t))_{t\geq 0}$ standard. Dann gilt für jedes $i\in I$*

$$\lim_{h\downarrow 0}\sup_{t\geq 0}\sum_{j\in I}\big|P_{i,j}(t+h)-P_{i,j}(t)\big|=0. \tag{3.3.3}$$

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

Beweis. Sei $h>0$. Wir schätzen ab:

$$\begin{aligned}\sum_{j\in I}|P_{i,j}(t+h)-P_{i,j}(t)| &= \sum_{j\in I}\Big|\sum_{k\in I}\big(P_{i,k}(h)P_{k,j}(t)-\delta_{i,k}P_{k,j}(t)\big)\Big|\\ &\leq \sum_{j\in I}\sum_{k\in I}|P_{i,k}(h)-\delta_{i,k}|P_{k,j}(t)=\sum_{k\in I}|P_{i,k}(h)-\delta_{i,k}|\sum_{j\in I}P_{k,j}(t)\\ &\leq \sum_{k\in I}|P_{i,k}(h)-\delta_{i,k}|=1-P_{i,i}(h)+\sum_{k\in I\setminus\{i\}}P_{i,k}(h)\leq 2(1-P_{i,i}(h)).\end{aligned} \tag{3.3.4}$$

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

Nun bringen wir endlich den Beweis der Existenz der Grenzwerte in (3.1.3). Die Matrix Q, die im folgenden Satz eingeführt wird, nennt man die *Q-Matrix* der Halbgruppe $(P(t))_{t\geq 0}$.

Satz 3.3.7. *Sei $(P(t))_{t\geq 0}$ standard. Dann existiert für alle $i,j\in I$ der Grenzwert*

$$q_{i,j}=\lim_{h\downarrow 0}\frac{P_{i,j}(h)-\delta_{i,j}}{h}. \tag{3.3.5}$$

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

$$\sum_{j\in I\colon j\neq i} q_{i,j}\leq -q_{i,i}=: q_i. \tag{3.3.6}$$

Beweis. (Siehe z. B. [KT81] oder [GS75]).

1. Teil: $i=j$. Definiere

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

Wir werden zeigen, dass $q_i'=-q_{i,i}=\lim_{h\downarrow 0}h^{-1}(1-P_{i,i}(h))$ gilt. Sei $c<q_i'$, dann kann man $t_0\in(0,\infty)$ so wählen, dass

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

gilt. Nun wählen wir $\tau\in(0,t_0)$ und $n\in\mathbb{N}$ so, dass $n\tau<t_0\leq(n+1)\tau$ gilt, und schätzen ab:

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

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 $a=(P_{i,i}(\tau))^n$ und $b=P_{ii}(t_0-n\tau)$ und beim vierten die Beziehung $1-a^n=(1-a)(1+a+a^2+\cdots+a^{n-1})\leq n(1-a)$ für $a=P_{i,i}(\tau)$ verwendet haben. Wir lassen nun τ gegen Null und damit $n\to\infty$ gehen, und somit konvergiert der letzte Summand wegen der “standard” Eigenschaft gegen Null. Es folgt

$$c<\liminf_{\tau\downarrow 0}\frac{1-P_{i,i}(\tau)}{\tau}\leq q_i'.$$

Da $c<q_i'$ beliebig war, haben wir sogar

$$q_i'=\lim_{\tau\downarrow 0}\frac{1-P_{i,i}(\tau)}{\tau},$$

wie behauptet.

2. Teil: $i\neq j$. Sei $\mathbb{X}=(X(t))_{t\geq 0}$ eine MKSZ mit sub-Markov’scher standard Halbgruppe $(P(t))_{t\geq 0}$ und Start in i. Definiere für $i,j\in I$, $n\in\mathbb{N}$ und $h>0$

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

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 $t \in [0, t_0]$ gelten: $1 - P_{i,i}(t) < \varepsilon$, $1 - P_{j,j}(t) < \varepsilon$ und $P_{i,j}(t) < \varepsilon$. Seien $n \in \mathbb{N}$ und $h > 0$ gegeben mit $nh \leq t_0$, und sei $u \in [nh, t_0]$. Dann haben wir

$$\varepsilon > P_{i,j}(u) \geq \sum_{r=1}^{n} f_{i,j}^{(r)}(h) P_{j,j}(u - rh) \geq (1-\varepsilon) \sum_{r=1}^{n} f_{i,j}^{(r)}(h), \qquad \text{also gilt } \sum_{r=1}^{n} f_{i,j}^{(r)}(h) \leq \frac{\varepsilon}{1-\varepsilon}. \tag{3.3.7}$$

Also folgt für $r = 1, \dots, n$:

$$\begin{aligned} {}_jP_{i,i}^{(r)}(h) &= P_{i,i}(rh) - \sum_{m=1}^{r-1} f_{i,j}^{(m)}(h) P_{j,i}((r-m)h) \\ &\geq P_{i,i}(rh) - \sum_{m=1}^{r-1} f_{i,j}^{(m)}(h) \geq 1 - \varepsilon - \frac{\varepsilon}{1-\varepsilon} \geq \frac{1-3\varepsilon}{1-\varepsilon}. \end{aligned} \tag{3.3.8}$$

Daher haben wir

$$P_{i,j}(u) \geq \sum_{r=0}^{n-1} {}_jP_{i,i}^{(r)}(h) P_{i,j}(h) P_{j,j}(u - (r+1)h) \geq n(1-3\varepsilon) P_{i,j}(h), \tag{3.3.9}$$

wobei wir ${}_jP_{i,i}^{(0)}(h) = 1$ setzten und bei der letzten Ungleichung (3.3.8) 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 (3.3.9):

$$\frac{P_{i,j}(u)}{u-h} \geq (1-3\varepsilon) \frac{P_{i,j}(h)}{h}. \tag{3.3.10}$$

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

$$\infty > \frac{1}{1-3\varepsilon} \frac{P_{i,j}(u)}{u} \geq \limsup_{h\downarrow 0} \frac{P_{i,j}(h)}{h}. \tag{3.3.11}$$

Nun geht $u \downarrow 0$:

$$\frac{1}{1-3\varepsilon} \liminf_{u\downarrow 0} \frac{P_{i,j}(u)}{u} \geq \limsup_{h\downarrow 0} \frac{P_{i,j}(h)}{h}. \tag{3.3.12}$$

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

3. Teil: Schluss. Sei $J \subset I$ endlich. Dann gilt

$$\frac{1 - P_{i,i}(h)}{h} \geq \sum_{j \in I \setminus \{i\}} \frac{P_{i,j}(h)}{h} \geq \sum_{j \in J \setminus \{i\}} \frac{P_{i,j}(h)}{h}. \tag{3.3.13}$$

Die linke Seite geht für $h \downarrow 0$ gegen $q_i = -q_{i,i}$, die rechte gegen $\sum_{j\in J\setminus\{i\}} q_{i,j}$. Da J beliebig war, folgt $-q_{i,i} \geq \sum_{j\in I\setminus\{i\}} q_{i,j}$. □

Bemerkung 3.3.8. Der Fall $-q_{i,i} > \sum_{j\in I\setminus\{i\}} q_{i,j}$ kann wirklich auftreten (siehe [Ch67, S. 275 f]). Die in Abschnitt 3.1 präsentierte Konstruktion funktioniert in diesem Fall nicht, da man in einem solchen Fall mit positiver Wahrscheinlichkeit unendlich viele Sprünge in endlicher Zeit vollführt.

Auch der Fall $-q_{i,i} = \infty$ 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. Bei praktischen Anwendungen sind solche eher pathologischen Fälle allerdings von geringer Bedeutung. ◇

3.4 Der erste Sprung

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

$$\tau = \inf\{s > 0 \colon X(s) \neq X(0)\} \tag{3.4.1}$$

bei Start in i bestimmen. Mit $\mathbb{P}_i$ bezeichnen wir die Wahrscheinlichkeit bei Start in i. Wir hatten uns anschaulich bereits überlegt, dass bei Start in i die Zeit τ exponentiell zum Parameter $q_i = \sum_{j \in I \setminus \{i\}} q_{i,j}$ verteilt sein müsste. Dies ist allerdings ohne weitere Voraussetzungen nicht richtig. Man braucht eine Bedingung an die Trajektorien, die sichert, dass diese nicht zu wild aussehen. Die Tatsache, dass $(P(t))_{t \geq 0}$ standard ist, garantiert dies nicht. Statt dessen benötigen wir den Begriff der Separabilität. Wir erinnern daran, dass unsere MKSZ $\mathbb{X} = (X(t))_{t \in [0,\infty)}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ definiert ist, und wir schreiben $\omega \mapsto X(t, \omega)$ für die Zufallsgröße $X(t)$.

Definition 3.4.1 (Separable MKSZ). *Eine MKSZ $\mathbb{X} = (X(t))_{t \in [0,\infty)}$ heißt* separabel, *wenn für jede abzählbare dichte Menge $M \subset [0, \infty)$ ein $N \in \mathcal{F}$ existiert 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 $\{X(s, \omega) \colon s \in [t-\varepsilon, t+\varepsilon] \cap M\}$ liegt. Dabei heißt eine Menge $A \subset I \cup \{\partial\}$ abgeschlossen, wenn A endlich ist oder ∂ enthält.*

Proposition 3.4.2. *Sei $(P(t))_{t \geq 0}$ standard und $\mathbb{X}$ eine zugehörige separable MKSZ. Dann gilt für die in* (3.4.1) *definierte Zeit τ*

$$\mathbb{P}_i(\tau \geq t) = e^{-q_i t} \qquad \text{für alle } t \geq 0, \tag{3.4.2}$$

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

$$\begin{aligned} \mathbb{P}_i(\tau \geq t) &= \mathbb{P}_i\big(X(s) = i \text{ für alle } s \in [0, t)\big) \\ &= \mathbb{P}_i\big(X(k2^{-n}t) = i \text{ für alle } n \in \mathbb{N} \text{ und alle } k = 0, \dots, 2^n - 1\big) \\ &= \lim_{n \to \infty} \mathbb{P}_i\big(X(k2^{-n}t) = i \text{ für alle } k = 0, \dots, 2^n - 1\big) \\ &= \lim_{n \to \infty} \big(P_{i,i}(t2^{-n})\big)^{2^n - 1} = \lim_{n \to \infty} \big(1 - q_i t 2^{-n} + o(t2^{-n})\big)^{2^n} = \mathrm{e}^{-q_i t}, \end{aligned} \tag{3.4.3}$$

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

Wir würden nun gerne zeigen, dass bei Start in i die Zufallsgrößen τ und $X(\tau)$ unabhängig sind und $\mathbb{P}_i(X(\tau) = j) = q_{i,j}/q_i$ 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 Matrix ist.

Definition 3.4.3. *Eine Matrix $Q = (q_{i,j})_{i,j \in I}$ heißt* konservativ, *wenn gelten*

(i) $q_{ij} \geq 0$ *für alle* $i, j \in I$ *mit* $i \neq j$,

(ii) $q_i := -q_{ii} = \sum_{j \in I \setminus \{i\}} q_{ij} < \infty$ *für alle* $i \in I$.

Nicht jede Q-Matrix einer Standardhalbgruppe $(P(t))_{t\geq 0}$ (siehe Satz 3.3.7) ist konservativ. Andererseits ist Konservativität eine notwendige Voraussetzung dafür, dass das in Abschnitt 3.1 beschriebene Simulationsverfahren funktioniert. Wir werden später sehen, dass es zu jeder konservativen Matrix mindestens eine Standardhalbgruppe $(P(t))_{t\geq 0}$ gibt, deren Q-Matrix sie ist.

Nun beweisen wir die Unabhängigkeit des ersten Sprungs und seines Zeitpunkts und identifizieren deren Verteilungen.

Satz 3.4.4. *Sei $(P(t))_{t\geq 0}$ standard mit Q-Matrix $Q = (q_{i,j})_{i,j\in I}$ und $\mathbb{X} = (X(t))_{t\geq 0}$ eine zugehörige separable MKSZ. Seien $i, j \in I$ mit $i \neq j$ sowie $q_i \in (0,\infty)$ und $q_{i,j} < \infty$. Sei τ wie in* (3.4.1)*. Dann gilt für alle $u \geq 0$:*

$$\mathbb{P}_i\big(\tau \geq u, \text{ und es existiert } s > 0\colon X(t) = j \text{ für alle } t \in (\tau, \tau+s]\big) = \mathrm{e}^{-q_i u}\frac{q_{ij}}{q_i}. \tag{3.4.4}$$

Ist Q konservativ, so existiert $\lim_{h\downarrow 0} X(\tau+h,\omega) = J(\omega)$ fast sicher, ist unabhängig von τ und erfüllt $\mathbb{P}_i(J = j) = q_{i,j}/q_i$ für jedes $j \in I \setminus \{i\}$.

Beweis. Wir bezeichnen das Ereignis auf der linken Seite mit

$$B = \big\{\tau \geq u, \text{ es existiert } s > 0\colon X(t) = j \text{ für alle } t \in (\tau, \tau+s]\big\}, \tag{3.4.5}$$

also ist unser Ziel zu zeigen, dass $\mathbb{P}_i(B) = \mathrm{e}^{-q_i u}\frac{q_{i,j}}{q_i}$ gilt. Wir approximieren B mit dem Ereignis

$$B_\gamma = \big\{\tau \geq u, X(t) = j \text{ für alle } t \in (\tau, \tau+\gamma)\big\}, \qquad \gamma > 0. \tag{3.4.6}$$

Dann gilt $B_\gamma \uparrow B$ für $\gamma \downarrow 0$. Wir werden zeigen, dass

$$\mathbb{P}_i(B_\gamma) = \mathrm{e}^{-q_i u}\frac{q_{i,j}}{q_i}\mathrm{e}^{-q_j\gamma}, \qquad \gamma \in 2^{-\mathbb{N}}, \tag{3.4.7}$$

gilt, woraus dann aufgrund der Stetigkeit von $\mathbb{P}_i$ folgt, dass $\mathbb{P}_i(B) = \mathrm{e}^{-q_i u} q_{i,j}/q_i$ gilt.

Sei für $n \in \mathbb{N}, m \in \mathbb{N}$ und $\gamma := 2^{-m}$

$$\begin{aligned} B_{n,\gamma} := \big\{ &\text{ es existiert } s \geq u, \text{ so dass } X(k2^{-n}) = i \text{ für alle } k2^{-n} < s,\\ &\text{und } X(k2^{-n}) = j \text{ für alle } s \leq k2^{-n} < s+\gamma\big\}. \end{aligned} \tag{3.4.8}$$

Dann existiert eine Menge B'_γ mit $B_{n,\gamma} \downarrow B'_\gamma$ für $n \to \infty$, und wegen der Separabilität von $\mathbb{X}$ gilt $\mathbb{P}_i(B'_\gamma) = \mathbb{P}_i(B_\gamma)$. Also genügt es zu zeigen, dass $\lim_{n\to\infty}\mathbb{P}_i(B_{n,\gamma}) = \mathrm{e}^{-q_i u}\frac{q_{i,j}}{q_i}\mathrm{e}^{-q_j\gamma}$ gilt. Für $n \geq m$ folgt:

$$\begin{aligned} \mathbb{P}_i(B_{n,\gamma}) &= \sum_{k=\lfloor u2^n\rfloor}^{\infty} \big(P_{i,i}(2^{-n})\big)^k P_{i,j}(2^{-n})\,\big(P_{j,j}(2^{-n})\big)^{2^{n-m}-1}\\ &= P_{i,j}(2^{-n})P_{j,j}(2^{-n})^{2^{n-m}-1}P_{i,i}(2^{-n})^{\lfloor u2^n\rfloor}\frac{1}{1-P_{i,i}(2^{-n})}, \end{aligned} \tag{3.4.9}$$

wobei $\lfloor x\rfloor$ die größte ganze Zahl $\leq x$ sei. Wegen $P_{i,i}(2^{-n}) < 1$ für alle hinreichend großen n konvergiert die Reihe. Die vier Faktoren sind für $n \to \infty$ in obiger Reihenfolge asymptotisch äquivalent zu

$$q_{i,j}2^{-n}, \qquad \mathrm{e}^{-q_j\gamma}, \qquad \mathrm{e}^{-q_i u}, \qquad \frac{1}{q_i 2^{-n}}.$$

Also ist die erste Behauptung, (3.4.7), bewiesen.

Ist Q konservativ, dann folgt die Existenz des Grenzwerts $J(\omega)$ und die Formel für die Verteilung, indem man in (3.4.4) $u = 0$ setzt und über alle $j \in I \setminus \{i\}$ summiert. Die Unabhängigkeit ist klar, da

$$\mathbb{P}_i(J = j \mid \tau \geq u) = \frac{q_{i,j}}{q_i}$$

nicht von u abhängt. □

Mit Satz 3.4.4 ist die Konstruktion aus Abschnitt 3.1 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 zum Beispiel durch eine Erweiterung von Satz 3.4.4 zeigen, indem man die gemeinsame Verteilung der ersten k Sprünge und der Sprungzeiten berechnet, was recht mühsam ist. Daher verzichten wir hier auf eine Durchführung dieser Argumente.

Bemerkung 3.4.5 (Starke Markoveigenschaft). Alternativ lässt sich die Unabhängigkeit des zweiten Sprunges von der Wartezeit und dem zweiten Zustand auch aus der starken Markoveigenschaft herleiten, die wir aber weder formulieren noch beweisen wollen. Für eine exakte Definition siehe z. B. [Ch67]; sie ist ähnlich dem diskreten Fall. Wir betonen, dass die starke Markoveigenschaft ohne Separabilität im Allgemeinen nicht gilt und selbst im separablen Fall nicht gelten muss; z. B. dann nicht, wenn man als Stoppzeit $\tau = \inf\{s \geq 0 \colon X(s) = \partial\}$ wählt und ein Rücksprung nach I möglich ist. Die Rücksprungverteilung kann nämlich explizit von dem Verhalten von $\mathbb{X}$ kurz zuvor abhängen, ohne dass dies die Markoveigenschaft zerstört. Die starke Markoveigenschaft kann daran scheitern, dass die "Kompaktifizierung" $I \cup \{\partial\}$ von I zu grob ist. Die Theorie der *Ray-Knight-Kompaktifizierung* zeigt, wie man, abhängig von der Standardhalbgruppe $(P(t))_{t\geq 0}$, die Menge I so kompaktifiziert, dass eine MKSZ mit Werten in der Kompaktifizierung existiert mit den geforderten endlich-dimensionalen Verteilungen, so dass sie rechtsstetige Pfade hat und die starke Markoveigenschaft erfüllt. Der interessierte Leser sei auf [Wi79] verwiesen. ◇

3.5 Vorwärts- und Rückwärtsgleichungen

Wenn $(P(t))_{t\geq 0}$ eine Standard-Sub-Markov'sche Halbgruppe ist, dann existiert nicht nur $P'(0)$, sondern auch $P'(t)$ für jedes t und kann mit Hilfe von Q ausgedrückt werden. Dies werden wir in Abschnitt 3.6 zum Anlass nehmen, zu gegebenem Q eine zugehörige Halbgruppe $(P(t))_{t\geq 0}$ zu konstruieren.

Satz 3.5.1 (Rückwärtsgleichung). *Sei $(P(t))_{t\geq 0}$ standard mit konservativer Q-Matrix Q. Dann ist $t \mapsto P_{i,j}(t)$ auf $[0,\infty)$ stetig differenzierbar für alle $i, j \in I$, und es gelten*

$$P'(t) = QP(t), \quad t \geq 0, \qquad \textit{und} \qquad P(0) = I, \tag{3.5.1}$$

wobei Id *die Identität auf I ist (d. h. $\mathrm{Id}_{i,j} = \delta_{i,j}$) und die Ableitung komponentenweise zu verstehen ist.*

Beweis. Für $h > 0$ und $t \geq 0$ gilt:

$$\begin{aligned}\frac{P_{i,j}(h+t) - P_{i,j}(t)}{h} &= \frac{1}{h}\Big[\sum_{k\in I} P_{i,k}(h)P_{k,j}(t) - P_{i,j}(t)\Big] \\ &= \frac{1}{h}\Big[\big(P_{i,i}(h) - 1\big)P_{i,j}(t) + \sum_{k\in I\setminus\{i\}} P_{i,k}(h)P_{k,j}(t)\Big].\end{aligned} \tag{3.5.2}$$

Falls die eventuell unendliche Summe mit dem Grenzwert $h \downarrow 0$ vertauscht werden darf, so erhalten wir aus (3.5.2) leicht:

$$\lim_{h\downarrow 0} \frac{P_{i,j}(h+t) - P_{i,j}(t)}{h} = q_{i,i}P_{i,j}(t) + \sum_{k\in I\setminus\{i\}} q_{i,k}P_{k,j}(t) = \sum_{k\in I} q_{i,k}P_{k,j}(t). \tag{3.5.3}$$

Wir rechtfertigen nun die Vertauschung. Seien $\varepsilon > 0$ und $J \subset I$ mit $i \in J$ und $|J| < \infty$, so dass $\sum_{k\in I\setminus J} q_{i,k} < \varepsilon/2$ (hier benutzen wir, dass Q konservativ ist). Dann gilt

$$\begin{aligned}\Big|\sum_{k\in I\setminus J}\Big(\frac{P_{i,k}(h)}{h} - q_{i,k}\Big)P_{k,j}(t)\Big| &\leq \Big|\sum_{k\in I\setminus J}\frac{P_{i,k}(h)}{h}\Big| + \sum_{k\in I\setminus J} q_{i,k} \\ &< \Big|\frac{1 - P_{i,i}(h)}{h} - \sum_{k\in J\setminus\{i\}}\frac{P_{i,k}(h)}{h}\Big| + \frac{\varepsilon}{2} \\ &\stackrel{h\downarrow 0}{\longrightarrow} -q_{i,i} - \sum_{k\in J\setminus\{i\}} q_{i,k} + \frac{\varepsilon}{2} = \sum_{k\in I\setminus J} q_{i,k} + \frac{\varepsilon}{2} < \varepsilon,\end{aligned} \tag{3.5.4}$$

wobei wir beim letzten Gleichheitszeichen die Konservativität von Q benutzten. Damit ist die Konvergenz in (3.5.3) gerechtfertigt, und es folgt $P'(t) = QP(t)$, wobei allerdings der Strich zunächst nur als rechtsseitige Ableitung zu verstehen ist, da $h > 0$ war.

Aus Lemma 3.3.6 wissen wir, dass $P_{i,j}(\cdot)$ gleichmäßig stetig ist. Wegen der Konservativität von Q und Beschränktheit der $P_{i,j}(\cdot)$ durch 1 konvergiert $\sum_{k\in I} q_{i,k}P_{k,j}(t)$ gleichmäßig für alle $t \in [0,\infty)$. Da gleichmäßige Grenzwerte stetiger Funktionen stetig sind, folgt, dass $P'_{i,j}(\cdot)$ 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 zu beweisen ist, wie man meinen könnte, zeigen wir dies im folgenden Lemma 3.5.2). Da $P(0) = \mathrm{Id}$ im Standardfall immer gilt, folgt die Behauptung. □

Lemma 3.5.2. *Sei $f\colon [0,t_0] \to \mathbb{R}$ 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. Setze $F(t) := f(0) + \int_0^t f^+(s)\,\mathrm{d}s$ 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 φ 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 φ 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

$$\gamma^+(s) = -\frac{\varphi(t^*)}{t^*} < 0 \quad \text{für } s \in [0,t^*) \qquad \text{und} \qquad \varphi(0) = 0 = \varphi(t^*). \tag{3.5.5}$$

Sei $s^* \in [0, t^*)$ so gewählt, dass $\gamma(s^*) = \min_{s\in[0,t^*]} \gamma(s)$. Dann folgt

$$\gamma^+(s^*) = \lim_{h\downarrow 0} \frac{\gamma(s^* + h) - \gamma(s^*)}{h} \geq 0 \tag{3.5.6}$$

im Widerspruch zu (3.5.5). □

Es gibt nicht viele Beispiele, in denen die Halbgruppe explizit bekannt ist, aber hier ist eines, das wir als Übungsaufgabe formulieren:

Beispiel 3.5.3 (Yule-Prozess). Wir betrachten eine Bakterienkultur unter der Annahme, dass sich die Bakterien unabhängig voneinander teilen, und zwar mit Rate $\lambda > 0$, und dass keine Bakterien sterben. Die Anzahl der Bakterien zum Zeitpunkt t sei bezeichnet mit $X(t) \in \mathbb{N}$, dann ist $(X(t))_{t\in[0,\infty)}$ eine MKSZ auf $\mathbb{N}$, der sogenannte *Yule-Prozess*, ein reiner Geburtsprozess. Man kann die Q-Matrix leicht explizit berechnen. Dann wird die Rückwärtsgleichung erfüllt durch

$$P_{i,j}(t) = \begin{cases} \binom{j-1}{j-i}(\mathrm{e}^{\lambda t} - 1)^{j-i}\mathrm{e}^{-j\lambda t}, & \text{falls } i \leq j, \\ 0 & \text{sonst.} \end{cases}$$

Als Übungsaufgabe leite man diese Formel mit Hilfe kombinatorischer Überlegungen aus der Beschreibung des Prozesses her. ◇

Im Beweis der Rückwärtsgleichungen hatten wir $P(t + h)$ als $P(h)P(t)$ geschrieben. Wenn man es als $P(t)P(h)$ schreibt, so endet man mit $P(t)Q$ statt $QP(t)$:

Satz 3.5.4 (Vorwärtsgleichung). *Sei $(P(t))_{t\geq 0}$ standard mit konservativer Q-Matrix Q. Weiter sei*

$$c := \sup_{i\in I} q_i < \infty. \tag{3.5.7}$$

Dann gelten

$$P'(t) = P(t)Q, \quad t \geq 0, \qquad \textit{und} \qquad P(0) = I. \tag{3.5.8}$$

Beweis. Seien $t \geq 0$ und $h > 0$, sowie $i, j \in I$. Es gilt

$$\Big|\frac{P_{k,j}(h) - \delta_{k,j}}{h}\Big| \leq \frac{1 - P_{k,k}(h)}{h} \leq q_k; \tag{3.5.9}$$

letzteres wurde im 1. Teil des Beweises von Satz 3.3.7 gezeigt. Somit gilt

$$\frac{P_{i,j}(t + h) - P_{i,j}(t)}{h} = \sum_{k\in I} P_{i,k}(t)\frac{P_{k,j}(h) - \delta_{k,j}}{h} \overset{h\downarrow 0}{\longrightarrow} \sum_{k\in I} P_{i,k}(t)q_{kj}, \tag{3.5.10}$$

denn wegen der Voraussetzung (3.5.7) gilt für endliches $J \subset I$ mit $\sum_{k\in I\setminus J} P_{i,k}(t) < \varepsilon/(2c)$:

$$\sum_{k\in I\setminus J} \Big|P_{i,k}(t)\Big(\frac{P_{k,j}(h) - \delta_{k,j}}{h} - q_{k,j}\Big)\Big| \leq \sum_{k\in I\setminus J} P_{i,k}(t)\Big(\Big|\frac{P_{k,j}(h) - \delta_{k,j}}{h}\Big| + q_{k,j}\Big) \leq \frac{\varepsilon}{2c}\,2c = \varepsilon. \tag{3.5.11}$$

Nun verläuft der Rest des Beweises wie im Beweis der Rückwärtsgleichungen, Satz 3.5.1. Dieser Satz sagt uns auch, dass $P_{i,j}(\cdot)$ stetig differenzierbar ist, was dabei benutzt wird. □

Beispiel 3.5.5 (Poissonprozess). Sei Q die (konservative) Q-Matrix des Poissonprozesses aus Beispiel 3.1.2. 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:

$$\begin{aligned} P'_{i,j}(t) &= \sum_{k\in I} q_{i,k}P_{k,j}(t) = \lambda P_{i+1,j}(t) - \lambda P_{i,j}(t), \quad t>0, i,j\in\mathbb{N}_0, && (3.5.12)\\ P_{i,j}(0) &= \delta_{i,j}, \qquad i,j\in\mathbb{N}_0. && (3.5.13) \end{aligned}$$

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, denn zur Berechnung von $P_{0,j}(t)$ muss man bereits $P_{1,j}(t)$ kennen usw. Die Vorwärtsgleichung lautet ausgeschrieben:

$$\begin{aligned} P'_{i,j}(t) &= \sum_{k\in I} P_{i,k}(t)q_{k,j} = \begin{cases} \lambda P_{i,j-1}(t) - \lambda P_{i,j}(t), & \text{falls } j\geq 1,\\ -\lambda P_{i,0}(t), & \text{falls } j=0, \end{cases} && (3.5.14)\\ P_{i,j}(0) &= \delta_{i,j}, \qquad i,j\in\mathbb{N}_0. && (3.5.15) \end{aligned}$$

Hieraus kann man für festes $i\in\mathbb{N}_0$ das System rekursiv für $j=0,1,2,\dots$ lösen. Für $j=0$ erhält man:

$$P_{i,0}(t) = \begin{cases} 0, & \text{falls } i\geq 1,\\ \mathrm{e}^{-\lambda t}, & \text{falls } i=0. \end{cases} \tag{3.5.16}$$

Dies setzt man in die Gleichung für $j=1$ ein. Mit dieser Methode zeigt man leicht, dass

$$P_{i,j}(t) = \begin{cases} 0, & \text{falls } i>j,\\ \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}, & \text{falls } i\leq j, \end{cases} \tag{3.5.17}$$

gilt. Die Anzahl der Sprünge des Poissonprozesses in einem Zeitintervall der Länge t ist also Poisson-verteilt mit Parameter λt. Man kann nun explizit nachrechnen, dass diese (einzige) Lösung der Vorwärtsgleichung wirklich eine Markov'sche Standardhalbgruppe ist. Dies wird sich allerdings gleich als unnötig herausstellen (siehe Satz 3.6.2). ◇

Beispiel 3.5.6 (Galton-Watson-Prozess). Wir benutzen die Rückwärtsgleichung, um die erwartete Populationsgröße eines Galton-Watson-Prozesses zu berechnen, siehe Beispiel 3.2.4. Wir setzen wiederum voraus, dass die erwartete Nachkommenschaft, $m=\sum_{j\in\mathbb{N}_0} jp_j$, endlich ist.

Zunächst sehen wir, dass für jedes $i\in\mathbb{N}_0$ gilt: $\mathbb{E}_i[X(t)] = i\mathbb{E}_1[X(t)]$. Dies ersieht man aus der Konstruktion des Prozesses oder rechnet es mühsam mit Hilfe der Q-Matrix nach. Die Rückwärtsgleichung ergibt daher

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_1[X(t)] = -\mathbb{E}_1[X(t)] + \sum_{j=2}^\infty p_j\mathbb{E}_j[X(t)] = -\mathbb{E}_1[X(t)] + \mathbb{E}_1[X(t)]\sum_{j=2}^\infty jp_j = (m-1)\mathbb{E}_1[X(t)].$$

Wegen der Startbedingung $\mathbb{E}_1[X(0)]=1$ haben wir also $\mathbb{E}_1[X(t)] = \mathrm{e}^{(m-1)t}$ für jedes $t\in[0,\infty)$. Der Prozess konvergiert also im Mittel exponentiell gegen ∞, bleibt konstant oder konvergiert exponentiell gegen Null, je nachdem ob $m>1$, $m=1$ oder $m<1$. Man nennt den Prozess *superkritisch*, falls $m>1$ und *subkritisch* sonst. ◇

3.6 Q-Matrizen und ihre Halbgruppen

Bislang wissen wir nicht, ob zu einer konservativen Matrix immer eine Standardhalbgruppe gehört und unter welchen Umständen diese eindeutig ist. Die Existenz wird im folgenden Satz sichergestellt. Dabei können wir die Konservativität von Q etwas abschwächen.

Definition 3.6.1. *Eine Matrix $Q = (q_{i,j})_{i,j\in I}$ heißt* schwach konservativ, *wenn gelten:*

(i) $q_{ij} \geq 0$ *für alle* $i, j \in I$ *mit* $i \neq j$,

(ii) $q_{ii} > -\infty$ *für alle* $i \in I$,

(iii) $\sum_{j\in I} q_{ij} \leq 0$ *für alle* $i \in I$.

Wieder definieren wir $q_i := -q_{ii}$ *für alle* $i \in I$.

Unser Ziel besteht darin, zu einer gegebenen schwach konservativen Matrix Q eine Standardhalbgruppe $(\overline{P}(t))_{t\geq 0}$ zu finden, die die Vorwärts- und Rückwärtsgleichung löst, für die also insbesondere (nach der Rückwärtsgleichung für $t = 0$) $\overline{P}'(0) = Q$ ist, d.h. Q die Q-Matrix von $(\overline{P}(t))_{t\geq 0}$ ist. Wir wissen dann insbesondere, dass, wenn Q schwach konservativ ist und die Vorwärtsgleichung, bzw. die Rückwärtsgleichung, eine eindeutige Lösung hat, diese automatisch eine Standardhalbgruppe sein muss (vgl. Bemerkung am Ende von Beispiel 3.5.5).

Satz 3.6.2. *Sei Q eine schwach konservative Matrix. Dann existiert eine Standardhalbgruppe $(\overline{P}(t))_{t\geq 0}$, die die Vorwärtsgleichung und die Rückwärtsgleichung löst und für die gilt*

$$\overline{P}_{i,j}(t) \leq Z_{i,j}(t), \qquad t \geq 0, i, j \in I, \tag{3.6.1}$$

für jede Standardhalbgruppe $(Z(t))_{t\geq 0}$ mit Q-Matrix Q. Diese Standardhalbgruppe $(\overline{P}(t))_{t\geq 0}$ heißt Minimallösung.

Beweisskizze. Wir zeigen nur die Konstruktion der Standardhalbgruppe $(\overline{P}(t))_{t\geq 0}$ (die uns beim Beweis von Satz 3.7.5 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 $i, j \in I$, $t \geq 0$

$$P^0_{i,j}(t) := \delta_{i,j} \mathrm{e}^{-q_i t} \tag{3.6.2}$$

und für $n \in \mathbb{N}_0$ induktiv

$$P^{n+1}_{i,j}(t) := \sum_{k\in I\setminus\{j\}} \int_0^t P^n_{i,k}(s) q_{k,j} \mathrm{e}^{-q_j(t-s)}\,\mathrm{d}s. \tag{3.6.3}$$

Induktiv zeigt man, dass $P^n_{i,j}(\cdot)$ stetig differenzierbar ist und

$$H^N_{i,j}(t) := \sum_{n=0}^{N} P^n_{i,j}(t) \leq 1 \tag{3.6.4}$$

ist. Schließlich definiert man für $t \geq 0$

$$\overline{P}_{i,j}(t) := \lim_{N\uparrow\infty} H^N_{i,j}(t) \tag{3.6.5}$$

(der Grenzwert existiert und ist ≤ 1). Es bleibt zu zeigen, dass $(\overline{P}(t))_{t\geq 0}$ standard ist und die Vorwärtsgleichung und die Rückwärtsgleichung löst, sowie die Minimalitätseigenschaft. Der Beweis der Gültigkeit der Vorwärtsgleichung ist mit der obigen Rekursion relativ leicht. Für den Nachweis der Rückwärtsgleichung zeigt man zunächst, dass die oben definierten $P^n_{i,j}$ auch der Rekursion

$$P^{n+1}_{i,j}(t) = \sum_{k\in I\setminus\{i\}} \int_0^t \mathrm{e}^{-q_i(t-s)} q_{i,k} P^n_{k,j}(s)\,\mathrm{d}s \tag{3.6.6}$$

genügen. Die letzte Gleichung lässt sich anschaulich wie folgt interpretieren: $P^n_{i,j}(t)$ ist die Wahrscheinlichkeit, bei Start in i mit genau n Sprüngen zur Zeit t in j zu landen. $\overline{P}_{i,j}(t)$ ist dann die Wahrscheinlichkeit, bei Start in i mit *endlich* vielen Sprüngen zur Zeit t in j zu landen. □

Bemerkung 3.6.3 (Interpretation der Minimallösung). Die Minimallösung $(\overline{P}(t))_{t\geq 0}$ ist die Halbgruppe derjenigen MKSZ $\mathbb{X}$ zu Q, die nach der ersten Explosion für immer in ∂ bleibt, d. h. dem ersten Zeitpunkt, zu dem $\mathbb{X}$ entweder in den Zustand ∂ springt, oder dem Infimum der Zeitpunkte, an denen $\mathbb{X}$ unendlich viele Zustände besucht hat. Jede MKSZ zu einer anderen Standardlösung Z verhält sich bis zur Explosion genauso wie $\mathbb{X}$, springt dann aber von ∂ aus auf irgendeine Weise zurück nach I, weswegen $\overline{P}_{i,j}(t) \leq Z_{i,j}(t)$ gilt. ◇

Korollar 3.6.4. *Sei Q schwach konservativ. Wenn die Minimallösung $(\overline{P}(t))_{t\geq 0}$ Markov'sch 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 3.6.2, dass $Z_{i,j}(t) \geq \overline{P}_{i,j}(t)$ für alle $i,j \in I$ und $t \geq 0$, also

$$1 \geq \sum_{j\in I} Z_{i,j}(t) \geq \sum_{j\in I} \overline{P}_{i,j}(t) = 1, \qquad i \in I, t \geq 0, \tag{3.6.7}$$

also folgt $Z_{i,j}(t) = \overline{P}_{i,j}(t)$ für alle $i,j \in I$ und $t \geq 0$. □

Proposition 3.6.5. *Sei Q konservativ und $c = \sup_{i\in I} q_i < \infty$. Dann ist die Minimallösung $(\overline{P}(t))_{t\geq 0}$ Markov'sch.*

Beweis. Unser Ziel besteht darin, zu zeigen, dass $\frac{\mathrm{d}}{\mathrm{d}t}\sum_{j\in I} \overline{P}_{i,j}(t) = 0$ für alle $t \geq 0$ gilt, woraus wegen $\sum_{j\in I} \overline{P}_{i,j}(0) = 1$ folgt, dass $(\overline{P}(t))_{t\geq 0}$ Markov'sch ist.

Sei $J \subset I$ endlich. Da $(\overline{P}(t))_{t\geq 0}$ die Vorwärtsgleichung erfüllt, gilt

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{j\in J} \overline{P}_{i,j}(t) = \sum_{j\in J} \overline{P}'_{i,j}(t) = \sum_{j\in J}\sum_{k\in I\setminus\{j\}} \overline{P}_{i,k}(t) q_{k,j} - \sum_{j\in J} \overline{P}_{i,j}(t) q_j. \tag{3.6.8}$$

Nun ist $\sum_{j\in J}\sum_{k\in I\setminus\{j\}} \overline{P}_{i,k}(\cdot) q_{k,j}$ stetig, denn $|J| < \infty$, und $\sum_{k\in I\setminus\{j\}} \overline{P}_{i,k}(t) q_{k,j} = \overline{P}'_{i,j}(t) + \overline{P}_{i,j}(t) q_j$ ist stetig in t nach Satz 3.5.1.

Weiter konvergiert mit $J \uparrow I$

$$\sum_{j\in J}\sum_{k\in I\setminus\{j\}} \overline{P}_{i,k}(\cdot) q_{k,j} \overset{\text{monoton}}{\longrightarrow} \sum_{j\in I}\sum_{k\in I\setminus\{j\}} \overline{P}_{i,k}(\cdot) q_{k,j} = \sum_{k\in I} \overline{P}_{i,k}(\cdot) q_k, \tag{3.6.9}$$

da Q konservativ ist, und

$$\sum_{j\in J}\overline{P}_{i,j}(t)q_j \overset{\text{monoton}}{\longrightarrow} \sum_{j\in I}\overline{P}_{i,j}(t)q_j. \tag{3.6.10}$$

Wir zeigen, dass $t\mapsto \sum_{j\in I}\overline{P}_{i,j}(t)q_j$ stetig ist:

$$\Big|\sum_{j\in I}\big(\overline{P}_{i,j}(t+h)-\overline{P}_{i,j}(t)\big)q_j\Big| \le c\sum_{j\in I}\big|\overline{P}_{i,j}(t+h)-\overline{P}_{i,j}(t)\big| \overset{h\downarrow 0}{\longrightarrow} 0 \tag{3.6.11}$$

gleichmäßig für alle $t\in[0,\infty)$ nach Lemma 3.3.6. Der *Satz von Dini* besagt, dass, wenn eine Folge stetiger Funktionen auf einem kompakten Intervall monoton gegen eine stetige Funktion punktweise konvergiert, die Konvergenz sogar gleichmäßig ist. Somit konvergiert die rechte Seite von (3.6.8) gleichmäßig auf kompakten Intervallen gegen

$$\sum_{k\in I}\overline{P}_{i,k}(t)q_k - \sum_{j\in I}\overline{P}_{i,j}(t)q_j = 0. \tag{3.6.12}$$

Da jede Folge stetig differenzierbarer Funktionen $f_n\colon[0,t_0]\to\mathbb{R}$ mit $f_n(0)\to a\in\mathbb{R}$, deren Ableitungen gleichmäßig konvergieren, gleichmäßig auf $[0,t_0]$ gegen eine stetig differenzierbare Funktion f konvergiert, deren Ableitung gleich dem Grenzwert der Ableitungen der f_n ist, folgt

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{j\in I}\overline{P}_{i,j}(t) = 0. \tag{3.6.13}$$

□

Aus den bisherigen Resultaten folgt: Ist Q schwach konservativ, und hat die Vorwärtsgleichung eine eindeutige Lösung $(P(t))_{t\ge 0}$ und erfüllt diese $\sum_{j\in I}P_{i,j}(t)=1$ für alle $i\in I$, dann ist $(P(t))_{t\ge 0}$ standard und die einzige Standardhalbgruppe mit Q-Matrix Q. Außerdem wissen wir auch, dass zu einer konservativen Matrix Q mit beschränkten Diagonaleinträgen genau eine Markov'sche Standardhalbgruppe gehört, deren Q-Matrix Q ist.

3.7 Langzeitverhalten und invariante Maße

Wir studieren nun die Existenz von Grenzwerten von $P_{i,j}(t)$ für $t\to\infty$. Interessanterweise ist dies weniger kompliziert als im zeitlich diskreten Fall, da das Problem der Periodizität bei MKSZ nicht auftaucht. Zunächst stellen wir eine Aussage über das Langzeitverhalten von MKDZ bereit, das den Satz 2.6.1 als Spezialfall enthält:

Satz 3.7.1. *Sei P eine stochastische Matrix auf I, so dass jeder Zustand für P aperiodisch ist. Dann existiert für jedes $i,j\in I$ der Grenzwert $\pi_{i,j}=\lim_{n\to\infty}P^{(n)}_{i,j}$ und erfüllt $\sum_{j\in I}\pi_{i,j}\le 1$. Hierbei ist $P^n=(P^{(n)}_{i,j})_{i,j\in I}$ die n-te Potenz von P.*

Im Spezialfall, dass P irreduzibel und positiv rekurrent ist, ist dies Satz 2.6.1, und der Grenzwert $\pi_{i,j}$ hängt nicht von i ab und ist ein Wahrscheinlichkeitsmaß in j, welches invariant unter P ist. Im Fall von irreduziblem P kann man sowohl im nullrekurrenten also auch im transienten Fall zeigen, dass $\lim_{n\to\infty}P^{(n)}_{i,j}=0$ gilt. Wenn P nicht irreduzibel ist, so wendet man das Bisherige auf die einzelnen Klassen an, aber dann entstehen Abhängigkeiten vom Startpunkt.

Nun übertragen wir Satz 3.7.1 auf den zeitlich stetigen Fall:

Satz 3.7.2. *Sei $(P(t))_{t\geq 0}$ standard. Dann existiert für alle $i, j \in I$*

$$\pi_{i,j} := \lim_{t\to\infty} P_{i,j}(t), \tag{3.7.1}$$

und es gilt $\sum_{j\in I} \pi_{i,j} \leq 1$.

Beweis. Wegen Bemerkung 3.3.3 können wir annehmen, dass $(P(t))_{t\geq 0}$ standard und Markov'sch ist. Für $j \in I$ existiert daher ein $h_0 > 0$, so dass $P_{j,j}(h) > 0$ 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:

$$P_{j,j}(t) \geq (P_{j,j}(h))^n > 0. \tag{3.7.2}$$

Für beliebiges $h > 0$ betrachte nun die MKDZ $(X(nh))_{n\in\mathbb{N}_0}$ mit Übergangsmatrix $P(h)$. Diese ist wegen $P_{j,j}(h) > 0$ für alle $j \in I$ aperiodisch, also existiert nach Satz 3.7.1 der Grenzwert

$$\pi_{i,j}(h) := \lim_{n\to\infty} P_{i,j}(nh). \tag{3.7.3}$$

Nun müssen wir zeigen, dass $\pi_{i,j}(h)$ nicht von h abhängt. Sei $\varepsilon > 0$ vorgegeben und $i, j \in I$. Da $P_{i,j}(\cdot)$ nach Lemma 3.3.6 gleichmäßig stetig ist, existiert ein h, so dass $|P_{i,j}(t+s) - P_{i,j}(t)| < \varepsilon/4$ für alle $t \geq 0$ und alle $s \leq h$. Wähle nun $n_0 = n_0(h)$, so dass

$$|P_{i,j}(nh) - \pi_{i,j}(h)| < \frac{\varepsilon}{4}, \qquad n \geq n_0. \tag{3.7.4}$$

Dann gilt für $t, t' \geq n_0 h$ mit $kh \leq t \leq (k+1)h$ und $mh \leq t' \leq (m+1)h$

$$\begin{aligned} |P_{i,j}(t) - P_{i,j}(t')| &\leq |P_{i,j}(t) - P_{i,j}(kh)| + |P_{i,j}(kh) - \pi_{i,j}(h)| \\ &\quad + |\pi_{i,j}(h) - P_{i,j}(mh)| + |P_{i,j}(mh) - P_{i,j}(t')| < 4\frac{\varepsilon}{4} = \varepsilon. \end{aligned} \tag{3.7.5}$$

Also existiert $\lim_{t\to\infty} P_{i,j}(t)$, und es ist $\pi_{i,j} := \lim_{t\to\infty} P_{i,j}(t) = \lim_{t\to\infty} P_{i,j}(nh) = \pi_{i,j}(h)$. Insbesondere hängt $\pi_{i,j}(h)$ gar nicht von h ab.

Sei nun $J \subset I$ endlich. Dann gilt

$$\sum_{j\in J} \pi_{i,j} = \sum_{j\in J} \lim_{t\to\infty} P_{i,j}(t) = \lim_{t\to\infty} \sum_{j\in J} P_{i,j}(t) \leq 1, \tag{3.7.6}$$

also folgt $\sum_{j\in I} \pi_{i,j} \leq 1$. (Diese Aussage folgt auch durch eine Anwendung des Lemmas von Fatou.) □

Wie im diskreten Fall wollen wir die Beziehung zwischen den Grenzwerten $\pi_{i,j}$ und invarianten Maßen näher studieren.

Definition 3.7.3 (Invariantes Maß). *Sei $(P(t))_{t\geq 0}$ eine sub-Markov'sche Halbgruppe und $\mathbb{X}$ eine zugehörige MKSZ. Eine Abbildung $\pi\colon I \to [0, \infty]$ heißt ein* invariantes Maß *für $(P(t))_{t>0}$ (oder für $\mathbb{X}$), wenn $\pi P(t) = \pi$ für jedes $t \geq 0$ gilt. Gilt zusätzlich $\sum_{i\in I} \pi_i = 1$, dann heißt π* invariantes Wahrscheinlichkeitsmaß *oder* invariante Verteilung *von $(P(t))_{t>0}$.*

Proposition 3.7.4. *Sei $(P(t))_{t\geq 0}$ standard und $\pi_{i,j}$ wie in Satz 3.7.2 definiert. Dann ist $(\pi_{i,j})_{j\in I}$ für jedes $i \in I$ ein invariantes Maß.*

Beweis. Wir setzen wie in Bemerkung 3.3.3 $(P(t))_{t\geq 0}$ zu einer Markovhalbgruppe auf $I \cup \{\partial\}$ fort. Dann gilt für $s, t \geq 0$ und $k \in I \cup \{\partial\}$

$$\sum_{j \in I \cup \{\partial\}} P_{i,j}(s) P_{j,k}(t) = P_{i,k}(t+s). \tag{3.7.7}$$

Lässt man $s \to \infty$ gehen, so folgt mit der üblichen Abschneidetechnik (d. h. nach Betrachtung zunächst endlicher Summationsbereiche), oder auch durch eine Anwednung des Lemmas von Fatou,

$$\sum_{j \in I \cup \{\partial\}} \pi_{i,j} P_{j,k}(t) \leq \pi_{i,k}. \tag{3.7.8}$$

Summiert man über alle $k \in I \cup \{\partial\}$, so folgt $\sum_{j \in I \cup \{\partial\}} \pi_{i,j} \leq \sum_{k \in I \cup \{\partial\}} \pi_{i,k}$, also – da die Summe endlich ist – die Gleichheit in (3.7.8) für alle $k \in I \cup \{\partial\}$. Für $k \in I$ ist $P_{\partial k}(t) = 0$ für alle $t \geq 0$, also ist $(\pi_{i,j})_{j \in I}$ ein invariantes Maß. □

Satz 3.7.5. *Sei Q konservativ und die Minimallösung $(P(t))_{t\geq 0}$ Markov'sch. Dann ist ein Wahrscheinlichkeitsmaß π auf I genau dann invariant, wenn $\pi Q = 0$ gilt.*

Beweis. (nach [Mi63]): '$\Longrightarrow$': Sei π ein invariantes Wahrscheinlichkeitsmaß, und sei $j \in I$, dann gilt

$$\pi_j(1 - P_{j,j}(t)) = \sum_{i \in I \setminus \{j\}} \pi_i P_{i,j}(t). \tag{3.7.9}$$

Dividiert man beidseitig durch $t > 0$ und lässt $t \downarrow 0$ gehen, so folgt

$$\infty > \pi_j q_j \geq \sum_{i \in I \setminus \{j\}} \pi_i q_{i,j} \geq 0. \tag{3.7.10}$$

(Um das mittlere "$\geq$" zu erhalten, summiere man zunächst über endliche Mengen oder benutze das Lemma von Fatou).

Sei $J \subset I$ endlich. Dann folgt mit der Vorwärtsgleichung (die ja von der Minimallösung erfüllt wird)

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_{i \in J} \pi_i P_{i,j}(t) = \sum_{i \in J} \pi_i \frac{\mathrm{d}}{\mathrm{d}t} P_{i,j}(t) = -q_j \sum_{i \in J} \pi_i P_{i,j}(t) + \sum_{i \in J} \pi_i \sum_{k \in I \setminus \{j\}} P_{i,k}(t) q_{k,j}. \tag{3.7.11}$$

Wie im Beweis von Satz 3.6.5 folgt die Stetigkeit der letzten Summe. Die beiden Terme auf der rechten Seite von (3.7.11) konvergieren mit $J \uparrow I$ jeweils monoton gegen $-q_j \sum_{i \in I} \pi_i P_{i,j}(t) = -q_j \pi_j$ bzw.

$$\sum_{i \in I} \pi_i \sum_{k \in I \setminus \{j\}} P_{i,k}(t) q_{k,j} = \sum_{k \in I \setminus \{j\}} q_{k,j} \sum_{i \in I} \pi_i P_{i,k}(t) = \sum_{k \in I \setminus \{j\}} q_{k,j} \pi_k,$$

da π invariant ist. Der Grenzwert ist jeweils konstant und endlich nach (3.7.10), insbesondere also stetig. Wie im Beweis von Satz 3.6.5 folgt mit dem Satz von Dini und dem Satz über die Vertauschbarkeit von Grenzwert und Ableitung

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} \pi_j = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{i \in I} \pi_i P_{i,j}(t) = -q_j \pi_j + \sum_{k \in I \setminus \{j\}} q_{k,j} \pi_k = \sum_{k \in I} \pi_k q_{k,j}. \tag{3.7.12}$$

Also gilt $\pi Q = 0$.

'$\Longleftarrow$': Es gelte $\pi Q = 0$. Definiere $P^n_{i,j}(t)$ und $H^N_{i,j}(t)$ für $n, N \in \mathbb{N}_0$, und $i, j \in I$ und $t \geq 0$ wie im Beweis von Satz 3.6.2. Wir zeigen per Induktion nach $N \in \mathbb{N}_0$, dass gilt:

$$\sum_{i\in I} \pi_i H^N_{i,j}(t) \leq \pi_j \qquad \text{für alle } j \in I, t \geq 0. \tag{3.7.13}$$

Für $N = 0$ gilt

$$\sum_{i\in I} \pi_i H^0_{i,j}(t) = \sum_{i\in I} \pi_i \delta_{i,j} \mathrm{e}^{-q_j t} = \pi_j \mathrm{e}^{-q_j t} \leq \pi_j. \tag{3.7.14}$$

Gelte also (3.7.13) für ein $N \in \mathbb{N}_0$. Dann folgt

$$\begin{aligned}\sum_{i\in I} \pi_i H^{N+1}_{i,j}(t) &= \pi_j \mathrm{e}^{-q_j t} + \sum_{k\in I\setminus\{j\}} \int_0^t \sum_{i\in I} \pi_i H^N_{i,k}(s) q_{k,j} \mathrm{e}^{-q_j(t-s)}\, \mathrm{d}s \\ &\leq \pi_j \mathrm{e}^{-q_j t} + \sum_{k\in I\setminus\{j\}} \int_0^t \pi_k q_{k,j} \mathrm{e}^{-q_j(t-s)}\, \mathrm{d}s \\ &= \pi_j \mathrm{e}^{-q_j t} + \pi_j q_j \int_0^t \mathrm{e}^{-q_j(t-s)}\, \mathrm{d}s = \pi_j,\end{aligned} \tag{3.7.15}$$

wobei in "$\leq$" die Induktionsvoraussetzung einging und beim vorletzten Gleichheitszeichen die Eigenschaft $\pi Q = 0$.

Da $P_{i,j}(t) = \overline{P}_{i,j}(t) = \lim_{N\uparrow\infty} H^N_{i,j}(t)$, folgt aus (3.7.13)

$$\sum_{i\in I} \pi_i P_{i,j}(t) \leq \pi_j \qquad \text{für alle } j \in I, t \geq 0. \tag{3.7.16}$$

Summiert man über alle $j \in I$, so sieht man, dass in (3.7.16) in Wirklichkeit Gleichheit gilt, d. h. π ist ein invariantes Wahrscheinlichkeitsmaß. □

Satz 3.7.5 zeigt, dass man unter den angegebenen Voraussetzungen alle invarianten Wahrscheinlichkeitsmaße durch Lösen von $\pi Q = 0$ erhält. Dies ist für die Berechnung von π 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 3.7.5. Miller gibt in [Mi63] ein Beispiel, bei dem für ein konservatives Q eine eindeutiges Wahrscheinlichkeitsmaß π mit $\pi Q = 0$ existiert, ohne dass π invariant ist. Die Minimallösung ist in seinem Beispiel nicht Markov'sch.

Wie im zeitlich diskreten Fall ist die Frage nach der Eindeutigkeit von invarianten Maßen eng mit dem Begriff der Irreduzibilität verknüpft.

Definition 3.7.6 (Irreduzibilität). *Eine Standardhalbgruppe $(P(t))_{t\geq 0}$ mit schwach konservativer Q-Matrix Q (oder Q selbst) heißt* irreduzibel, *wenn für alle $i, j \in I$ mit $i \neq j$ ein $n \in \mathbb{N}_0$ und $i = i_0, i_1, \dots, i_n = j$ aus I existieren mit $\prod_{m=0}^{n-1} q_{i_m,i_{m+1}} > 0$.*

Korollar 3.7.7. *Sei Q konservativ und irreduzibel und die Minimallösung Markov'sch. Wenn es ein Wahrscheinlichkeitsmaß π mit $\pi Q = 0$ gibt, dann ist dieses eindeutig, und es gilt $\lim_{t\to\infty} P_{i,j}(t) = \pi_j$ für alle $i, j \in I$.*

Beweis. Nach Satz 3.7.5 ist die Lösung π invariant unter $(P(t))_{t\geq 0}$, also ist π auch ein invariantes Wahrscheinlichkeitsmaß der MKDZ $(X(n))_{n\in\mathbb{N}_0}$ mit Übergangsmatrix $P(1)$. Diese MKDZ ist irreduzibel: Im Beweis von Satz 3.7.2 sahen wir, dass $P_{i,i}(t) > 0$ für alle $t \geq 0$. Für $i \neq j$ seien $i_0, \dots, i_n$ wie bei der Definition 3.7.6 gegeben. Es genügt zu zeigen, dass $i \rightsquigarrow i_1$ für die MKDZ $(X(n))_{n\in\mathbb{N}_0}$ gilt, wobei wir an Definition 2.3.1 erinnern. 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 $P_{i,i_1}(t) \geq P_{i,i}(t-h_0)P_{i,i_1}(h_0) > 0$, womit insbesondere die Irreduzibilität von $P(1)$ gezeigt ist. Wegen $P_{i,i}(1) > 0$ ist die Aperiodizität klar. Also folgt nach Satz 3.7.1 für $j \in I$, dass $\pi_j = \lim_{n\to\infty} P_{i,j}(n)$, und wegen Satz 3.7.2 gilt sogar $\pi_j = \lim_{t\to\infty} P_{i,j}(t)$. Dies zeigt auch die Eindeutigkeit des Wahrscheinlichkeitsmaßes π mit $\pi Q = 0$ in Satz 3.7.5. □

Korollar 3.7.8. *Sei Q konservativ und die Minimallösung Markov'sch. Wenn es kein Wahrscheinlichkeitsmaß π mit $\pi Q = 0$ gibt, dann folgt*

$$\lim_{t\to\infty} P_{ij}(t) = 0 \qquad \textit{für alle } i,j \in I. \tag{3.7.17}$$

Beweis. Nach Satz 3.7.2 existiert $\pi_{i,k} = \lim_{t\to\infty} P_{i,k}(t)$ und es gilt $\sum_{k\in I} \pi_{i,k} \leq 1$. Angenommen, es existieren $i,j \in I$ mit $\lim_{t\to\infty} P_{i,j}(t) = \pi_{i,j} > 0$, dann definiere

$$\overline{\pi}_{i,k} := \frac{\pi_{i,k}}{\sum_{r\in I} \pi_{i,r}}. \tag{3.7.18}$$

Nach Satz 3.7.4 ist $\overline{\pi} = (\overline{\pi}_{i,k})_{k\in I}$ ein invariantes Wahrscheinlichkeitsmaß und erfüllt $\overline{\pi}Q = 0$ im Widerspruch zur Voraussetzung. Also gilt $\pi_{i,j} = 0$ für alle $i,j \in I$. □

3.8 Beispiele invarianter Maße von MKSZ

Beispiel 3.8.1 (Geburts- und Todesprozesse). Für Geburts- und Todesprozesse (siehe Beispiel 3.2.1) 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

$$b_j = \frac{\lambda_0 \dots \lambda_{j-1}}{\mu_1 \dots \mu_j}, \qquad j \in \mathbb{N}, \tag{3.8.1}$$

so kann man zeigen, dass genau im Fall

$$\sum_{i=0}^{\infty} \Big(\frac{1}{\lambda_i b_i} \sum_{j=0}^{i} b_j \Big) = \infty \tag{3.8.2}$$

der Prozess nicht explodiert (d. h. die Minimallösung Markov'sch ist). Wir setzen dies im Folgenden voraus. Solche expliziten notwendige und hinreichende Nichtexplosionskriterien sind übrigens für allgemeine MKSZ nicht bekannt.

Der Geburts- und Todesprozess ist wegen der Voraussetzung der Positivität der λ_i und μ_i offenbar irreduzibel. Anstatt das Gleichungssystem in Satz 3.7.5 direkt zu lösen, führt auch die folgende Überlegung zu einer Lösung:

Wenn ein invariantes Wahrscheinlichkeitsmass π 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:

$$\pi_i \lambda_i = \pi_{i+1} \mu_{i+1} \qquad \text{für alle } i \in \mathbb{N}_0. \tag{3.8.3}$$

Man zeigt leicht, dass (3.8.3) äquivalent zu (i) in Satz 3.7.5 ist. Die Aussage in (3.8.3) erhält man, indem man die ersten $i + 1$ Gleichungen von $\pi Q = 0$ addiert. Aus (3.8.3) folgt für $k \in \mathbb{N}$

$$\pi_k = \pi_{k-1} \frac{\lambda_{k-1}}{\mu_k} = \pi_{k-2} \frac{\lambda_{k-2}\lambda_{k-1}}{\mu_{k-1}\mu_k} = \cdots = \pi_0 b_k. \tag{3.8.4}$$

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 (3.8.4) $\pi_k = 0$ 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

$$\pi_j = \frac{1}{\sum_{k=0}^{\infty} b_k} b_j, \qquad j \in \mathbb{N}_0, \tag{3.8.5}$$

eine Lösung von (3.8.3) 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 3.7.7 folgt dann sogar $\pi_j = \lim_{t\to\infty} P_{ij}(t)$ für alle $i, j \in \mathbb{N}_0$. ◇

Beispiel 3.8.2 (M/M/c-Warteschlange). Wir betrachten die M/M/c-Warteschlange aus Beispiel 3.2.3. Wie wir dort gesehen haben, ist die Anzahl der Kunden insbesondere ein Geburts- und Todesprozess, also können wir Beispiel 3.8.1 anwenden. Mit den dort definierten b_j gilt:

$$\sum_{j=0}^{\infty} b_j = \sum_{j=0}^{c-1} \frac{\lambda^j}{j!\mu^j} + \sum_{j=c}^{\infty} \frac{\lambda^j}{c!c^{j-c}\mu^j}. \tag{3.8.6}$$

Die erste Summe ist immer endlich, die zweite ist endlich genau dann, wenn $\lambda < c\mu$ ist. Also existiert eine invariante Verteilung genau dann, wenn die *Verkehrsdichte* $\rho = \lambda/(c\mu)$ kleiner als Eins ist. In diesem Fall haben wir

$$\sum_{j=0}^{\infty} b_j = \sum_{j=0}^{c-1} \frac{\lambda^j}{j!\mu^j} + \frac{c^c \rho^c}{c!(1-\rho)}, \tag{3.8.7}$$

woraus sich $\pi_j = b_j \big(\sum_{k=0}^{\infty} b_k\big)^{-1}$ explizit berechnen lässt:

$$\pi_j = \pi_0 \Big(\frac{\lambda}{\mu}\Big)^j \times \begin{cases} \frac{1}{j!}, & \text{falls } j \le c, \\ \frac{1}{c!c^{j-c}}, & \text{falls } j \ge c. \end{cases} \tag{3.8.8}$$

Im Spezialfall $c = 1$ erhält man $\pi_j = (1-\rho)\rho^j$ für alle $j \in \mathbb{N}_0$, also eine geometrische Verteilung. ◇

Beispiel 3.8.3 (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 als Eckpunkt eines endlichen Graphen auffasst. Man verbindet i mit j durch einen gerichteten Pfeil und beschriftet ihn mit $\pi_{i,j}$, wenn $\pi_{i,j} > 0$ 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 Bedienzeitverteilung. Nimmt man an, dass alle Bedienzeiten, 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 divergierende Zeit, aber auch die Verweildauerverteilung eines Kunden im System. Wenn alle Bedien- und Zwischenankunftsverteilungen endliche Mischungen von E_k-Verteilungen sind, so kann man mit der in Beispiel 3.2.6 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-Netzwerke* (vgl. [HS82, S. 456 ff]). Wir machen die folgenden Annahmen:

1. Es gibt $N \in \mathbb{N}$ Knoten - durchnumeriert von 1 bis N.
2. Im Knoten i befinden sich c_i Bediener. Die Bedienzeiten seien $\mathrm{Exp}(\mu_i)$-verteilt.
3. Kunden, die von außen (und nicht von einem anderen Knoten) bei Knoten i eintreffen, haben unabhängige $\mathrm{Exp}(\lambda_i)$-verteilte Zwischenankunftszeiten.
4. Gegeben ist eine substochastische $N \times N$-Matrix $P = (p_{i,j})_{i,j=1,\dots,N}$, wobei $p_{i,j}$ die Wahrscheinlichkeit sei, dass ein Kunde nach der Bedienung im Knoten i sich (sofort!) am Knoten j anstellt. Dann interpretieren wir $p_{i,0} := 1 - \sum_{j=1}^{N} p_{i,j}$ als 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 herein- 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:

$$\alpha_i = \lambda_i + \sum_{j=1}^{N} \alpha_j p_{j,i}, \qquad i = 1, \dots, N, \tag{3.8.9}$$

denn was abfließt (α_i) muss auch reinfließen (rechte Seite). Wir nehmen nun zusätzlich an, dass gilt

$$P^n \overset{n\to\infty}{\longrightarrow} 0 \qquad \text{komponentenweise,} \tag{3.8.10}$$

was besagt, dass Kunden von jedem Knoten aus irgendwann das System verlassen – sicher keine allzu starke und eine leicht überprüfbare Voraussetzung an P.

Lemma 3.8.4. *Unter obigen Annahmen hat* (3.8.9) *genau eine Lösung* α, *und zwar*

$$\alpha^{\mathrm{T}} = \lambda^{\mathrm{T}}(I - P)^{-1} = \lambda^{\mathrm{T}} \sum_{k=0}^{\infty} P^k. \tag{3.8.11}$$

Beweis. (3.8.9) kann man in der Form $\alpha^{\mathrm{T}} = \lambda^{\mathrm{T}} + \alpha^{\mathrm{T}} P$, also $\alpha^{\mathrm{T}}(I - P) = \lambda^{\mathrm{T}}$ schreiben. Nun gilt

$$I - P^n = (I - P)(I + P + P^2 + \cdots + P^{n-1}). \tag{3.8.12}$$

Wegen unserer Voraussetzung (3.8.10) ist $I - P^n$ für hinreichend große n invertierbar und daher $\det(I - P^n) \neq 0$. Also ist nach (3.8.12) auch $\det(I - P) \neq 0$, d. h., $I - P$ ist invertierbar, woraus das erste Gleichheitszeichen der Behauptung folgt. Weiter folgt aus (3.8.12) nach Grenzübergang $n \to \infty$ die Gleichung $(I - P)^{-1} = \sum_{k=0}^{\infty} P^k$. □

Wir machen nun noch die weitere Annahme

$$\alpha_i = \sum_{j=1}^{N} \lambda_j \Big(\sum_{k=0}^{\infty} P^k \Big)_{ji} > 0 \qquad \text{für alle } i, \tag{3.8.13}$$

was zum Beispiel aus der stärkeren Forderung $\lambda_j > 0$ für alle j folgt. Nun wählen wir – naheliegenderweise – als Zustandsraum des Netzwerks $I = \mathbb{N}_0^N$, 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 3.8.5 (Jackson). *Unter den obigen Annahmen hat die MKSZ $(X(t))_{t\geq 0}$ genau dann ein invariantes Wahrscheinlichkeitsmaß π, wenn für alle $i = 1, \dots, N$ gilt: $\alpha_i < c_i \mu_i$.*

In diesem Fall ist π eindeutig, und es gilt

$$\pi_{n_1,\dots,n_N} = \Psi_1(n_1) \cdot \dots \cdot \Psi_N(n_N), \tag{3.8.14}$$

wobei $\Psi_i(\cdot)$ das invariante Wahrscheinlichkeitsmaß einer $M/M/c_i$-Schlange mit Ankunftsrate α_i und Bedienrate μ_i ist, also

$$\Psi_i(n) = \Psi_i(0) \Big(\frac{\alpha_i}{\lambda_i} \Big)^n \times \begin{cases} \frac{1}{n!}, & \textit{falls } n \leq c_i, \\ \frac{1}{c_i!} \frac{1}{c_i^{n-c_i}}, & \textit{falls } n \geq c_i. \end{cases} \tag{3.8.15}$$

Bemerkung 3.8.6. 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 3.8.5. Offenbar hat die Q-Matrix folgende Gestalt (mit $\mathfrak{e}_i$ bezeichen wir das Element aus $\mathbb{N}_0^N$ mit einer Eins an der i-ten Stelle und Nullen sonst). Für alle $\mathfrak{n} \in E = \mathbb{N}_0^N$ gilt

$$\begin{aligned} q_{\mathfrak{n},\mathfrak{n}+\mathfrak{e}_i} &= \lambda_i, \\ q_{\mathfrak{n},\mathfrak{n}-\mathfrak{e}_i} &= (n_i \wedge c_i)\mu_i p_{i0}, \\ q_{\mathfrak{n},\mathfrak{n}-\mathfrak{e}_i+\mathfrak{e}_j} &= (n_i \wedge c_i)\mu_i p_{ij}, \end{aligned} \tag{3.8.16}$$

alle anderen q_{ij} mit $i \neq j$ sind Null. Weiter ist Q irreduzibel.

Wir setzen nun voraus, dass $\alpha_i < c_i\mu_i$ für alle i (den Beweis im umgekehrten Fall ersparen wir uns). Definiere π durch (3.8.14). Nach Satz 3.7.5 ist zu zeigen, dass $\pi Q = 0$; die Eindeutigkeit von π folgt dann aus Satz 3.7.7. Wir zeigen sogar, dass für jedes $\mathfrak{n} = (n_1, \dots, n_N)$ gilt

$$\sum_{j=1}^{N} \pi_{\mathfrak{n}+\mathfrak{e}_j} q_{\mathfrak{n}+\mathfrak{e}_j,\mathfrak{n}} = \pi_{\mathfrak{n}} \sum_{j=1}^{N} q_{\mathfrak{n},\mathfrak{n}+\mathfrak{e}_j}, \tag{3.8.17}$$

$$\sum_{j=1}^{N} \pi_{\mathfrak{n}-\mathfrak{e}_j} q_{\mathfrak{n}-\mathfrak{e}_j,\mathfrak{n}} = \pi_{\mathfrak{n}} \sum_{j=1}^{N} q_{\mathfrak{n},\mathfrak{n}-\mathfrak{e}_j}, \tag{3.8.18}$$

$$\sum_{\substack{i=1\\i\neq j}}^{N} \pi_{\mathfrak{n}-\mathfrak{e}_i+\mathfrak{e}_j} q_{\mathfrak{n}-\mathfrak{e}_i+\mathfrak{e}_j,\mathfrak{n}} = \pi_{\mathfrak{n}} \sum_{\substack{i=1\\i\neq j}}^{N} q_{\mathfrak{n},\mathfrak{n}-\mathfrak{e}_i+\mathfrak{e}_j}, \tag{3.8.19}$$

woraus sofort $\pi Q = 0$ folgt.

Tatsächlich zeigen wir nur (3.8.17), da (3.8.18) und (3.8.19) analog bewiesen werden. Die linke Seite von (3.8.17) ist

$$\begin{aligned}
&\sum_{j=1}^{N} \Big(\prod_{\substack{i=1\\i\neq j}}^{N} \Psi_i(n_i)\Big) \Psi_j(n_j+1)((n_j+1)\wedge c_j)\mu_j p_{j0} \\
&= \prod_{i=1}^{N} \Psi_i(n_i) \sum_{j=1}^{N} \frac{\Psi_j(n_j+1)}{\Psi_j(n_j)} \big((n_j+1)\wedge c_j\big)\mu_j p_{j0} \\
&= \prod_{i=1}^{N} \Psi_i(n_i) \sum_{j=1}^{N} \alpha_j p_{j0} = \prod_{i=1}^{N} \Psi_i(n_i) \sum_{j=1}^{N} \alpha_j \Big(1 - \sum_{k=1}^{N} p_{jk}\Big) \\
&= \prod_{i=1}^{N} \Psi_i(n_i) \sum_{j=1}^{N} \lambda_j,
\end{aligned} \tag{3.8.20}$$

was gleich der rechten Seite von (3.8.17) ist, wobei wir beim letzten Gleichheitszeichen die Verkehrsgleichung (3.8.9) und beim vorletzten die explizite Formel für $\Psi_j(n_j+1)/\Psi_j(n_j)$ benutzten. $\square$

$\diamond$

Bibliographie

[Ba91] H. Bauer, *Maß- und Integrationstheorie,* 2. Auflage, Walter de Gruyter, Berlin–New York, 1991.

[Ba02] H. Bauer, *Wahrscheinlichkeitstheorie*, 5. Auflage, Walter de Gruyter, Berlin–New York, 2002.

[Bi86] P. Billingsley, *Probability and Measure*, Wiley, New York 1986.

[Ch78] K. L. Chung, *Elementare Wahrscheinlichkeitstheorie und stochastische Prozesse*, Springer, Berlin, 1978.

[Br68] L. Breiman, *Probability*, Addison-Wesley, Reading, 1968.

[Ch67] K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, Springer, Berlin 1967.

[Co80] D.L. Cohn, *Measure Theory,* Birkhäuser, Basel 1980.

[Fe68] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd ed., Wiley, New York, 1968.

[Ge02] H.-O. Georgii, *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Walter de Gruyter, 2002.

[GS75] I. I. Gikhman und A. V. Skorohod, *The Theory of Stochastic Processes II*, Springer, Berlin, 1975.

[H02] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, 2002.

[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.

[Kr02] U. Krengel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 6. Auflage, Vieweg, 2002.

[Mi63] R. Miller, Stationary equations in continuous time Markov chains, *Trans. Amer. Math. Soc.* **109**, 35-44, 1963.

[No97] J. Norris, *Markov Chains*, Cambridge University Press, 1997.

[Pi90] B. Pittel, On a Daley-Kendall model of random rumours, *Journal of Applied Probability* **27**, 1990.

[Sch98] K. Schürger, *Wahrscheinlichkeitstheorie*, Oldenbourg, München, 1998.

[Se81] E. Seneta, *Non-Negative Matrices and Markov Chains*, Springer, New York, 1981.

[Wi79] D. Williams, *Diffusions, Markov Processes and Martingales*, Wiley, 1979.

Gut lesbare Einführungen in die Theorie der Markovketten sind [Ch67], [Se81], [H02] und [No97], wobei [Ch67] hauptsächlich die Theorie der stationären Markovketten in stetiger Zeit sehr umfassend behandelt, [Se81] stochastische Matrizen hauptsächlich als Spezialfall nichtnegativer Matrizen auffasst und deren Frobenius-Eigenwerttheorie zum Hauptthema hat, und [H02] und [No97] sich auf eine elementare Einführung von Markovketten in diskreter Zeit konzentrieren mit vielen Beispielen (insbesondere algorithmischen Anwendungen in [H02]), aber nicht ganz so viel Material. Die Lehrbücher [Ch78], [Ge02] und [Kr02] sind hauptsächlich der Elementaren Wahrscheinlichkeitstheorie gewidmet und dienen zum Nacharbeiten von wichtigen Beispielen und Anwendungen der bekannten diskreten und absolutstetigen Verteilungen. Die Lehrbücher [Fe68], [Bi86], [Ba02], [Sch98] dienen dem Nachschlagen benötigter Hilfsmittel der Wahrscheinlichkeitstheorie, und [Co80] und [Ba91] informieren über die benötigte Maßtheorie.