

Studiengang Diplom-Mathematik mit Schwerpunkt Biowissenschaften

DIPLOMARBEIT

Abschätzungen der Konvergenzgeschwindigkeit von Markov-Ketten gegen die Gleichgewichtsverteilung

von: Christina Boll geb. Wolf

eingereicht am: 24.03.2009

Erstgutachterin: Prof. Dr. Christine Müller

Zweitgutachter: Dr. Peter Harmand

Inhaltsverzeichnis

1	Einleitung	4
2	Einführung in Markov-Ketten	6
2.1	Allgemeines über Markov-Ketten	6
2.1.1	Homogene Markov-Ketten und Markov-Eigenschaft	6
2.1.2	Übergangsmatrix und Startverteilung	8
2.1.3	Verteilung zum Zeitpunkt k und k -stufige Übergangsmatrix . .	11
2.2	Eigenschaften von Markov-Ketten	13
2.2.1	Irreduzible Markov-Ketten und Erneuerungsgleichung	13
2.2.2	Aperiodische Markov-Ketten	16
2.2.3	Ergodische Markov-Ketten	17
3	Konvergenz gegen die Gleichgewichtsverteilung	20
3.1	Die Gleichgewichtsverteilung	20
3.1.1	Definition und Beispiele	20
3.1.2	Zentrale Fragestellungen	22
3.2	Die Gleichgewichtsverteilung ergodischer Markov-Ketten	23
3.2.1	Existenz und Eindeutigkeit	23
3.2.2	Direkte Konvergenzabschätzung Teil 1	28
3.2.3	Direkte Konvergenzabschätzung Teil 2	32
3.2.4	Konvergenzabschätzung mittels Eigenwerten	35
3.3	Die Gleichgewichtsverteilung reversibler Markov-Ketten	39
3.3.1	Reversible Markov-Ketten	39
3.3.2	Präzisierte Konvergenzabschätzung mittels Eigenwerten	41
3.3.3	χ^2 -Kontrast	45
3.3.4	Die multiplikativ reversible Version der Übergangsmatrix . . .	50
3.4	Charakterisierung und Abschätzung der Eigenwerte	55
3.4.1	Das Theorem von Rayleigh	55
3.4.2	Der Poincaré-Koeffizient und die Poincaré-Ungleichung	62

3.4.3	Abschätzung des kleinsten Eigenwertes	66
3.5	Berechnung der Gleichgewichtsverteilung	71
3.5.1	Direkte Berechnungsmethode	72
3.5.2	Iterative Berechnungsmethode	75
4	Anhang	77
4.1	Grundlagen aus der linearen Algebra	77
4.1.1	Eigenwerte und Eigenvektoren	77
4.1.2	Spektraldarstellung	78
4.1.3	Skalarprodukt und Norm	81
4.1.4	Symmetrische Matrizen	83

1 Einleitung

Diese Diplomarbeit beschäftigt sich mit Markov-Ketten und ihrem Konvergenzverhalten. Der Schwerpunkt liegt dabei auf der Abschätzung der Konvergenzgeschwindigkeit von Markov-Ketten gegen die sogenannte Gleichgewichtsverteilung.

Im Allgemeinen sind Markov-Ketten Zufallsprozesse, die die zeitliche Entwicklung von bestimmten Modellen, Systemen u.s.w. beschreiben. Diese Modelle bzw. Systeme müssen dabei so aufgebaut sein, dass sie jeweils nur eine Ausprägung je Zeitpunkt annehmen können. Auf diese Weise finden Markov-Ketten eine Vielzahl von Anwendungen in sämtlichen Bereichen. Zum Beispiel sind die täglich in einer Wetterstation gemessene Temperatur eines Ortes oder die aktuelle Anzahl von Personen, die an einer Infektionskrankheit erkrankt sind zwei Gegebenheiten, die durch eine Markov-Kette dargestellt werden können.

Diese Diplomarbeit beginnt mit dem Kapitel Einführung in Markov-Ketten. In diesem Kapitel 2 werden die Grundkenntnisse über Markov-Ketten vermittelt. In 2.1 wird die Markov-Eigenschaft, sowie die Homogenität einer Markov-Kette zur Sprache gebracht. Des Weiteren werden in 2.1 die Übergangsmatrix und die Startverteilung einer Markov-Kette eingeführt und die Verteilung einer Markov-Kette zum Zeitpunkt $k \in \mathbb{N}$ beschrieben. In 2.2 werden dann bestimmte Arten von Markov-Ketten vorgestellt, die irreduziblen, aperiodischen und ergodischen Markov-Ketten. Die Einführung in Markov-Ketten wird bewusst so kurz wie möglich gehalten. Dort werden nur die Kenntnisse über Markov-Ketten vermittelt, die für das Verständnis von Kapitel 3, der Konvergenz gegen die Gleichgewichtsverteilung, vonnöten sind.

Kapitel 3 ist das Kernstück dieser Diplomarbeit. Hier wird sich mit dem Konvergenzverhalten von Markov-Ketten gegen die Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ beschäftigt. Zu Beginn dieses Kapitels wird in 3.1 der Begriff der Gleichgewichtsverteilung geklärt. Des Weiteren sind dort die in diesem Zusammenhang zentralen Fragen angegeben. Diese Fragen werden im Laufe des Kapitels 3 beantwortet.

Das Teilkapitel 3.2 geht auf die Gleichgewichtsverteilung ergodischer Markov-Ketten ein. Es werden die Existenz und Eindeutigkeit einer Gleichgewichtsverteilung ergo-

discher Markov-Ketten gezeigt und drei Abschätzungen der Konvergenzgeschwindigkeit ergodischer Markov-Ketten gegen die Gleichgewichtsverteilung angegeben. Zwei dieser Abschätzungen sind sogenannte direkte Abschätzungen und die dritte Konvergenzabschätzung wird über den betragsmäßig zweitgrößten Eigenwert der in Kapitel 2 eingeführten Übergangsmatrix einer Markov-Kette gemacht.

In 3.3 werden reversible Markov-Ketten eingeführt. Nachdem geklärt wird, was reversible Markov-Ketten sind, wird wieder auf die Abschätzung der Konvergenzgeschwindigkeit, in diesem Fall ergodisch reversibler Markov-Ketten, eingegangen. Die zwei Schranken der Konvergenzgeschwindigkeit, die dort angegeben sind, nutzen beide einen Eigenwert der Übergangsmatrix der Markov-Kette. Die erste Konvergenzabschätzung ist eine Präzisierung der Konvergenzabschätzung mittels Eigenwerten aus 3.2 und die zweite Abschätzung der Konvergenzgeschwindigkeit wird mit Hilfe des χ^2 -Kontrasts gemacht. Zum Abschluss von 3.3 wird die multiplikativ reversible Version der Übergangsmatrix eingeführt. Mit diesen Kenntnissen kann die Übergangsmatrix einer Markov-Kette, die nur ergodisch und nicht reversibel ist, so umgeformt werden, dass sich Reversibilität einstellt.

Im nächsten Teilkapitel 3.4 werden die Eigenwerte der Übergangsmatrix einer Markov-Kette charakterisiert und der zweitgrößte, sowie der kleinste Eigenwert abgeschätzt. Die Charakterisierung gibt das Theorem von Rayleigh an. Damit erfolgt dann auch eine erste Abschätzung des zweitgrößten Eigenwertes. Eine weitere Abschätzung des zweitgrößten Eigenwertes einer Markov-Kette, die insbesondere die vorherige präzisiert, ist mit Hilfe des Poincaré-Koeffizienten in der Poincaré-Ungleichung angegeben. Auf ähnliche Weise wird außerdem eine Abschätzung des kleinsten Eigenwertes gemacht.

Das letzte, in Kapitel 3 enthaltene, Teilkapitel befasst sich mit der Berechnung der Gleichgewichtsverteilung. Dort wird eine direkte, so wie eine iterative Berechnungsmethode für die Gleichgewichtsverteilung π angegeben.

2 Einführung in Markov-Ketten

In diesem Kapitel werden die für die weitere Arbeit nötigen Grundkenntnisse über Markov-Ketten vermittelt.

2.1 Allgemeines über Markov-Ketten

2.1.1 Homogene Markov-Ketten und Markov-Eigenschaft

Bevor eine Markov-Kette definiert wird, sei hier eine dafür wichtige Definition aus der Wahrscheinlichkeitstheorie angegeben.

Definition 1 Eine Sequenz $(X_0, X_1, X_2, \dots) = \{X_k\}_{k \in \mathbb{N}}$ von Zufallsvariablen mit Werten in einer Menge S wird als stochastischer Prozess (oder auch Zufallsprozess) mit Zustandsraum S bezeichnet. (Vergl. [Häg02] S. 3.) \square

Hierbei und im Folgenden wird die Null immer als ein Element der natürlichen Zahlen betrachtet.

Insbesondere sind Markov-Ketten eine spezielle Klasse von stochastischen Prozessen. In dieser Arbeit wird stets angenommen, dass es sich bei dem Zustandsraum S um eine endliche Menge handelt. Es sei $S := \{1, 2, \dots, N\}$. Oft werden im Folgenden die Elemente aus S mit i, j, k oder l bezeichnet. Wenn $X_k = i$ ($i \in S, k \in \mathbb{N}$) ist, dann spricht man davon, dass sich der stochastische Prozess zum Zeitpunkt k im Zustand i befindet bzw. Zustand i zum Zeitpunkt k besucht wird.

Damit kann nun eine Markov-Kette definiert werden.

Definition 2 (Markov-Kette) Ein stochastischer Prozess $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum $S = \{1, \dots, N\}$ wird als Markov-Kette bezeichnet, wenn für alle $k \in \mathbb{N}$ und alle Zustände $i_0, i_1, \dots, i_{k-1}, i, j \in S$

$$\mathbb{P}(X_{k+1} = j \mid X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{k+1} = j \mid X_k = i) \quad (2.1)$$

ist, wobei beide Seiten wohl definiert sein müssen. (Vergl. [Häg02] S. 10.) \square

Bei beiden Ausdrücken in (2.1) handelt es sich um bedingte Wahrscheinlichkeiten. Im Allgemeinen bezeichnen bedingte Wahrscheinlichkeiten das Eintreten eines Er-

eignisses unter der Bedingung, dass ein anderes Ereignis bereits eingetreten ist. Für die beliebigen Ereignisse A und B mit $\mathbb{P}(B) > 0$ gilt

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} . \quad (2.2)$$

Durch Umformung erhält man

$$\mathbb{P}(A \cap B) = \mathbb{P}(A | B)\mathbb{P}(B) . \quad (2.3)$$

Eine Markov-Kette wird als homogene Markov-Kette bezeichnet, wenn die rechte Seite in (2.1) außerdem unabhängig von k ist. Im Folgenden wird stets davon ausgegangen, dass es sich bei den betrachteten Markov-Ketten um homogene Markov-Ketten handelt.

Die Eigenschaft (2.1) wird in der Literatur auch als Markov-Eigenschaft bzw. „memoryless property“ bezeichnet. Besonders die letzte Bezeichnung ergibt sich daraus, dass die bedingte Wahrscheinlichkeit von X_{k+1} bei gegebenen X_0, \dots, X_k nur von X_k abhängt. Um also eine Vorhersage zu treffen, was „morgen“ (Zeitpunkt $k+1$) passiert, muss nur betrachtet werden, was „heute“ (Zeitpunkt k) geschieht. Die „Vergangenheit“ (Zeitpunkte 0 bis $k-1$) gibt keine weiteren brauchbaren Informationen.

Nun wird ein erstes sehr einfaches aber anschauliches Beispiel einer homogenen Markov-Kette angegeben.

Beispiel 1 [Läufer] Es sei ein Rechteck wie in Abbildung 2.1 gegeben. Man stelle sich einen Läufer vor, der sich beispielsweise in der Ecke 1 befindet. Von dort aus bewegt er sich nun jeweils mit Wahrscheinlichkeit $\frac{1}{2}$ entweder zur Ecke 2 oder zur Ecke 4. Dort angekommen entscheidet er sich für seinen nächsten Lauf ebenfalls zufällig (mit Wahrscheinlichkeit $\frac{1}{2}$) für eine der benachbarten Ecken u.s.w.. Für jeden Zeitpunkt $k \in \mathbb{N}$ bezeichne X_k die Ecke, in der sich der Läufer zur Zeit k befindet. Folglich ist $\{X_k\}_{k \in \mathbb{N}}$ ein stochastischer Prozess mit Werten in $S = \{1, 2, 3, 4\}$. Nehme man beispielsweise an, der Läufer befinde sich zum Zeitpunkt $k \in \mathbb{N}$ in der Ecke 2. Von dort aus hat er die Möglichkeit, sich im nächsten Schritt zur Ecke 1 oder Ecke 3 zu begeben. Damit erhält man die bedingten Wahrscheinlichkeiten

$$\mathbb{P}(X_{k+1} = 1 | X_k = 2) = \frac{1}{2}$$

und

$$\mathbb{P}(X_{k+1} = 3 | X_k = 2) = \frac{1}{2} .$$

Die gleichen bedingten Wahrscheinlichkeiten erhält man, wenn man die Stand-

orte des vorhergehenden Prozesses mit in die Bedingung aufnimmt. Für beliebige i_0, \dots, i_{k-1} gilt nämlich

$$\mathbb{P}(X_{k+1} = 1 \mid X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1}, X_k = 2) = \frac{1}{2}$$

und

$$\mathbb{P}(X_{k+1} = 3 \mid X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1}, X_k = 2) = \frac{1}{2}.$$

Dies liegt daran, dass die zufällige Entscheidung zum Zeitpunkt $k + 1$ unabhängig von den vorherigen Entscheidungen und damit auch unabhängig von X_0, \dots, X_k ist. Des Weiteren ist die bedingte Wahrscheinlichkeit von X_{k+1} bei gegebenem X_k für alle Zeitpunkte $k \in \mathbb{N}$ die gleiche. Insgesamt handelt es sich somit bei $\{X_k\}_{k \in \mathbb{N}}$ um eine homogene Markov-Kette. (Vergl. [Häg02] S. 8 und 9.) \square

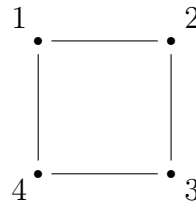


Abbildung 2.1: Lauf auf einem Rechteck

2.1.2 Übergangsmatrix und Startverteilung

Definition 3 (Übergangsmatrix) Die $N \times N$ Matrix $P = (p_{ij})_{i,j \in S}$ der Übergangswahrscheinlichkeiten

$$p_{ij} := \mathbb{P}(X_{k+1} = j \mid X_k = i) \text{ mit} \\ p_{ij} \geq 0 \text{ und } \sum_{j=1}^N p_{ij} = 1 \text{ für alle } i, j \in S \quad (2.4)$$

wird als Übergangsmatrix der Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ bezeichnet. (Vergl. [Häg02] S. 10.) \square

Die Bezeichnung Übergangswahrscheinlichkeit bzw. Übergangsmatrix rührt daher, dass p_{ij} die Wahrscheinlichkeit angibt, vom Zustand i im nächsten Schritt zum Zustand j zu gelangen.

Definition 4 (stochastische Matrix) Jede quadratische Matrix, die den Bedingungen in (2.4) genügt, wird als stochastische Matrix bezeichnet. \square

Neben der Übergangsmatrix P spielt die sogenannte Startverteilung („initial distribution“) auch eine wichtige Rolle beim Thema Markov-Ketten. Sie gibt an, wie eine Markov-Kette gestartet wird.

Definition 5 (Startverteilung) Die Startverteilung einer Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ ist ein Vektor $\alpha^{(0)} \in \mathbb{R}^N$, der gegeben ist durch

$$\begin{aligned} \alpha^{(0)} &:= (\alpha_1^{(0)}, \dots, \alpha_N^{(0)})^T \\ &:= (\mathbb{P}(X_0 = 1), \dots, \mathbb{P}(X_0 = N))^T . \end{aligned}$$

(Vergl. [Häg02] S. 10 und 11.) □

Die Zufallsvariable X_0 wird hierbei als Start-Zustand bezeichnet. Da $\alpha^{(0)}$ eine Wahrscheinlichkeitsverteilung darstellt, gilt $\sum_{i=1}^N \alpha_i^{(0)} = 1$.

Mit diesen Kenntnissen ist es auch möglich, eine Markov-Kette über bzw. als das Tripel $(S, \alpha^{(0)}, P)$ zu definieren (siehe hierzu [Beh00] S. 4).

Nun werden noch weitere Beispiele sowie eine Ergänzung zu Beispiel 1 zur Veranschaulichung der Thematik angegeben.

Beispiel 2 (Ergänzung zum Beispiel des Läufers (Beispiel 1)) Die zur Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ aus Beispiel 1 gehörende 4×4 Übergangsmatrix P hat folgende Gestalt

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} .$$

Wenn außerdem davon ausgegangen wird, dass der Läufer auf jeden Fall in Ecke 1 startet, dann ist $\alpha^{(0)} = (1, 0, 0, 0)$ die Startverteilung der Markov-Kette. (Vergl. [Häg02] S. 10 und 11.) □

Beispiel 3 (Wettervorhersage) 1. Zunächst betrachte man Regionen, in denen sich typischerweise längere Regen- bzw. Trockenperioden abwechseln. Regentage sowie Sonnentage sollen dabei im Mittel etwa gleich oft vorkommen. Zur Vereinfachung werden nur die zwei Wetterlagen v_1 :=“Regen“ und v_2 :=“Sonne“ betrachtet. Eine mögliche Wettervorhersage für den nächsten Tag wäre nun, dass man zum Beispiel annimmt, das Wetter würde beständig sein und sich nicht ändern. Diese Prognose sei in 75% der Fälle richtig, unabhängig davon, ob am gegenwärtigen Tag die Sonne scheint oder ob es regnet. Auf diese Weise bildet die Wettervorhersage

eine Markov-Kette mit Zustandsraum $S = \{v_1, v_2\}$ und Übergangsmatrix

$$P = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}.$$

2. Für Regionen, in denen keine Symmetrie zwischen Regen- und Sonnentagen wie in 1. vorliegt, sondern beispielsweise erheblich mehr Regentage als Sonnentage vorkommen, sollte eine andere Übergangsmatrix gewählt werden. Zum Beispiel könnte in diesem Fall

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$$

sein. (Vergl. [Häg02] S. 12.) □

Beispiel 4 (Zufällige Irrfahrten („Random Walks“)) Zufällige Irrfahrten sind klassische Beispiele für Markov-Ketten. Das Basismodell ist auf folgende Weise gegeben:

Sei $Z_0, Z_1, Z_2, \dots : \Omega \rightarrow S = \{1, \dots, N\}$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen. Sei außerdem $X_0 : \Omega \rightarrow S$ eine beliebige Zufallsvariable, die unabhängig von Z_1, Z_2, \dots ist, und sei

$$X_k = X_{k-1} + Z_k \text{ für alle } k \geq 1 \in \mathbb{N}.$$

Damit bilden die Zufallsvariablen X_0, X_1, X_2, \dots eine Markov-Kette mit Zustandsraum S , Startverteilung $\alpha^{(0)} = (\alpha_1^{(0)}, \alpha_2^{(0)}, \dots)^T$, wobei $\alpha_i^{(0)} = \mathbb{P}(X_0 = i)$ ist, und Übergangswahrscheinlichkeiten $p_{ij} = \mathbb{P}(Z_0 = j - i)$ ($i, j \in S$). (Man gelangt somit von jedem Punkt $i \in S$ zu jedem Punkt $j \in S$ mit der Wahrscheinlichkeit p_{ij} , wobei p_{ij} im Übrigen auch Null sein kann.) (Vergl. [Sch03] S. 8.) Es soll ein Sprung innerhalb von S mit Wahrscheinlichkeit a_i ($i \in S$) zum linken Nachbarn und mit Wahrscheinlichkeit c_i zum rechten Nachbarn gemacht werden. Mit Wahrscheinlichkeit b_i wird in dem Punkt verblieben, in dem man sich gerade befindet. Es gelte $a_i + b_i + c_i = 1$ für alle $i \in S$.

Je nachdem, wie bei den Randpunkten 1 und N verfahren wird oder ob spezielle Regeln bei der Irrfahrt gelten, gibt es verschiedene Arten von Irrfahrten auf $S = \{1, \dots, N\}$. Nun werden drei solcher Arten von Irrfahrten auf S vorgestellt. Um nicht all zu große Übergangsmatrizen angeben zu müssen wird $N = 4$ gesetzt. Natürlich lässt sich das Folgende für jede Wahl von N analog darstellen.

Stets sei $\alpha^{(0)}$ beliebig und a_i, b_i und c_i ($i \in S$) wie im obigen Text beschrieben.

1. *reflektierende Irrfahrt*

Bei der reflektierenden Irrfahrt auf $S = \{1, 2, 3, 4\}$ sei

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ a_2 & b_2 & c_2 & 0 \\ 0 & a_3 & b_3 & c_3 \\ 0 & 0 & 1 & 0 \end{pmatrix} .$$

2. *absorbierende Irrfahrt*

Bei der absorbierenden Irrfahrt auf $S = \{1, 2, 3, 4\}$ sei

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a_2 & b_2 & c_2 & 0 \\ 0 & a_3 & b_3 & c_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

3. *zyklische Irrfahrt*

Bei der zyklischen Irrfahrt auf $S = \{1, 2, 3, 4\}$ sei

$$P = \begin{pmatrix} b_1 & c_1 & 0 & a_1 \\ a_2 & b_2 & c_2 & 0 \\ 0 & a_3 & b_3 & c_3 \\ c_4 & 0 & a_4 & b_4 \end{pmatrix} .$$

Eine zyklische Irrfahrt heißt insbesondere symmetrisch, wenn $a_i = \frac{1}{2}$, $c_i = \frac{1}{2}$ und dementsprechend $b_i = 0$ für alle $i \in S$ ist. (Bei Beispiel 1 handelt es sich somit um eine symmetrische zyklische Irrfahrt.) (Vergl. [Beh00] S. 12 und 13.) \square

2.1.3 Verteilung zum Zeitpunkt k und k -stufige Übergangsmatrix

Analog zur Startverteilung (also der Verteilung der Markov-Kette zum Zeitpunkt 0) kann auch die Verteilung der Markov-Kette zum Zeitpunkt k definiert werden. Es sei

$$\begin{aligned} \alpha^{(k)} &:= (\alpha_1^{(k)}, \dots, \alpha_N^{(k)})^T \\ &:= (\mathbb{P}(X_k = 1), \dots, \mathbb{P}(X_k = N))^T \text{ für alle } k \in \mathbb{N} . \end{aligned}$$

Der nun folgende Satz gibt einen Zusammenhang zwischen Startverteilung, Verteilung zum Zeitpunkt k und Übergangsmatrix einer Markov-Kette an.

Satz 1 Für eine Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum S , Startverteilung $\alpha^{(0)} \in \mathbb{R}^N$ und $N \times N$ Übergangsmatrix P , gilt für alle $k \in \mathbb{N}$

$$\alpha^{(k)T} = \alpha^{(0)T} P^k . \quad (2.5)$$

$\alpha^{(k)}$ bezeichnet hierbei die Verteilung der Markov-Kette zum Zeitpunkt $k \in \mathbb{N}$ und $P^k = (p_{ij}^{(k)})_{i,j=1,\dots,N}$ die k -te Potenz der Übergangsmatrix P . (Vergl. [Häg02] S. 11.) \square

Mit anderen Worten besagt Satz 1, dass sämtliche Verteilungen $\alpha^{(k)}$ ($k \in \mathbb{N}$) einer Markov-Kette allein schon durch die Startverteilung $\alpha^{(0)}$ und die Übergangsmatrix P gegeben sind und demnach dadurch auch bestimmt werden können.

Nun wird Satz 1 per Induktion nach k bewiesen.

BEWEIS (VON SATZ 1) Für den Fall $k = 1$ gilt für alle $j = 1, \dots, N$, dass

$$\begin{aligned} \alpha_j^{(1)} &= \mathbb{P}(X_1 = j) = \sum_{i \in S} \mathbb{P}(X_0 = i, X_1 = j) \\ &\stackrel{(2.3)}{=} \sum_{i \in S} \mathbb{P}(X_0 = i) \mathbb{P}(X_1 = j \mid X_0 = i) \\ &= \sum_{i \in S} \alpha_i^{(0)} p_{ij} = (\alpha^{(0)T} P)_j \end{aligned}$$

ist. $(\alpha^{(0)T} P)_j$ ist hierbei der j -te Eintrag des Vektors $\alpha^{(0)T} P$. Folglich gilt damit $\alpha^{(1)T} = \alpha^{(0)T} P$.

Für den Fall $k + 1$ gilt ebenso

$$\begin{aligned} \alpha_j^{(k+1)} &= \mathbb{P}(X_{k+1} = j) = \sum_{i \in S} \mathbb{P}(X_k = i, X_{k+1} = j) \\ &\stackrel{(2.3)}{=} \sum_{i \in S} \mathbb{P}(X_k = i) \mathbb{P}(X_{k+1} = j \mid X_k = i) = \sum_{i \in S} \alpha_i^{(k)} p_{ij} = (\alpha^{(k)T} P)_j \end{aligned}$$

und somit ist $\alpha^{(k+1)T} = \alpha^{(k)T} P$. Mit der Induktionsannahme, dass (2.5) für ein beliebiges $k \in \mathbb{N}$ gilt, folgt insgesamt für $k + 1$

$$\alpha^{(k+1)T} = \alpha^{(k)T} P = \alpha^{(0)T} P^k P = \alpha^{(0)T} P^{k+1} .$$

Nach dem Prinzip der vollständigen Induktion ist damit die Aussage für alle $k \in \mathbb{N}$ gültig und damit der Satz bewiesen. (Vergl. [Häg02] S. 11 und 12.) \blacksquare

P^k wird oft als die k -stufige Übergangsmatrix der Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ bezeichnet. Für die einzelnen Elemente von P^k gilt

$$p_{ij}^{(k)} = \mathbb{P}(X_k = j \mid X_0 = i) \text{ für alle } k \in \mathbb{N} \text{ und } i, j \in S , \quad (2.6)$$

falls $\mathbb{P}(X_0 = i) > 0$ ist. Dies ist äquivalent mit

$$p_{ij}^{(k)} = \mathbb{P}(X_{m+k} = j \mid X_m = i) \text{ für alle } k, m \in \mathbb{N} \text{ und } i, j \in S, \quad (2.7)$$

wobei $\mathbb{P}(X_m = i) > 0$ sein muss.

Eine übliche Schreibweise für die Wahrscheinlichkeit $\mathbb{P}(\cdot \mid X_0 = i)$ ($i \in S$) ist auch $\mathbb{P}_i(\cdot)$. Dabei verdeutlicht das $i \in S$, dass die Markov-Kette auf jeden Fall im Zustand i startet. Damit gilt

$$p_{ij}^{(k)} = \mathbb{P}_i(X_k = j) \text{ für alle } k \in \mathbb{N} \text{ und } i, j \in S. \quad (2.8)$$

Im Kontext der k -stufigen Übergangsmatrizen ist die sogenannte Chapman-Kolmogorov-Gleichung (C.-K.-G.) von großer Bedeutung. Sie besagt, dass für beliebige $k, m \in \mathbb{N}$

$$P^{k+m} = P^k P^m \quad (2.9)$$

ist bzw. dass damit für die einzelnen Einträge der Matrix P^{k+m}

$$p_{ij}^{(k+m)} = \sum_{l \in S} p_{il}^{(k)} p_{lj}^{(m)} \text{ für alle } i, j \in S \quad (2.10)$$

gilt. Hierbei ist (2.9) im Rahmen der allgemeinen Multiplikations- und Potenzierungsregeln quadratischer Matrizen gültig.

2.2 Eigenschaften von Markov-Ketten

Es gibt im allgemeinen verschiedene Arten von Markov-Ketten. Damit ist gemeint, dass eine Markov-Kette ganz spezifische Eigenschaften aufweisen bzw. besonderen Bedingungen genügen kann. Einige dieser besonderen Bedingungen spielen z.B. eine Schlüsselrolle bei der Betrachtung von Gleichgewichtsverteilungen, die in dieser Arbeit von zentraler Bedeutung sind. Im Folgenden werden nur diese speziellen Markov-Ketten vorgestellt, die für die weiteren Ausführungen von Interesse sind.

2.2.1 Irreduzible Markov-Ketten und Erneuerungsgleichung

Eine Art von Markov-Ketten sind die irreduziblen Markov-Ketten. Lapidar kann man sagen, dass eine Markov-Kette irreduzibel ist, wenn jeder Zustand, in dem sie sich befindet, von wiederum jedem Zustand aus erreicht werden kann. Um dies nun mathematisch korrekt auszudrücken, werden vorab einige dafür nötige Bezeichnungen eingeführt.

Für $i, j \in S$ schreibt man $i \rightarrow j$, falls die Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum S und $N \times N$ Übergangsmatrix P mit echt positiver Wahrscheinlichkeit irgendwann von Zustand i nach Zustand j gelangt. Mit anderen Worten bedeutet dies, dass ein $k \in \mathbb{N}$ existiert, so dass $p_{ij}^{(k)} > 0$ ist. Des Weiteren sagt man $i, j \in S$ kommunizieren ($i \leftrightarrow j$), wenn sowohl $i \rightarrow j$ als auch $j \rightarrow i$ gilt.

Damit kann nun eine Definition von Irreduzibilität angegeben werden.

Definition 6 Eine Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum S heißt irreduzibel, wenn alle Zustände $i, j \in S$ kommunizieren. (Vergl. [Häg02] S. 24.) \square

Dies bedeutet somit, dass im Falle irreduzibler Markov-Ketten für jedes $i, j \in S$ ein $k \in \mathbb{N}$ gefunden werden kann, so dass $p_{ij}^{(k)} > 0$ ist.

Dazu einige Beispiele.

Beispiel 5 Aus der Definition der Irreduzibilität ergibt sich direkt, dass die Markov-Kette mit der 2×2 Übergangsmatrix

$$P_1 = \begin{pmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{pmatrix},$$

sowie die Markov-Kette mit der 2×2 Übergangsmatrix

$$P_2 = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}$$

irreduzibel ist. Jeder Zustand kann hier von jedem Zustand aus erreicht werden. Im Gegensatz dazu ist die Markov-Kette, deren Übergangsmatrix die aus P_1 und P_2 gebildete 4×4 Blockmatrix

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$$

ist, nicht irreduzibel. Zu keinem Zeitpunkt $k \in \mathbb{N}$ ist es möglich, von Zustand 1 oder 2 zum Zustand 3 oder 4 überzugehen und umgekehrt. Dies liegt daran, dass P^k für alle $k \in \mathbb{N}$ eine Blockstruktur beibehält und die Blöcke nicht wieder verlassen werden können. (Vergl. [Sch03] S. 26.) \square

Beispiel 6 (Ergänzung zum Beispiel des Läufers (Beispiel 1)) Die Markov-Kette aus Beispiel 1 ist irreduzibel. Dies verdeutlicht allein schon die Abbildung 2.1. Egal wo der Läufer startet, er kann sich gegebenenfalls über andere Ecken hinweg zu jeder beliebigen Ecke begeben. \square

Beispiel 7 (Ergänzung zu den zufälligen Irrfahrten (Beispiel 4)) Die reflektierende und die zyklische Irrfahrt sind irreduzibel. Dem entgegen ist die absorbie-

rende Irrfahrt nicht irreduzibel, da die Zustände 1 und N zu keiner Zeit verlassen werden können. \square

Nun folgen zwei weitere Definitionen, durch die anschließend eine Eigenschaft irreduzibler Markov-Ketten angegeben werden kann.

Definition 7 Für die Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum S , $N \times N$ Übergangsmatrix P , $k \in \mathbb{N}$ und $i, j \in S$ sei

$$(a) f_{ij}^{(k)} := \mathbb{P}(X_1 \neq j, \dots, X_{k-1} \neq j, X_k = j \mid X_0 = i) \\ = \mathbb{P}_i(X_1 \neq j, \dots, X_{k-1} \neq j, X_k = j) ,$$

d.h. $f_{ij}^{(k)}$ ist die Wahrscheinlichkeit, dass die Markov-Kette bei Start in i , den Zustand j zum ersten Mal im k -ten Schritt erreicht.

$$(b) f_{ij}^* := \sum_{k=1}^{\infty} f_{ij}^{(k)} .$$

f_{ij}^* bezeichnet die Wahrscheinlichkeit, dass die Markov-Kette bei Start in i irgendwann j erreicht.

$$(c) \mu_{ij} := \begin{cases} \sum_{k=1}^{\infty} k f_{ij}^{(k)} + \infty (1 - f_{ij}^*) & \text{falls } f_{ij}^* < 1 \\ \sum_{k=1}^{\infty} k f_{ij}^{(k)} & \text{falls } f_{ij}^* = 1 . \end{cases}$$

μ_{ij} bezeichnet die mittlere Anzahl von Schritten um von i nach j zu kommen.

(Vergl. [Beh00] S. 29.) \square

Definition 8 (a) $i \in S$ heißt rekurrent, falls $f_{ii}^* = 1$ ist.

(b) $i \in S$ heißt positiv rekurrent, falls $\mu_{ii} < \infty$ ist. (Vergl. [Beh00] S. 29.) \square

In Bemerkung 1 ist die angekündigte Eigenschaft irreduzibler Markov-Ketten angegeben.

Bemerkung 1 Alle Zustände einer irreduziblen Markov-Kette sind positiv rekurrent. (Vergl. [Beh00] S. 49.) \square

(Diese Eigenschaft wird ausschließlich für einen Beweis in Kapitel 2.2 genutzt. Um die Aussage zu beweisen, muss auf weitere Informationen über rekurrente Zustände und irreduzible Markov-Ketten zurückgegriffen werden. Da aber in dieser Arbeit keine derlei weiteren Informationen benötigt werden und dieses Kapitel auch nur eine kurze Einführung darstellt, werden diese Informationen nicht angegeben womit die Aussage nicht bewiesen werden kann. Bei Interesse ist ein Widerspruchsbeweis der Aussage zum Beispiel in [Beh00] auf Seite 49 nachzulesen.)

Mit Definition 7 (a) kann des Weiteren eine andere Darstellung des ii 'ten Eintrags $p_{ii}^{(k)}$ der k -stufigen Übergangsmatrix P^k angegeben werden.

Satz 2 (Erneuerungsgleichung) Für die $N \times N$ Übergangsmatrix P der Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum S gilt

$$p_{ii}^{(k)} = \sum_{l=1}^k p_{ii}^{(k-l)} f_{ii}^{(l)} = \sum_{t=0}^{k-1} f_{ii}^{(k-t)} p_{ii}^{(t)} \text{ für alle } k \in \mathbb{N} \text{ und } i = 1, \dots, N .$$

(Vergl. [Kön05] S. 16.) □

BEWEIS Gemäß (2.8) ist $p_{ii}^{(k)} = \mathbb{P}_i(X_k = i)$. Wenn man nun das Ereignis $\{X_k = i\}$ nach dem ersten Zeitpunkt, an dem die Kette i erreicht, aufspaltet, so erhält man

$$\begin{aligned} p_{ii}^{(k)} &= \sum_{l=1}^k \mathbb{P}_i(X_1 \neq i, \dots, X_{l-1} \neq i, X_l = i, X_k = i) \\ &\stackrel{(2.3)}{=} \sum_{l=1}^k \mathbb{P}_i(X_k = i \mid X_1 \neq i, \dots, X_{l-1} \neq i, X_l = i) \mathbb{P}_i(X_1 \neq i, \dots, X_{l-1} \neq i, X_l = i) \\ &\stackrel{(2.1)}{=} \sum_{l=1}^k \mathbb{P}_i(X_k = i \mid X_l = i) f_{ii}^{(l)} \stackrel{(2.1)}{=} \sum_{l=1}^k \mathbb{P}(X_k = i \mid X_l = i) f_{ii}^{(l)} \stackrel{(2.7)}{=} \sum_{l=1}^k p_{ii}^{(k-l)} f_{ii}^{(l)} \\ &= f_{ii}^{(1)} p_{ii}^{(k-1)} + \dots + f_{ii}^{(k)} p_{ii}^{(0)} = \sum_{t=0}^{k-1} f_{ii}^{(k-t)} p_{ii}^{(t)} . \end{aligned}$$

(Angelehnt an [Kön05] S. 16.) ■

Genau wie im obigen Beweis kann man die Beziehung

$$p_{ji}^{(l)} = \sum_{k=1}^l p_{ii}^{(l-k)} f_{ji}^{(k)} = \sum_{t=0}^{l-1} f_{ji}^{(l-t)} p_{ii}^{(t)} \tag{2.11}$$

für alle $l \in \mathbb{N}$ und $j, i \in S$ herleiten.

2.2.2 Aperiodische Markov-Ketten

Eine weitere besondere Art von Markov-Ketten sind die sogenannten aperiodischen Markov-Ketten. Im Allgemeinen bezeichnet man die Periode eines Zustands $i \in S$ mit $d_i := \text{ggT} \left\{ k \geq 1 \in \mathbb{N} \mid p_{ii}^{(k)} > 0 \right\}$. Dies bedeutet somit, dass die Periode von $i \in S$ der größte gemeinsame Teiler der Menge der Zeitpunkte ist, in denen die Markov-Kette, bei Start in i , nach i zurückkehrt. Der Zustand $i \in S$ heißt aperiodisch, wenn $d_i = 1$ ist.

Definition 9 Wenn alle Zustände $i \in S$ die gleiche existente Periode d haben, dann heißt d Periode der Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ bzw. der zur Markov-Kette gehörenden $N \times N$ Übergangsmatrix P . Eine Markov-Kette (bzw. P) wird als aperiodisch bezeichnet, wenn alle Zustände von $\{X_k\}_{k \in \mathbb{N}}$ aperiodisch sind. (Vergl. [Sch03] S. 26.) \square

Zur Veranschaulichung der Periode einer Markov-Kette sowie der Aperiodizität wird wiederum auf vorhergehende Beispiele zurückgegriffen.

Beispiel 8 (Ergänzung zur Wettervorhersage (Beispiel 3)) Bei dem Beispiel der Wettervorhersage (Beispiel 3) ist die Markov-Kette sowohl im ersten, als auch im zweiten Fall aperiodisch. Dies liegt daran, dass die Wahrscheinlichkeit, dass in k Tagen das gleiche Wetter vorherrscht wie am derzeitigen Tag echt positiv für alle $k \in \mathbb{N}$ ist, egal ob es regnet ($i = v_1$) oder die Sonne scheint ($i = v_2$). Mit anderen Worten bedeutet dies $p_{ii}^{(k)} > 0$ für alle $k \in \mathbb{N}$ und $i = v_1$ oder $i = v_2$. Dementsprechend gilt für alle $i \in S$, $d_i = ggT \left\{ k \geq 1 \in \mathbb{N} \mid p_{ii}^{(k)} > 0 \right\} = 1$ und die Markov-Kette ist damit aperiodisch. \square

Beispiel 9 (Ergänzung zu Beispiel 1 (Läufer)) Der Läufer aus Beispiel 1 braucht stets eine gerade Anzahl an Läufen um, von einer bestimmten Ecke zur selben zurückzukehren. Demnach ist $p_{ii}^{(k)}$ für alle $i = 1, \dots, 4$ nur größer als Null, wenn $k \in \mathbb{N}$ eine gerade Zahl ist. Daraus folgt

$$d_i = ggT \left\{ k \geq 1 \in \mathbb{N} \mid p_{ii}^{(k)} > 0 \right\} = ggT \{2, 4, 6, 8, \dots\} = 2 \text{ für } i = 1, 2, 3, 4 .$$

Die Markov-Kette besitzt somit die Periode 2. (Vergl. [Häg02] S. 25.) \square

Bemerkung 2 Im allgemeinen gilt, dass bei rekurrentem $i \in S$ und $d_i = 1$, $\lim_{k \rightarrow \infty} p_{ii}^{(k)} = \frac{1}{\mu_{ii}}$ ist, wobei in diesem Fall nach Definition 7 (c) und Definition 8 (a) $\mu_{ii} = \sum_{k=1}^{\infty} k f_{ii}^{(k)}$ ist. (Vergl. [Beh00] S. 48.) \square

(Aus den selben Beweggründen wie oben wird auch hierzu kein Beweis angegeben. Nachzulesen ist ein Beweis wiederum in [Beh00] auf den Seiten 48 und 49.)

2.2.3 Ergodische Markov-Ketten

Markov-Ketten, die sowohl irreduzibel als auch aperiodisch sind, kann man in einer gemeinsamen Art von Markov-Ketten vereinen, den sogenannten ergodischen Markov-Ketten. Dementsprechend ergibt sich die Definition ergodischer Markov-Ketten.

Definition 10 Eine irreduzible und aperiodische Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ heißt ergodisch. (Vergl. [Beh00] S. 50.) \square

Für den folgenden Satz 3 bedarf es vorab einer weiteren Definition.

Definition 11 Eine beliebige Matrix A bezeichnet man als echt positiv, wenn alle Einträge (a_{ij}) der Matrix A echt größer als Null sind. Man schreibt dann

$$A > 0 . \quad \square$$

Satz 3 Eine Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum S und $N \times N$ Übergangsmatrix P ist genau dann ergodisch, wenn es ein $k \in \mathbb{N}$ gibt, so dass $P^k > 0$ ist. (Vergl. [Beh00] S. 50.) □

BEWEIS „ $P^k > 0 \Rightarrow \{X_k\}_{k \in \mathbb{N}}$ ergodisch“: Sei $k \in \mathbb{N}$, so dass $P^k > 0$ ist. Daraus ergibt sich unmittelbar die Irreduzibilität der Markov-Kette.

Nun muss noch gezeigt werden, dass die Markov-Kette auch aperiodisch ist. Nach der Chapman-Kolmogorov-Gleichung (siehe (2.10)) gilt für alle $i, j \in S$ und $k \in \mathbb{N}$ $p_{ij}^{(k+1)} = \sum_{l \in S} p_{il} p_{lj}^{(k)}$. Da $P^k > 0$ vorausgesetzt wird, ist also $p_{lj}^{(k)} > 0$ für alle $l \in S$. Des Weiteren weiß man, dass P eine stochastische Matrix ist, weswegen mindestens einer der Summanden aus $\sum_{l \in S} p_{il}$ echt größer als Null sein muss. Insgesamt gilt damit $p_{ij}^{(k+1)}$ ist echt größer als Null für alle $i, j \in S$. Nun folgt, dass für alle $i \in S$ und für alle $k' \geq k + 1$, $p_{ii}^{(k')}$ auch größer als Null sein muss. Da dies wie gesagt für alle $k' \geq k + 1$ der Fall ist (damit auch alle Primzahlen größer oder gleich $k + 1$ einschließend), ist der größte gemeinsame Teiler der Menge $\{k' \geq k + 1 \in \mathbb{N} \mid p_{ii}^{(k')} > 0\}$ gleich 1. Die Markov-Kette ist also aperiodisch und insgesamt somit ergodisch.

„ $\{X_k\}_{k \in \mathbb{N}}$ ergodisch $\Rightarrow P^k > 0$ “: Nach Voraussetzung ist die Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ ergodisch. Aus Bemerkung 1 weiß man, dass alle Zustände $i \in S$ positiv rekurrent sind. Damit gilt dann nach Bemerkung 2: $\lim_{k \rightarrow \infty} p_{ii}^{(k)} = \frac{1}{\mu_{ii}}$. Nach Definition 8 (a) sind in diesem Fall $\mu_{ii} = \sum_{k=1}^{\infty} k f_{ii}^{(k)}$ und $\sum_{k=1}^{\infty} f_{ii}^{(k)} = 1$. Damit ist μ_{ii} größer als Null (bzw. größer oder gleich Eins) und $\frac{1}{\mu_{ii}}$ ist somit echt größer als Null. Nun kann ein $k' \in \mathbb{N}$ gefunden werden, so dass $p_{ii}^{(k')} > 0$ für alle $i \in S$ gilt. Da P nach Voraussetzung irreduzibel ist, kommunizieren alle Zustände $i, j \in S$, d.h. es gibt ein $k(i, j) \in \mathbb{N}$, so dass $p_{ij}^{(k(i,j))} > 0$ ist. Mit $k'' \geq k_0 := \max_{i,j \in S} k(i, j)$ folgt nun aus der Chapman-Kolmogorov-Gleichung (siehe (2.10)), dass $p_{ij}^{(k'+k'')} \geq p_{ii}^{(k')} p_{ij}^{(k'')}$ für alle $i, j \in S$ sein muss. Dies ist wiederum echt größer als Null, da $p_{ii}^{(k')} > 0$ war (siehe oben) und k'' auch größer oder gleich $k(i, j)$ ist, wobei $p_{ij}^{(k(i,j))} > 0$ gilt. Insgesamt bedeutet dies somit, dass $P^{k'+k''} > 0$ für alle $k'' \geq k_0$ ist. Mit $k \geq k' + k''$ gilt dann die Behauptung. (In Anlehnung an [Beh00] S. 50.) ■

Zwei Beispiele sollen dies veranschaulichen.

Beispiel 10 (Ergänzung zu Beispiel 3 (Wettervorhersage)) In beiden Fällen des Beispiels der Wettervorhersage sind die Markov-Ketten ergodisch. Alle Einträge der Übergangsmatrizen sind echt größer als Null. \square

Beispiel 11 (Ergänzung zu Beispiel 1 (Läufer)) Für die 4×4 Übergangsmatrix P aus dem Beispiel des Läufers (siehe dazu Beispiel 2) gilt

$$P^{2k+1} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} \quad \text{und} \quad P^{2k} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \quad \text{für alle } k \in \mathbb{N} .$$

Es gibt somit keinen Zeitpunkt $k \in \mathbb{N}$, so dass $P^k > 0$ ist. Die Markov-Kette aus Beispiel 1 ist somit nach Satz 3 nicht ergodisch. \square

Das Ziel bei dieser Einführung in Markov-Ketten war es, nur die Kenntnisse näher zu bringen, von denen im Folgenden auch Gebrauch gemacht wird. Natürlich gibt es noch viel mehr interessante und wissenswerte Dinge über Markov-Ketten zu erfahren. Bei Bedarf kann sich der interessierte Leser in der zahlreichen Literatur, die es über Markov-Ketten gibt, weitergehend informieren.

3 Konvergenz gegen die Gleichgewichtsverteilung

Dieses Kapitel beschäftigt sich mit dem Langzeitverhalten bzw. Konvergenzverhalten von Markov-Ketten. Im Grunde genommen wird hierbei nicht das Konvergenzverhalten einer Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ an sich betrachtet, sondern das Konvergenzverhalten der Verteilung einer Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$. Eine Markov-Kette, die gemäß Kapitel 2 eine besondere Art eines stochastischen Prozesses ist, kann im allgemeinen nicht konvergieren. Die Werte, die eine Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ annimmt variieren für $k \in \mathbb{N}$ gegen unendlich stark, so dass kein Grenzwert von $\{X_k\}_{k \in \mathbb{N}}$ anzugeben ist. Es kann jedoch, wiederum nach Kapitel 2, eine endliche Markov-Kette als Matrix angesehen werden bzw. durch die Übergangsmatrix P verkörpert werden. Nun können Aussagen über das Konvergenzverhalten von P^k für $k \in \mathbb{N}$ gegen unendlich gemacht werden. Auf Grund dieser Gegebenheiten spricht man von der Konvergenz von Markov-Ketten, wobei hiermit sinngemäß die Konvergenz der Verteilung einer Markov-Kette gemeint ist.

Wie schon in Kapitel 1 wird stets davon ausgegangen, dass es sich bei dem Zustandsraum S um die endliche Menge $\{1, \dots, N\}$ handelt und dass die betrachteten Markov-Ketten homogen sind.

3.1 Die Gleichgewichtsverteilung

Es wird mit der Definition der Gleichgewichtsverteilung begonnen, um dann konkrete Fragestellungen formulieren zu können.

3.1.1 Definition und Beispiele

Definition 12 (Gleichgewichtsverteilung) Sei $\{X_k\}_{k \in \mathbb{N}}$ eine Markov-Kette mit Zustandsraum $S = \{1, \dots, N\}$ und Übergangsmatrix $P = (p_{ij})_{i,j \in S}$, dann heißt der Vektor $\pi = (\pi_1, \dots, \pi_N)^T \in \mathbb{R}^N$, falls er existiert, Gleichgewichtsverteilung (oder Grenzverteilung bzw. Stationäre-Verteilung) der Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$, falls fol-

gendes erfüllt ist:

$$(i) \quad \pi_i \geq 0 \text{ für alle } i \in S . \quad (3.1)$$

$$(ii) \quad \sum_{i=1}^N \pi_i = 1 . \quad (3.2)$$

$$(iii) \quad \pi^T P = \pi^T, \text{ d.h. } \sum_{i=1}^N \pi_i p_{ij} = \pi_j \text{ für alle } j \in S . \quad (3.3)$$

(Vergl. [Bré99] S. 75.) □

Da die Definition der Gleichgewichtsverteilung nur von der Übergangsmatrix P abhängig ist, kann man auch Gleichgewichtsverteilung von P sagen.

(3.1) und (3.2) der Definition 12 bedeuten, dass es sich bei π um eine Wahrscheinlichkeitsverteilung auf S handelt. Die Gleichung (3.3) ist von zentraler Bedeutung. Sie wird in der Literatur auch als „globale Gleichgewichtsgleichung“ („global balance equation“) bezeichnet.

Nun folgen einige Beispiele zur Veranschaulichung.

Beispiel 12 (Ergänzung zu Beispiel 3 (Wettersvorhersage)) Beim ersten Fall im Beispiel der Wettervorhersage ist $\pi = (0.5, 0.5)^T$ die Gleichgewichtsverteilung der Markov-Kette. Es sind damit nämlich (3.1) und (3.2) erfüllt und es gilt (3.3), da

$$\pi^T P = (0.5, 0.5) \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix} = (0.5, 0.5) = \pi^T$$

ist.

Im zweiten Fall des Wettervorhersagen-Beispiels ist die Gleichgewichtsverteilung durch $\pi = (\frac{5}{6}, \frac{1}{6})^T$ gegeben. Wiederum sind mit diesem π die Voraussetzungen (3.1) und (3.2) erfüllt und es gilt (3.3), da

$$\pi^T P = \left(\frac{5}{6}, \frac{1}{6}\right) \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix} = \left(\frac{5}{6}, \frac{1}{6}\right) = \pi^T$$

ist. □

Beispiel 13 (Irrfahrten) 1. Die symmetrische, zyklische Irrfahrt auf $S = \{1, 2, 3, 4\}$

mit der Übergangsmatrix $P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$ (vergleiche Beispiel 2 und Beispiel 4, 3.) besitzt die Gleichgewichtsverteilung $\pi = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^T$. Damit sind (3.1)

und (3.2) erfüllt und es gilt

$$\pi^T P = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right) \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right) = \pi^T,$$

womit dann auch (3.3) erfüllt ist.

2. Der Vektor $\pi = (\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8})^T$ ist die Gleichgewichtsverteilung der reflektierenden

Irrfahrt auf $S = \{1, 2, 3, 4\}$ mit Übergangsmatrix $P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{pmatrix}$. Mit

diesem π sind wiederum (3.1) und (3.2) erfüllt und es gilt (3.3) der Definition der Gleichgewichtsverteilung, da

$$\pi^T P = \left(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8} \right) \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{pmatrix} = \left(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8} \right) = \pi^T$$

ist. □

(Wie man jeweils auf die Gleichgewichtsverteilungen aus den Beispielen 12 und 13 kommt, wird später betrachtet.)

Wenn die Markov-Kette mit einer Anfangsverteilung $\alpha^{(0)}$, die π entspricht, gestartet wird, dann ergibt sich nach k Iterationsschritten, dass

$$\alpha^{(k)T} \stackrel{(2.5)}{=} \alpha^{(0)T} P^k = \pi^T P^k \stackrel{(3.3)}{=} \pi^T$$

für alle $k \in \mathbb{N}$ ist. Dies bedeutet somit, dass eine Markov-Kette, die mit einer Gleichgewichtsverteilung startet, diese Verteilung zu jedem Zeitpunkt behält.

3.1.2 Zentrale Fragestellungen

Es stellen sich nun folgende Fragen :

Gibt es stets eine Gleichgewichtsverteilung π ? Ist π eindeutig? Wenn die Markov-Kette mit einer beliebigen Anfangsverteilung $\alpha^{(0)}$ gestartet wird, konvergiert dann $\alpha^{(0)T} P^k$ für $k \in \mathbb{N}$ gegen unendlich immer gegen π^T ? Wie schnell ist die Konvergenz und um was für eine Art von Konvergenz handelt es sich? Wie lässt sich π berechnen?

Im Folgenden wird nun versucht, diese Fragen zu beantworten. Da dies im allgemeinen für beliebige Markov-Ketten nicht bzw. nur teilweise möglich ist, werden die Fragen für bestimmte Arten von Markov-Ketten beantwortet. Auf die letzte Frage, wie sich π berechnen lässt, wird in einem gesonderten Abschnitt eingegangen.

3.2 Die Gleichgewichtsverteilung ergodischer Markov-Ketten

3.2.1 Existenz und Eindeutigkeit

Satz 4 Sei $\{X_k\}_{k \in \mathbb{N}}$ eine ergodische Markov-Kette mit Zustandsraum $S = \{1, \dots, N\}$ und sei P die zur Markov-Kette gehörende $N \times N$ Übergangsmatrix, dann gilt

- (a) $\{X_k\}_{k \in \mathbb{N}}$ besitzt eine eindeutig bestimmte Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$.
 (b) P^k konvergiert komponentenweise für $k \in \mathbb{N}$ gegen unendlich gegen die $N \times N$

$$\text{Matrix } W := \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix}.$$

- (c) $\pi_i = \frac{1}{\mu_{ii}}$ für alle $i \in S$.

(Vergl. [Beh00] S. 50.) □

BEWEIS Der Beweis von (a) gliedert sich in zwei Teile, einmal den Existenzbeweis einer Gleichgewichtsverteilung und den Eindeutigkeitsbeweis dieser. Begonnen wird mit dem Existenzbeweis. Im Rahmen dieses Existenzbeweises werden, wie man sehen wird, auch (b) und (c) bewiesen.

Um nun zu beweisen, dass ergodische Markov-Ketten stets eine Gleichgewichtsverteilung haben, wird zuerst gezeigt, dass P^k komponentenweise, für $k \in \mathbb{N}$ gegen un-

endlich, gegen die Matrix $W := \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix}$ konvergiert, wobei $\pi^T = (\pi_1, \dots, \pi_N) \in \mathbb{R}^N$,

$\pi_i > 0$ für alle $i \in S$ und $\sum_{i=1}^N \pi_i = 1$ ist.

Da die Markov-Kette nach Voraussetzung ergodisch und damit irreduzibel ist, weiß man aus Bemerkung 1 und Bemerkung 2, dass alle $i \in S$ positiv rekurrent sind und $\lim_{k \rightarrow \infty} p_{ii}^{(k)} = \frac{1}{\mu_{ii}}$ ist. μ_{ii} ist hierbei größer als Eins, da bei rekurrentem i nach den Definitionen 7 und 8 $\mu_{ii} = \sum_{k=1}^{\infty} k f_{ii}^{(k)}$ und $\sum_{k=1}^{\infty} f_{ii}^{(k)} = 1$ sind. Damit gilt die

Abschätzung

$$1 \geq \frac{1}{\mu_{ii}} > 0 . \quad (3.4)$$

Es wird nun gezeigt, dass $\lim_{k \rightarrow \infty} p_{ji}^{(k)} = \frac{f_{ji}^*}{\mu_{ii}}$ für alle $i, j \in S$ ist.

Sei $\epsilon > 0$ und sei außerdem $t_0 \in \mathbb{N}$ so gewählt, dass

$$\left| p_{ii}^{(t)} - \frac{1}{\mu_{ii}} \right| \leq \frac{\epsilon}{2} \text{ für alle } t \geq t_0 \in \mathbb{N} \quad (3.5)$$

gilt. Da nach Definition 7 (a) $f_{ji}^{(t)}$ die Wahrscheinlichkeit angibt, dass die Markov-Kette bei Start in j den Zustand i zum ersten mal im t -ten Schritt erreicht, kann ein $k \geq t_0 + 1 \in \mathbb{N}$ gewählt werden, so dass für alle $l \geq k \in \mathbb{N}$ gilt

$$\sum_{t=l-t_0+1}^{\infty} f_{ji}^{(t)} = \sum_{t=-\infty}^{t_0-1} f_{ji}^{(l-t)} \leq \frac{\epsilon}{2(1 + \frac{1}{\mu_{ii}})} . \quad (3.6)$$

Nun ist für alle $l \geq k \in \mathbb{N}$

$$\begin{aligned} \left| p_{ji}^{(l)} - \frac{f_{ji}^*}{\mu_{ii}} \right| &\stackrel{(2.11)}{=} \left| \sum_{t=0}^{l-1} f_{ji}^{(l-t)} p_{ii}^{(t)} - \frac{1}{\mu_{ii}} \sum_{t=1}^{\infty} f_{ji}^{(t)} \right| = \left| \sum_{t=0}^{l-1} f_{ji}^{(l-t)} p_{ii}^{(t)} - \frac{1}{\mu_{ii}} \sum_{t=-\infty}^{l-1} f_{ji}^{(l-t)} \right| \\ &= \left| \sum_{t=t_0}^{l-1} f_{ji}^{(l-t)} p_{ii}^{(t)} + \sum_{t=0}^{t_0-1} f_{ji}^{(l-t)} p_{ii}^{(t)} - \left(\frac{1}{\mu_{ii}} \sum_{t=t_0}^{l-1} f_{ji}^{(l-t)} + \frac{1}{\mu_{ii}} \sum_{t=-\infty}^{t_0-1} f_{ji}^{(l-t)} \right) \right| \\ &= \left| \sum_{t=t_0}^{l-1} f_{ji}^{(l-t)} \left(p_{ii}^{(t)} - \frac{1}{\mu_{ii}} \right) + \sum_{t=0}^{t_0-1} f_{ji}^{(l-t)} p_{ii}^{(t)} - \frac{1}{\mu_{ii}} \sum_{t=-\infty}^{t_0-1} f_{ji}^{(l-t)} \right| . \end{aligned} \quad (3.7)$$

Wegen (3.5) und der Tatsache, dass $\sum_{t=t_0}^{l-1} f_{ji}^{(l-t)} \leq 1$ ist (siehe Definition 7), gilt hierbei

$$\sum_{t=t_0}^{l-1} f_{ji}^{(l-t)} \left(p_{ii}^{(t)} - \frac{1}{\mu_{ii}} \right) \leq \frac{\epsilon}{2} .$$

Des Weiteren ist mit (3.6) und Bemerkung 2 bzw. (3.4)

$$\sum_{t=0}^{t_0-1} f_{ji}^{(l-t)} p_{ii}^{(t)} \leq \frac{\epsilon}{2(1 + \frac{1}{\mu_{ii}})}$$

und wiederum mit (3.6) folgt unmittelbar

$$\frac{1}{\mu_{ii}} \sum_{t=-\infty}^{t_0-1} f_{ji}^{(l-t)} \leq \frac{\epsilon}{2\mu_{ii}(1 + \frac{1}{\mu_{ii}})} .$$

Damit gilt insgesamt

$$\begin{aligned}
 \left| p_{ji}^{(l)} - \frac{f_{ji}^*}{\mu_{ii}} \right| &\stackrel{(3.7)}{=} \left| \sum_{t=t_0}^{l-1} f_{ji}^{(l-t)} \left(p_{ii}^{(t)} - \frac{1}{\mu_{ii}} \right) + \sum_{t=0}^{t_0-1} f_{ji}^{(l-t)} p_{ii}^{(t)} - \frac{1}{\mu_{ii}} \sum_{t=-\infty}^{t_0-1} f_{ji}^{(l-t)} \right| \\
 &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2(1 + \frac{1}{\mu_{ii}})} + \frac{\epsilon}{2\mu_{ii}(1 + \frac{1}{\mu_{ii}})} \\
 &= \frac{\epsilon\mu_{ii}(1 + \frac{1}{\mu_{ii}}) + \epsilon\mu_{ii} + \epsilon}{2\mu_{ii}(1 + \frac{1}{\mu_{ii}})} = \frac{\epsilon\mu_{ii} + \epsilon + \epsilon\mu_{ii} + \epsilon}{2\mu_{ii}(1 + \frac{1}{\mu_{ii}})} \\
 &= \frac{2\epsilon\mu_{ii} + 2\epsilon}{2\mu_{ii}(1 + \frac{1}{\mu_{ii}})} = \frac{\epsilon 2\mu_{ii}(1 + \frac{1}{\mu_{ii}})}{2\mu_{ii}(1 + \frac{1}{\mu_{ii}})} = \epsilon .
 \end{aligned}$$

Nun muss noch gezeigt werden, dass $f_{ji}^* = 1$ ist.

Ohne Einschränkungen sei $i \neq j$. Des Weiteren wird $k' := \min \{ k \in \mathbb{N} \mid p_{ij}^{(k)} > 0 \}$ gesetzt. Nun gibt $1 - f_{ji}^*$ die Wahrscheinlichkeit an, dass die Markov-Kette bei Start in j niemals i erreicht und $p_{ij}^{(k')}(1 - f_{ji}^*)$ ist die Wahrscheinlichkeit, dass die Markov-Kette bei Start in i im Schritt k' den Zustand j besucht und danach nicht mehr zu i zurückkehrt. Es folgt somit unmittelbar

$$p_{ij}^{(k')}(1 - f_{ji}^*) \leq 1 - f_{ii}^* .$$

Da i rekurrent ist und somit $f_{ii}^* = 1$ ist (Definition 7 (a)), gilt dabei

$$1 - f_{ii}^* = 1 - 1 = 0 .$$

Insgesamt folgt somit

$$p_{ij}^{(k')}(1 - f_{ji}^*) = 0 .$$

Da $p_{ij}^{(k')}$ per Definition von k' nicht Null sein kann, muss $(1 - f_{ji}^*)$ gleich Null sein, womit dann $f_{ji}^* = 1$ ist.

Insgesamt hat sich somit ergeben

$$p_{ji}^{(k)} \xrightarrow{k \rightarrow \infty} \frac{1}{\mu_{ii}} =: \pi_i \text{ für alle } i, j \in S$$

und damit

$$P^k \xrightarrow{k \rightarrow \infty} W := \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix} .$$

Die Aussagen (b) und (c) sind insbesondere bewiesen.

Es ist nun mit (3.4) $\pi_i > 0$ und es gilt

$$\sum_{i=1}^N \pi_i = \sum_{i=1}^N \lim_{k \rightarrow \infty} p_{ji}^{(k)} = \lim_{k \rightarrow \infty} \sum_{i=1}^N p_{ji}^{(k)} = 1 .$$

Mit dem erhaltenen $W = \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix}$ folgt des Weiteren

$$WP = \left(\lim_{k \rightarrow \infty} P^k \right) P = \lim_{k \rightarrow \infty} P^{k+1} = W ,$$

wobei es sich um komponentenweise Konvergenz handelt. Klar ist, dass $\pi^T P = \pi^T$ mit obigem π äquivalent mit $WP=W$ ist. Jede ergodische Markov-Kette besitzt somit eine Gleichgewichtsverteilung.

Nun wird die Eindeutigkeit von π bewiesen.

Sei π^* ein weiterer Vektor aus \mathbb{R}^N , so dass $\pi^* > 0$, $\sum_{i=1}^N \pi_i^* = 1$ und $\pi^{*T} P^k = \pi^{*T}$ für alle $k \in \mathbb{N}$ gilt.

Dann gilt für alle $k \in \mathbb{N}$, dass aus $\pi^{*T} P^k = \pi^{*T}$

$$\begin{aligned} \pi^{*T} &= \lim_{k \rightarrow \infty} \pi^{*T} P^k = \pi^{*T} \left(\lim_{k \rightarrow \infty} P^k \right) = \pi^{*T} W \\ &= \left(\sum_{i=1}^N \pi_i^* \pi_1, \dots, \sum_{i=1}^N \pi_i^* \pi_N \right) = \left(\pi_1 \sum_{i=1}^N \pi_i^*, \dots, \pi_N \sum_{i=1}^N \pi_i^* \right) = \pi^T \end{aligned}$$

folgt. Die letzte Gleichung ergibt sich, da $\sum_{i=1}^N \pi_i^* = 1$ ist.

Die Gleichgewichtsverteilung π ist somit eindeutig. (In Anlehnung an [Beh00] S. 50 und 51.) ■

Wenn man nun eine ergodische Markov-Kette mit beliebiger Anfangsverteilung $\alpha^{(0)} \in \mathbb{R}^N$ startet, dann folgt aus Satz 4 und der Tatsache, dass $\sum_{i=1}^N \alpha_i^{(0)} = 1$ ist

$$\lim_{k \rightarrow \infty} \alpha^{(0)T} P^k \stackrel{\text{Satz 4}}{=} \alpha^{(0)T} W = \left(\sum_{i=1}^N \alpha_i^{(0)} \pi_1, \dots, \sum_{i=1}^N \alpha_i^{(0)} \pi_N \right) = (\pi_1, \dots, \pi_N) = \pi^T .$$

Bei beliebiger Anfangsverteilung konvergiert eine ergodische Markov-Kette somit gegen ihre Gleichgewichtsverteilung.

Die folgende Verallgemeinerung von Beispiel 3 soll diese Gegebenheit veranschaulichen.

Beispiel 14 (Verallgemeinerte Wettervorhersage) Das Beispiel der Wettervorhersage (Beispiel 3) kann auf einfache Weise verallgemeinert werden. Bei dem ver-

allgemeinerten Modell ist die Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ durch den Zustandsraum $S = \{1, 2\}$ und die 2×2 Übergangsmatrix

$$P = \begin{pmatrix} 1-p & p \\ p' & 1-p' \end{pmatrix} \text{ mit } 0 < p, p' < 1$$

gegeben. Mit dieser Wahl von p und p' gilt stets, dass die Markov-Kette ergodisch ist. Die nach Satz 4 eindeutig bestimmte Gleichgewichtsverteilung von $\{X_k\}_{k \in \mathbb{N}}$ ist nun durch $\pi = \left(\frac{p'}{p+p'}, \frac{p}{p+p'}\right)^T$ gegeben. Mit diesem π sind nämlich die Bedingungen (3.1), (3.2) und (3.3) der Definition der Gleichgewichtsverteilung erfüllt, da

$$\begin{aligned} \pi_1 &= \frac{p'}{p+p'} > 0, \quad \pi_2 = \frac{p}{p+p'} > 0, \quad \pi_1 + \pi_2 = \frac{p'}{p+p'} + \frac{p}{p+p'} = 1 \text{ und} \\ \pi^T P &= \left(\frac{p'}{p+p'}, \frac{p}{p+p'}\right) \begin{pmatrix} 1-p & p \\ p' & 1-p' \end{pmatrix} = \left(\frac{p' - pp' + pp'}{p+p'}, \frac{pp' + p - pp'}{p+p'}\right) = \pi^T \end{aligned}$$

gilt. Die k -stufige Übergangsmatrix ist dann gegeben durch

$$P^k = \frac{1}{p+p'} \begin{pmatrix} p' & p \\ p' & p \end{pmatrix} + \frac{(1-p-p')^k}{p+p'} \begin{pmatrix} p & -p \\ -p' & p' \end{pmatrix}.$$

(Die Berechnung dieser k -stufigen Übergangsmatrix bzw. wie man auf die obige Gestalt von P^k kommt, kann in [Beh00] auf Seite 75 nachgelesen werden.)

Wenn man nun die Markov-Kette mit irgendeiner beliebigen Anfangsverteilung $\alpha^{(0)} = (a, b)^T$ mit $0 \leq a, b \leq 1$ und $a + b = 1$ startet, dann gilt

$$\begin{aligned} \lim_{k \rightarrow \infty} \alpha^{(0)T} P^k &= \alpha^{(0)T} \lim_{k \rightarrow \infty} P^k = (a, b) \frac{1}{p+p'} \begin{pmatrix} p' & p \\ p' & p \end{pmatrix} \\ &= \frac{1}{p+p'} (ap' + bp', ap + bp) = \frac{1}{p+p'} (p'(a+b), p(a+b)) \\ &= \left(\frac{p'}{p+p'}, \frac{p}{p+p'}\right) = \pi^T. \end{aligned}$$

(Vergl. [Sch03] S. 17.) □

Anmerkend sei an dieser Stelle noch erwähnt, dass Satz 4 nicht für beliebige Markov-Ketten gilt. Es kann zwar gezeigt werden, dass jede Markov-Kette mit endlichem Zustandsraum mindestens eine Gleichgewichtsverteilung besitzt, aber beliebige Markov-Ketten, die mit einer beliebigen Verteilung $\alpha^{(0)}$ starten, nicht notwendig konvergieren.

In der Literatur finden sich viele Ansätze, über die Abschätzungen für die Konver-

genzgeschwindigkeit einer ergodischen Markov-Kette hin zu ihrer Gleichgewichtsverteilung gefunden werden können. Im Folgenden werden zwei dieser Ansätze detailliert dargestellt und außerdem eine weitere Grenze der Konvergenzgeschwindigkeit zur Anschauung angegeben.

3.2.2 Direkte Konvergenzabschätzung Teil 1

Der Erste Ansatz, der genauer darstellt wird, gibt eine direkte Abschätzung der Konvergenzgeschwindigkeit an. Direkt bedeutet hierbei, dass die Schranke der Konvergenzgeschwindigkeit durch konkrete Einträge der Übergangsmatrix P beziehungsweise P^k gebildet wird.

Zuerst benötigt man allerdings zwei Hilfssätze.

Lemma 1 Sei $P=(p_{ij})_{i,j \in S}$ eine beliebige stochastische Matrix. Für $k \in \mathbb{N}$ definiere man die minimalen und maximalen Einträge der j -ten Spalte von P^k mit $m_j^{(k)} := \min_{i \in S} p_{ij}^{(k)}$ und $M_j^{(k)} := \max_{i \in S} p_{ij}^{(k)}$ ($j \in S$).

Damit gilt für alle $k, h \in \mathbb{N}$ und $i, j \in S$

$$(a) \quad m_j^{(k+1)} \geq m_j^{(k)}, \quad M_j^{(k+1)} \leq M_j^{(k)}, \quad m_j^{(k+h)} \leq M_j^{(h)},$$

$$d.h. \quad m_j^{(1)} \leq m_j^{(2)} \leq \dots \leq M_j^{(2)} \leq M_j^{(1)}.$$

(b) Sei P ergodisch. Setze $\delta := \min_{i,j \in S} p_{ij}$ und $T := 1 - N\delta$, dann ist

$$(i) \quad \left| p_{ij}^{(k)} - \pi_j \right| \leq M_j^{(k)} - m_j^{(k)}.$$

$$(ii) \quad T \in [0, 1].$$

$$(iii) \quad M_j^{(k+1)} - m_j^{(k+1)} \leq T^k (M_j^{(1)} - m_j^{(1)}).$$

(Vergl. [Beh00] S. 83 und 84.) □

BEWEIS (a) Für alle $k \in \mathbb{N}$ und $j \in S$ gilt

$$m_j^{(k+1)} = \min_{i \in S} p_{ij}^{(k+1)} \stackrel{(2.10)}{=} \min_{i \in S} \sum_{l=1}^N p_{il} p_{lj}^{(k)}$$

$$\geq \min_{i \in S} \left(\sum_{l=1}^N p_{il} \right) \min_{l \in S} p_{lj}^{(k)} \stackrel{(2.4)}{=} 1 \cdot \min_{l \in S} p_{lj}^{(k)} = m_j^{(k)}.$$

Die zweite Ungleichung in (a) kann analog dazu bewiesen werden. Auch der Beweis der letzten Ungleichung aus (a) ist ähnlich.

Für alle $k, h \in \mathbb{N}$ und $j \in S$ gilt, wiederum mit der Chapman-Kolmogorov-Gleichung

(siehe (2.10)), dass

$$\begin{aligned} m_j^{(k+h)} &= \min_{i \in S} p_{ij}^{(k+h)} \stackrel{(2.10)}{=} \min_{i \in S} \sum_{l=1}^N p_{il}^{(k)} p_{lj}^{(h)} \\ &\leq \min_{i \in S} \left(\sum_{l=1}^N p_{il}^{(k)} \right) \max_{l \in S} p_{lj}^{(h)} \stackrel{(2.4)}{=} 1 \cdot \max_{l \in S} p_{lj}^{(h)} = M_j^{(h)} \end{aligned}$$

ist.

(b)(i) Da P ergodisch ist, weiß man aus Satz 4, dass $p_{ij}^{(k)}$ für alle $i, j \in S$ und $k \in \mathbb{N}$ gegen unendlich gegen π_j konvergiert. Außerdem ist mit den Voraussetzungen bekannt, dass $p_{ij}^{(k)}$ im Intervall $[m_j^{(k)}, M_j^{(k)}]$ liegt. Damit muss auch π_j im Intervall $[m_j^{(k)}, M_j^{(k)}]$ liegen und es gilt

$$\left| p_{ij}^{(k)} - \pi_j \right| \leq M_j^{(k)} - m_j^{(k)} .$$

(ii) Für alle $i \in S$ ist $\sum_{j=1}^N p_{ij} = 1$ (Definition 3). Dies bedeutet, dass mindestens ein Summand der Summe $\sum_{j=1}^N p_{ij}$ kleiner oder gleich $\frac{1}{N}$ sein muss. Somit gilt

$$\delta = \min_{i,j \in S} p_{ij} \leq \frac{1}{N} . \quad (3.8)$$

Damit liegt dann $T = 1 - N\delta$ im Intervall $[0,1]$.

(iii) Seien i_0, j_0 zwei beliebige Werte aus S und sei $S' := \{j \in S \mid p_{i_0j} \geq p_{j_0j}\}$ und $S'' := \{j \in S \mid p_{i_0j} < p_{j_0j}\}$.

Mit der Chapman-Kolmogorov-Gleichung (siehe (2.10)) gilt nun für alle $s \in S$

$$\begin{aligned} p_{i_0s}^{(k+1)} - p_{j_0s}^{(k+1)} &\stackrel{(2.10)}{=} \sum_{j=1}^N p_{i_0j} p_{js}^{(k)} - \sum_{j=1}^N p_{j_0j} p_{js}^{(k)} = \sum_{j=1}^N (p_{i_0j} - p_{j_0j}) p_{js}^{(k)} \\ &= \sum_{j \in S'} (p_{i_0j} - p_{j_0j}) p_{js}^{(k)} + \sum_{j \in S''} (p_{i_0j} - p_{j_0j}) p_{js}^{(k)} . \end{aligned}$$

Da $(p_{i_0j} - p_{j_0j})$ für alle $j \in S'$ stets größer oder gleich Null und $(p_{i_0j} - p_{j_0j})$ für alle $j \in S''$ stets kleiner oder gleich Null ist folgt hierbei

$$\begin{aligned} &\sum_{j \in S'} (p_{i_0j} - p_{j_0j}) p_{js}^{(k)} + \sum_{j \in S''} (p_{i_0j} - p_{j_0j}) p_{js}^{(k)} \\ &\leq \sum_{j \in S'} (p_{i_0j} - p_{j_0j}) M_s^{(k)} + \sum_{j \in S''} (p_{i_0j} - p_{j_0j}) m_s^{(k)} . \end{aligned}$$

Ferner gilt $\sum_{j \in S'} (p_{i_0j} - p_{j_0j}) + \sum_{j \in S''} (p_{i_0j} - p_{j_0j})$ ist gleich

$\sum_{j \in S'} p_{i_0j} - \sum_{j \in S'} p_{j_0j} + \sum_{j \in S''} p_{i_0j} - \sum_{j \in S''} p_{j_0j} = \sum_{j=1}^N p_{i_0j} - \sum_{j=1}^N p_{j_0j} \stackrel{(2.4)}{=} 1 - 1 = 0$
 und demnach $-\sum_{j \in S'} (p_{i_0j} - p_{j_0j}) = \sum_{j \in S''} (p_{i_0j} - p_{j_0j})$. Damit ist nun

$$\begin{aligned} & \sum_{j \in S'} (p_{i_0j} - p_{j_0j}) M_s^{(k)} + \sum_{j \in S''} (p_{i_0j} - p_{j_0j}) m_s^{(k)} \\ &= \sum_{j \in S'} (p_{i_0j} - p_{j_0j}) M_s^{(k)} - \sum_{j \in S'} (p_{i_0j} - p_{j_0j}) m_s^{(k)} \\ &= \left(\sum_{j \in S'} (p_{i_0j} - p_{j_0j}) \right) (M_s^{(k)} - m_s^{(k)}) . \end{aligned}$$

Angenommen $\sum_{j \in S''} p_{i_0j}$ enthält $l \in S$ Summanden, dann enthält $\sum_{j \in S'} p_{j_0j}$ genau $N - l$ Summanden und damit gilt $\sum_{j \in S''} p_{i_0j} + \sum_{j \in S'} p_{j_0j} \geq N \min_{i,j \in S} p_{ij} = N\delta$. Außerdem ist $\sum_{j \in S'} p_{i_0j} + \sum_{j \in S''} p_{i_0j} = 1$, womit dann $\sum_{j \in S'} (p_{i_0j} - p_{j_0j}) = \sum_{j \in S'} p_{i_0j} - \sum_{j \in S'} p_{j_0j} + \sum_{j \in S''} p_{i_0j} - \sum_{j \in S''} p_{i_0j} = 1 - \sum_{j \in S''} p_{i_0j} - \sum_{j \in S'} p_{j_0j} \leq 1 - N\delta = T$ ist. Es gilt nun

$$\begin{aligned} & \left(\sum_{j \in S'} (p_{i_0j} - p_{j_0j}) \right) (M_s^{(k)} - m_s^{(k)}) \\ & \leq T (M_s^{(k)} - m_s^{(k)}) . \end{aligned}$$

Da zu Beginn angenommen wurde, $i_0, j_0 \in S$ seien beliebig, wurde somit insgesamt gezeigt

$$M_j^{(k+1)} - m_j^{(k+1)} \leq T (M_j^{(k)} - m_j^{(k)}) .$$

Es gilt damit

$$M_j^{(k+1)} - m_j^{(k+1)} \leq T (M_j^{(k)} - m_j^{(k)}) \leq T^2 (M_j^{(k-1)} - m_j^{(k-1)}) \leq \dots \leq T^k (M_j^{(1)} - m_j^{(1)}) ,$$

womit (iii) bewiesen ist. (In Anlehnung an [Beh00] S. 83 und 84.) ■

Lemma 2 beinhaltet eine allgemeine Aussage über fallende Folgen.

Lemma 2 Sei $(c_k)_{k \in \mathbb{N}}$ eine Folge mit $c_k \in [0, \infty)$, $c_{k+1} \leq c_k$ für alle $k \in \mathbb{N}$ und $c_{k' \cdot k_0} \leq T^{k'}$ für ein festes $k_0 \in \mathbb{N}$, $T \in [0, 1]$ und alle $k' \in \mathbb{N}$. Dann gilt

$$c_k \leq \frac{T^{\frac{k}{k_0}}}{T} \text{ für alle } k \in \mathbb{N} . \quad (3.9)$$

(Vergl. [Beh00] S. 85.) □

BEWEIS Für jedes $k \in \mathbb{N}$ gibt es ein $k' \in \mathbb{N}$ und ein $s \in \{0, \dots, k_0 - 1\}$, so dass $k = k'k_0 + s$ ist. Somit gilt

$$c_k = c_{k' \cdot k_0 + s} \leq c_{k' \cdot k_0} \leq T^{k'} = T^{\frac{k-s}{k_0}} = (T^{\frac{k}{k_0}}) T^{-\frac{s}{k_0}} \leq \frac{T^{\frac{k}{k_0}}}{T} .$$

(Siehe [Beh00] S. 85.) ■

Nun die angekündigte direkte Abschätzung der Konvergenzgeschwindigkeit ergodischer Markov-Ketten gegen die Gleichgewichtsverteilung:

Satz 5 Sei $\{X_k\}_{k \in \mathbb{N}}$ eine ergodische Markov-Kette mit Zustandsraum S , $N \times N$ Übergangsmatrix P . Des Weiteren sei $k_0 \in \mathbb{N}$, so dass $P^{k_0} > 0$ ist, (solch ein k_0 existiert nach Satz 3), und es definiere $\delta := \min_{i,j \in S} p_{ij}^{(k_0)}$ und $T := 1 - N\delta$. Dann gilt:

(a) $T \in [0, 1]$.

(b) Falls $k_0 = 1$ ist, dann gilt $\left| p_{ij}^{(k)} - \pi_j \right| \leq T^k$ für alle $i, j \in S$, $k \in \mathbb{N}$.

(c) Falls $k_0 > 1$ ist, dann gilt $\left| p_{ij}^{(k)} - \pi_j \right| \leq \frac{T^{\frac{k}{k_0}}}{T}$ für alle $i, j \in S$, $k \in \mathbb{N}$.

(Vergl. [Beh00] S. 84, 85.) □

BEWEIS (a) kann analog zu Lemma 1 (b)(ii) bewiesen werden.

(b) Nach Voraussetzung ist $k_0 = 1$ und man setze $M_j^{(1)} = \max_{i \in S} p_{ij} := p_{i_0j}$. Da im allgemeinen p_{i_0s} für alle $s \in S$ stets größer oder gleich $\delta := \min_{i,j \in S} p_{ij}$ ist, gilt

$$1 = M_j^{(1)} + \sum_{s=1, s \neq j}^N p_{i_0s} \geq M_j^{(1)} + (N-1) \min_{i,j \in S} p_{ij} = M_j^{(1)} + (N-1)\delta . \quad (3.10)$$

Umformen von (3.10) ergibt

$$M_j^{(1)} \leq 1 - (N-1)\delta .$$

Subtrahiert man mit $m_j^{(1)} = \min_{i \in S} p_{ij}$, so erhält man

$$M_j^{(1)} - m_j^{(1)} \leq 1 - (N-1)\delta - m_j^{(1)} = 1 - (N-1)\delta - \delta = 1 - N\delta = T .$$

Nach Lemma 1 (b)(i) und (iii) gilt nun für alle $k \in \mathbb{N}$ und $i, j \in S$

$$\left| p_{ij}^{(k)} - \pi_j \right| \leq M_j^{(k)} - m_j^{(k)} \leq T^{k-1} (M_j^{(1)} - m_j^{(1)}) \leq T^k , \quad (3.11)$$

womit (b) bewiesen ist.

(c) In dem Beweis von (c) wird Lemma 2 angewendet.

Man setze $c_k := M_j^{(k)} - m_j^{(k)}$ mit $M_j^{(k)} = \max_{i \in S} p_{ij}^{(k)}$ und $m_j^{(k)} = \min_{i \in S} p_{ij}^{(k)}$ für alle $k \in \mathbb{N}$. Es folgt $c_k \in [0, \infty)$, c_k ist eine nach Lemma 1 (a) bezüglich $k \in \mathbb{N}$ fallende Folge und es gilt $c_{k' \cdot k_0} = M_j^{(k' \cdot k_0)} - m_j^{(k' \cdot k_0)} \stackrel{(3.11)}{\leq} T^{k' \cdot k_0} \leq T^{k'}$.

Die Voraussetzungen von Lemma 2 sind somit erfüllt, und es gilt insgesamt mit Lemma 1 (b) (i) und (3.9)

$$\left| p_{ij}^{(k)} - \pi_j \right| \leq M_j^{(k)} - m_j^{(k)} \leq \frac{T^{\frac{k}{k_0}}}{T} \text{ für alle } i, j \in S, k \in \mathbb{N} .$$

(In Anlehnung an [Beh00] S. 85.) ■

3.2.3 Direkte Konvergenzabschätzung Teil 2

Eine alternative direkte Konvergenzabschätzung kann über das Theorem von Doeblin gemacht werden. Das Theorem von Doeblin gilt nicht nur für ergodische Markov-Ketten sondern für alle Markov-Ketten, deren Übergangsmatrix allerdings kleinen Einschränkungen unterliegen muss. Für derlei Markov-Ketten gibt das Theorem von Doeblin außerdem Auskunft über die Existenz und Eindeutigkeit einer Gleichgewichtsverteilung. Die aus dem Theorem von Doeblin resultierende Abschätzung der Konvergenzgeschwindigkeit, sowie das Theorem an sich werden im Folgenden zur Anschauung und zu Diskussions- und Vergleichszwecken ohne Beweis angegeben.

Theorem 1 (Theorem von Doeblin) *Sei $\{X_k\}_{k \in \mathbb{N}}$ eine Markov-Kette mit Zustandsraum S und sei $P = (p_{ij})_{i,j \in S}$ die zur Markov-Kette gehörende $N \times N$ Übergangsmatrix, für die $p_{ij_0} \geq \epsilon$ für ein $j_0 \in S$, $\epsilon > 0$ und alle $i \in S$ ist. Die Markov-Kette besitzt dann eine eindeutig bestimmte Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$, mit $\pi_{j_0} \geq \epsilon$ und für alle Startverteilungen $\alpha^{(0)} \in \mathbb{R}^N$ gilt*

$$\|\alpha^{(0)T} P^k - \pi^T\|_1 \leq 2(1 - \epsilon)^k \text{ für alle } k \in \mathbb{N} . \quad (3.12)$$

(Vergl. [Str05] S. 28.) □

Bei der verwendeten Norm handelt es sich um die Betragssummennorm. Hierbei gilt für einen beliebigen Vektor $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$, dass $\|x\|_1 = \sum_{i=1}^N |x_i|$ ist.

Durch Anwendung und Erweiterung des Theorems von Doeblin kann folgende Schranke der Konvergenzgeschwindigkeit ergodischer Markov-Ketten hin zu ihrer Gleich-

gewichtsverteilung angegeben werden

$$\|\alpha^{(0)T} P^k - \pi^T\|_1 \leq 2(1 - \min_{i,j \in S} p_{ij}^{(k_0)})^{\lfloor \frac{k}{k_0} \rfloor}, \quad (3.13)$$

für alle $k \in \mathbb{N}$ und eine beliebige Startverteilung $\alpha^{(0)} \in \mathbb{R}^N$, wobei $P^{k_0} > 0$ ist und mit $\lfloor \frac{k}{k_0} \rfloor$ der ganzzahlige Teil von $\frac{k}{k_0}$ bezeichnet wird.

Der Beweis sowohl des Theorems von Doeblin als auch der obigen Konvergenzabschätzung kann in [Str05] auf den Seiten 28 bis 30 und Seite 53 nachgelesen werden.

Bei den bisherigen sowie den noch folgenden Abschätzungen der Konvergenzgeschwindigkeit einer Markov-Kette hin zu ihrer Gleichgewichtsverteilung ist es stets wünschenswert, dass die angegebenen Schranken nahe bei Null liegen. In diesem Fall kann dann davon ausgegangen werden, dass die Konvergenzgeschwindigkeit hoch ist. Aufgrund dieser Tatsache sind die bisher angegebenen direkten Schranken der Konvergenzgeschwindigkeit unter Umständen nur von geringem Nutzen. Wenn zum Beispiel $\min_{i,j \in S} p_{ij}$ sehr klein ist, dann liegt T aus Satz 5 nahe bei Eins und T^k aus Satz 5 (b) wird in dem Fall erst zu einem sehr späten Zeitpunkt $k \in \mathbb{N}$ wünschenswert klein (d.h. nahe Null). Dies ist sehr unschön, da sehr lange iteriert werden muss, damit die Abschätzung nicht zu grob gerät.

Bei Satz 5 (c) und (3.13) treten eventuell schon Probleme bei der Bestimmung von $\min_{i,j \in S} p_{ij}^{(k_0)}$ auf. Gegebenenfalls ist es mit einem hohen Rechenaufwand verbunden, bis ein k_0 gefunden wird, so dass $P^{k_0} > 0$ ist. Des Weiteren gilt im Allgemeinen auch, dass $\frac{T^{\frac{k}{k_0}}}{T}$ verhältnismäßig groß wird, je größer k_0 ist. Dies soll die folgende Bemerkung verdeutlichen.

Bemerkung 3 Für $T \in [0, 1]$ und $k, k_0 \in \mathbb{N}$ gilt

$$\frac{T^{\frac{k}{k_0}}}{T} \leq \frac{T^{\frac{k}{k_0+1}}}{T}. \quad \square$$

BEWEIS Mit $T \in [0, 1]$ gilt

$$\begin{aligned} \sqrt[k_0]{T} &\leq \sqrt[k_0+1]{T} \\ \Leftrightarrow T^{\frac{1}{k_0}} &\leq T^{\frac{1}{k_0+1}} \\ \Leftrightarrow T^{\frac{k}{k_0}} &\leq T^{\frac{k}{k_0+1}} \\ \Leftrightarrow \frac{T^{\frac{k}{k_0}}}{T} &\leq \frac{T^{\frac{k}{k_0+1}}}{T}. \quad \blacksquare \end{aligned}$$

Auch $2(1 - \min_{i,j \in S} p_{ij}^{(k_0)})^{\lfloor \frac{k}{k_0} \rfloor}$ aus (3.13) kann bei kleinem $\min_{i,j \in S} p_{ij}^{(k_0)}$ recht groß werden. Generell lässt sich sagen, dass die Schranke der Konvergenzgeschwindigkeit aus Satz 5 (b) (d.h. im Fall $k_0 = 1$) stets kleiner und dementsprechend besser ist als die Schranke, die sich nach Doeblin für $k_0 = 1$ ergibt. Veranschaulicht und bewiesen wird dies in Bemerkung 4.

Bemerkung 4 Sei $k \in \mathbb{N}$, $P = (p_{ij})_{i,j \in S}$ eine stochastische, echt positive $N \times N$ Matrix und $\delta := \min_{i,j \in S} p_{ij} (>0)$, dann gilt

$$(1 - N\delta)^k < 2(1 - \delta)^k . \quad \square$$

BEWEIS Es gilt

$$\begin{aligned} N\delta &> \delta \\ \Leftrightarrow N\delta - 1 &> \delta - 1 \\ \Leftrightarrow 1 - N\delta &< 1 - \delta \\ \Leftrightarrow (1 - N\delta)^k &< (1 - \delta)^k \\ \Leftrightarrow (1 - N\delta)^k &< 2(1 - \delta)^k . \quad \blacksquare \end{aligned}$$

Aufgrund dieser Tatsache wurde Satz 5 detailliert bearbeitet und die Konvergenzabschätzung nach Doeblin nur angegeben.

Das folgende Zahlenbeispiel soll die Unterschiede zwischen den direkten Konvergenzabschätzungen verdeutlichen.

Beispiel 15 Sei $S = \{1, 2\}$, $P = \frac{1}{4} \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$ die zur Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ gehörende 2×2 Übergangsmatrix, $\alpha^{(0)} = (1, 0)^T$ die Startverteilung und $\pi = (\frac{1}{3}, \frac{2}{3})^T$ die zur Markov-Kette gehörende Gleichgewichtsverteilung (bei Bedarf sind die Bedingungen (3.1), (3.2) und (3.3) der Definition einer Gleichgewichtsverteilung nachzuprüfen). Diese Markov-Kette ist nach Satz 3 ergodisch, und es gilt $k_0 = 1$, $\delta := \min_{i,j \in S} p_{ij}^{(k_0)} = \frac{1}{4}$ und $T = 1 - N\delta = 1 - 2 \cdot \frac{1}{4} = \frac{1}{2}$.

Nach Satz 5 ist

$$\max\{|p_{ij}^{(k)} - \pi_j| \mid i, j \in S\} \leq T^k = \left(\frac{1}{2}\right)^k$$

und der Satz von Doeblin bzw. (3.13) liefert

$$\|(1, 0)P^k - (\pi_1, \pi_2)\|_1 \leq 2(1 - \delta)^k = 2 \left(\frac{3}{4}\right)^k .$$

Man betrachte beispielsweise den Fall $k = 4$. Damit gilt $P^4 = \frac{1}{256} \begin{pmatrix} 86 & 170 \\ 85 & 171 \end{pmatrix}$ und

$$\begin{aligned} |p_{11}^{(4)} - \pi_1| &= \left| \frac{86}{256} - \frac{1}{3} \right| = \frac{1}{384}, \\ |p_{12}^{(4)} - \pi_2| &= \left| \frac{170}{256} - \frac{2}{3} \right| = \frac{1}{384}, \\ |p_{21}^{(4)} - \pi_1| &= \left| \frac{85}{256} - \frac{1}{3} \right| = \frac{1}{768}, \\ |p_{22}^{(4)} - \pi_2| &= \left| \frac{171}{256} - \frac{2}{3} \right| = \frac{1}{768} \end{aligned}$$

bzw.

$$\begin{aligned} \left\| \alpha^{(0)T} P^4 - \pi^T \right\|_1 &= \left\| \frac{1}{256} (1, 0) \begin{pmatrix} 86 & 170 \\ 85 & 171 \end{pmatrix} - \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \right\|_1 \\ &= \left\| \begin{pmatrix} \frac{43}{128} & \frac{85}{128} \end{pmatrix} - \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \right\|_1 = \left\| \begin{pmatrix} \frac{1}{384} & -\frac{1}{384} \end{pmatrix} \right\|_1 = \frac{1}{192}. \end{aligned}$$

Mit Satz 5 ergibt sich

$$\max\{|p_{ij}^{(4)} - \pi_j| \mid i, j \in S\} \leq \frac{1}{16}$$

und (3.13) liefert

$$\left\| \alpha^{(0)T} P^4 - \pi^T \right\|_1 \leq \frac{81}{128}.$$

Wie man sehr leicht nachrechnen kann, liegt hierbei $\frac{1}{16}$ näher an den genauen Werten $\frac{1}{384}$ und $\frac{1}{768}$, als $\frac{81}{128}$ an $\frac{1}{192}$. Satz 5 liefert somit eine zwar ungenaue, jedoch trotzdem bessere Schranke als (3.13). (Vergl. [Beh00] S. 85.) \square

3.2.4 Konvergenzabschätzung mittels Eigenwerten

Ein anderer Ansatz, die Konvergenzgeschwindigkeit abzuschätzen, kann über die Eigenwerte der Übergangsmatrix P gemacht werden. Dazu wird ein Theorem von zentraler Bedeutung angegeben. Aus diesem Theorem von Perron-Frobenius können dann Schlussfolgerungen gezogen werden, die Informationen über die Konvergenzgeschwindigkeit einer Markov-Kette gegen ihre Gleichgewichtsverteilung liefern.

Theorem 2 (Perron-Frobenius Theorem) *Sei P eine stochastische $N \times N$ Matrix. Des Weiteren gebe es ein $k \in \mathbb{N}$, so dass $P^k > 0$ ist. P habe als Eigenwerte $\lambda_1, \dots, \lambda_N$ mit $|\lambda_1| \geq \dots \geq |\lambda_N|$. Dann gilt*

(a) Der Eigenwert λ_1 ist reellwertig und positiv.

(b) $\lambda_1 > |\lambda_i|$ für alle $i = 1, \dots, N$.

(c) Der zu λ_1 gehörende N -dimensionale linke Eigenvektor $u_1 := (u_{11}, \dots, u_{1N})^T$ und N -dimensionale rechte Eigenvektor $v_1 := (v_{11}, \dots, v_{1N})^T$ kann so gewählt werden, dass alle Einträge positiv sind und $u_1^T v_1 = 1$ ist. (Vergl. [Sch03] S. 22.) \square

Der Beweis dieses Theorems ist recht lang und für diese Arbeit im Weiteren nicht von Bedeutung. Deswegen wird der Beweis hier nicht geführt. Nachzulesen ist er zum Beispiel in [Sen81] auf den Seiten 4 bis 6.

Korollar 1 Sei $\{X_k\}_{k \in \mathbb{N}}$ eine ergodische Markov-Kette mit Zustandsraum S , $N \times N$ Übergangsmatrix P und Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$. Des Weiteren habe P die Eigenwerte $\lambda_1, \dots, \lambda_N$ mit $|\lambda_1| \geq \dots \geq |\lambda_N|$, v_i sei der rechte N -dimensionale Eigenvektor zum Eigenwert λ_i und u_i der linke N -dimensionale Eigenvektor zum Eigenwert λ_i ($i \in S$). Dann gilt $\lambda_1 = 1$, $v_1 = e := (1, \dots, 1)^T$ und $u_1 = \pi$. Außerdem ist $1 > |\lambda_i|$ für alle $i = 2, \dots, N$. (Vergl. [Sch03] S. 22.) \square

BEWEIS Es gilt $Pe = \begin{pmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^N p_{1j} \\ \vdots \\ \sum_{j=1}^N p_{Nj} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1e$.

Außerdem gilt nach Definition 12 (iii), dass $\pi^T P = \pi^T = \pi^T 1$ ist. Somit ist e ein rechter Eigenvektor und π ein linker Eigenvektor zum Eigenwert 1 von P .

Nun muss noch bewiesen werden, dass 1 der größte Eigenwert von P ist. Dazu betrachte man einen beliebigen Eigenwert λ_i von P und den zu λ_i gehörenden rechten Eigenvektor $v_i = (v_{i1}, \dots, v_{iN})^T$ ($i \in S$), außerdem wird $|v_{i0}| := \max_{j \in S} |v_{ij}|$ definiert. Dann ergibt sich aus der Definitionsgleichung von Eigenwert und Eigenvektor (siehe (4.1)), dass $\lambda_i v_{i0} = \sum_{j=1}^N p_{i0j} v_{ij}$ ist und

$$\begin{aligned} |\lambda_i| |v_{i0}| &= |\lambda_i v_{i0}| = \left| \sum_{j=1}^N p_{i0j} v_{ij} \right| \leq \sum_{j=1}^N p_{i0j} |v_{ij}| \\ &\leq \sum_{j=1}^N p_{i0j} |v_{i0}| = |v_{i0}| \sum_{j=1}^N p_{i0j} = |v_{i0}| \end{aligned}$$

gilt. Es folgt, dass jeder beliebige Eigenwert λ_i von P kleiner oder gleich Eins ist. Insgesamt ist damit $\lambda_1 = 1$ und des Weiteren ist, mit dem Theorem von Perron-Frobenius, λ_1 echt größer als alle anderen Eigenwerte, womit dann alle Eigenwerte λ_i ($i = 2, \dots, N$) echt kleiner als 1 sind. (Angelehnt an [Sch03] S. 22.) \blacksquare

Mit Hilfe von Korollar 1 kann nun in Satz 6 eine Abschätzung der Konvergenzgeschwindigkeit einer Markov-Kette hin zu ihrer Gleichgewichtsverteilung über den

betragsmäßig zweitgrößten Eigenwert von P angegeben werden. Um diesen Satz beweisen zu können, muss allerdings als Einschränkung angenommen werden, die Eigenwerte von P seien disjunkt. Außerdem bedarf es für den Beweis von Satz 6 vorab einer Definition.

Definition 13 Seien A und B beliebige $N \times N$ Matrizen, dann schreibt man

$$A \leq B, \text{ wenn } a_{ij} \leq b_{ij} \text{ für alle } i, j = 1, \dots, N$$

ist. □

Außerdem wird in Satz 6 das sogenannte Landau Symbol O verwendet. Im Allgemeinen schreibt man $f(n) = O(g(n))$, wenn $|f(n)| \leq c|g(n)|$ ist, wobei f und g beispielsweise reellwertige Funktionen sind (siehe dazu [Win03] Seite 126).

Satz 6 Sei $\{X_k\}_{k \in \mathbb{N}}$ eine ergodische Markov-Kette mit Zustandsraum S , $N \times N$ Übergangsmatrix P und Gleichgewichtsverteilung $\pi = (\pi_1, \dots, \pi_N)^T \in \mathbb{R}^N$. Man nehme an, alle Eigenwerte $\lambda_1, \dots, \lambda_N$ von P seien voneinander verschieden und $|\lambda_1| \geq \dots \geq |\lambda_N|$, dann gilt

$$|p_{ij}^{(k)} - \pi_j| = O(|\lambda_2|^k) \text{ für alle } i, j \in S, k \in \mathbb{N}. \quad \square$$

BEWEIS Da nach Voraussetzung alle Eigenwerte von P verschieden sind, kann P^k ($k \in \mathbb{N}$) in Spektraldarstellung (siehe dazu Satz 14 und Lemma 6 im Anhang) angegeben werden. Somit gilt (4.4), d. h. $P^k = \sum_{i=1}^N \lambda_i^k v_i u_i^T$. Nach Korollar 1 weiß man, dass für den größten Eigenwert und seinen rechten bzw. linken Eigenvektor gilt $\lambda_1 = 1$, $v_1 = (1, \dots, 1)^T (= e)$ und $u_1 = \pi$. Da die Eigenwerte als der Größe ihrer Beträge nach geordnet betrachtet werden, ist $|\lambda_2|$ der betragsmäßig zweitgrößte Eigenwert von P und es gilt

$$\begin{aligned} P^k - \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix} &= P^k - 1e\pi^T = P^k - \lambda_1^k v_1 u_1^T \stackrel{(4.4)}{=} \sum_{i=1}^N \lambda_i^k v_i u_i^T - \lambda_1^k v_1 u_1^T = \sum_{i=2}^N \lambda_i^k v_i u_i^T \\ &\leq \sum_{i=2}^N |\lambda_2|^k v_i u_i^T \leq |\lambda_2|^k \sum_{i=2}^N v_i u_i^T. \end{aligned}$$

Außerdem weiß man nach Korollar 1, dass $|\lambda_2|$ echt kleiner als 1 ist und somit $|\lambda_2|^k$ für $k \in \mathbb{N}$ gegen unendlich gegen Null konvergiert.

Abschließend gilt dann

$$|p_{ij}^{(k)} - \pi_j| \leq |\lambda_2|^k \sum_{l=2}^N v_{li} u_{lj}$$

$$\Rightarrow |p_{ij}^{(k)} - \pi_j| = O(|\lambda_2|^k) \text{ für alle } i, j \in S, k \in \mathbb{N} .$$

(v_{li} bezeichnet hierbei den i -ten Eintrag des l -ten Eigenvektors von P . Analog dazu u_{lj} .) (Angelehnt an [Sch03] S. 22 und 23). ■

Das nachfolgende Beispiel 16 gibt eine Konvergenzabschätzung mittels Satz 6 für die Markov-Kette aus Beispiel 15 an.

Beispiel 16 Es seien S , P und π wie in Beispiel 15 gegeben. Nach Korollar 1 ist $\lambda_1 = 1$. λ_2 lässt sich ganz einfach über die charakteristische Gleichung (siehe (4.2)) berechnen. Man erhält $\lambda_2 = \frac{1}{4}$. Nach Satz 6 ergibt sich damit folgende Konvergenzabschätzung

$$|p_{ij}^{(k)} - \pi_j| = O\left(\left(\frac{1}{4}\right)^k\right) \text{ für alle } i, j \in S, k \in \mathbb{N}.$$

Im Gegensatz dazu lieferte Satz 5

$$\max\{|p_{ij}^{(k)} - \pi_j| \mid i, j \in S\} \leq T^k = \left(\frac{1}{2}\right)^k . \quad \square$$

Die Problematik bei derlei Konvergenzabschätzungen liegt darin, den Eigenwert λ_2 überhaupt zu bestimmen. Bei sehr großem Zustandsraum S ist dies mit einem beträchtlichen Rechenaufwand verbunden.

Bei allen bisher angegebenen Abschätzungen handelt es sich um geometrische Schranken der Konvergenzgeschwindigkeit. Allgemein wird unter geometrischer Konvergenzgeschwindigkeit verstanden, dass eine Folge, die geometrisch konvergiert, für große Werte $k \in \mathbb{N}$ wie $c\beta^k$ konvergiert, wobei $0 \leq \beta \leq 1$ und $c \in \mathbb{R}_{>0}$ ist. (Für nähere Informationen dazu siehe [Vol] Seite 6 und 7.) Somit gilt in allen bisher betrachteten Fällen, dass $p_{ij}^{(k)}$ für alle $i, j \in S$ geometrisch schnell gegen π_j konvergiert.

3.3 Die Gleichgewichtsverteilung reversibler Markov-Ketten

3.3.1 Reversible Markov-Ketten

In diesem Abschnitt werden ergodische Markov-Ketten betrachtet, die als zusätzliche Eigenschaft Reversibilität besitzen. Dazu zuerst die Definition von reversibel.

Definition 14 Eine Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum S bzw. das zur Markov-Kette gehörende Paar (P, π) bezeichnet man als reversibel, wenn

$$\pi_i p_{ij} = \pi_j p_{ji} \text{ für alle } i, j \in S \quad (3.14)$$

ist. P ist dabei wie gehabt die zur Markov-Kette gehörende $N \times N$ Übergangsmatrix und $\pi \in \mathbb{R}^N$ die zur Markov-Kette gehörende Gleichgewichtsverteilung. \square

Die Gleichung in (3.14) wird als „Detailed-Balance-Equation“ oder auch „Detailed-Balance-Bedingung“ bezeichnet.

Nun wird ein Beispiel einer reversiblen Markov-Kette und einer Markov-Kette, die nicht reversibel ist, angegeben.

Beispiel 17 (Zufällige Irrfahrten auf Graphen) Man betrachte einen Graph $G = (V, K)$, wobei $V = \{v_1, \dots, v_N\}$ die Menge aller Eckpunkte ist und K die Menge aller Kanten darstellt, die jeweils zwei Eckpunkte miteinander verbindet. Zwei durch eine Kante verbundene Eckpunkte werden als Nachbarn bezeichnet. Für jedes Paar $v_i, v_j \in V$ von Eckpunkten soll es außerdem einen „Pfad“ geben, der die Eckpunkte v_i und v_j miteinander verbindet.

In Abbildung 3.1 ist ein solcher Graph dargestellt. Dabei ist $V = \{v_1, \dots, v_7\}$ und $K = \{(v_1, v_7), (v_1, v_2), (v_2, v_3), (v_2, v_6), (v_2, v_5), (v_3, v_4), (v_3, v_5), (v_4, v_5), (v_5, v_6), (v_6, v_7)\}$ eine Menge mit 10 Kanten.

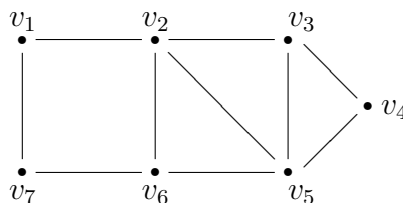


Abbildung 3.1: Zufällige Irrfahrt auf einem Graph

Im Allgemeinen ist nun die zufällige Irrfahrt auf dem Graph $G = (V, K)$ eine Markov-Kette mit Zustandsraum $S = \{1, \dots, N\}$ und dem Übergangsmechanismus, dass, wenn man sich zum Zeitpunkt $k \in \mathbb{N}$ in dem Eckpunkt v_i ($i \in S$) befindet, man zum

Zeitpunkt $k + 1$ zu einem der Nachbarn übergeht, wobei alle Nachbarn mit gleicher Wahrscheinlichkeit besucht werden. Wenn man nun die Anzahl der Nachbarn der Ecke v_i mit d_i bezeichnet, dann hat somit die Übergangsmatrix $P = (p_{ij})_{i,j \in S}$ die Einträge

$$p_{ij} = \begin{cases} \frac{1}{d_i} & \text{falls die Eckpunkte } v_i \text{ und } v_j \text{ Nachbarn sind,} \\ 0 & \text{sonst.} \end{cases} \quad (3.15)$$

Mit $d := \sum_{i=1}^N d_i$ ist die Gleichgewichtsverteilung der Markov-Kette gegeben durch

$$\pi = \left(\frac{d_1}{d}, \dots, \frac{d_N}{d} \right)^T .$$

Dieses π erfüllt die Eigenschaften (3.1) und (3.2) aus der Definition der Gleichgewichtsverteilung, da

$$\begin{aligned} \frac{d_i}{d} &\geq 0 \text{ für } i \in S \text{ und} \\ \sum_{i=1}^N \frac{d_i}{d} &= \frac{1}{d} \sum_{i=1}^N d_i = \frac{d}{d} = 1 \end{aligned}$$

ist. Des Weiteren gilt

$$\begin{aligned} \pi^T P &= \left(\frac{d_1}{d}, \dots, \frac{d_N}{d} \right) \begin{pmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{pmatrix} = \left(\sum_{i=1}^N \frac{d_i}{d} p_{i1}, \dots, \sum_{i=1}^N \frac{d_i}{d} p_{iN} \right) \\ &= \frac{1}{d} \left(\sum_{i=1}^N d_i p_{i1}, \dots, \sum_{i=1}^N d_i p_{iN} \right) . \end{aligned}$$

Dabei ist p_{ij} bei festem j und $i = 1, \dots, N$ gemäß (3.15) genau dann gleich $\frac{1}{d_i}$, wenn v_i ein Nachbar von v_j ist. Dies ist so oft der Fall, wie v_j Nachbarn hat, also genau d_j mal, so dass damit $\sum_{i=1}^N d_i p_{ij} = d_j$ ist. Es gilt somit

$$\frac{1}{d} \left(\sum_{i=1}^N d_i p_{i1}, \dots, \sum_{i=1}^N d_i p_{iN} \right) = \frac{1}{d} (d_1, \dots, d_N) = \left(\frac{d_1}{d}, \dots, \frac{d_N}{d} \right) = \pi^T$$

und auch die Eigenschaft (3.3) aus der Definition der Gleichgewichtsverteilung ist erfüllt.

Bei dem obigen Zahlenbeispiel ist $\pi = \left(\frac{2}{20}, \frac{4}{20}, \frac{3}{20}, \frac{2}{20}, \frac{4}{20}, \frac{3}{20}, \frac{2}{20} \right)^T$.

Da nun für beliebige $i, j \in S$ gilt, dass

$$\pi_i p_{ij} = \begin{cases} \frac{d_i}{d} \frac{1}{d_i} = \frac{1}{d} = \frac{d_j}{d} \frac{1}{d_j} = \pi_j p_{ji} & \text{falls die Eckpunkte } v_i \text{ und } v_j \text{ Nachbarn sind,} \\ 0 = \pi_j p_{ji}, & \text{sonst} \end{cases}$$

ist, ist die Markov-Kette reversibel. (Vergl. [Sch03] S. 41.) □

(Beispiel 17 kann als Verallgemeinerung von Beispiel 1 angesehen werden.)

Beispiel 18 Es sei eine Markov-Kette durch den Zustandsraum $S = \{1, 2, 3\}$, eine beliebige Anfangsverteilung und die Übergangsmatrix

$$P = \begin{pmatrix} a & a-b & 2b \\ a+b & b & a-b \\ 0 & a+b & a \end{pmatrix} \text{ mit } a, b > 0, b \leq a \text{ und } 2a + b = 1$$

gegeben. Durch einfaches nachrechnen kann überprüft werden, dass $\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^T$ die zur Markov-Kette gehörende Gleichgewichtsverteilung ist. Diese Markov-Kette ist nicht reversibel, denn

$$\pi_1 p_{12} = \frac{1}{3}(a-b) \neq \frac{1}{3}(a+b) = \pi_2 p_{21} .$$

(Vergl. [Beh00] S. 80.) □

3.3.2 Präzisierte Konvergenzabschätzung mittels Eigenwerten

Da die Ergebnisse aus Abschnitt 3.2 allgemein für ergodische Markov-Ketten gültig sind, gelten sie natürlich auch für ergodische Markov-Ketten, die zusätzlich noch reversibel sind. Die Fragen, ob es im Fall ergodischer, reversibler Markov-Ketten eine Gleichgewichtsverteilung π gibt, ob diese eindeutig bestimmt ist und ob bei beliebiger Anfangsverteilung $\alpha^{(0)}, \alpha^{(0)T} P^k$ für $k \in \mathbb{N}$ gegen unendlich immer gegen π^T konvergiert, wurden somit schon beantwortet. Auch die Abschätzungen der Konvergenzgeschwindigkeit verlieren nicht ihre Gültigkeit. Es lassen sich allerdings im Fall der zusätzlichen Reversibilität ergodischer Markov-Ketten noch weitere, präzisere Konvergenzabschätzungen machen.

Der nun folgende Satz 7 ist eine Präzisierung von Satz 6. Wie man im Beweis des Satzes sehen wird, ist es für reversible, ergodische Markov-Ketten auch nicht mehr nötig vorauszusetzen, dass alle Eigenwerte der zur Markov-Kette gehörenden Übergangsmatrix P voneinander verschieden sind.

Satz 7 Sei $\{X_k\}_{k \in \mathbb{N}}$ eine reversible und ergodische Markov-Kette mit Zustandsraum $S = \{1, \dots, N\}$, $N \times N$ Übergangsmatrix P und Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$. Des

Weiteren seien $\lambda_1 > |\lambda_2| \geq \dots \geq |\lambda_N|$ die nach der Größe ihrer Beträge geordneten zu P gehörenden Eigenwerte. Dann gilt

$$|p_{ij}^{(k)} - \pi_j| \leq \frac{1}{\sqrt{\min_{i \in S} \pi_i}} |\lambda_2|^k \text{ für alle } i, j \in S, k \in \mathbb{N}. \quad (3.16)$$

(Vergl. [Sch03] S. 45.) □

BEWEIS Man setze $D := \text{diag}(\sqrt{\pi_i})_{i \in S}$ und $A := DPD^{-1}$. Dann gilt

$$\begin{aligned} A = DPD^{-1} &= \begin{pmatrix} \sqrt{\pi_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\pi_N} \end{pmatrix} \begin{pmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\pi_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\pi_N}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\sqrt{\pi_1}}{\sqrt{\pi_1}} p_{11} & \cdots & \frac{\sqrt{\pi_1}}{\sqrt{\pi_N}} p_{1N} \\ \vdots & \ddots & \vdots \\ \frac{\sqrt{\pi_N}}{\sqrt{\pi_1}} p_{N1} & \cdots & \frac{\sqrt{\pi_N}}{\sqrt{\pi_N}} p_{NN} \end{pmatrix} = \left(\frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} p_{ij} \right)_{i,j \in S}. \end{aligned}$$

Da P reversibel ist, ergibt sich nun aus der Detailed-Balance-Bedingung (siehe (3.14))

$$\begin{aligned} \pi_i p_{ij} = \pi_j p_{ji} &\Rightarrow \frac{\pi_i}{\sqrt{\pi_j}} p_{ij} = \frac{\pi_j}{\sqrt{\pi_j}} p_{ji} = \sqrt{\pi_j} p_{ji} \\ &\Rightarrow \frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} p_{ij} = \frac{\sqrt{\pi_j}}{\sqrt{\pi_i}} p_{ji} \end{aligned}$$

für alle $i, j \in S$. Bei A handelt es sich somit um eine symmetrische Matrix. Nach Bemerkung 9 (4) sind die Eigenvektoren zu den Eigenwerten einer symmetrischen Matrix (also A) orthogonal. Daraus folgt (siehe [Fis02] S. 304), dass es eine Orthonormalbasis des \mathbb{R}^N aus den Eigenvektoren w_1, \dots, w_N von A gibt. w_1, \dots, w_N sind damit insbesondere linear unabhängig. Nach Bemerkung 6 sind des Weiteren die rechten Eigenvektoren der symmetrischen Matrix A gleich den linken Eigenvektoren von A und nach Bemerkung 9 (3) sind alle Eigenwerte von A reell.

Sei nun λ_i der Eigenwert zum Eigenvektor w_i von A . Wenn man dies in die Definitionsgleichung von Eigenwert und Eigenvektor (siehe (4.1)) von A einsetzt, gilt

$$\begin{aligned} w_i^T A = w_i^T \lambda_i &\Rightarrow w_i^T DPD^{-1} = w_i^T \lambda_i \\ &\Rightarrow w_i^T DP = w_i^T D \lambda_i. \end{aligned} \quad (3.17)$$

Mit $u_i^T := w_i^T D$ folgt dann aus (3.17), dass $u_i^T P = u_i^T \lambda_i$ ist. Damit stimmen die Eigenwerte von P und A überein.

Aufgrund dieser Eigenschaften von A und der damit folgenden Tatsache, dass A

in Spektraldarstellung (siehe Satz 14) dargestellt werden kann, ergibt sich für alle $k \in \mathbb{N}$

$$\begin{aligned}
 P^k &= (D^{-1}DPD^{-1}D)^k = (D^{-1}AD)^k = (D^{-1}AD)(D^{-1}AD)\cdots(D^{-1}AD) \\
 &= D^{-1}A^kD \stackrel{(4.4)}{=} D^{-1}\left(\sum_{l=1}^N \lambda_l^k w_l w_l^T\right)D \\
 &= \sum_{l=1}^N \lambda_l^k D^{-1}w_l w_l^T D. \tag{3.18}
 \end{aligned}$$

Gemäß den obigen Ausführungen sind u_1, \dots, u_N die linken Eigenvektoren von P und aus Korollar 1 ist bekannt, dass $\lambda_1 = 1$ und $u_1 = \pi$ ist. Damit ist

$$w_1^T = u_1^T D^{-1} = \pi^T D^{-1} = (\pi_1, \dots, \pi_N) \begin{pmatrix} \frac{1}{\sqrt{\pi_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\pi_N}} \end{pmatrix} = (\sqrt{\pi_1}, \dots, \sqrt{\pi_N}) \tag{3.19}$$

und der erste Summand ($l = 1$) in (3.18) hat folgende Gestalt

$$\begin{aligned}
 &\lambda_1^k D^{-1}w_1 w_1^T D \\
 &\stackrel{(3.19)}{=} 1^k \begin{pmatrix} \frac{1}{\sqrt{\pi_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\pi_N}} \end{pmatrix} \begin{pmatrix} \sqrt{\pi_1} \\ \vdots \\ \sqrt{\pi_N} \end{pmatrix} (\sqrt{\pi_1}, \dots, \sqrt{\pi_N}) \begin{pmatrix} \sqrt{\pi_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\pi_N} \end{pmatrix} \\
 &= \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} (\pi_1, \dots, \pi_N) = \begin{pmatrix} \pi_1 \cdots \pi_N \\ \vdots \cdots \vdots \\ \pi_1 \cdots \pi_N \end{pmatrix}.
 \end{aligned}$$

Mit $w_l = (w_{l_1}, \dots, w_{l_N})$ gilt für alle weiteren Summanden ($l \geq 2 \in S$) in (3.18)

$$\begin{aligned}
 \lambda_l^k D^{-1}w_l w_l^T D &= \lambda_l^k \begin{pmatrix} \frac{1}{\sqrt{\pi_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\pi_N}} \end{pmatrix} \begin{pmatrix} w_{l_1} \\ \vdots \\ w_{l_N} \end{pmatrix} (w_{l_1}, \dots, w_{l_N}) \begin{pmatrix} \sqrt{\pi_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\pi_N} \end{pmatrix} \\
 &= \lambda_l^k \begin{pmatrix} \frac{1}{\sqrt{\pi_1}} w_{l_1} \\ \vdots \\ \frac{1}{\sqrt{\pi_N}} w_{l_N} \end{pmatrix} (\sqrt{\pi_1} w_{l_1}, \dots, \sqrt{\pi_N} w_{l_N})
 \end{aligned}$$

$$= \begin{pmatrix} \lambda_l^k \frac{\sqrt{\pi_1}}{\sqrt{\pi_1}} w_{l_1} w_{l_1} & & \lambda_l^k \frac{\sqrt{\pi_N}}{\sqrt{\pi_1}} w_{l_1} w_{l_N} \\ & \ddots & \\ \lambda_l^k \frac{\sqrt{\pi_1}}{\sqrt{\pi_N}} w_{l_N} w_{l_1} & & \lambda_l^k \frac{\sqrt{\pi_N}}{\sqrt{\pi_N}} w_{l_N} w_{l_N} \end{pmatrix} = (\lambda_l^k \frac{\sqrt{\pi_j}}{\sqrt{\pi_i}} w_{l_i} w_{l_j})_{i,j \in S} .$$

Insgesamt folgt somit, dass

$$p_{ij}^{(k)} = \pi_j + \frac{\sqrt{\pi_j}}{\sqrt{\pi_i}} \sum_{l=2}^N \lambda_l^k w_{l_i} w_{l_j}$$

für beliebige $i, j \in S$ und $k \in \mathbb{N}$ ist. Da λ_2 nach Korollar 1 der betragsmäßig größte Eigenwert ungleich Eins von A bzw. P ist und da nach Definition 12 für alle $j \in S$ gilt $\sqrt{\pi_j} \leq 1$, folgt nun

$$\begin{aligned} |p_{ij}^{(k)} - \pi_j| &\leq \frac{\sqrt{\pi_j}}{\sqrt{\pi_i}} |\lambda_2|^k \sum_{l=2}^N |w_{l_i}| |w_{l_j}| \leq \frac{1}{\sqrt{\pi_i}} |\lambda_2|^k \sum_{l=2}^N |w_{l_i}| |w_{l_j}| \\ &\leq \frac{1}{\sqrt{\min_{i \in S} \pi_i}} |\lambda_2|^k \sum_{l=2}^N |w_{l_i}| |w_{l_j}| \\ &= \frac{\sum_{l=2}^N |w_{l_i}| |w_{l_j}|}{\sqrt{\min_{i \in S} \pi_i}} |\lambda_2|^k . \end{aligned} \tag{3.20}$$

Nach der Cauchy-Schwarzschen-Ungleichung (siehe (4.8)) ist allerdings

$$\sum_{l=2}^N |w_{l_i}| |w_{l_j}| \leq \left(\sum_{l=1}^N w_{l_i}^2 \right)^{\frac{1}{2}} \left(\sum_{l=1}^N w_{l_j}^2 \right)^{\frac{1}{2}} \tag{3.21}$$

und die rechte Seite in (3.21) ist gleich Eins, da die Eigenvektoren w_i ($i \in S$) von A orthonormal zueinander sind. Insgesamt folgt somit für alle $i, j \in S$, $k \in \mathbb{N}$

$$|p_{ij}^{(k)} - \pi_j| \stackrel{(3.20)}{\leq} \frac{\sum_{l=2}^N |w_{l_i}| |w_{l_j}|}{\sqrt{\min_{i \in S} \pi_i}} |\lambda_2|^k \leq \frac{1}{\sqrt{\min_{i \in S} \pi_i}} |\lambda_2|^k .$$

(Angelehnt an [Sch03] S. 44 und 45.) ■

Nun ein Beispiel dazu.

Beispiel 19 Es sei eine Markov-Kette durch den Zustandsraum $S = \{1, 2, 3\}$, die

Übergangsmatrix $P = \begin{pmatrix} \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{4}{9} & \frac{1}{9} & \frac{4}{9} \\ \frac{4}{9} & \frac{4}{9} & \frac{1}{9} \end{pmatrix}$ und eine beliebige Startverteilung gegeben.

Durch einfaches Überprüfen von (3.1), (3.2) und (3.3) sieht man, dass die Gleichgewichtsverteilung der Markov-Kette durch $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$ gegeben ist. Da P echt

größer als Null ist, ist die Markov-Kette nach Satz 3 ergodisch. Des Weiteren ist mit diesem P die Detailed-Balance-Equation erfüllt. Dies ist unmittelbar daran zu erkennen, dass P symmetrisch ist und π nur gleiche Einträge hat. Die Markov-Kette ist damit reversibel. Mit einem geeigneten Computerprogramm oder von Hand können nun die Eigenwerte der Matrix P berechnet werden. Man erhält $\lambda_1 = 1$, $\lambda_2 = -\frac{1}{3}$ und $\lambda_3 = -\frac{1}{3}$, so dass $\lambda_1 > \lambda_2 \geq \lambda_3$ ist. Die Anwendung von Satz 7 ergibt

$$|p_{ij}^{(k)} - \pi_j| \leq \frac{1}{\sqrt{\frac{1}{3}}} \left| -\frac{1}{3} \right|^k = \frac{\sqrt{3}}{3^k} \text{ für alle } i, j \in S, k \in \mathbb{N}. \quad \square$$

3.3.3 χ^2 -Kontrast

Eine alternative Abschätzung der Konvergenzgeschwindigkeit einer ergodischen und reversiblen Markov-Kette hin zu ihrer Gleichgewichtsverteilung, die auch den betragsmäßig zweitgrößten Eigenwert von P nutzt, kann über den so genannten χ^2 -Kontrast gemacht werden. Um diese Abschätzung angeben und anschließend beweisen zu können, werden einige Definitionen und Vorbemerkungen benötigt. Zuerst wird ein Skalarprodukt mit zugehöriger Norm eingeführt, die beide mit der Wahrscheinlichkeitsverteilung π verbunden sind.

Definition 15 ($\frac{1}{\pi}$ -Skalarprodukt und $\frac{1}{\pi}$ -Norm) Sei $\pi = (\pi_1, \dots, \pi_N)^T$ eine beliebige echt positive Wahrscheinlichkeitsverteilung aus \mathbb{R}^N . Für beliebige Vektoren $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ und $y = (y_1, \dots, y_N)^T \in \mathbb{R}^N$ bezeichne

$$\langle x, y \rangle_{\frac{1}{\pi}} := \sum_{i=1}^N \frac{x_i y_i}{\pi_i} \quad (3.22)$$

das $\frac{1}{\pi}$ -Skalarprodukt und

$$\|x\|_{\frac{1}{\pi}} := \left(\sum_{i=1}^N \frac{x_i^2}{\pi_i} \right)^{\frac{1}{2}} \quad (3.23)$$

die zugehörige $\frac{1}{\pi}$ -Norm. Insbesondere gilt damit

$$\|x\|_{\frac{1}{\pi}}^2 = \sum_{i=1}^N \frac{x_i^2}{\pi_i} = \langle x, x \rangle_{\frac{1}{\pi}}. \quad (3.24)$$

(Vergl. [Bré99] S. 201 und 202.) □

Für den Beweis des noch folgenden Satzes 8 ist es hilfreich $x \in \mathbb{R}^N$ mit Hilfe des $\frac{1}{\pi}$ -Skalarproduktes darzustellen. Um diese Darstellung von $x \in \mathbb{R}^N$ zu finden, ist

es vorerst wichtig, sich die symmetrische Matrix $A = DPD^{-1}$ aus dem Beweis von Satz 7 ins Gedächtnis zu rufen. Hierbei ist P wie gehabt die zur Markov-Kette gehörende $N \times N$ Übergangsmatrix und $D = \text{diag}(\sqrt{\pi_i})_{i \in S}$, wobei $\pi = (\pi_1, \dots, \pi_N)^T$ die Gleichgewichtsverteilung der Markov-Kette ist. Gemäß dem Beweis von Satz 8 gibt es Eigenvektoren w_1, \dots, w_N der Matrix A , die eine Orthonormalbasis des \mathbb{R}^N bilden, und es ist $w_i^T = u_i^T D^{-1}$, wobei u_1, \dots, u_N die linken Eigenvektoren von P sind. Daraus ergibt sich, dass (u_1, \dots, u_N) auch orthonormal in \mathbb{R}^N ist und dann insbesondere eine Basis des \mathbb{R}^N bildet. Somit kann nun jeder beliebige Vektor $x \in \mathbb{R}^N$ als Linearkombination von u_1, \dots, u_N dargestellt werden. Mit anderen Worten gilt: $x = \sum_{i=1}^N \beta_i u_i$, wobei $\beta_i \in \mathbb{R}$ ($i = 1, \dots, N$) ist. Weiterhin gilt im \mathbb{R}^N

$$\delta_{ij} := \langle u_i, u_j \rangle_{\frac{1}{\pi}} = \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{falls } i \neq j \end{cases} . \quad (3.25)$$

Somit gilt für das $\frac{1}{\pi}$ -Skalarprodukt $\langle x, u_j \rangle_{\frac{1}{\pi}}$

$$\begin{aligned} \langle x, u_j \rangle_{\frac{1}{\pi}} &= \left\langle \sum_{i=1}^N \beta_i u_i, u_j \right\rangle_{\frac{1}{\pi}} = \langle \beta_1 u_1, u_j \rangle_{\frac{1}{\pi}} + \dots + \langle \beta_N u_N, u_j \rangle_{\frac{1}{\pi}} \\ &= \beta_1 \langle u_1, u_j \rangle_{\frac{1}{\pi}} + \dots + \beta_N \langle u_N, u_j \rangle_{\frac{1}{\pi}} \stackrel{(3.25)}{=} \beta_j . \end{aligned}$$

Insgesamt ist somit für jedes $x \in \mathbb{R}^N$,

$$x = \sum_{i=1}^N \beta_i u_i = \sum_{j=1}^N \langle x, u_j \rangle_{\frac{1}{\pi}} u_j . \quad (3.26)$$

Des Weiteren werden die folgenden zwei Begriffe eingeführt.

Definition 16 (Abstand in Variation und χ^2 -Kontrast) Seien

$\alpha = (\alpha_1, \dots, \alpha_N)^T$ und $\beta = (\beta_1, \dots, \beta_N)^T$ zwei beliebige Wahrscheinlichkeitsverteilungen auf der Menge $S = \{1, \dots, N\}$, dann bezeichne

$$d_V(\alpha, \beta) = \frac{1}{2} |\alpha - \beta| = \frac{1}{2} \sum_{i=1}^N |\alpha_i - \beta_i| \quad (3.27)$$

den Abstand in Variation („distance in variation“) von α und β .

Falls β_i für jedes $i \in S$ echt größer als Null ist, dann sei

$$\chi^2(\alpha; \beta) = \sum_{i=1}^N \frac{(\alpha_i - \beta_i)^2}{\beta_i}$$

der χ^2 -Kontrast (oder χ^2 -Abstand) von α und β . (Vergl. [Bré99] S. 207 und 209.) \square

Man sieht sofort, dass

$$\chi^2(\alpha; \pi) = \sum_{i=1}^N \frac{(\alpha_i - \pi_i)^2}{\pi_i} \stackrel{(3.22)}{=} \langle \alpha - \pi, \alpha - \pi \rangle_{\frac{1}{\pi}} \stackrel{(3.24)}{=} \|\alpha - \pi\|_{\frac{1}{\pi}}^2 \quad (3.28)$$

ist.

Der Abstand in Variation kann durch den χ^2 -Kontrast abgeschätzt werden. Dieser Zusammenhang zwischen Abstand in Variation und χ^2 -Kontrast ist im folgenden Lemma angegeben.

Lemma 3 *Seien $\alpha = (\alpha_1, \dots, \alpha_N)^T$ und $\beta = (\beta_1, \dots, \beta_N)^T$ zwei beliebige Wahrscheinlichkeitsverteilungen auf der Menge $S = \{1, \dots, N\}$. Falls $\beta_i > 0$ für jedes $i \in S$ ist, dann gilt*

$$4d_V^2(\alpha, \beta) \leq \chi^2(\alpha; \beta) .$$

(Vergl. [Bré99] S. 209.) \square

BEWEIS Mit der Cauchy-Schwarzschen Ungleichung (siehe (4.8)) und da $\sum_{i=1}^N \beta_i = 1$ ist gilt

$$\begin{aligned} 4d_V^2(\alpha, \beta) &\stackrel{(3.27)}{=} \left(\sum_{i=1}^N |\alpha_i - \beta_i| \right)^2 = \left(\sum_{i=1}^N \frac{1}{\beta_i^{\frac{1}{2}}} |\alpha_i - \beta_i| \beta_i^{\frac{1}{2}} \right)^2 \\ &\stackrel{(4.8)}{\leq} \sum_{i=1}^N \frac{1}{\beta_i} (\alpha_i - \beta_i)^2 \sum_{i=1}^N \beta_i = \sum_{i=1}^N \frac{1}{\beta_i} (\alpha_i - \beta_i)^2 = \chi^2(\alpha; \beta) . \end{aligned}$$

(In Anlehnung an [Bré99] S. 209.) \blacksquare

Nun kann, basierend auf dem Vorhergehenden, eine Abschätzung über die Konvergenzgeschwindigkeit von $\alpha^{(0)T} P^k$ zur Gleichgewichtsverteilung π^T gemacht werden.

Satz 8 *Sei $\{X_k\}_{k \in \mathbb{N}}$ eine reversible und ergodische Markov-Kette mit Zustandsraum S , $N \times N$ Übergangsmatrix P und der Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$. Dann gilt mit jeder beliebigen Anfangsverteilung $\alpha^{(0)} \in \mathbb{R}^N$ und für jedes $k \geq 1 \in \mathbb{N}$, dass*

$$2d_V(\alpha^{(0)T} P^k, \pi) \leq C |\lambda_2|^k$$

ist, wobei $C := (\chi^2(\alpha^{(0)}; \pi))^{\frac{1}{2}}$ und λ_2 der betragsmäßig zweitgrößte Eigenwert von P ist. (Vergl. [Win03] S. 201.) \square

BEWEIS Durch Quadrierung kann die Behauptung wie folgt umformuliert werden:

$$4d_V^2(\alpha^{(0)T} P^k, \pi) \leq (C|\lambda_2|^k)^2 = \chi^2(\alpha^{(0)}; \pi) |\lambda_2|^{2k} \stackrel{(3.28)}{=} \|\alpha^{(0)} - \pi\|_{\frac{1}{\pi}}^2 |\lambda_2|^{2k} .$$

Nach Lemma 3 weiß man, dass $4d_V^2(\alpha^{(0)T} P^k, \pi) \leq \chi^2(\alpha^{(0)T} P^k; \pi)$ ist, wobei hier wiederum mit (3.28) $\chi^2(\alpha^{(0)T} P^k; \pi) = \left\| \alpha^{(0)T} P^k - \pi^T \right\|_{\frac{1}{\pi}}^2$ ist. Somit ist die Behauptung bewiesen, wenn gezeigt werden kann, dass gilt

$$\left\| \alpha^{(0)T} P^k - \pi^T \right\|_{\frac{1}{\pi}}^2 \leq |\lambda_2|^{2k} \|\alpha^{(0)} - \pi\|_{\frac{1}{\pi}}^2 . \quad (3.29)$$

Man setze $\rho_j := \langle \alpha^{(0)} - \pi, u_j \rangle_{\frac{1}{\pi}}$, wobei u_j wie gehabt der linke Eigenvektor von P^k zum Eigenwert λ_j^k ist. Man erhält nun

$$\begin{aligned} \left\| \alpha^{(0)T} P^k - \pi^T \right\|_{\frac{1}{\pi}}^2 &\stackrel{(3.3)}{=} \left\| \alpha^{(0)T} P^k - \pi^T P^k \right\|_{\frac{1}{\pi}}^2 = \left\| (\alpha^{(0)} - \pi)^T P^k \right\|_{\frac{1}{\pi}}^2 \\ &\stackrel{(3.26)}{=} \left\| \sum_{j=1}^N \langle \alpha^{(0)} - \pi, u_j \rangle_{\frac{1}{\pi}} u_j^T P^k \right\|_{\frac{1}{\pi}}^2 \stackrel{(4.1)}{=} \left\| \sum_{j=1}^N \lambda_j^k \langle \alpha^{(0)} - \pi, u_j \rangle_{\frac{1}{\pi}} u_j^T \right\|_{\frac{1}{\pi}}^2 \\ &= \left\| \sum_{j=1}^N \lambda_j^k \rho_j u_j^T \right\|_{\frac{1}{\pi}}^2 \stackrel{(3.24)}{=} \left\langle \sum_{j=1}^N \lambda_j^k \rho_j u_j^T, \sum_{j=1}^N \lambda_j^k \rho_j u_j^T \right\rangle_{\frac{1}{\pi}} \\ &= \left\langle \lambda_1^k \rho_1 u_1^T, \sum_{j=1}^N \lambda_j^k \rho_j u_j^T \right\rangle_{\frac{1}{\pi}} + \dots + \left\langle \lambda_N^k \rho_N u_N^T, \sum_{j=1}^N \lambda_j^k \rho_j u_j^T \right\rangle_{\frac{1}{\pi}} \\ &= \langle \lambda_1^k \rho_1 u_1^T, \lambda_1^k \rho_1 u_1^T \rangle_{\frac{1}{\pi}} + \dots + \langle \lambda_1^k \rho_1 u_1^T, \lambda_N^k \rho_N u_N^T \rangle_{\frac{1}{\pi}} + \dots + \\ &\quad \langle \lambda_N^k \rho_N u_N^T, \lambda_1^k \rho_1 u_1^T \rangle_{\frac{1}{\pi}} + \dots + \langle \lambda_N^k \rho_N u_N^T, \lambda_N^k \rho_N u_N^T \rangle_{\frac{1}{\pi}} \\ &\stackrel{(3.25)}{=} \lambda_1^{2k} \rho_1^2 + \lambda_2^{2k} \rho_2^2 + \dots + \lambda_N^{2k} \rho_N^2 . \end{aligned}$$

Mit Korollar 1 ergibt sich, dass

$$\begin{aligned} \lambda_1^{2k} \rho_1^2 &= \lambda_1^{2k} \langle \alpha^{(0)} - \pi, u_1 \rangle_{\frac{1}{\pi}}^2 = 1 \cdot \langle \alpha^{(0)} - \pi, \pi \rangle_{\frac{1}{\pi}}^2 \\ &\stackrel{(3.22)}{=} \left(\sum_{i=1}^N \frac{(\alpha_i^{(0)} - \pi_i) \pi_i}{\pi_i} \right)^2 = \left(\sum_{i=1}^N (\alpha_i^{(0)} - \pi_i) \right)^2 \\ &= \left(\sum_{i=1}^N \alpha_i^{(0)} - \sum_{i=1}^N \pi_i \right)^2 = 1 - 1 = 0 \end{aligned}$$

ist. Damit gilt

$$\begin{aligned}
 & \lambda_1^{2k} \rho_1^2 + \lambda_2^{2k} \rho_2^2 + \dots + \lambda_N^{2k} \rho_N^2 = 0 + \lambda_2^{2k} \rho_2^2 + \dots + \lambda_N^{2k} \rho_N^2 = \sum_{j=2}^N \lambda_j^{2k} \rho_j^2 \\
 & \leq |\lambda_2|^{2k} \sum_{j=2}^N \rho_j^2 = |\lambda_2|^{2k} \sum_{j=1}^N \rho_j^2 \\
 & \stackrel{(3.25)}{=} |\lambda_2|^{2k} (\langle \rho_1 u_1, \rho_1 u_1 \rangle_{\frac{1}{\pi}} + \dots + \langle \rho_1 u_1, \rho_N u_N \rangle_{\frac{1}{\pi}} \\
 & \quad + \dots + \langle \rho_N u_N, \rho_1 u_1 \rangle_{\frac{1}{\pi}} + \dots + \langle \rho_N u_N, \rho_N u_N \rangle_{\frac{1}{\pi}}) \\
 & = |\lambda_2|^{2k} \left(\left\langle \rho_1 u_1, \sum_{j=1}^N \rho_j u_j \right\rangle_{\frac{1}{\pi}} + \dots + \left\langle \rho_N u_N, \sum_{j=1}^N \rho_j u_j \right\rangle_{\frac{1}{\pi}} \right) \\
 & = |\lambda_2|^{2k} \left\langle \sum_{j=1}^N \rho_j u_j, \sum_{j=1}^N \rho_j u_j \right\rangle_{\frac{1}{\pi}} \stackrel{(3.26)}{=} |\lambda_2|^{2k} \langle \alpha^{(0)} - \pi, \alpha^{(0)} - \pi \rangle_{\frac{1}{\pi}} \\
 & \stackrel{(3.24)}{=} |\lambda_2|^{2k} \|\alpha^{(0)} - \pi\|_{\frac{1}{\pi}}^2 .
 \end{aligned}$$

Nun ist (3.29) bewiesen und damit auch die Behauptung. (Angelehnt an [Bré99] S. 210 und [Win03] S. 201 und 202.) ■

Jetzt wird ein Zahlenbeispiel angegeben. Dazu wird die Markov-Kette aus Beispiel 19 genutzt, so dass die Schranken der Konvergenzgeschwindigkeit von Satz 7 und Satz 8 gegenüber gestellt werden können.

Beispiel 20 Es sei die reversible und ergodische Markov-Kette aus Beispiel 19 ge-

geben. Demnach ist $S = \{1, 2, 3\}$, $P = \begin{pmatrix} \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{4}{9} & \frac{1}{9} & \frac{4}{9} \\ \frac{4}{9} & \frac{4}{9} & \frac{1}{9} \end{pmatrix}$ und $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$. Des

Weiteren ist nach Beispiel 19 der betragsmäßig zweitgrößte Eigenwert von P gleich $-\frac{1}{3}$. Im Allgemeinen gilt in $S = \{1, 2, 3\}$, dass

$$\begin{aligned}
 \alpha^{(0)T} P^k &= (\alpha_1^{(0)}, \alpha_2^{(0)}, \alpha_3^{(0)}) \begin{pmatrix} p_{11}^{(k)} & p_{12}^{(k)} & p_{13}^{(k)} \\ p_{21}^{(k)} & p_{22}^{(k)} & p_{23}^{(k)} \\ p_{31}^{(k)} & p_{32}^{(k)} & p_{33}^{(k)} \end{pmatrix} \\
 &= \left(\sum_{t=1}^3 \alpha_t^{(0)} p_{t1}, \sum_{t=1}^3 \alpha_t^{(0)} p_{t2}, \sum_{t=1}^3 \alpha_t^{(0)} p_{t3} \right) \tag{3.30}
 \end{aligned}$$

ist. Man setze nun $\alpha^{(0)} = (1, 0, 0)^T$. Die Anwendung von Satz 8 liefert damit für alle $k \in \mathbb{N}$ die folgende Abschätzung

$$\begin{aligned}
 2d_v(\alpha^{(0)T} P^k, \pi) &= \sum_{i=1}^3 \left| \sum_{t=1}^3 \alpha_t^{(0)} p_{ti} - \pi_i \right| \\
 &\leq (\chi^2(\alpha^{(0)}; \pi))^{\frac{1}{2}} |\lambda_2|^k = \left(\sum_{i=1}^3 \frac{(\alpha_i^{(0)} - \pi_i)^2}{\pi_i} \right)^{\frac{1}{2}} |\lambda_2|^k \\
 &= \left(\frac{4}{3} + \frac{1}{3} + \frac{1}{3} \right)^{\frac{1}{2}} \left| -\frac{1}{3} \right|^k = \frac{\sqrt{2}}{3^k} .
 \end{aligned}$$

Im Gegensatz dazu lieferte Satz 7 für alle $i, j \in S$ und $k \in \mathbb{N}$

$$|p_{ij}^{(k)} - \pi_j| \leq \frac{1}{\sqrt{\frac{1}{3}}} \left| -\frac{1}{3} \right|^k = \frac{\sqrt{3}}{3^k} . \quad \square$$

3.3.4 Die multiplikativ reversible Version der Übergangsmatrix

Wenn nicht davon ausgegangen werden kann, dass die ergodische Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ reversibel ist, dann kann trotzdem eine ähnliche Konvergenzabschätzung wie in Satz 8 gemacht werden. Und zwar kann man die Übergangsmatrix $P = (p_{ij})_{i,j \in S}$ der Markov-Kette so transformieren, dass sich die Eigenschaft der Reversibilität einstellt. Dazu wird die $N \times N$ Matrix $\tilde{P} = (\tilde{p}_{ij})_{i,j \in S}$ mit $\tilde{p}_{ij} = \frac{\pi_j p_{ji}}{\pi_i}$ betrachtet. $\pi \in \mathbb{R}^N$ ist hierbei wie gehabt die nach Satz 4 eindeutig bestimmte, zur ergodischen Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ gehörende Gleichgewichtsverteilung. Anders ausgedrückt gilt

$$\tilde{P} = D^{-2} P^T D^2 \text{ mit } D = \text{diag}(\sqrt{\pi_i})_{i \in S} .$$

\tilde{P} ist eine stochastische Matrix, da mit der Definition der Übergangsmatrix (Definition 3) und der Definition der Gleichgewichtsverteilung (Definition 12)

$$\begin{aligned}
 \tilde{p}_{ij} &= \frac{\pi_j p_{ji}}{\pi_i} \geq 0 \text{ und} \\
 \sum_{j=1}^N \tilde{p}_{ij} &= \sum_{j=1}^N \frac{\pi_j p_{ji}}{\pi_i} = \frac{1}{\pi_i} \sum_{j=1}^N \pi_j p_{ji} \stackrel{(3.3)}{=} \frac{1}{\pi_i} \pi_i = 1
 \end{aligned}$$

für alle $i, j \in S$ ist. Somit kann \tilde{P} als Übergangsmatrix der Markov-Kette $\{\tilde{X}_k\}_{k \in \mathbb{N}}$ angesehen werden.

Des Weiteren gilt

$$\begin{aligned} \pi^T \tilde{P} &= (\pi_1, \dots, \pi_N) \begin{pmatrix} \frac{\pi_1 p_{11}}{\pi_1} & \dots & \frac{\pi_N p_{N1}}{\pi_1} \\ \vdots & \ddots & \vdots \\ \frac{\pi_1 p_{1N}}{\pi_N} & \dots & \frac{\pi_N p_{NN}}{\pi_N} \end{pmatrix} = \left(\sum_{i=1}^N \pi_i \frac{\pi_1 p_{1i}}{\pi_i}, \dots, \sum_{i=1}^N \pi_i \frac{\pi_N p_{Ni}}{\pi_i} \right) \\ &= \left(\pi_1 \sum_{i=1}^N p_{1i}, \dots, \pi_N \sum_{i=1}^N p_{Ni} \right) \stackrel{(2.4)}{=} \pi^T, \end{aligned}$$

so dass insbesondere $\{\tilde{X}_k\}_{k \in \mathbb{N}}$ und $\{X_k\}_{k \in \mathbb{N}}$ dieselbe Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ haben. Außerdem ist $\{\tilde{X}_k\}_{k \in \mathbb{N}}$ bei ergodischer Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ gemäß Satz 3 und der Definition von \tilde{P} auch ergodisch. Nun setze man $M := P\tilde{P} = (\sum_{l=1}^N p_{il}\tilde{p}_{lj})_{i,j \in S}$. Als Produkt zweier stochastischer Matrizen ist damit M auch eine stochastische Matrix und kann ebenfalls als Übergangsmatrix einer Markov-Kette $\{X_k^*\}_{k \in \mathbb{N}}$ angesehen werden. Wiederum hat hierbei die Markov-Kette $\{X_k^*\}_{k \in \mathbb{N}}$ dieselbe Gleichgewichtsverteilung wie die Markov-Ketten $\{X_k\}_{k \in \mathbb{N}}$ und $\{\tilde{X}_k\}_{k \in \mathbb{N}}$, da

$$\pi^T M = \pi^T P\tilde{P} = \pi^T \tilde{P} = \pi^T$$

ist. Des Weiteren ist offensichtlich gemäß der Definition von M die Markov-Kette $\{X_k^*\}_{k \in \mathbb{N}}$ auch ergodisch. Das Paar (M, π) ist nun reversibel, denn für die $N \times N$ Matrix M gilt

$$\pi_i m_{ij} = \pi_i \sum_{l=1}^N p_{il} \frac{\pi_j p_{jl}}{\pi_l} = \pi_j \sum_{l=1}^N p_{jl} \frac{\pi_i p_{il}}{\pi_l} = \pi_j m_{ji} \text{ für alle } i, j \in S.$$

Die Matrix M wird als die multiplikativ reversible Version der Übergangsmatrix P bezeichnet. (Siehe dazu [Sch03] auf Seite 45 und 46.)

In Beispiel 18 wurde eine Markov-Kette angegeben, die nicht reversibel ist. Zur Veranschaulichung der obigen Thematik wird nun die Übergangsmatrix dieser Markov-Kette auf die beschriebene Weise umgeformt, so dass sich Reversibilität einstellt.

Beispiel 21 Es sei die Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ und damit auch S , P und π aus Beispiel 18 gegeben. Bei dieser Markov-Kette handelt es sich um eine ergodische Markov-Kette, was schon durch die Betrachtung von P^2 klar wird. P^2 hat folgende Gestalt

$$P^2 = PP = \begin{pmatrix} a & a-b & 2b \\ a+b & b & a-b \\ 0 & a+b & a \end{pmatrix} \begin{pmatrix} a & a-b & 2b \\ a+b & b & a-b \\ 0 & a+b & a \end{pmatrix}$$

$$= \begin{pmatrix} 2a^2 - b^2 & (a+b)^2 & (a+b)^2 \\ (a+b)^2 & 2a^2 - b^2 & (a+b)^2 \\ (a+b)^2 & (a+b)^2 & 2a^2 + b^2 \end{pmatrix}$$

mit $a, b > 0$, $b \leq a$ und $2a + b = 1$. Durch diese Wahl von a und b ist P^2 echt größer als Null, und die Markov-Kette ist damit nach Satz 3 ergodisch.

Aus Beispiel 18 weiß man, dass (P, π) nicht reversibel ist. Mit obiger Anleitung bilde man nun die multiplikativ reversible Version von P . Es gilt

$$\tilde{p}_{ij} = \frac{\pi_j p_{ji}}{\pi_i} = \frac{\frac{1}{3} p_{ji}}{\frac{1}{3}} = p_{ji} .$$

Damit ist $\tilde{P} = P^T$ und

$$\begin{aligned} M = P\tilde{P} = PP^T &= \begin{pmatrix} a & a-b & 2b \\ a+b & b & a-b \\ 0 & a+b & a \end{pmatrix} \begin{pmatrix} a & a+b & 0 \\ a-b & b & a+b \\ 2b & a-b & a \end{pmatrix} \\ &= \begin{pmatrix} a^2 + (a-b)^2 + (2b)^2 & a^2 + 4ab - 3b^2 & a^2 - b^2 + 2ab \\ a^2 + 4ab - 3b^2 & (a+b)^2 + b^2 + (a-b)^2 & a^2 + b^2 \\ a^2 - b^2 + 2ab & a^2 + b^2 & (a+b)^2 + a^2 \end{pmatrix} . \end{aligned}$$

Da die Matrix M offensichtlich symmetrisch ist und der Vektor $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$ drei identische Einträge hat, ist sofort klar, dass das Paar (M, π) reversibel ist. \square

Nun kann mit Hilfe des zweitgrößten Eigenwertes von M eine Abschätzung der Konvergenzgeschwindigkeit einer beliebigen ergodischen Markov-Kette gegen ihre Gleichgewichtsverteilung durch den χ^2 -Kontrast gemacht werden.

Diese Konvergenzabschätzung gehört rein formal eigentlich in den Abschnitt über ergodische Markov-Ketten. Da sie allerdings die Eigenschaft der Reversibilität nutzt, wurde sie an dieser Stelle angegeben. Des Weiteren soll der folgende Satz 9 nur der Vollständigkeit halber ohne Beweis zur Kenntnis genommen werden.

Satz 9 Sei $\{X_k\}_{k \in \mathbb{N}}$ eine ergodische Markov-Kette mit Zustandsraum S , $N \times N$ Übergangsmatrix P , einer beliebigen Anfangsverteilung $\alpha^{(0)} \in \mathbb{R}^N$ und Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$, dann gilt

$$d_V^2((\alpha^{(0)T} P^k)^T, \pi) \leq \frac{\chi^2(\alpha^{(0)}; \pi)}{4} \lambda_{M,2}^k \text{ für alle } k \in \mathbb{N} .$$

$\lambda_{M,2}^k$ bezeichnet hierbei den zweitgrößten Eigenwert der multiplikativ reversiblen Version von P . (Vergl. [Bré99] S. 211.) \square

Der Beweis dieses Satzes kann bei Interesse in [Bré99] auf den Seiten 211 und 212 nachgelesen werden.

In Satz 9 wird insbesondere der zweitgrößte Eigenwert von in diesem Fall M und, nicht wie bisher, der betragsmäßig zweitgrößte Eigenwert als Schranke der Konvergenzgeschwindigkeit genutzt. Die folgende Bemerkung liefert den Grund dafür.

Bemerkung 5 Sei M die multiplikativ reversible Version der Matrix P , wobei P die $N \times N$ Übergangsmatrix der ergodischen Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum S ist. Sämtliche Eigenwerte $\lambda_{M,1}, \dots, \lambda_{M,N}$ von M sind dann reell und liegen im Intervall $[0, 1]$. \square

BEWEIS Zuerst wird bewiesen, dass alle Eigenwerte von M reell sind. Dazu wird gezeigt, dass M die gleichen Eigenwerte wie die Matrix

$$\begin{aligned} M^* = DMD^{-1} &= \begin{pmatrix} \sqrt{\pi_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\pi_N} \end{pmatrix} \begin{pmatrix} m_{11} & \cdots & m_{1N} \\ \vdots & \ddots & \vdots \\ m_{N1} & \cdots & m_{NN} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\pi_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\pi_N}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\sqrt{\pi_1}m_{11}}{\sqrt{\pi_1}} & \cdots & \frac{\sqrt{\pi_1}m_{1N}}{\sqrt{\pi_N}} \\ \vdots & \ddots & \vdots \\ \frac{\sqrt{\pi_N}m_{N1}}{\sqrt{\pi_1}} & \cdots & \frac{\sqrt{\pi_N}m_{NN}}{\sqrt{\pi_N}} \end{pmatrix} = \begin{pmatrix} m_{11} & \cdots & \frac{\sqrt{\pi_1}m_{1N}}{\sqrt{\pi_N}} \\ \vdots & \ddots & \vdots \\ \frac{\sqrt{\pi_N}m_{N1}}{\sqrt{\pi_1}} & \cdots & m_{NN} \end{pmatrix} \end{aligned}$$

hat.

Es sei $v = (v_1, \dots, v_N)^T$ ein rechter Eigenvektor zum Eigenwert λ von M^* , das heißt $M^*v = \lambda v$. Man definiere nun $w := D^{-1}v$, so dass man durch Umformung $v = Dw$ erhält. Es gilt damit

$$\begin{aligned} M^*v = \lambda v &\Rightarrow DMD^{-1}v = \lambda v \Rightarrow DMw = \lambda Dw \Rightarrow D^{-1}DMw = \lambda D^{-1}Dw \\ &\Rightarrow Mw = \lambda w . \end{aligned}$$

M^* und M haben somit die selben Eigenwerte. Da das Paar (M, π) reversibel ist, ist M^* offensichtlich symmetrisch. Nach den Eigenschaften symmetrischer Matrizen (Bemerkung 9) hat damit M^* und somit auch M nur reelle Eigenwerte.

Nun wird der zweite Teil der Bemerkung bewiesen:

Für die Einträge der Matrix M^* gilt

$$m_{ij}^* = \frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} m_{ij} = \frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} \sum_{l=1}^N p_{il} \tilde{p}_{jl} = \frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} \sum_{l=1}^N p_{il} \frac{\pi_j p_{jl}}{\pi_l}$$

$$= \sum_{l=1}^N \frac{\sqrt{\pi_i} \pi_j}{\sqrt{\pi_j} \pi_l} p_{il} p_{jl} = \sum_{l=1}^N \frac{\sqrt{\pi_i} \sqrt{\pi_j} \sqrt{\pi_j}}{\sqrt{\pi_j} \sqrt{\pi_l} \sqrt{\pi_l}} p_{il} p_{jl} = \sum_{l=1}^N \left(\frac{\sqrt{\pi_i}}{\sqrt{\pi_l}} p_{il} \right) \left(\frac{\sqrt{\pi_j}}{\sqrt{\pi_l}} p_{jl} \right).$$

Damit kann M^* auch wie folgt dargestellt werden $M^* = (DPD^{-1})(DPD^{-1})^T$. Da nun aber $DPD^{-1} =: A$ nach dem Beweis von Satz 7 eine symmetrische Matrix ist, gilt $M^* = (DPD^{-1})^2$. Des Weiteren ist wiederum aus dem Beweis von Satz 7 bekannt, dass A die gleichen Eigenwerte wie P hat. Außerdem liegen nach Korollar 1 die Eigenwerte von P im Intervall $(-1, 1]$. Mit diesen Kenntnissen zeigt man nun ganz leicht, dass die Eigenwerte von M^* und somit auch die von M im Intervall $[0, 1]$ liegen.

Mit $Aw = \lambda w$ gilt $M^*w = AA^T w = AA w = \lambda Aw$, wobei λAw wiederum gleich $\lambda \lambda w = \lambda^2 w$ ist.

Somit gilt $M^*w = \lambda^2 w$ und damit hat M^* bzw. M nur Eigenwerte in $[0, 1]$. (Vergl. [Sch03] S. 46.) ■

Nun wird ein Beispiel zu Satz 9 angegeben. Dazu wird wiederum die Markov-Kette aus Beispiel 18 genutzt.

Beispiel 22 Es sei die Markov-Kette und damit auch P und π aus Beispiel 18 gegeben. Man setze nun $a = \frac{5}{12}$ und $b = \frac{2}{12}$, so dass damit die Bedingungen $a, b > 0$, $a \geq b$ und $2a + b = 1$ erfüllt sind. Mit diesen Festlegungen für a und b ist $P = \frac{1}{12} \begin{pmatrix} 5 & 3 & 4 \\ 7 & 2 & 3 \\ 0 & 7 & 5 \end{pmatrix}$ und für die multiplikativ reversible Version von P gilt nach Beispiel

21 $M = \frac{1}{144} \begin{pmatrix} 50 & 53 & 41 \\ 53 & 62 & 29 \\ 41 & 29 & 74 \end{pmatrix}$. Nach Korollar 1 ist $\lambda_{M,1} = 1$. Mit einem geeigneten

Computerprogramm können alle weiteren Eigenwerte von M ermittelt werden. Es gilt $\lambda_{M,2} = \frac{679}{2340}$ und $\lambda_{M,3} = \frac{65}{43456}$. Man setze $\alpha^{(0)} = (1, 0, 0)^T$. Damit ergibt die Anwendung von Satz 9 folgende Abschätzung

$$\begin{aligned} d_V^2((\alpha^{(0)T} P^k)^T, \pi) &\stackrel{(3.30)}{=} \left(\frac{1}{2} \sum_{i=1}^3 \left| \sum_{t=1}^3 \alpha_t^{(0)} p_{ti} - \pi_i \right| \right)^2 \\ &\leq \frac{\chi^2(\alpha^{(0)}; \pi)}{4} \lambda_{M,2}^k = \frac{\sum_{i=1}^3 \frac{(\alpha_i^{(0)} - \pi_i)^2}{\pi_i}}{4} \left(\frac{679}{2340} \right)^k \\ &= \frac{\frac{4}{3} + \frac{1}{3} + \frac{1}{3}}{4} \left(\frac{679}{2340} \right)^k = \frac{1}{2} \left(\frac{679}{2340} \right)^k. \quad \square \end{aligned}$$

Auch bei den in Satz 7, Satz 8 und Satz 9 angegebenen Abschätzungen handelt es sich um geometrische Schranken der Konvergenzgeschwindigkeit.

Alle in diesem Abschnitt gemachten Abschätzungen beinhalten einen Eigenwert von P (bzw. von M). Dies kann, wie schon in 3.2 erwähnt wurde, problematisch bezüglich der praktischen Nutzbarkeit dieser Konvergenzabschätzungen sein. Zumeist besteht das größte Problem darin, $|\lambda_2|$ bzw. $\lambda_{M,2}$ zu ermitteln. Um dies zu umgehen, wird sich in 3.4 mit der Charakterisierung und Abschätzung der interessierenden Eigenwerte befasst. Damit wird es dann möglich sein, auf die konkrete Berechnung der Eigenwerte zu verzichten.

3.4 Charakterisierung und Abschätzung der Eigenwerte

Aus dem Beweis von Satz 7 weiß man, dass sämtliche Eigenwerte der $N \times N$ Übergangsmatrix P einer ergodischen und reversiblen Markov-Kette reell sind. Des Weiteren hat man in Theorem 2, dem Theorem von Perron-Frobenius, und dem darauf folgenden Korollar 1 erfahren, dass alle Eigenwerte von P im Intervall $(-1, 1]$ liegen. Eins ist hierbei der größte Eigenwert von P , und alle anderen Eigenwerte von P sind betragsmäßig kleiner als Eins. Bezüglich der Eigenwerte $\lambda_1, \dots, \lambda_N$ von P wurde stets von der Anordnung $\lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_N|$ ausgegangen. Die Eigenwerte wurden somit nicht der Größe nach geordnet, sondern nach der Größe ihrer Beträge.

Im Folgenden erweist es sich als sinnvoll, die Eigenwerte $\lambda_1, \dots, \lambda_N$ als der Größe nach geordnet aufzufassen. Mit anderen Worten bedeutet dies

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_N > -1 . \quad (3.31)$$

Bei dem nach der alten Anordnung betrachteten $|\lambda_2|$ handelt es sich nunmehr um das λ_2 der neuen Anordnung oder um den Betrag von λ_N aus der neuen Anordnung.

Begonnen wird dieses Teilkapitel mit einer Darstellungsform des zweitgrößten Eigenwertes von P und einer daraus resultierenden Abschätzung von λ_2 .

3.4.1 Das Theorem von Rayleigh

Für die oben angekündigten Ausführungen wird zuerst ein weiteres Skalarprodukt mit zugehöriger Norm benötigt. Dieses Skalarprodukt steht in Analogie zu dem Skalarprodukt aus Definition 15. Außerdem wird die Definition des Mittels und der Varianz eines Vektors angegeben.

Definition 17 (π -Skalarprodukt, π -Norm und Varianz) Sei $\pi = (\pi_1, \dots, \pi_N)^T$ eine beliebige echt positive Wahrscheinlichkeitsverteilung aus \mathbb{R}^N . Für beliebige Vektoren $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ und $y = (y_1, \dots, y_N)^T \in \mathbb{R}^N$ ist dann

$$\langle x, y \rangle_\pi := \sum_{i=1}^N x_i y_i \pi_i \quad (3.32)$$

das π -Skalarprodukt und

$$\|x\|_\pi := \left(\sum_{i=1}^N x_i^2 \pi_i \right)^{\frac{1}{2}} \quad (3.33)$$

die dazu gehörige π -Norm. Des Weiteren seien

$$\langle x \rangle_\pi := \sum_{i=1}^N x_i \pi_i = \langle x, e \rangle_\pi \quad \text{und} \quad (3.34)$$

$$\text{Var}_\pi(x) := \|x\|_\pi^2 - \langle x \rangle_\pi^2 \quad (3.35)$$

das Mittel bzw. die Varianz von $x \in \mathbb{R}^N$. e bezeichnet hierbei einen Vektor aus \mathbb{R}^N , der nur aus Einsen besteht. (Siehe [Bré99] S. 201.) \square

Analog zu den Ausführungen nach Definition 15 seien die rechten Eigenvektoren v_1, \dots, v_N von P^k eine Orthonormalbasis des \mathbb{R}^N . Damit gilt

$$\langle v_i, v_j \rangle_\pi = \delta_{ij} = \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{falls } i \neq j \end{cases} \quad (3.36)$$

und jedes $x \in \mathbb{R}^N$ lässt sich wie folgt darstellen

$$x = \sum_{j=1}^N \langle x, v_j \rangle_\pi v_j \quad (3.37)$$

(vergleiche (3.26)). Da $Pv_j = \lambda_j v_j$ für alle $j \in S$ ist, gilt damit insbesondere

$$Px = P \sum_{j=1}^N \langle x, v_j \rangle_\pi v_j = \sum_{j=1}^N \langle x, v_j \rangle_\pi Pv_j = \sum_{j=1}^N \lambda_j \langle x, v_j \rangle_\pi v_j \quad \text{für alle } j \in S. \quad (3.38)$$

(Siehe dazu [Bré99] Seite 203.) Die versprochene Charakterisierung von λ_2 kann nun durch das Theorem von Rayleigh über die Varianz und die so genannte Dirichlet-Form angegeben werden. Dazu vorweg noch die Definition der Dirichlet-Form und ein Hilfssatz.

Definition 18 (Dirichlet-Form) Sei $x = (x_1, \dots, x_N)^T$ ein beliebiger Vektor aus \mathbb{R}^N und $\{X_k\}_{k \in \mathbb{N}}$ eine reversible Markov-Kette mit Zustandsraum S , $N \times N$ Übergangsmatrix P und Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$. Es definiere nun

$$D_{(P,\pi)}(x, x) := \langle (I - P)x, x \rangle_\pi \quad (3.39)$$

die Dirichlet-Form des Paares (P, π) von x . I bezeichnet hierbei die $N \times N$ Einheitsmatrix. (Vergl. [Bré99] S. 204.) \square

In der Literatur wird $I - P$ auch oft als „die Laplacesche“ der Markov-Kette bezeichnet.

Lemma 4 *Es sei eine reversible Markov-Kette und S , P und π wie in Definition 18 gegeben. Für jedes $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ gilt dann*

$$D_{(P,\pi)}(x, x) = \frac{1}{2} \sum_{i,j=1}^N \pi_i p_{ij} (x_j - x_i)^2 . \quad (3.40)$$

(Vergl. [Bré99] S. 204.) \square

BEWEIS Es gilt

$$\begin{aligned} (I - P)x &= \left(\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} - \begin{pmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{pmatrix} \right) \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \\ &= \begin{pmatrix} (1 - p_{11}) \cdots - p_{1N} \\ \vdots \\ -p_{N1} \cdots (1 - p_{NN}) \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} = \begin{pmatrix} x_1 - \sum_{j=1}^N p_{1j} x_j \\ \vdots \\ x_N - \sum_{j=1}^N p_{Nj} x_j \end{pmatrix} . \end{aligned} \quad (3.41)$$

Es folgt nun

$$\begin{aligned} 2D_{(P,\pi)}(x, x) &\stackrel{(3.39)}{=} 2 \langle (I - P)x, x \rangle_\pi \stackrel{(3.41)}{=} 2 \left\langle \begin{pmatrix} x_1 - \sum_{j=1}^N p_{1j} x_j \\ \vdots \\ x_N - \sum_{j=1}^N p_{Nj} x_j \end{pmatrix}, \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \right\rangle_\pi \\ &\stackrel{(3.32)}{=} 2 \sum_{i=1}^N (x_i - \sum_{j=1}^N p_{ij} x_j) x_i \pi_i = 2 \sum_{i=1}^N (x_i x_i \pi_i - \sum_{j=1}^N p_{ij} x_j x_i \pi_i) . \end{aligned}$$

Es kann mit $\sum_{i=1}^N p_{ij}$ erweitert werden, da $\sum_{i=1}^N p_{ij} = 1$ nach Definition 3 ist. Damit

gilt

$$\begin{aligned}
 2 \sum_{i=1}^N (x_i x_i \pi_i - \sum_{j=1}^N p_{ij} x_j x_i \pi_i) &= 2 \sum_{i=1}^N (x_i x_i \pi_i \sum_{j=1}^N p_{ij} - \sum_{j=1}^N p_{ij} x_j x_i \pi_i) \\
 &= 2 \sum_{i=1}^N (\sum_{j=1}^N \pi_i p_{ij} x_i x_i - \sum_{j=1}^N p_{ij} x_j x_i \pi_i) \\
 &= 2 \sum_{i,j=1}^N (\pi_i p_{ij} x_i x_i - \pi_i p_{ij} x_i x_j) = 2 \sum_{i,j=1}^N \pi_i p_{ij} x_i (x_i - x_j) \\
 &\stackrel{i \leftrightarrow j}{=} \sum_{i,j=1}^N \pi_i p_{ij} x_i (x_i - x_j) + \sum_{i,j=1}^N \pi_j p_{ji} x_j (x_j - x_i) \\
 &\stackrel{(3.14)}{=} \sum_{i,j=1}^N \pi_i p_{ij} x_i (x_i - x_j) + \sum_{i,j=1}^N \pi_i p_{ij} x_j (x_j - x_i) \\
 &= \sum_{i,j=1}^N \pi_i p_{ij} x_i (x_i - x_j) + \pi_i p_{ij} x_j (x_j - x_i) \\
 &= \sum_{i,j=1}^N \pi_i p_{ij} (x_j^2 - 2x_i x_j + x_i^2) = \sum_{i,j=1}^N \pi_i p_{ij} (x_j - x_i)^2 .
 \end{aligned}$$

Dividiert man dies durch zwei, so erhält man insgesamt

$$D_{(P,\pi)}(x, x) = \frac{1}{2} \sum_{i,j=1}^N \pi_i p_{ij} (x_j - x_i)^2 .$$

(In Anlehnung an [Bré99] S. 205.) ■

Aus Lemma 4 ergibt sich mit einem $c \in \mathbb{R}$, dass $D_{(P,\pi)}(x - ce, x - ce) = \frac{1}{2} \sum_{i,j=1}^N \pi_i p_{ij} (x_j - c - x_i + c)^2$ ist. e ist dabei wiederum ein Vektor aus \mathbb{R}^N , der nur aus Einsen besteht. Des Weiteren ist $\frac{1}{2} \sum_{i,j=1}^N \pi_i p_{ij} (x_j - c - x_i + c)^2 = \frac{1}{2} \sum_{i,j=1}^N \pi_i p_{ij} (x_j - x_i)^2 = D_{(P,\pi)}(x, x)$. Damit gilt für beliebige $c \in \mathbb{R}$ und $x \in \mathbb{R}^N$

$$D_{(P,\pi)}(x, x) = D_{(P,\pi)}(x - ce, x - ce) . \quad (3.42)$$

Theorem 3 (Theorem von Rayleigh) Sei $\{X_k\}_{k \in \mathbb{N}}$ eine irreduzible Markov-Kette mit Zustandsraum S , Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ und $N \times N$ Übergangsmatrix P . Außerdem sei das Paar (P, π) reversibel, λ_2 sei der zweitgrößte Eigenwert von P und v_1 sei der rechte Eigenvektor zum größten Eigenwert λ_1 von P . Dann gilt

$$\lambda_2 = 1 - \inf \left\{ \frac{D_{(P,\pi)}(x, x)}{\text{Var}_\pi(x)}; \langle x, v_1 \rangle_\pi = 0, x \neq 0 \in \mathbb{R}^N \right\} . \quad (3.43)$$

(Vergl. [Beh00] S. 205.) □

Da insbesondere $v_1 = e$ ist (siehe Korollar 1), gilt damit und mit (3.34), dass (3.43) wie folgt umformuliert werden kann

$$\lambda_2 = 1 - \inf \left\{ \frac{D_{(P,\pi)}(x, x)}{\text{Var}_\pi(x)}; \langle x \rangle_\pi = 0; x \neq 0 \in \mathbb{R}^N \right\} . \quad (3.44)$$

BEWEIS (DES THEOREMS VON RAYLEIGH) Es sei v_j ein rechter N -dimensionaler Eigenvektor zum Eigenwert λ_j von P . Mit der $N \times N$ Einheitsmatrix I sei des Weiteren $Iv_j = v_j$. Damit gilt

$$Pv_j = \lambda_j v_j \Leftrightarrow Iv_j - Pv_j = 1v_j - \lambda_j v_j \Leftrightarrow (I - P)v_j = (1 - \lambda_j)v_j .$$

Somit ist nun

$$(I - P)x \stackrel{(3.37)}{=} \sum_{j=1}^N \langle x, v_j \rangle_\pi (I - P)v_j = \sum_{j=1}^N (1 - \lambda_j) \langle x, v_j \rangle_\pi v_j . \quad (3.45)$$

Es folgt

$$\begin{aligned} D_{(P,\pi)}(x, x) &\stackrel{(3.39)}{=} \langle (I - P)x, x \rangle_\pi \stackrel{(3.45)}{=} \left\langle \sum_{j=1}^N (1 - \lambda_j) \langle x, v_j \rangle_\pi v_j, \sum_{j=1}^N \langle x, v_j \rangle_\pi v_j \right\rangle_\pi \\ &= \left\langle (1 - \lambda_1) \langle x, v_1 \rangle_\pi v_1, \sum_{j=1}^N \langle x, v_j \rangle_\pi v_j \right\rangle_\pi \\ &\quad + \dots + \left\langle (1 - \lambda_N) \langle x, v_N \rangle_\pi v_N, \sum_{j=1}^N \langle x, v_j \rangle_\pi v_j \right\rangle_\pi \\ &= \langle (1 - \lambda_1) \langle x, v_1 \rangle_\pi v_1, \langle x, v_1 \rangle_\pi v_1 \rangle_\pi \\ &\quad + \dots + \langle (1 - \lambda_1) \langle x, v_1 \rangle_\pi v_1, \langle x, v_N \rangle_\pi v_N \rangle_\pi \\ &\quad + \dots + \langle (1 - \lambda_N) \langle x, v_N \rangle_\pi v_N, \langle x, v_1 \rangle_\pi v_1 \rangle_\pi \\ &\quad + \dots + \langle (1 - \lambda_N) \langle x, v_N \rangle_\pi v_N, \langle x, v_N \rangle_\pi v_N \rangle_\pi \\ &= (1 - \lambda_1) |\langle x, v_1 \rangle_\pi|^2 \langle v_1, v_1 \rangle_\pi + \dots + (1 - \lambda_1) \langle x, v_1 \rangle_\pi \langle x, v_N \rangle_\pi \langle v_1, v_N \rangle_\pi \\ &\quad + \dots + (1 - \lambda_N) \langle x, v_N \rangle_\pi \langle x, v_1 \rangle_\pi \langle v_N, v_1 \rangle_\pi \\ &\quad + \dots + (1 - \lambda_N) |\langle x, v_N \rangle_\pi|^2 \langle v_N, v_N \rangle_\pi \\ &\stackrel{(3.36)}{=} (1 - \lambda_1) |\langle x, v_1 \rangle_\pi|^2 + \dots + (1 - \lambda_N) |\langle x, v_N \rangle_\pi|^2 \end{aligned}$$

$$= \sum_{j=1}^N (1 - \lambda_j) |\langle x, v_j \rangle_\pi|^2 . \quad (3.46)$$

Aus (3.36) und (3.37) folgt außerdem

$$\|x\|_\pi^2 = \langle x, x \rangle_\pi = \left\langle \sum_{j=1}^N \langle x, v_j \rangle_\pi v_j, \sum_{j=1}^N \langle x, v_j \rangle_\pi v_j \right\rangle_\pi = \sum_{j=1}^N |\langle x, v_j \rangle_\pi|^2 . \quad (3.47)$$

Wenn nun $\langle x, v_1 \rangle_\pi \stackrel{Kor.1}{=} \langle x, e \rangle_\pi = \langle x \rangle_\pi = 0$ mit $x \neq 0 \in \mathbb{R}^N$ ist, dann ist $Var_\pi(x) = \|x\|_\pi^2 - \langle x \rangle_\pi^2 = \|x\|_\pi^2 - 0 = \|x\|_\pi^2$ und somit identisch mit der Formel (3.47). Somit ergibt sich zusammen mit (3.46), dass

$$\frac{D_{(P,\pi)}(x, x)}{Var_\pi(x)} = \frac{\sum_{j=1}^N (1 - \lambda_j) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=1}^N |\langle x, v_j \rangle_\pi|^2}$$

ist. Wiederum mit $v_1 = e$ und der Voraussetzung $\langle x, v_1 \rangle_\pi = \langle x, e \rangle_\pi = \langle x \rangle_\pi = 0$ gilt

$$\begin{aligned} & \frac{\sum_{j=1}^N (1 - \lambda_j) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=1}^N |\langle x, v_j \rangle_\pi|^2} \\ &= \frac{\sum_{j=2}^N (1 - \lambda_j) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=2}^N |\langle x, v_j \rangle_\pi|^2} \\ &= (1 - \lambda_2) + \frac{\sum_{j=2}^N (1 - \lambda_j) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=2}^N |\langle x, v_j \rangle_\pi|^2} - (1 - \lambda_2) \\ &= (1 - \lambda_2) + \frac{\sum_{j=2}^N (1 - \lambda_j) |\langle x, v_j \rangle_\pi|^2 - \sum_{j=2}^N (1 - \lambda_2) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=2}^N |\langle x, v_j \rangle_\pi|^2} \\ &= (1 - \lambda_2) + \frac{\sum_{j=2}^N (1 - \lambda_j - 1 + \lambda_2) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=2}^N |\langle x, v_j \rangle_\pi|^2} \\ &= (1 - \lambda_2) + \frac{\sum_{j=3}^N (\lambda_2 - \lambda_j) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=2}^N |\langle x, v_j \rangle_\pi|^2} . \end{aligned}$$

Auf Grund der zu Beginn dieses Teilkapitels vorausgesetzten Anordnung der Eigenwerte (3.31), ist λ_2 stets größer oder gleich λ_j für $j = 3, \dots, N$. Damit ist $\frac{\sum_{j=3}^N (\lambda_2 - \lambda_j) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=2}^N |\langle x, v_j \rangle_\pi|^2}$ größer oder gleich Null und es ergibt sich

$$\frac{D_{(P,\pi)}(x, x)}{Var_\pi(x)} = (1 - \lambda_2) + \frac{\sum_{j=3}^N (\lambda_2 - \lambda_j) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=2}^N |\langle x, v_j \rangle_\pi|^2} \geq (1 - \lambda_2) . \quad (3.48)$$

Durch Umformung erhält man

$$\lambda_2 \geq 1 - \frac{D_{(P,\pi)}(x, x)}{\text{Var}_\pi(x)} .$$

Wenn man nun $x = v_2$ wählt, dann ist $\frac{\sum_{j=3}^N (\lambda_2 - \lambda_j) |\langle x, v_j \rangle_\pi|^2}{\sum_{j=2}^N |\langle x, v_j \rangle_\pi|^2} = \frac{\sum_{j=3}^N (\lambda_2 - \lambda_j) |\langle v_2, v_j \rangle_\pi|^2}{\sum_{j=2}^N |\langle v_2, v_j \rangle_\pi|^2} \stackrel{(3.36)}{=} 0$ und es gilt im Hinblick auf (3.48)

$$\lambda_2 = 1 - \frac{D_{(P,\pi)}(x, x)}{\text{Var}_\pi(x)} .$$

Insgesamt folgt nun

$$\lambda_2 = 1 - \inf \left\{ \frac{D_{(P,\pi)}(x, x)}{\text{Var}_\pi(x)} ; \langle x, v_1 \rangle_\pi = 0, x \neq 0 \in \mathbb{R}^N \right\} .$$

(Angelehnt an [Bré99] S. 205 und 206.) ■

Mit Hilfe des Theorems von Rayleigh kann eine erste Abschätzung des zweitgrößten Eigenwertes λ_2 von P gemacht werden.

Korollar 2 *Es sei $\{X_k\}_{k \in \mathbb{N}}$ eine irreduzible Markov-Kette mit Zustandsraum S , $N \times N$ Übergangsmatrix P und Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$. Das Paar (P, π) sei reversibel und es sei $A > 0$ so gegeben, dass für alle $x \in \mathbb{R}^N$*

$$\text{Var}_\pi(x) \leq A D_{(P,\pi)}(x, x) \tag{3.49}$$

ist. Dann gilt für den zweitgrößten Eigenwert λ_2 von P

$$\lambda_2 \leq 1 - \frac{1}{A} .$$

(Vergl. [Bré99] S. 206.) □

BEWEIS Umformung von (3.49) ergibt: $\frac{D_{(P,\pi)}(x, x)}{\text{Var}_\pi(x)} \geq \frac{1}{A}$. Damit gilt

$$\lambda_2 = 1 - \inf \left\{ \frac{D_{(P,\pi)}(x, x)}{\text{Var}_\pi(x)} ; \langle x, v_1 \rangle_\pi = 0, x \neq 0 \in \mathbb{R}^N \right\} \leq 1 - \frac{1}{A} .$$

([Bré99] S. 206.) ■

Diese Abschätzung des zweitgrößten Eigenwertes von P ist vorerst nur von geringem praktischen Nutzen. Das Problem besteht darin, solch ein A zu finden, so dass für alle $x \in \mathbb{R}^N$ die Bedingung (3.49) erfüllt ist. Korollar 2 gibt keinen Aufschluss darüber, wie man solch ein A finden kann bzw. welchen Wert genau A annehmen

könnte, um (3.49) für alle $x \in \mathbb{R}^N$ zu erfüllen. Eine Lösung diese Problems ist in den folgenden Abschnitten 3.4.2 und 3.4.3 angegeben. Dort werden mehrere Vorschläge für die Wahl von A gemacht. Damit ist es dann auch möglich, ein konkretes Beispiel zu der Abschätzung in Korollar 2 anzugeben.

3.4.2 Der Poincaré-Koeffizient und die Poincaré-Ungleichung

Mit den bisherigen Erkenntnissen aus 3.4 ist es möglich, weitere bzw. präzisere Grenzen für den Eigenwert λ_2 der $N \times N$ Übergangsmatrix P einer Markov-Kette, sowie Grenzen für den Eigenwert λ_N zu finden. Diese nun folgenden Abschätzungen für λ_2 und λ_N werden in Bezug auf die Geometrie der $N \times N$ Übergangsmatrix $P = (p_{ij})_{i,j \in S}$ der Markov-Kette gemacht. Auf Grund dessen bezeichnet man diese Abschätzungen auch als geometrische Grenzen der Eigenwerte.

Zuvor müssen wieder einige Begriffe und Bezeichnungen eingeführt werden. Es sei dabei stets vorausgesetzt, dass die in diesem Abschnitt betrachteten Markov-Ketten irreduzibel sind. Für jedes Paar $i, j \in S$ mit $i \neq j$ bezeichne nun $s = s_{ij} = (i, j)$ eine gerichtete Kante von i nach j und $\Theta := \{s = s_{ij} = (i, j) \mid i, j \in S, i \neq j\}$ die Menge aller gerichteter Kanten. Durch die vorausgesetzte Irreduzibilität ist insbesondere gewährleistet, dass der Übergang von i nach j stets möglich ist. Es seien $s^- = i$ und $s^+ = j$ die Anfangsecke und die Endecke von $s = s_{ij}$. Für jede solche Kante s definiere man des Weiteren $Q(s) := \pi_i p_{ij}$. Für jedes Paar $i, j \in S$ mit $i \neq j$ wird ein Pfad γ_{ij} von i nach j betrachtet. Dieser Pfad stellt insbesondere einen Vektor von Zuständen dar, so dass $\gamma_{ij} = (i_0, i_1, \dots, i_m, i_{m+1})$ mit $i_0 = i$ und $i_{m+1} = j$ ist. Dabei sei $p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_m i_{m+1}} > 0$. Bezeichne nun Γ die Menge aller so gewählter Pfade. Im allgemeinen gibt es mehrere Pfade γ_{ij} pro Paar $i, j \in S$. Man betrachte aber nur genau einen Pfad pro Zustandspaar, so dass die Menge Γ nur genau einen Pfad pro Paar $i, j \in S$ enthält. Man definiere $\Theta_{ij} := \{s_{i_t i_{t+1}} = (i_t, i_{t+1}) \mid t = 0, \dots, m, i = i_0 \text{ und } j = i_{m+1}\}$ eine Menge von Kanten des Pfades γ_{ij} , wobei keine der Kanten $s_{i_t i_{t+1}}$ mehrfach auftreten soll. Damit definiere man außerdem für jeden Pfad $\gamma_{ij} = (i, i_1, \dots, i_m, j) \in \Gamma$

$$|\gamma_{ij}|_Q := \sum_{s \in \Theta_{ij}} \frac{1}{Q(s)} = \frac{1}{\pi_i p_{i i_1}} + \frac{1}{\pi_{i_1} p_{i_1 i_2}} + \dots + \frac{1}{\pi_{i_m} p_{i_m j}} \quad (3.50)$$

$$= \sum_{t=0}^m \frac{1}{\pi_{i_t} p_{i_t i_{t+1}}} \text{ mit } i_0 = i \text{ und } i_{m+1} = j . \quad (3.51)$$

Der Poincaré-Koeffizient κ der Pfadmenge Γ ist dann gegeben durch

$$\kappa := \kappa(\Gamma) = \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} |\gamma_{ij}|_Q \pi_i \pi_j, \quad (3.52)$$

wobei das Maximum über alle gerichteten Kanten gebildet wird und die Summe über alle Pfade γ_{ij} geht, die s überqueren. (Siehe dazu [DS91] auf Seite 37.)

Mit diesen Bezeichnungen kann nun Satz 10, die Poincaré-Ungleichung, formuliert werden. Sie gibt eine Abschätzung des zweitgrößten Eigenwertes λ_2 von P an.

Satz 10 (Poincaré-Ungleichung) *Sei $\{X_k\}_{k \in \mathbb{N}}$ eine irreduzibel Markov-Kette mit Zustandsraum S , $N \times N$ Übergangsmatrix P und Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$. Das Paar (P, π) sei außerdem reversibel und λ_2 bezeichne den zweitgrößten Eigenwert von P . Dann gilt*

$$\lambda_2 \leq 1 - \frac{1}{\kappa}, \quad (3.53)$$

wobei κ wie in (3.52) definiert ist. (Vergl. [DS91] S. 37.) □

BEWEIS Nach Korollar 2 genügt es zu zeigen, dass $\text{Var}_\pi(x) \leq \kappa D_{(P,\pi)}(x, x)$ für alle $x \in \mathbb{R}^N$ ist. Es gilt

$$\begin{aligned} \text{Var}_\pi(x) &\stackrel{(3.35)}{=} \|x\|_\pi^2 - \langle x \rangle_\pi^2 = \frac{1}{2} (2(\|x\|_\pi^2 - \langle x \rangle_\pi \langle x \rangle_\pi)) \\ &= \frac{1}{2} (2(\sum_{i \in S} x_i^2 \pi_i - \sum_{i \in S} x_i \pi_i \sum_{i \in S} x_i \pi_i)) \\ &= \frac{1}{2} (\sum_{i \in S} x_i^2 \pi_i - \sum_{i \in S} x_i \pi_i \sum_{i \in S} x_i \pi_i + \sum_{i \in S} x_i^2 \pi_i - \sum_{i \in S} x_i \pi_i \sum_{i \in S} x_i \pi_i) \\ &= \frac{1}{2} (\sum_{i \in S} x_i^2 \pi_i - \sum_{i \in S} x_i \pi_i \sum_{j \in S} x_j \pi_j + \sum_{j \in S} x_j^2 \pi_j - \sum_{i \in S} x_i \pi_i \sum_{j \in S} x_j \pi_j) \end{aligned}$$

Da $\sum_{t \in S} \pi_t = 1$ für alle $t \in S$ ist, kann wie folgt erweitert werden

$$\begin{aligned} &\frac{1}{2} (\sum_{i \in S} x_i^2 \pi_i - \sum_{i \in S} x_i \pi_i \sum_{j \in S} x_j \pi_j + \sum_{j \in S} x_j^2 \pi_j - \sum_{i \in S} x_i \pi_i \sum_{j \in S} x_j \pi_j) \\ &= \frac{1}{2} (\sum_{i \in S} x_i^2 \pi_i \sum_{j \in S} \pi_j + \sum_{j \in S} x_j^2 \pi_j \sum_{i \in S} \pi_i - 2 \sum_{i, j \in S} x_i x_j \pi_i \pi_j) \\ &= \frac{1}{2} \sum_{i, j \in S} (x_i^2 \pi_i \pi_j - 2x_i x_j \pi_i \pi_j + x_j^2 \pi_i \pi_j) \\ &= \frac{1}{2} \sum_{i, j \in S} (x_i^2 - 2x_i x_j + x_j^2) \pi_i \pi_j = \frac{1}{2} \sum_{i, j \in S} (x_i - x_j)^2 \pi_i \pi_j \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2} \sum_{i,j \in S} (x_i - x_{i_1} + x_{i_1} - x_{i_2} + \dots + x_{i_m} - x_j)^2 \pi_i \pi_j \\
 &= \frac{1}{2} \sum_{i,j \in S} \left(\sum_{s \in \Theta_{ij}} (x_{s^-} - x_{s^+}) \right)^2 \pi_i \pi_j \\
 &= \frac{1}{2} \sum_{i,j \in S} \left(\sum_{s \in \Theta_{ij}} \frac{1}{\sqrt{Q(s)}} \sqrt{Q(s)} (x_{s^-} - x_{s^+}) \right)^2 \pi_i \pi_j .
 \end{aligned}$$

Mit der Cauchy-Schwarzschen Ungleichung (siehe dazu (4.8)) folgt nun insgesamt

$$\begin{aligned}
 \text{Var}_\pi(x) &\leq \frac{1}{2} \sum_{i,j \in S} \left(\sum_{s \in \Theta_{ij}} \frac{1}{Q(s)} \sum_{s \in \Theta_{ij}} Q(s) (x_{s^-} - x_{s^+})^2 \right) \pi_i \pi_j \\
 &\stackrel{(3.50)}{=} \frac{1}{2} \sum_{i,j \in S} (|\gamma_{ij}|_Q \sum_{s \in \Theta_{ij}} Q(s) (x_{s^-} - x_{s^+})^2) \pi_i \pi_j \\
 &= \frac{1}{2} \sum_{s \in \Theta} Q(s) (x_{s^-} - x_{s^+})^2 \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} |\gamma_{ij}|_Q \pi_i \pi_j \\
 &= \frac{1}{2} \sum_{i,j \in S} \pi_i p_{ij} (x_i - x_j)^2 \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} |\gamma_{ij}|_Q \pi_i \pi_j \\
 &\leq \sum_{i,j \in S} \pi_i p_{ij} (x_i - x_j)^2 \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} |\gamma_{ij}|_Q \pi_i \pi_j \\
 &\stackrel{(3.40)(3.52)}{=} \kappa D_{(P,\pi)}(x, x) .
 \end{aligned}$$

Die Poincaré-Ungleichung ist damit bewiesen. (Angelehnt an [DS91] S. 38.) ■

Satz 10 kann als Präzisierung von Korollar 2 aufgefasst werden. Aus dem Beweis von Satz 10 ergibt sich nämlich, dass mit $\kappa (> 0)$ (3.49) erfüllt ist. Für den Wert A aus Korollar 2 kann somit der Poincaré-Koeffizient κ gewählt werden.

Es wird nun ein Beispiel für die Abschätzung aus Satz 10 und damit auch ein Beispiel für die Abschätzung aus Korollar 2 angegeben. Dazu wird auf die zufällige Irrfahrt (Beispiel 17) zurückgegriffen.

Beispiel 23 Es sei eine zufällige Irrfahrt auf $G = (V, K)$ gegeben. Dabei ist $G = (V, K)$ ein verbundener Graph mit der Menge $V = \{v_1, \dots, v_N\}$ von N Eckpunkten und der Menge K , die die Menge von Kanten darstellt, die jeweils zwei Eckpunkte miteinander verbinden. Zwei durch eine Kante verbundene Eckpunkte werden als Nachbarn bezeichnet. v_i und v_{i+1} seien insbesondere Nachbarn. Für jedes Paar $v_i, v_j \in V$ von Eckpunkten soll es einen Pfad γ_{ij} geben, der die Eckpunkte v_i und v_j miteinander verbindet. Natürlich kann es wiederum mehrere solcher Pfade γ_{ij} geben. In diesem Fall betrachte man jedoch nur einen Pfad pro Paar v_i und v_j , der

insbesondere am kürzesten von all den v_i und v_j verbindenden Pfaden ist.

In Beispiel 17 wurde gezeigt, dass die zufällige Irrfahrt auf $G = (V, K)$ eine Markov-Kette mit Zustandsraum $S = \{1, \dots, N\}$ und der Übergangsmatrix $P = (p_{ij})_{i,j \in S}$ mit den Einträgen

$$p_{ij} = \begin{cases} \frac{1}{d_i} & \text{falls die Eckpunkte } v_i \text{ und } v_j \text{ Nachbarn sind,} \\ 0 & \text{sonst} \end{cases}$$

ist. d_i ($i \in S$) bezeichnet hierbei die Anzahl der Nachbarn von v_i . Außerdem wurde gezeigt, dass die Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ der Markov-Kette durch

$$\pi = \left(\frac{d_1}{d}, \dots, \frac{d_N}{d} \right)^T \quad \text{mit } d := \sum_{i=1}^N d_i$$

gegeben ist und dass die Markov-Kette reversibel ist. Auf Grund der oben genannten Voraussetzung, jedes Eckpunktepaar sei durch einen Pfad $\gamma_{ij} = (i, i_1, \dots, i_m, j)$ verbunden, ist die Markov-Kette des Weiteren irreduzibel.

Für den Poincaré-Koeffizienten κ gilt

$$\begin{aligned} \kappa &\stackrel{(3.52)}{=} \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} |\gamma_{ij}|_Q \pi_i \pi_j \stackrel{(3.50)}{=} \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} \left(\sum_{s \in \Theta_{ij}} \frac{1}{Q(s)} \right) \pi_i \pi_j \\ &\stackrel{(3.51)}{=} \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} \left(\sum_{t=0}^m \frac{1}{\pi_{i_t} p_{i_t i_{t+1}}} \right) \frac{d_i}{d} \frac{d_j}{d}. \end{aligned}$$

Da nach Voraussetzung v_{i_t} und $v_{i_{t+1}}$ Nachbarn sind, gilt

$$\begin{aligned} &\max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} \left(\sum_{t=0}^m \frac{1}{\pi_{i_t} p_{i_t i_{t+1}}} \right) \frac{d_i}{d} \frac{d_j}{d} = \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} \left(\sum_{t=0}^m \frac{1}{\frac{d_{i_t}}{d} \frac{1}{d_{i_t}}} \right) \frac{d_i}{d} \frac{d_j}{d} \\ &= \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} \left(\sum_{t=0}^m d \right) \frac{d_i}{d} \frac{d_j}{d} = \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} (m+1) d \frac{d_i}{d} \frac{d_j}{d}. \end{aligned} \quad (3.54)$$

$m+1$ ist hierbei die Länge des Pfades γ_{ij} . Im Folgenden wird die Länge eines Pfades γ_{tl} mit $\Delta(\gamma_{tl})$ bezeichnet.

Man definiere nun die maximale Pfadlänge $\Delta := \max_{\gamma \in \Gamma} \Delta(\gamma)$, $\delta := \max_{i \in S} d_i$ und den sogenannten Bottleneck-Koeffizienten $\beta := \max_{s \in \Theta} \#\{\gamma_{ij} \in \Gamma \mid s \in \Theta_{ij}\}$, der die maximale Anzahl der Pfade angibt, die jeweils durch eine einzelne Kante laufen.

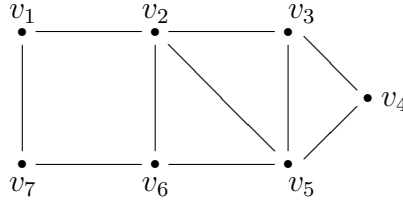
Damit gilt nun

$$\begin{aligned} \kappa &\stackrel{(3.54)}{=} \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} \Delta(\gamma_{ij}) d \frac{d_i}{d} \frac{d_j}{d} \leq \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} \Delta \frac{d_i d_j}{d} \\ &\leq \max_{s \in \Theta} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} \Delta \frac{\delta^2}{d} \leq \frac{\Delta \delta^2 \beta}{d} . \end{aligned}$$

Es sind nun alle Voraussetzungen von Satz 10 erfüllt, und die Anwendung des Satzes ergibt für den zweitgrößten Eigenwert der Übergangsmatrix P

$$\lambda_2 \leq 1 - \frac{1}{\kappa} \leq 1 - \frac{d}{\delta^2 \Delta \beta} .$$

Für das bereits in Beispiel 17 betrachtete Zahlenbeispiel



gilt

$$d = 20, \delta = 4, \Delta = 3 \text{ und } \beta = 5 .$$

Damit ergibt sich

$$\lambda_2 \leq 1 - \frac{20}{16 \cdot 3 \cdot 5} = 1 - \frac{1}{12} = \frac{11}{12} .$$

(Vergl. [Sch03] S. 56 und 57.) □

3.4.3 Abschätzung des kleinsten Eigenwertes

Mit ähnlichen Bezeichnungen wie in 3.4.2 kann auch eine Abschätzung des kleinsten Eigenwertes λ_N der $N \times N$ Übergangsmatrix P einer Markov-Kette gemacht werden. Für jedes $i \in S$ wird ein Pfad $\gamma_i = (i_0, i_1, \dots, i_n, i_{n+1})$ mit $i_0 = i_{n+1} = i$ von i nach i betrachtet. Die Pfadlänge soll dabei ungerade sein. Des Weiteren sei $s = s_{ii} = (i, i)$ eine gerichtete Kante von i nach i und $\Theta_i = \{s_{i_t i_{t+1}} = (i_t, i_{t+1}) \mid t = 0, \dots, n \text{ und } i_0 = i_{n+1} = i\}$ sei die Menge aller Kanten des Pfades γ_i , wobei keine Kante $s_{i_t i_{t+1}}$ mehrfach in Θ_i enthalten sein soll. Genau wie in 3.4.2 bezeichne $s^+ = i_t$ und $s^- = i_{t+1}$ die Anfangsecke und die Endecke der Kante $s = s_{i_t i_{t+1}}$ und für jede Kante $s = s_{i_t i_{t+1}}$ sei $Q(s) := \pi_{i_t} p_{i_t i_{t+1}}$. Außerdem wird wieder vorausgesetzt, dass es sich bei

den in diesem Abschnitt betrachteten Markov-Ketten um irreduzible Markov-Ketten handelt. Dadurch ist gewährleistet, dass die Übergänge von i nach i für alle $i \in S$ möglich sind. Da außerdem festgelegt wurde, die Länge des Pfades γ_i sei ungerade, gilt $ggT\{k \geq 1 \in \mathbb{N} \mid p_{ii}^{(k)} > 0\} = ggT\{1, 3, 5, 7, \dots\} = 1$, so dass die Markov-Ketten nicht nur irreduzibel sondern auch aperiodisch und somit ergodisch sind. Man bezeichne außerdem mit Σ die Menge aller so gewählter Pfade γ_i . Natürlich kann es wieder mehrere Pfade γ_i pro Zustand $i \in S$ geben. Es wird aber nur genau ein Pfad pro Zustand betrachtet, so dass Σ nur einen Pfad je $i \in S$ enthält. Für einen Pfad $\gamma_i \in \Sigma$ definiere man

$$|\gamma_i|_Q := \sum_{s \in \Theta_i} \frac{1}{Q(s)} \quad (3.55)$$

$$= \sum_{l=0}^n \frac{1}{\pi_{i_l} p_{i_l i_{l+1}}} \text{ mit } i_0 = i_{n+1} = i . \quad (3.56)$$

Der geometrische Koeffizient ζ der Pfadmenge Σ ist nun gegeben durch

$$\zeta := \zeta(\Sigma) = \max_{s \in \Theta} \sum_{\substack{\gamma_i \\ s \in \Theta_i}} |\gamma_i|_Q \pi_i , \quad (3.57)$$

wobei hier wiederum das Maximum über alle gerichteten Kanten s gebildet wird und die Summe über alle Pfade γ_i geht, die s überqueren.

Damit ergibt sich nun folgende Abschätzung für den kleinsten Eigenwert λ_N von P .

Satz 11 *Sei $\{X_k\}_{k \in \mathbb{N}}$ eine ergodische Markov-Kette mit Zustandsraum S , $N \times N$ Übergangsmatrix P und Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$. Des Weiteren sei das Paar (P, π) reversibel. Dann gilt für den kleinsten Eigenwert λ_N von P*

$$\lambda_N \geq -1 + \frac{2}{\zeta} , \quad (3.58)$$

wobei ζ wie in (3.57) definiert ist. (Vergl. [DS91] S. 40.) □

Für den Beweis von Satz 11 wird der folgende Hilfssatz benötigt.

Lemma 5 *Es sei eine ergodische Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ mit Zustandsraum S , $N \times N$ Übergangsmatrix P und Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ gegeben und das Paar (P, π) sei reversibel. Dann gilt für jedes $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$*

$$\frac{1}{2} \sum_{i,j \in S} (x_i + x_j)^2 \pi_i p_{ij} = \langle Px, x \rangle_\pi + \|x\|_\pi^2 . \quad (3.59)$$

□

BEWEIS Aus der Reversibilität von (P, π) ergibt sich

$$\begin{aligned} \frac{1}{2} \sum_{i,j \in S} (x_i + x_j)^2 \pi_i p_{ij} &= \frac{1}{2} \left(\sum_{i,j \in S} x_i^2 \pi_i p_{ij} + \sum_{i,j \in S} 2x_i x_j \pi_i p_{ij} + \sum_{i,j \in S} x_j^2 \pi_i p_{ij} \right) \\ &\stackrel{(3.14)}{=} \frac{1}{2} \sum_{i \in S} x_i^2 \pi_i \sum_{j \in S} p_{ij} + \sum_{i,j \in S} x_i x_j \pi_i p_{ij} + \frac{1}{2} \sum_{j \in S} x_j^2 \pi_j \sum_{i \in S} p_{ji} . \end{aligned}$$

Da $\sum_{j \in S} p_{ij} = 1$ für alle $j \in S$ ist, gilt

$$\begin{aligned} &\frac{1}{2} \sum_{i \in S} x_i^2 \pi_i \sum_{j \in S} p_{ij} + \sum_{i,j \in S} x_i x_j \pi_i p_{ij} + \frac{1}{2} \sum_{j \in S} x_j^2 \pi_j \sum_{i \in S} p_{ji} \\ &= \frac{1}{2} \sum_{i \in S} x_i^2 \pi_i + \sum_{i,j \in S} x_i x_j \pi_i p_{ij} + \frac{1}{2} \sum_{j \in S} x_j^2 \pi_j = \sum_{i \in S} x_i^2 \pi_i + \sum_{i,j \in S} x_i x_j \pi_i p_{ij} \\ &= \|x\|_\pi^2 + \left\langle \left(\sum_{j \in S} x_j p_{ij} \right)_{i \in S}, x \right\rangle_\pi = \|x\|_\pi^2 + \langle Px, x \rangle_\pi . \end{aligned}$$

Damit gilt insgesamt

$$\frac{1}{2} \sum_{i,j \in S} (x_i + x_j)^2 \pi_i p_{ij} = \langle Px, x \rangle_\pi + \|x\|_\pi^2$$

und die Behauptung ist bewiesen. ■

Nun kann Satz 11 bewiesen werden.

BEWEIS (VON SATZ 11) Sei $\gamma_i = (i, i_1, \dots, i_n, i)$ ein Pfad von i nach i mit ungerader Länge. $\Theta_i = \{s_{it_{t+1}} = (i_t, i_{t+1}) \mid t = 0, \dots, n \text{ und } i_0 = i_{n+1} = i\}$ sei die Menge aller Kanten des Pfades γ_i , wobei keine Kante $s_{it_{t+1}}$ mehrfach auftritt. Dann lässt sich x_i darstellen als

$$\begin{aligned} x_i &= \frac{1}{2} \left((x_i + x_{i_1}) - (x_{i_1} + x_{i_2}) + (x_{i_2} + x_{i_3}) - \dots + (x_{i_n} + x_i) \right) \\ &= \frac{1}{2} \sum_{s \in \Theta_i} (-1)^{g(s)} (x_{s^+} + x_{s^-}) , \end{aligned} \tag{3.60}$$

mit $g(s) = t$, wenn $s = (i_t, i_{t+1}) \in \Theta_i$ ist. Damit und aus der Cauchy-Schwarzschen-Ungleichung (4.8) ergibt sich nun für alle $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$

$$\begin{aligned} \|x\|_\pi^2 &\stackrel{(3.60)}{=} \sum_{i \in S} \left(\frac{1}{2} \sum_{s \in \Theta_i} (-1)^{g(s)} (x_{s^+} + x_{s^-}) \right)^2 \pi_i \\ &= \sum_{i \in S} \frac{\pi_i}{4} \left(\sum_{s \in \Theta_i} \frac{1}{\sqrt{Q(s)}} \sqrt{Q(s)} (-1)^{g(s)} (x_{s^+} + x_{s^-}) \right)^2 \end{aligned}$$

$$\begin{aligned}
 &\stackrel{(4.8)}{\leq} \sum_{i \in S} \frac{\pi_i}{4} \left(\sum_{s \in \Theta_i} \frac{1}{Q(s)} \sum_{s \in \Theta_i} Q(s) (-1)^{2g(s)} (x_{s^+} + x_{s^-})^2 \right) \\
 &\stackrel{(3.55)}{=} \frac{1}{4} \sum_{i \in S} (\pi_i |\gamma_i|_Q \sum_{s \in \Theta_i} (x_{s^+} + x_{s^-})^2 Q(s)) \\
 &= \frac{1}{4} \sum_{s \in \Theta} (x_{s^+} + x_{s^-})^2 Q(s) \sum_{\substack{\gamma_i \\ s \in \Theta_i}} |\gamma_i|_Q \pi_i \\
 &\leq \frac{1}{4} \sum_{s \in \Theta} (x_{s^+} + x_{s^-})^2 Q(s) \max_{\substack{\gamma_i \\ s \in \Theta_i}} \sum_{\gamma_i} |\gamma_i|_Q \pi_i \\
 &\stackrel{(3.57)}{=} \frac{\zeta}{4} \sum_{s \in \Theta} (x_{s^+} + x_{s^-})^2 Q(s) = \frac{\zeta}{4} \sum_{i,j \in S} (x_i + x_j)^2 \pi_i p_{ij} .
 \end{aligned}$$

Im Hinblick auf (3.59) ergibt sich,

$$\|x\|_\pi^2 \leq \frac{\zeta}{4} \sum_{i,j \in S} (x_i + x_j)^2 \pi_i p_{ij} \stackrel{(3.59)}{=} \frac{\zeta}{2} (\langle Px, x \rangle_\pi + \|x\|_\pi^2) . \quad (3.61)$$

Man wähle nun $x := v_N$, wobei v_N rechter Eigenvektor von λ_N sein soll. v_N ist insbesondere orthonormal zu den rechten Eigenvektoren von λ_i ($i = 1, \dots, N-1$) (siehe dazu 3.3.3). Es gilt damit und mit (3.38)

$$\begin{aligned}
 1 &= \|v_N\|_\pi^2 = \|x\|_\pi^2 \stackrel{(3.61)}{\leq} \frac{\zeta}{2} (\langle Pv_N, v_N \rangle_\pi + \|v_N\|_\pi^2) \\
 \Rightarrow 1 &\stackrel{(3.38)}{\leq} \frac{\zeta}{2} \left(\left\langle \sum_{i=1}^N \lambda_i \langle v_N, v_i \rangle_\pi v_i, v_N \right\rangle_\pi + 1 \right) \stackrel{(3.36)}{=} \frac{\zeta}{2} (\lambda_N + 1) . \quad (3.62)
 \end{aligned}$$

Nun muss nur noch die Ungleichung in (3.62) nach λ_N aufgelöst werden und man erhält

$$\lambda_N \geq -1 + \frac{2}{\zeta} ,$$

was zu beweisen war. (Angelehnt an [DS91] S. 40 und 41.) ■

Es wird nun Satz 11 auf die Markov-Kette aus Beispiel 23 angewendet.

Beispiel 24 Es sei die Markov-Kette der zufälligen Irrfahrt auf Graphen mit allen Voraussetzungen und Definitionen aus Beispiel 23 (bzw. Beispiel 17) gegeben. Des Weiteren setze man voraus, die Markov-Kette sei aperiodisch, und es definiere $\Delta' := \max_{\gamma \in \Sigma} \Delta(\gamma)$ die maximale Pfadlänge und $\beta' := \max_{s \in \Theta} \#\{\gamma_i \in \Sigma \mid s \in \Theta_i\}$ die maximale Anzahl der Pfade, die jeweils durch eine einzelne Kante laufen. Außerdem

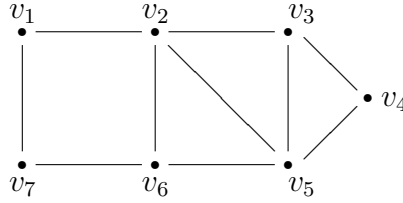
ist $\delta := \max_{i \in S} d_i$ (siehe dazu Beispiel 23). Damit gilt

$$\begin{aligned}
 \zeta &\stackrel{(3.57)}{=} \max_{s \in \Theta} \sum_{\substack{\gamma_i \\ s \in \Theta_i}} |\gamma_i|_Q \pi_i \stackrel{(3.55)}{=} \max_{s \in \Theta} \sum_{\substack{\gamma_i \\ s \in \Theta_i}} \sum_{s \in \Theta_i} \frac{1}{Q(s)} \frac{d_i}{d} \\
 &\stackrel{(3.56)}{=} \max_{s \in \Theta} \sum_{\substack{\gamma_i \\ s \in \Theta_i}} \sum_{t=0}^n \frac{1}{\pi_{i_t} \mathcal{P}_{i_t i_{t+1}}} \frac{d_i}{d} = \max_{s \in \Theta} \sum_{\substack{\gamma_i \\ s \in \Theta_i}} \sum_{t=0}^n \frac{1}{\frac{d_t}{d} \frac{1}{d}} \frac{d_i}{d} \\
 &= \max_{s \in \Theta} \sum_{\substack{\gamma_i \\ s \in \Theta_i}} \sum_{t=0}^n d \frac{d_i}{d} = \max_{s \in \Theta} \sum_{\substack{\gamma_i \\ s \in \Theta_i}} (n+1) d_i \leq \max_{s \in \Theta} \sum_{\substack{\gamma_i \\ s \in \Theta_i}} \Delta' d_i \leq \max_{s \in \Theta} \sum_{\substack{\gamma_i \\ s \in \Theta_i}} \Delta' \delta \\
 &\leq \max_{s \in \Theta} \beta' \Delta' \delta = \beta' \Delta' \delta .
 \end{aligned}$$

Somit gilt mit Satz 11

$$\lambda_N \geq -1 + \frac{2}{\zeta} \geq -1 + \frac{2}{\beta' \Delta' \delta} .$$

Für das in Beispiel 23 und Beispiel 17 betrachtete Zahlenbeispiel



gilt $\Delta' = 4$, $\beta' = 4$ und $\delta = 4$, so dass

$$\lambda_7 = -1 + \frac{2}{4 \cdot 4 \cdot 4} = -1 + \frac{1}{32} = -\frac{31}{32}$$

ist. (Vergl. [Sch03] S. 57.) □

Wenn man Satz 10 und Satz 11 zusammenfasst, hat sich somit ergeben

$$\lambda_* := \max \{ \lambda_2, |\lambda_N| \} \leq 1 - \min \left\{ \frac{1}{\kappa}, \frac{2}{\zeta} \right\} .$$

Natürlich sind die in Satz 10 und Satz 11 gemachten Abschätzungen für λ_2 bzw. λ_N nur zwei von sehr vielen verschiedenen Abschätzungen, die in der Literatur zu finden sind. Zum Beispiel sind in dem Artikel von Diaconis und Stroock ([DS91]) noch zwei weitere, Satz 10 ähnelnde Abschätzungen von λ_2 angegeben. Dort heißt

es zum Einen, dass mit den Voraussetzungen aus Satz 10

$$\lambda_2 \leq 1 - \frac{1}{K}$$

ist. Dabei ist K wie folgt definiert

$$K := \max_{s \in \Theta} Q(s)^{-1} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} |\Theta_{ij}| \pi_i \pi_j$$

und $|\Theta_{ij}|$ gibt die Anzahl der Ecken im Pfad γ_{ij} an. (Siehe [DS91] auf Seite 38.)

Zum Anderen kann man ebenfalls mit den selben Voraussetzungen wie in Satz 10 λ_2 z.B. so abschätzen

$$\lambda_2 \leq 1 - \frac{1}{8\eta^2} ,$$

mit

$$\eta := \max_{s \in \Theta} Q(s)^{-1} \sum_{\substack{\gamma_{ij} \\ s \in \Theta_{ij}}} \pi_i \pi_j .$$

(Vergl. [DS91] Seite 54.)

Ebenso können Abschätzungen von $\lambda_* = \max \{ \lambda_2, |\lambda_N| \}$ gemacht werden, bei denen die Betrachtung von Pfaden und Ecken ganz entfällt. Auch solch eine Abschätzung wird zur Anschauung ohne Beweis angegeben.

Für den betragsmäßig zweitgrößten Eigenwert λ_* der $N \times N$ Übergangsmatrix P einer ergodischen und reversiblen Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ gilt

$$\lambda_* \leq \max_{ij \in S} d_v(p_i, p_j) .$$

Dabei ist p_i und analog p_j die Wahrscheinlichkeitsverteilung, die durch die i -te Zeile von P gegeben ist und $d_v(p_i, p_j)$ ist der Abstand in Variation von p_i und p_j (siehe Definition 16). Diese Abschätzung wird als Dobrushin Grenze bezeichnet (siehe [Sen81] Seite 63).

3.5 Berechnung der Gleichgewichtsverteilung

In diesem Abschnitt wird auf die Berechnung der Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ einer Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ eingegangen. Generell gibt es mehrere Wege, mit denen sich die Gleichgewichtsverteilung einer Markov-Kette ermitteln lässt. Im Folgenden wird sowohl auf eine direkte als auch auf eine iterative Berechnungsmöglichkeit eingegangen.

3.5.1 Direkte Berechnungsmethode

Wenn der Zustandsraum $S = \{1, \dots, N\}$ der ergodischen Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ nicht zu groß ist, lässt sich die zur Markov-Kette gehörende Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ recht einfach direkt berechnen.

Satz 12 *Es sei $\{X_k\}_{k \in \mathbb{N}}$ eine ergodische Markov-Kette mit Zustandsraum $S = \{1, \dots, N\}$ und $N \times N$ Übergangsmatrix P . Dann ist die $N \times N$ Matrix $I - P + E$ invertierbar und die Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ der Markov-Kette ist gegeben durch*

$$\pi^T = e^T (I - P + E)^{-1} . \quad (3.63)$$

Dabei ist $e = (1, \dots, 1)^T \in \mathbb{R}^N$, sämtliche Einträge der $N \times N$ Matrix E sind gleich Eins und I ist die $N \times N$ Einheitsmatrix. (Vergl. [Sch03] S. 36.) \square

BEWEIS Zuerst wird die Invertierbarkeit der Matrix $I - P + E$ bewiesen. Aus der linearen Algebra ist bekannt, dass eine $N \times N$ Matrix genau dann invertierbar ist, wenn ihre Zeilen bzw. Spalten linear unabhängig sind (siehe [Fis02] S. 150). Somit muss nun gemäß Definition 19 gezeigt werden, dass die Gleichung

$$(I - P + E)x = 0 , x \in \mathbb{R}^N \quad (3.64)$$

nur gegeben ist, wenn $x = 0$ gewählt wird.

Nach Definition 12 (iii) weiß man, dass für die Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ der Markov-Kette $\pi^T = \pi^T P$ gilt. Einfache Umformung dieser Gleichung ergibt $\pi^T - \pi^T P = 0$ bzw.

$$\pi^T (I - P) = 0 . \quad (3.65)$$

Somit folgt

$$0 = \pi^T 0 \stackrel{(3.64)}{=} \pi^T (I - P + E)x = \pi^T (I - P)x + \pi^T E x \stackrel{(3.65)}{=} 0 + \pi^T E x , x \in \mathbb{R}^N . \quad (3.66)$$

Des Weiteren folgt unmittelbar aus der Definition der Gleichgewichtsverteilung (Definition 12) $\pi^T E = e^T$ und damit

$$0 \stackrel{(3.66)}{=} \pi^T E x = e^T x \text{ bzw. } E x = 0 , x \in \mathbb{R}^N . \quad (3.67)$$

Man sieht nun $0 \stackrel{(3.64)}{=} (I - P + E)x = (I - P)x + E x \stackrel{(3.67)}{=} (I - P)x + 0$ und somit $Px = x$. Damit gilt auch für alle $k \leq 1 \in \mathbb{N}$, dass $x = P^k x$ ($x \in \mathbb{R}^N$) ist.

Aus Satz 7 ist bekannt, dass P^k gegen W konvergiert. W ist dabei die $N \times N$ Matrix,

die aus den N identischen Zeilenvektoren π^T besteht. Dies bedeutet

$$x = P^k x \xrightarrow{k \rightarrow \infty} Wx$$

bzw.

$$\lim_{k \rightarrow \infty} x_i = \sum_{j=1}^N \pi_j x_j \text{ für alle } i \in S, x \in \mathbb{R}^N . \quad (3.68)$$

Weil in (3.68) die rechte Seite nicht von i abhängig ist, muss x ein Vektor sein, bei dem alle Einträge gleich sind d.h. $x = ce$ für eine Konstante $c \in \mathbb{R}$. Es folgt nun $0 \stackrel{(3.67)}{=} e^T x = ce^T e = cN$ und somit $c = 0$ bzw. $x = 0$. Die Matrix $I - P + E$ ist damit invertierbar.

Dass π^T die im Satz angegebene Gestalt hat, ergibt sich nun ganz einfach. Aus den vorherigen Überlegungen weiß man, dass

$$\pi^T(I - P + E) = \pi^T(I - P) + \pi^T E \stackrel{(3.65)}{=} 0 + \pi^T E \stackrel{Def.12(ii)}{=} e^T$$

ist. Löst man dies nach π^T auf, so ergibt sich

$$\pi^T = e^T(I - P + E)^{-1} .$$

(Vergl. [Sch03] S. 36 und 37.) ■

Es wird wieder auf das Beispiel der Wettervorhersage (Beispiel 3) und der Irrfahrten (Beispiel 13) zurückgegriffen, um Satz 12 zu veranschaulichen.

Beispiel 25 (Ergänzung zu Beispiel 3 (Wettervorhersage)) Im ersten Fall des Beispiels der Wettervorhersage war die Übergangsmatrix der dortigen Markov-Kette durch $P = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$ gegeben. In Beispiel 12 wurde dann bewiesen, dass $\pi = (0.5, 0.5)^T$ die zur Markov-Kette gehörende Gleichgewichtsverteilung ist. Hierbei wurde nicht gezeigt, wie man auf $\pi = (0.5, 0.5)^T$ kommt, sondern nur, dass dieses π den Bedingungen (3.1), (3.2) und (3.3) der Definition der Gleichgewichtsverteilung (Definition 12) genügt. Mit Hilfe von Satz 12 kann dies nun nachgeholt werden. Es gilt

$$\begin{aligned} \pi^T &= e^T(I - P + E)^{-1} = (1, 1) \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right)^{-1} \\ &= (1, 1) \begin{pmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{pmatrix} = (0.5, 0.5) . \end{aligned}$$

Die Markov-Kette im zweiten Fall des Wettervorhersagenbeispiels hatte die Übergangsmatrix $P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$. Damit gilt gemäß Satz 12

$$\begin{aligned} \pi^T &= e^T(I - P + E)^{-1} = (1, 1) \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right)^{-1} \\ &= (1, 1) \begin{pmatrix} 1.25 & -0.75 \\ -\frac{5}{12} & \frac{11}{12} \end{pmatrix} = \left(\frac{5}{6}, \frac{1}{6} \right). \end{aligned}$$

Wiederum Beispiel 12 bestätigt, dass dies die Gleichgewichtsverteilung der dortigen Markov-Kette ist. \square

Beispiel 26 (Ergänzung zu Beispiel 13 (Irrfahrten)) 1. Die Übergangsmatrix der symmetrischen, zyklischen Irrfahrt auf $S = \{1, 2, 3, 4\}$ hat die Form

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}. \text{ Es gilt nun}$$

$$\begin{aligned} \pi^T &= e^T(I - P + E)^{-1} \\ &= \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}^T \left(\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \right)^{-1} \\ &= (1, 1, 1, 1) \frac{1}{16} \begin{pmatrix} 11 & -1 & -5 & -1 \\ -1 & 11 & -1 & -5 \\ -5 & -1 & 11 & -1 \\ -1 & -5 & -1 & 11 \end{pmatrix} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right). \end{aligned}$$

Dies ist nach Beispiel 13 die Gleichgewichtsverteilung der Markov-Kette der symmetrischen, zyklischen Irrfahrt.

2. Die reflektierende Irrfahrt auf $S = \{1, 2, 3, 4\}$ hat die Übergangsmatrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{pmatrix}. \text{ Es gilt}$$

$$\begin{aligned}
 \pi^T &= e^T(I - P + E)^{-1} \\
 &= \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}^T \left(\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \right)^{-1} \\
 &= (1, 1, 1, 1) \frac{1}{32} \begin{pmatrix} 31 & 21 & -27 & -17 \\ -3 & 33 & -15 & -13 \\ -13 & -15 & 33 & 3 \\ 17 & -27 & 21 & 31 \end{pmatrix} = \left(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8} \right) .
 \end{aligned}$$

Auch dies stimmt mit der in Beispiel 13 angegebenen Gleichgewichtsverteilung der Markov-Kette der reflektierenden Irrfahrt überein. \square

3.5.2 Iterative Berechnungsmethode

Die Gleichgewichtsverteilung $\pi \in \mathbb{R}^N$ einer Markov-Kette $\{X_k\}_{k \in \mathbb{N}}$ lässt sich neben der oben angegebenen direkten Berechnungsmethode auch iterativ ermitteln.

Eine Möglichkeit, π iterativ zu berechnen, gibt zum Beispiel Satz 4 an. Satz 4 besagt nämlich (unter anderem), dass P^k für $k \in \mathbb{N}$ gegen unendlich gegen eine Matrix W konvergiert, deren Zeilen aus π^T besteht. (P war hierbei die zur ergodischen Markov-Kette gehörende Übergangsmatrix.) Dies ist am besten durch ein Beispiel zu veranschaulichen. Wiederum wird das Wettervorhersagenbeispiel (Beispiel 3) dazu genutzt.

Beispiel 27 Wie gehabt war im ersten Fall des Beispiels der Wettervorhersage die Übergangsmatrix der dortigen Markov-Kette durch $P = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$ gegeben. Mit einem geeigneten Computerprogramm (oder von Hand) kann nun P potenziert werden, und es ergibt sich

$$P^l = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} \text{ für } l \geq 14 .$$

Nach 13 Iterationsschritten ist somit die Gleichgewichtsverteilung π ermittelt. (Ein Vergleich mit Beispiel 12 zeigt, dass die Zeilen der Matrix P^l für alle $l \geq 14 \in \mathbb{N}$ auch wirklich der Gleichgewichtsverteilung π^T der Markov-Kette entsprechen.)

Im zweiten Fall des Beispiels der Wettervorhersage ist die Gleichgewichtsverteilung

nach 11 Iterationsschritten gegeben. □

Für sehr kleine Matrizen, wie im obigen Beispiel, mag diese Berechnung mit Hilfe eines Computerprogramms oftmals sehr schnell und leicht zum Ziel führen. Ist jedoch der Zustandsraum S erheblich größer und dementsprechend die Übergangsmatrix P der Markov-Kette sehr groß, dann kann diese Art, die Gleichgewichtsverteilung zu berechnen, recht ungeeignet sein. Sie nimmt dann unter Umständen sehr viel Rechenzeit in Anspruch.

4 Anhang

4.1 Grundlagen aus der linearen Algebra

Hier werden die für diese Arbeit wichtigen Grundkenntnisse aus der linearen Algebra vermittelt.

4.1.1 Eigenwerte und Eigenvektoren

Sei A eine (nicht notwendig stochastische) $N \times N$ Matrix mit Elementen aus \mathbb{R} oder \mathbb{C} , des Weiteren seien u und v zwei Vektoren ungleich Null aus \mathbb{R}^N oder \mathbb{C}^N und λ eine beliebige reelle oder komplexe Zahl. Falls

$$Av = \lambda v \text{ bzw. } u^T A = \lambda u^T \quad (4.1)$$

ist, dann ist λ ein Eigenwert von A und v ein rechter zu λ gehörender Eigenvektor von A bzw. u ein linker zu λ gehörender Eigenvektor von A .

Bemerkung 6 Da insbesondere für die $N \times N$ Matrix A

$$\begin{aligned} (u^T A)^T &= (\lambda u^T)^T \\ \Leftrightarrow A^T u &= \lambda u \end{aligned}$$

gilt, sind die linken Eigenvektoren von A gleich den rechten Eigenvektoren von A^T . \square

Satz 13 λ ist genau dann ein Eigenwert der $N \times N$ Matrix A , wenn λ eine Lösung der charakteristischen Gleichung

$$\det(A - xI) = 0, \quad x \in \mathbb{R} \text{ oder } \mathbb{C} \quad (4.2)$$

ist. I sei hierbei die $N \times N$ Einheitsmatrix. (Vergl. [Fis02] S. 228.) \square

BEWEIS „ λ ist Eigenwert von $A \Rightarrow \lambda$ ist Lösung der charakteristischen Gleichung“:
Es sei λ ein Eigenwert von A . Mit dem N -dimensionalen zu λ gehörenden Eigenvek-

tor $v \neq 0$ gilt dann

$$\begin{aligned} & Av = \lambda v \\ \Leftrightarrow & (A - \lambda I)v = 0 \\ \Leftrightarrow & \text{Ker}(A - \lambda I) \neq \{0\}, \text{ da nach Voraussetzung } v \neq 0 \text{ ist} \\ \Leftrightarrow & \text{Lös}(A - \lambda I, 0) \neq \{0\}, \text{ da } \text{Lös}(A - \lambda I, 0) = \{v \in \mathbb{C}^N \mid (A - \lambda I)v = 0\} \\ & = \text{Ker}(A - \lambda I) \text{ ist} \\ \Leftrightarrow & \dim(\text{Lös}(A - \lambda I, 0)) > 0 \\ \Leftrightarrow & \text{rang}(A - \lambda I) < N, \text{ da } \dim(\text{Lös}(A - \lambda I, 0)) = N - \text{rang}(A - \lambda I) \text{ ist} \\ \Leftrightarrow & \det(A - \lambda I) = 0. \end{aligned}$$

„ λ ist Lösung der charakteristischen Gleichung $\Rightarrow \lambda$ ist Eigenwert von A “: Hier kann analog argumentiert werden. (Angelehnt an [Fis02] S. 228.) ■

(Die Definitionen und Eigenschaften der im Beweis benutzten Ausdrücke Kern, Rang und Determinante, so wie die Dimensionsformel können bei Bedarf in [Fis02] nachgelesen werden.)

(4.2) ist eine algebraische Gleichung der Ordnung N . Sie besitzt somit N gegebenenfalls komplexe Lösungen $\lambda_1, \dots, \lambda_N$, die nicht alle von einander verschieden sein müssen. Für jeden Eigenwert λ_i ($i = 1, \dots, N$) gibt es linke bzw. rechte Eigenvektoren u_i bzw. v_i .

4.1.2 Spektraldarstellung

Wenn man $N \times N$ Matrizen bestehend aus den rechten Eigenvektoren v_1, \dots, v_N und den linken Eigenvektoren u_1, \dots, u_N der $N \times N$ Matrix A wie folgt bildet

$$V := (v_1, \dots, v_N) \text{ und } U := \begin{pmatrix} u_1^T \\ \vdots \\ u_N^T \end{pmatrix},$$

dann gilt mit (4.1)

$$AV = V \text{diag}(\lambda) \text{ bzw. } UA = \text{diag}(\lambda)U. \quad (4.3)$$

Hierbei sei $\lambda := (\lambda_1, \dots, \lambda_N)^T$ und $\text{diag}(\lambda)$ die Diagonalmatrix mit den Diagonalelementen $\lambda_1, \dots, \lambda_N$.

Damit kann nun der folgende Satz bewiesen werden, der die so genannte Spektraldarstellung von A angibt.

Satz 14 (Spektraldarstellung) Falls die rechten Eigenvektoren v_1, \dots, v_N der $N \times N$ Matrix A linear unabhängig sind, dann kann A in Spektraldarstellung dargestellt werden. Es gilt dann

$$A^k = \sum_{i=1}^N \lambda_i^k v_i u_i^T . \quad (4.4)$$

u_i ($i = 1, \dots, N$) sei hierbei ein linker N -dimensionaler Eigenvektor von A zum Eigenwert λ_i . (Vergl. [Sch03] S. 15.) \square

BEWEIS Da die rechten Eigenvektoren v_1, \dots, v_N nach Voraussetzung linear unabhängig sind, existiert die zur Matrix $V = (v_1, \dots, v_N)$ inverse Matrix V^{-1} . Aus (4.3) folgt

$$\begin{aligned} V^{-1}AVV^{-1} &= V^{-1}V \text{diag}(\lambda)V^{-1} \\ \Rightarrow V^{-1}A &= \text{diag}(\lambda)V^{-1} . \end{aligned}$$

Die Matrix U , die aus den linken Eigenvektoren von A besteht, kann also durch den Ansatz $U := V^{-1}$ gewählt werden. Damit und wiederum mit (4.3) gilt dann

$$\begin{aligned} AVU &= V \text{diag}(\lambda)U \\ \Rightarrow AVV^{-1} &= V \text{diag}(\lambda)U \\ \Rightarrow A &= V \text{diag}(\lambda)U \end{aligned}$$

und somit ist

$$\begin{aligned} A^k &= V(\text{diag}(\lambda))^k U \\ &= \begin{pmatrix} v_{1_1} \cdots v_{1_N} \\ \vdots \cdots \vdots \\ v_{1_N} \cdots v_{N_N} \end{pmatrix} \begin{pmatrix} \lambda_1^k & & 0 \\ & \cdots & \\ 0 & & \lambda_N^k \end{pmatrix} \begin{pmatrix} u_{1_1} \cdots u_{1_N} \\ \vdots \cdots \vdots \\ u_{N_1} \cdots u_{N_N} \end{pmatrix} \\ &= \begin{pmatrix} v_{1_1} \lambda_1^k \cdots v_{1_N} \lambda_N^k \\ \vdots \cdots \vdots \\ v_{1_N} \lambda_1^k \cdots v_{N_N} \lambda_N^k \end{pmatrix} \begin{pmatrix} u_{1_1} \cdots u_{1_N} \\ \vdots \cdots \vdots \\ u_{N_1} \cdots u_{N_N} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^N \lambda_i^k v_{i_1} u_{i_1} \cdots \sum_{i=1}^N \lambda_i^k v_{i_N} u_{i_N} \\ \vdots \cdots \vdots \\ \sum_{i=1}^N \lambda_i^k v_{i_N} u_{i_1} \cdots \sum_{i=1}^N \lambda_i^k v_{i_N} u_{i_N} \end{pmatrix} \\ &= \sum_{i=1}^N \lambda_i^k \begin{pmatrix} v_{i_1} u_{i_1} \cdots v_{i_1} u_{i_N} \\ \vdots \cdots \vdots \\ v_{i_N} u_{i_1} \cdots v_{i_N} u_{i_N} \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^N \lambda_i^k \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} (u_1 \cdots u_N) \\
&= \sum_{i=1}^N \lambda_i^k v_i u_i^T.
\end{aligned}$$

(In Anlehnung an [Sch03] S. 15.) ■

Für Satz 14 wird vorausgesetzt, dass alle Eigenvektoren v_1, \dots, v_N linear unabhängig sind. Eine hinreichende Bedingung dafür liefert das folgende Lemma.

Lemma 6 *Falls alle Eigenwerte $\lambda_1, \dots, \lambda_N$ der $N \times N$ Matrix A voneinander verschieden sind, dann sind die rechten Eigenvektoren v_1, \dots, v_N von A linear unabhängig. (Vergl. [Sch03] S. 15.)* □

Um Lemma 6 zu beweisen, wird vorab eine allgemeine Definition der linearen Unabhängigkeit angegeben.

Definition 19 Die N -dimensionalen Vektoren v_1, \dots, v_N bezeichnet man als linear unabhängig, wenn es $a_1, \dots, a_N \in \mathbb{R}$ gibt, derart dass $a_1 v_1 + \dots + a_N v_N = 0$ nur gilt, wenn alle a_i ($i = 1, \dots, N$) gleich Null sind. (Vergl. [Fis02] S. 81.) □

Das Lemma 6 wird nun mit vollständiger Induktion bewiesen.

BEWEIS (VON LEMMA 6) Der Eigenvektor v_1 zum Eigenwert λ_1 von A hat per Definition mindestens einen Eintrag ungleich Null. Demnach gilt $a_1 v_1 = 0$ nur, wenn $a_1 = 0$ ist.

Es seien alle Eigenwerte $\lambda_1, \dots, \lambda_N$ von A voneinander verschieden und für ein gewisses $l \leq N$ seien die Eigenvektoren v_1, \dots, v_{l-1} nach der Induktionsannahme linear unabhängig. Um zu zeigen, dass dann auch die Eigenvektoren v_1, \dots, v_l für alle $l \leq N$ linear unabhängig sind, muss gemäß Definition 19 gezeigt werden, dass

$$\sum_{j=1}^l a_j v_j = 0 \tag{4.5}$$

nur gilt, wenn $a_1 = \dots = a_l = 0$ ist.

Es seien a_1, \dots, a_l so gewählt, dass (4.5) gilt. Damit und mit (4.1) gilt dann ebenfalls

$$0 = A0 \stackrel{(4.5)}{=} \sum_{j=1}^l a_j A v_j \stackrel{(4.1)}{=} \sum_{j=1}^l a_j \lambda_j v_j$$

und

$$0 = \lambda_l 0 \stackrel{(4.5)}{=} \lambda_l \sum_{j=1}^l a_j v_j = \sum_{j=1}^l \lambda_l a_j v_j .$$

Zusammengefasst ergibt dies

$$0 = \sum_{j=1}^l \lambda_l a_j v_j - \sum_{j=1}^l a_j \lambda_j v_j = \sum_{j=1}^l (\lambda_l - \lambda_j) a_j v_j = \sum_{j=1}^{l-1} (\lambda_l - \lambda_j) a_j v_j .$$

Gemäß der Induktionsannahme folgt hieraus

$$(\lambda_l - \lambda_1) a_1 = \dots = (\lambda_l - \lambda_{l-1}) a_{l-1} = 0 ,$$

bzw. da $\lambda_l \neq \lambda_j$ für $1 \leq j \leq l-1$ ist, dass $a_1 = a_2 = \dots = a_{l-1} = 0$ gilt. Wegen (4.5) muss damit auch $a_l = 0$ sein, womit der Beweis beendet ist. ([Sch03] S. 16.) ■

4.1.3 Skalarprodukt und Norm

In Kapitel 2 werden zwei spezielle Skalarprodukte mit dazugehöriger Norm betrachtet. Hier die allgemeinen Definitionen von Skalarprodukt und Norm.

Definition 20 (Skalarprodukt) Ein Skalarprodukt auf dem reellen Vektorraum V ist eine Abbildung

$$\begin{aligned} \langle \cdot, \cdot \rangle : V \times V &\rightarrow \mathbb{R} \\ (u, v) &\mapsto \langle u, v \rangle \end{aligned}$$

mit den folgenden Eigenschaften:

- (a) $\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$,
- $\langle u, \alpha v + \beta w \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle$ (Bilinearität) ,
- (b) $\langle u, v \rangle = \langle v, u \rangle$ (Symmetrie) ,
- (c) $\langle u, u \rangle > 0$ falls $u \neq 0$ ist (positive Definitheit)

für alle $u, v, w \in V$ und alle $\alpha, \beta \in \mathbb{R}$. (Vergl. [Fis02] S. 274.) □

Definition 21 (Norm) Sei V ein reeller oder komplexer Vektorraum. Eine Funktion

$$\|\cdot\| : V \rightarrow \mathbb{R}_{\geq 0}$$

heißt Norm auf V , wenn für alle Vektoren $u, v \in V$ und alle $\alpha \in \mathbb{R}$ oder \mathbb{C} gilt

- (a) $\|u\| = 0 \Leftrightarrow u = 0$ (Definitheit) ,
- (b) $\|\alpha u\| = |\alpha| \|u\|$ (Homogenität) ,
- (c) $\|u + v\| \leq \|u\| + \|v\|$ (Dreiecksungleichung) .

(Vergl. [Fis02] S. 274 und 275.) □

Bemerkung 7 In reellen Vektorräumen wird die Norm auch oft über das Skalarprodukt definiert. Für alle Vektoren u aus dem reellen Vektorraum gilt dann

$$\|u\| := \sqrt{\langle u, u \rangle} . \quad (4.6)$$

Man spricht in diesem Fall von der zum Skalarprodukt gehörigen Norm. (Vergl. [Fis02] S. 274.) □

Die nun folgende Cauchy-Schwarzsche-Ungleichung ist von zentraler Bedeutung. Sie ist ein viel genutztes Beweis-Hilfsmittel.

Satz 15 (Cauchy-Schwarzsche-Ungleichung) Sei $\langle \cdot, \cdot \rangle$ ein Skalarprodukt auf dem reellen Vektorraum V , dann gilt für alle $u, v \in V$

$$\langle u, v \rangle^2 \leq \langle u, u \rangle \langle v, v \rangle$$

bzw. mit (4.6)

$$\langle u, v \rangle \leq \|u\| \|v\| .$$

(Vergl. [Fis02] S. 275.) □

BEWEIS Für den Fall $v = 0$ (oder analog $u = 0$ bzw. u und $v = 0$) ist die Cauchy-Schwarzsche-Ungleichung trivialerweise klar. Es gilt dann nämlich $\langle 0, v \rangle = 0 \langle 0, v \rangle = 0$.

Nun wird der Fall $u, v \neq 0$ betrachtet. Für alle $\alpha \in \mathbb{R}$ gilt

$$\begin{aligned} 0 &\leq \langle u - \alpha v, u - \alpha v \rangle = \langle u - \alpha v, u \rangle - \alpha \langle u - \alpha v, v \rangle \\ &= \langle u, u \rangle - 2\alpha \langle u, v \rangle + \alpha^2 \langle v, v \rangle . \end{aligned}$$

Mit $\alpha := \frac{\langle u, v \rangle}{\langle v, v \rangle}$ folgt

$$0 \leq \langle u, u \rangle - 2 \frac{\langle u, v \rangle}{\langle v, v \rangle} \langle u, v \rangle + \frac{\langle u, v \rangle^2}{\langle v, v \rangle^2} \langle v, v \rangle$$

$$\begin{aligned} &= \langle u, u \rangle - 2 \frac{\langle u, v \rangle^2}{\langle v, v \rangle} + \frac{\langle u, v \rangle^2}{\langle v, v \rangle} \\ &= \langle u, u \rangle - \frac{\langle u, v \rangle^2}{\langle v, v \rangle} . \end{aligned} \tag{4.7}$$

Damit und mit (4.6) gilt wiederum

$$\begin{aligned} 0 &\stackrel{(4.7)}{\leq} \langle u, u \rangle - \frac{\langle u, v \rangle^2}{\langle v, v \rangle} = \|u\|^2 - \langle u, v \rangle^2 \|v\|^{-2} \\ &\Rightarrow \langle u, v \rangle^2 \|v\|^{-2} \leq \|u\|^2 \\ &\Rightarrow \langle u, v \rangle^2 \leq \|u\|^2 \|v\|^2 , \end{aligned}$$

womit der Beweis beendet ist. (Vergl. [Def05].) ■

Bemerkung 8 Wendet man die Cauchy-Schwarzsche-Ungleichung auf euklidische Vektorräume \mathbb{R}^N an, wobei auf \mathbb{R}^N das Standardskalarprodukt definiert sei, so gilt für alle Vektoren $u = (u_1, \dots, u_N) \in \mathbb{R}^N$ und $v = (v_1, \dots, v_N) \in \mathbb{R}^N$

$$\left(\sum_{i=1}^N u_i v_i \right)^2 \leq \left(\sum_{i=1}^N u_i^2 \right) \left(\sum_{i=1}^N v_i^2 \right) . \tag{4.8}$$

□

4.1.4 Symmetrische Matrizen

An dieser Stelle werden die Definition, sowie die allgemeinen Eigenschaften symmetrischer Matrizen angegeben.

Definition 22 Eine beliebige Matrix A heißt symmetrisch, wenn $A = A^T$ gilt. (Vergl. [Fis02] S. 104.) □

Symmetrische Matrizen weisen spezielle Eigenschaften auf, die in dieser Arbeit, besonders bei Beweisen (z.B. von Satz 10 und Satz 11), von großem Nutzen sind.

Bemerkung 9 (Eigenschaften symmetrischer Matrizen) Reelle symmetrische Matrizen haben folgende Eigenschaften:

1. Sie sind quadratisch.
2. Sie sind diagonalisierbar (und besitzen somit nur linear unabhängige Eigenvektoren).
3. Sie haben nur reelle Eigenwerte.

4. Ihre Eigenvektoren zu den verschiedenen Eigenwerten sind orthogonal zueinander.

(Bemerkung 9 ist eine Zusammenstellung von Erkenntnissen aus [Fis02]). □

Wem die Begriffe diagonalisierbar oder orthogonal fremd sind, der kann sich z.B. in [Fis02] darüber informieren. Der Beweis der Bemerkung 9 wird dem Leser überlassen. Teilweise sind es sehr lange algebraische Beweise, die weiterführende algebraische Kenntnisse voraussetzen. Wer trotzdem das Vorhaben hat, Bemerkung 9 nachzuweisen, der kann [Fis02] dazu nutzen.

Literaturverzeichnis

- [Beh00] BEHRENDTS, Erhard: *Introduction to Markov Chains*. vieweg, 2000
- [Bré99] BRÉMAUD, Pierre: *Markov Chains, Gibbs Fields, Monte Carlo Simulation and Queues*. Springer, New York, 1999
- [Def05] DEFANT, Andreas: *Funktionalanalysis 1*. 2004/2005
- [DS91] DIACONIS, Persi ; STROOCK, Daniel: Geometric bounds for eigenvalues of Markov chains. In: *The Annals of Applied Probability* (1991), S. 36–61
- [Fis02] FISCHER, Gerd: *Lineare Algebra*. vieweg, 2002
- [Häg02] HÄGGSTRÖM, Olle: *Finite Markov Chains and Algorithmic Applications*. Cambridge, 2002
- [Kön05] KÖNIG, Wolfgang: *Stochastische Prozesse I*. www.math.uni-leipzig.de/~koenig/www/StPrI.pdf. Version: 2005
- [Sch03] SCHMIDT, Volker: *Markov-Ketten und Monte-Carlo-Simulation*. www.mathematik.uni-ulm.de/stochastik/lehre/ss03/markov/skript/skript.pdf. Version: 2003
- [Sen81] SENETA, Eugene: *Non-negative matrices and Markov chains*. Springer, Berlin, 1981
- [Str05] STROOCK, Daniel W.: *An introduction to Markov processes*. Springer, Berlin, 2005
- [Vol] VOLKWEIN, Stefan: *Grundlagen der Optimierung*. www.uni-graz.at/imawww/volkwein/opt.pdf
- [Win03] WINKLER, Gerhard: *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Springer, Berlin Heidelberg, 2003

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.