



Markov-Ketten und Monte-Carlo-Simulation

Universität Ulm
Abteilung Stochastik

Vorlesungsskript
Prof. Dr. Volker Schmidt
Stand: Sommersemester 2003

ULM, IM SEPTEMBER 2003

Inhaltsverzeichnis

1	Einleitung	4
2	Markov-Ketten	5
2.1	Modellbeschreibung und Beispiele	5
2.1.1	Zustandsraum, Anfangsverteilung und Übergangswahrscheinlichkeiten	5
2.1.2	Beispiele	8
2.1.3	Rekursive Darstellung	10
2.1.4	Matrix der n -stufigen Übergangswahrscheinlichkeiten	13
2.2	Ergodizität und Stationarität	16
2.2.1	Grundlegende Definitionen und quasi-positive Übergangsmatrizen	16
2.2.2	Abschätzung der Konvergenzgeschwindigkeit; Perron–Frobenius–Theorem	21
2.2.3	Irreduzibilität und Aperiodizität	25
2.2.4	Stationäre Anfangsverteilungen	31
2.2.5	Direkte und iterative Berechnungsmethoden	36
2.3	Abschätzung der Konvergenzgeschwindigkeit; Reversibilität	39
2.3.1	Definition und Beispiele	39
2.3.2	Rekursive Konstruktion der „Vergangenheit“	43
2.3.3	Bestimmung der Konvergenzgeschwindigkeit bei Reversibilität	44
2.3.4	Multiplikativ reversible Version der Übergangsmatrix; Spektraldarstellung	45
2.3.5	Alternative Abschätzung der Konvergenzgeschwindigkeit; χ^2 -Kontrast	47
2.3.6	Dirichlet–Formen und Rayleigh–Theorem	51
2.3.7	Schranken für die Eigenwerte λ_2 und λ_ℓ	53
3	Monte-Carlo-Simulation	58
3.1	Erzeugung von Pseudozufallszahlen	59
3.1.1	Einfache Anwendungsbeispiele; Monte-Carlo-Schätzer	59
3.1.2	Lineare Kongruenzgeneratoren	61
3.1.3	Tests von Güteeigenschaften	63
3.2	Transformation gleichverteilter Pseudozufallszahlen	69
3.2.1	Inversionsmethode	69
3.2.2	Transformationsalgorithmen für diskrete Verteilungen	72
3.2.3	Akzeptanz- und Verwerfungsmethode	76
3.2.4	Quotienten von gleichverteilten Zufallsvariablen	80
3.3	Simulationsmethoden mit Markov-Ketten	83
3.3.1	Beispiel: Hard-Core-Modell	83
3.3.2	Gibbs-Sampler	86

3.3.3	Metropolis-Hastings-Algorithmus	90
3.4	Fehleranalyse bei MCMC-Simulation	93
3.4.1	Abschätzung der Konvergenzgeschwindigkeit	93
3.4.2	MCMC-Schätzer; Bias und Fundamentalmatrix	97
3.4.3	Asymptotische Schätzvarianz; mittlerer quadratischer Fehler	100
3.5	Kopplungsalgorithmen; perfekte MCMC-Simulation	104
3.5.1	Vorwärtskopplung; Gegenbeispiel	105
3.5.2	Propp-Wilson-Algorithmus; Coupling-from-the-Past	107
3.5.3	Monotone Kopplungsalgorithmen	109
3.5.4	Beispiele: Geburts- und Todesprozesse; Ising-Modell	111
3.5.5	Read-Once-Modifikation des CFTP-Algorithmus	115

1 Einleitung

- Markov-Ketten sind eine grundlegende Klasse stochastischer Modelle
 - für Folgen von Zufallsvariablen, die nicht unabhängig sind, das heißt, gewisse *Abhängigkeitsstrukturen* aufweisen.
 - Markov-Ketten besitzen zahlreiche *Anwendungen in der Finanz- und Versicherungsmathematik*.
 - Sie spielen aber auch eine wichtige Rolle bei der mathematischen Modellierung und Analyse weiterer Problemstellungen mit sozio-ökonomischer Bedeutung, beispielsweise in den *Lebenswissenschaften*.
 - Oftmals sind jedoch die interessierenden Fragestellungen und damit auch die entsprechenden mathematischen Modelle so komplex, dass sie nicht ausschließlich mit analytischen Formeln untersucht werden können.
 - Markov-Ketten bieten dann ein alternatives Analyse-Tool, und zwar im Zusammenhang mit der Konstruktion von *Computer-Algorithmen* zur Markov-Chain-Monte-Carlo-Simulation (MCMC) der betrachteten mathematischen Modelle.
- Die Vorlesung „Markov-Ketten und Monte-Carlo-Simulation“ vertieft Methoden und Modelle, die in der Vorlesung „Wahrscheinlichkeitsrechnung“ behandelt wurden. Kenntnisse aus den Vorlesungen „Statistik I“ bzw. „Statistik II“ sind nützlich, werden jedoch nicht vorausgesetzt.
- Schwerpunkte der Vorlesung sind die folgenden Themen:
 - Markov-Ketten mit diskreter Zeit und endlichem Zustandsraum
 - Stationarität und Ergodizität
 - Markov-Chain-Monte-Carlo (MCMC)
 - Reversibilität und Kopplungsalgorithmen
- Dabei werden wir insbesondere Begriffe und Ergebnisse nutzen, die in der Vorlesung „Wahrscheinlichkeitsrechnung“ eingeführt worden sind, vgl. das Skript zur Vorlesung „Wahrscheinlichkeitsrechnung“ im Wintersemester 2001/02:

http://www.mathematik.uni-ulm.de/stochastik/lehre/ws01_02/wr/vs1/vs1.html
- Verweise auf dieses Vorlesungsmanuskript werden wir mit dem Zusatz „WR“ vor der Nummer der zitierten Abschnitte, Lemmata, Theoreme, Korollare bzw. Formeln kennzeichnen.
- Die folgende Liste von einführenden Lehrbüchern umfasst lediglich eine kleine Auswahl von Texten, die neben dem Vorlesungsmanuskript für ein ergänzendes und vertiefendes Studium empfohlen werden können.
 - D. Aldous, J.A. Fill (2002) *Reversible Markov Chains and Random Walks on Graphs*. Buchmanuskript
 - E. Behrends (2000) *Introduction to Markov Chains*. Vieweg, Braunschweig
 - P. Bremaud (1999) *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, New York
 - B. Chalmond (2003) *Modeling and Inverse Problems in Image Analysis*. Springer, New York
 - O. Häggström (2002) *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press, Cambridge
 - U. Krengel (2002) *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, Braunschweig
 - S.I. Resnick (1992) *Adventures in Stochastic Processes*. Birkhäuser, Boston
 - T. Rolski, H. Schmidli, V. Schmidt, J. Teugels (2002) *Stochastic Processes for Insurance and Finance*. Wiley, Chichester
 - H. Thorisson (2002) *Coupling, Stationarity, and Regeneration*. Springer, New York
 - G. Winkler (2003) *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Springer, Berlin

2 Markov-Ketten

- Markov-Ketten können die (zeitliche) Entwicklung von Objekten, Sachverhalten, Systemen etc. beschreiben,
 - die zu jedem Zeitpunkt jeweils nur eine von *endlich* (oder *abzählbar unendlich*) vielen Ausprägungen annehmen können,
 - die wir die *Zustände* des betrachteten Objektes, Sachverhaltes bzw. Systems nennen werden.
- Beispiele solcher Objekte, Sachverhalte bzw. Systeme sind
 - die jeweils aktuellen Preise von Produkten, wie Versicherungsprämien, Aktienkurse, Zinsraten, etc., die in einer diskreten (beispielsweise ganzzahligen) Werte-Skala gemessen werden,
 - der monatliche Nettogewinn eines Unternehmens,
 - die aktuellen Längen der Warteschlangen vor den Kassen in Supermärkten,
 - der Vektor von Temperatur, Luftdruck, Niederschlagsmenge und Windgeschwindigkeit, die stündlich an der Wetterstation Ulm-Kuhberg gemessen werden,
 - digitale Landkarten, die beispielsweise das aktuelle Ausbreitungsgebiet einer bestimmten Infektionskrankheit beschreiben können,
 - mikroskopische 2D- bzw. 3D-Bilder, die den aktuellen „Zustand“ (d.h. geometrische Struktureigenschaften) von biologischen Geweben, aber auch von keramischen bzw. metallischen Materialien darstellen.

Beachte

- In dieser Vorlesung werden wir *zeitdiskrete* Markov-Ketten untersuchen, d.h., die zeitliche Entwicklung der betrachteten Objekte, Sachverhalte bzw. Systeme wird *schrittweise*, beispielsweise zu ganzzahligen Zeitpunkten beobachtet.
- Die Algorithmen der Markov-Ketten-Monte-Carlo-Simulation, die wir in der zweiten Hälfte der Vorlesung diskutieren, beruhen gerade auf solchen zeitdiskreten Markov-Ketten.
- Die Anzahl der potentiell möglichen Zustände kann dabei sehr groß sein.
- Aus mathematischen Gründen ist es deshalb manchmal zweckmäßiger, auch den Fall von unendlich vielen Zuständen zu betrachten, wobei sich die Untersuchungsmethoden nur relativ wenig ändern, wenn *abzählbar* unendlich viele Zustände betrachtet werden.

2.1 Modellbeschreibung und Beispiele

2.1.1 Zustandsraum, Anfangsverteilung und Übergangswahrscheinlichkeiten

- Das stochastische Modell der zeitdiskreten Markov-Ketten (mit endlich vielen Zuständen) besteht aus den drei Komponenten: Zustandsraum, Anfangsverteilung und Übergangsmatrix.
 - Den Ausgangspunkt bildet die (endliche) Menge aller möglichen Zustände, die *Zustandsraum* der Markov-Kette genannt wird und die o.B.d.A. mit der Menge $E = \{1, 2, \dots, \ell\}$ identifiziert werden kann, wobei $\ell \in \mathbb{N} = \{1, 2, \dots\}$ eine beliebige, jedoch fest vorgegebene natürliche Zahl ist.
 - Für jedes $i \in E$ sei α_i die Wahrscheinlichkeit, dass sich das betrachtete Objekt, Sachverhalt bzw. System zum „Zeitpunkt“ $n = 0$ im Zustand i befindet, wobei

$$\alpha_i \in [0, 1], \quad \sum_{i=1}^{\ell} \alpha_i = 1 \quad (1)$$

vorausgesetzt wird. Der Vektor $\alpha = (\alpha_1, \dots, \alpha_\ell)^\top$ der (Einzel-) Wahrscheinlichkeiten $\alpha_1, \dots, \alpha_\ell$ bildet die *Anfangsverteilung* der Markov-Kette.

- Außerdem betrachten wir für jedes Paar $i, j \in E$ die (bedingte) Wahrscheinlichkeit $p_{ij} \in [0, 1]$, dass das betrachtete Objekt, Sachverhalt bzw. System in einem Schritt aus dem Zustand i in den Zustand j übergeht.
- Die $\ell \times \ell$ Matrix $\mathbf{P} = (p_{ij})_{i,j=1,\dots,\ell}$ der *Übergangswahrscheinlichkeiten* p_{ij} mit

$$p_{ij} \geq 0, \quad \sum_{j=1}^{\ell} p_{ij} = 1, \quad (2)$$

heißt (einstufige) *Übergangsmatrix* der Markov-Kette.

- Für jede Menge $E = \{1, 2, \dots, \ell\}$, für jeden Vektor $\alpha = (\alpha_1, \dots, \alpha_\ell)^\top$ bzw. jede Matrix $\mathbf{P} = (p_{ij})$, die den Bedingungen (1) und (2) genügen, kann nun der Begriff der zugehörigen Markov-Kette wie folgt eingeführt werden.

Definition

- Sei $X_0, X_1, \dots : \Omega \rightarrow E$ eine Folge von Zufallsvariablen, die über einunddemselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ definiert sind und ihre Werte in der Menge $E = \{1, 2, \dots, \ell\}$ annehmen.
- Dann heißt $X_0, X_1, \dots$ (homogene) *Markov-Kette* mit der Anfangsverteilung $\alpha = (\alpha_1, \dots, \alpha_\ell)^\top$ und der Übergangsmatrix $\mathbf{P} = (p_{ij})$, falls

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \alpha_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \quad (3)$$

für beliebige $n = 0, 1, \dots$ und $i_0, i_1, \dots, i_n \in E$ gilt.

Beachte

- Eine quadratische Matrix $\mathbf{P} = (p_{ij})$, die der Bedingung (2) genügt, wird *stochastische Matrix* genannt.
- Durch das folgende Theorem 2.1 wird die intuitive Bedeutung der Bedingung (3) deutlich, insbesondere die Wortwahlen „Anfangsverteilung“ und „Übergangsmatrix“.
- Außerdem ergibt sich aus Theorem 2.1 noch eine andere (äquivalente) Definitionsmöglichkeit von Markov-Ketten, die in der Literatur oft verwendet wird.

Theorem 2.1 Die Folge $\{X_n\}$ von E -wertigen Zufallsvariablen ist genau dann eine Markov-Kette, wenn es eine stochastische Matrix $\mathbf{P} = (p_{ij})$ gibt, so dass

$$P(X_n = i_n \mid X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p_{i_{n-1} i_n} \quad (4)$$

für beliebige $n = 1, 2, \dots$ und $i_0, i_1, \dots, i_n \in E$ mit $P(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$.

Beweis

- Die Notwendigkeit der Bedingung (4) ist offensichtlich, denn die Gültigkeit von (4) ergibt sich unmittelbar aus (3) und aus der Definition bedingter Wahrscheinlichkeiten; vgl. Abschnitt WR-2.6.1.
- Wir nehmen nun umgekehrt an, dass $\{X_n\}$ eine Folge von E -wertigen Zufallsvariablen ist, so dass es eine stochastische Matrix $\mathbf{P} = (p_{ij})$ gibt, für die die Bedingung (4) erfüllt ist.
- Wir setzen $\alpha_i = P(X_0 = i)$ für jedes $i \in E$ und erkennen, dass die Bedingung (3) für $n = 0$ offenbar erfüllt ist.
- Außerdem gilt:
 - Aus $P(X_0 = i_0) = 0$ folgt, dass $P(X_0 = i_0, X_1 = i_1) = 0$,
 - und falls $P(X_0 = i_0) > 0$, dann ergibt sich aus (4), dass

$$P(X_0 = i_0, X_1 = i_1) = P(X_0 = i_0)P(X_1 = i_1 \mid X_0 = i_0) \stackrel{(4)}{=} \alpha_{i_0} p_{i_0 i_1}.$$

- Insgesamt gilt also

$$P(X_0 = i_0, X_1 = i_1) = \begin{cases} 0, & \text{falls } \alpha_{i_0} = 0, \\ \alpha_{i_0} p_{i_0 i_1}, & \text{falls } \alpha_{i_0} > 0, \end{cases}$$

d.h., die Gültigkeit von (3) ist auch für den Fall $n = 1$ bewiesen.

- Es gelte (3) nun für ein gewisses $n = k - 1 \geq 1$.
 - Aus den elementaren Monotonieeigenschaften von Wahrscheinlichkeitsmaßen folgt (vgl. Teilaussage 2 in Theorem WR-2.1), dass $P(X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1}) = 0$ die Gültigkeit von $P(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) = 0$ impliziert.
 - Falls andererseits $P(X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1}) > 0$, dann gilt

$$\begin{aligned} P(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) \\ &= P(X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1}) P(X_k = i_k \mid X_{k-1} = i_{k-1}, \dots, X_0 = i_0) \\ &= \alpha_{i_0} p_{i_0 i_1} \cdots p_{i_{k-2} i_{k-1}} p_{i_{k-1} i_k}. \end{aligned}$$

- Hieraus folgt, dass (3) auch für $n = k$ und somit für jedes $n \in \mathbb{N}$ gilt. □

Korollar 2.1 Sei $\{X_n\}$ eine Markov-Kette. Dann gilt

$$P(X_n = i_n \mid X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_n = i_n \mid X_{n-1} = i_{n-1}), \quad (5)$$

falls $P(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$.

Beweis

- Sei $P(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$ und somit auch $P(X_{n-1} = i_{n-1}) > 0$.
- Aus (3) ergibt sich dann, dass

$$\begin{aligned} P(X_n = i_n \mid X_{n-1} = i_{n-1}) &= \frac{P(X_n = i_n, X_{n-1} = i_{n-1})}{P(X_{n-1} = i_{n-1})} \\ &= \frac{\sum_{i_0, \dots, i_{n-2} \in E} P(X_n = i_n, \dots, X_0 = i_0)}{\sum_{i_0, \dots, i_{n-2} \in E} P(X_{n-1} = i_{n-1}, \dots, X_0 = i_0)} \\ &\stackrel{(3)}{=} \frac{\sum_{i_0, \dots, i_{n-2} \in E} \alpha_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} p_{i_{n-1} i_n}}{\sum_{i_0, \dots, i_{n-2} \in E} \alpha_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}}} = p_{i_{n-1} i_n}. \end{aligned}$$

- Hieraus und aus (4) folgt (5). □

Beachte

- Die Aussage von Korollar 2.1 bedeutet, dass
 - die bedingte Verteilung des (zufälligen) Zustandes X_n der Markov-Kette $\{X_n\}$ zum „Zeitpunkt“ n vollständig durch den Zustand $X_{n-1} = i_{n-1}$ zum vorhergehenden Zeitpunkt $n - 1$ bestimmt wird,
 - wobei diese Verteilung *nicht* von den Zuständen $X_{n-2} = i_{n-2}, \dots, X_1 = i_1, X_0 = i_0$ abhängt, die die Markov-Kette in der weiter zurückliegenden Vergangenheit angenommen hat.
- Unmittelbar aus der Definition bedingter Wahrscheinlichkeiten ergibt sich,
 - dass (5) äquivalent ist mit

$$\begin{aligned} P(X_n = i_n, X_{n-2} = i_{n-2}, \dots, X_0 = i_0 \mid X_{n-1} = i_{n-1}) \\ &= P(X_n = i_n \mid X_{n-1} = i_{n-1}) P(X_{n-2} = i_{n-2}, \dots, X_0 = i_0 \mid X_{n-1} = i_{n-1}), \end{aligned} \quad (6)$$

- wobei die bedingte Unabhängigkeitseigenschaft (6) die *Markov-Eigenschaft* von $\{X_n\}$ genannt wird.
- Die Definitionen und Aussagen des Abschnittes 2.1.1 bleiben praktisch unverändert,
 - wenn anstelle des endlichen Zustandsraumes $E = \{1, 2, \dots, \ell\}$ ein (abzählbar) unendlicher Zustandsraum betrachtet wird, beispielsweise die Menge $E = \mathbb{N}$ sämtlicher natürlichen Zahlen bzw. die Menge $E = \{\dots, -1, 0, 1, \dots\}$ der ganzen Zahlen.
 - Es ist lediglich zu beachten, dass dann α bzw. $\mathbf{P}$ unendlich viele Komponenten bzw. Eintragungen haben.

2.1.2 Beispiele

1. Wettervorhersage

(vgl. O. Häggström (2002) *Finite Markov Chains and Algorithmic Applications*. CU Press, Cambridge)

- Wir betrachten zunächst Regionen, in denen sich typischerweise längere Regen- bzw. Trockenperioden abwechseln, wobei Regentage bzw. Sonnentage im Mittel etwa gleich oft vorkommen.
 - Dann ist es relativ einfach, das Wetter des folgenden Tages vorherzusagen, falls dabei nur die beiden „Zustände“ 1 = „Regen“ bzw. 2 = „Sonnenschein“ betrachtet werden.
 - Wenn wir annehmen, dass die Prognose in 75% aller Fälle richtig ist (und zwar unabhängig davon, ob es zum gegenwärtigen Tag regnet oder die Sonne scheint),
 - dann kann die Wettervorhersage durch eine Markov-Kette mit der folgenden Übergangsmatrix modelliert werden:

$$\mathbf{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}. \quad (7)$$

- In Regionen, in denen diese Symmetrie zwischen „Regen“ bzw. „Sonnenschein“ nicht vorliegt, sondern Sonnentage wesentlich öfter als Regentage vorkommen, sollte
 - die Wettervorhersage *nicht* durch die in (7) betrachtete Übergangsmatrix modelliert werden.
 - In diesem Fall könnte beispielsweise die Übergangsmatrix

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 \\ 0.1 & 0.9 \end{pmatrix} \quad (8)$$

ein geeignetes Modell sein.

2. Zufällige Irrfahrten; Risikoprozesse

- Klassische Beispiele für Markov-Ketten sind durch sogenannte *zufällige Irrfahrten* gegeben, die auch in der deutschsprachigen Literatur *Random Walk* genannt werden, wobei das (unbeschränkte) Basismodell auf die folgende Weise gegeben ist.
 - Sei $Z, Z_1, Z_2, \dots : \Omega \rightarrow \mathbb{Z}$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen, die nur Werte in der Menge $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ der ganzen Zahlen annehmen.
 - Sei $X_0 : \Omega \rightarrow \mathbb{Z}$ eine beliebige Zufallsvariable, die von den „Zuwächsen“ $Z_1, Z_2, \dots$ unabhängig ist, und sei

$$X_n = X_{n-1} + Z_n, \quad \forall n \geq 1. \quad (9)$$

- Dann bilden die Zufallsvariablen $X_0, X_1, \dots$ eine Markov-Kette mit dem (abzählbar unendlichen) Zustandsraum $E = \mathbb{Z}$, mit der Anfangsverteilung $\alpha = (\alpha_1, \alpha_2, \dots)^\top$, wobei $\alpha_i = P(X_0 = i)$, und mit den Übergangswahrscheinlichkeiten $p_{ij} = P(Z = j - i)$.

- Beachte

- Durch die in (9) gegebene Markov-Kette kann die *Risikoreserve* von versicherungs- bzw. finanz-technischen Bilanzierungsprozessen modelliert werden, wobei X_0 als (zufällige) Anfangsreserve aufgefasst wird und der „Zuwachs“ Z_n beispielsweise die Differenz $Z_n = a - Z'_n$ aus risikofreien Einnahmen $a > 0$ und zufallsbedingten Ausgaben/Verlusten Z'_n sein kann.
- Ein anderer Spezialfall ist das bereits in Abschnitt WR-1.3 betrachtete Beispiel des zufälligen Gesamtgewinns X_n aus n Roulette-Spielen, bei dem $X_0 = 0$ angenommen wird und
- die Verteilung des zufälligen „Zuwachses“ Z für jedes einzelne Spiel durch $P(Z = i) = 1/2$ für $i = -1, 1$ und $P(Z = i) = 0$ für $i \in \mathbb{Z} \setminus \{-1, 1\}$ gegeben ist.

3. Warteschlangen

- Die Anzahl der Kunden, die vor einer beliebigen, jedoch fest vorgegebenen Kasse eines Supermarktes warten, lässt sich wie folgt durch eine Markov-Kette modellieren.
 - Sei $X_0 = 0$ die Anzahl der Kunden, die bei Öffnung des Supermarktes vor der Kasse warten.
 - Mit Z_n bezeichnen wir die zufällige Anzahl derjenigen Kunden, die sich vor der Kasse anstellen, während der n -te Kunde bedient wird; $n = 1, 2, \dots$
 - Dabei setzen wir voraus, dass die Zufallsvariablen $Z, Z_1, Z_2, \dots : \Omega \rightarrow \{0, 1, \dots\}$ unabhängig und identisch verteilt sind.
- Die rekursiv definierte Folge $X_0, X_1, \dots : \Omega \rightarrow \{0, 1, \dots\}$ von Zufallsvariablen mit

$$X_n = \max\{0, X_{n-1} + Z_n - 1\}, \quad \forall n \geq 1, \quad (10)$$

bildet dann eine Markov-Kette, deren Übergangsmatrix $\mathbf{P} = (p_{ij})$ gegeben ist durch

$$p_{ij} = \begin{cases} P(Z = j + 1 - i), & \text{falls } j + 1 \geq i > 0 \text{ oder } j > i = 0, \\ P(Z = 0) + P(Z = 1), & \text{falls } j = i = 0, \\ 0, & \text{sonst.} \end{cases}$$

- Dabei ist X_n die zufällige Anzahl von Kunden in der Warteschlange, unmittelbar *nachdem* die Bedienung des n -ten Kunden beendet wurde (d.h. *ohne* die Berücksichtigung desjenigen Kunden, mit dessen Bedienung gegebenenfalls gerade begonnen wurde und der die Warteschlange deshalb bereits verlassen hat).

4. Verzweigungsprozesse

- Wir betrachten den Fortpflanzungsprozess einer bestimmten Population, wobei X_n die (zufällige) Gesamtanzahl der Nachkommen in der n -ten Generation sei; $X_0 = 1$.
- Wir nehmen an, dass

$$X_n = \sum_{i=1}^{X_{n-1}} Z_{n,i}, \quad (11)$$

wobei $\{Z_{n,i}, n, i \in \mathbb{N}\}$ eine Familie von unabhängigen und identisch verteilten Zufallsvariablen ist, die ihre Werte in der Menge $E = \{0, 1, \dots\}$ annehmen.

- Die Zufallsvariable $Z_{n,i}$ ist dabei die zufällige Anzahl der (unmittelbaren) Nachkommen des i -ten Individuums der $(n - 1)$ -ten Generation.
- Die Folge $X_0, X_1, \dots : \Omega \rightarrow \{0, 1, \dots\}$ von Zufallsvariablen, die durch $X_0 = 1$ und durch die Rekursionsgleichung (11) gegeben ist, wird *Verzweigungsprozess* genannt.
- Man kann zeigen (vgl. Abschnitt 2.1.3), dass $X_0, X_1, \dots$ eine Markov-Kette ist mit den Übergangswahrscheinlichkeiten

$$p_{ij} = \begin{cases} P\left(\sum_{k=1}^i Z_{1,k} = j\right), & \text{falls } i > 0, \\ 1, & \text{falls } i = j = 0, \\ 0, & \text{sonst.} \end{cases}$$

5. Zyklische zufällige Irrfahrten

- Weitere Beispiele von Markov-Ketten können wie folgt konstruiert werden (vgl. E. Behrends (2000) *Introduction to Markov Chains*. Vieweg, Braunschweig, S.4).

– Wir betrachten den endlichen Zustandsraum $E = \{0, 1, \dots, 999\}$, die Anfangsverteilung

$$\alpha = (1/16, 4/16, 6/16, 4/16, 1/16, 0, \dots, 0)^\top \quad (12)$$

und die Übergangswahrscheinlichkeiten

$$p_{ij} = \begin{cases} \frac{1}{6}, & \text{falls } (j + 1000 - i) \bmod (1000) \in \{1, \dots, 6\}, \\ 0, & \text{sonst.} \end{cases}$$

– Seien $X_0, Z_1, Z_2, \dots : \Omega \rightarrow \{0, 1, \dots, 999\}$ unabhängige Zufallsvariablen, wobei die Verteilung von X_0 durch (12) gegeben sei und

$$P(Z_1 = i) = P(Z_2 = i) = \dots = 1/6, \quad \forall i = 1, \dots, 6.$$

– Die rekursiv definierte Folge $X_0, X_1, \dots : \Omega \rightarrow \{0, 1, \dots, 999\}$ von Zufallsvariablen mit

$$X_n = (X_{n-1} + Z_n) \bmod (1000) \quad (13)$$

für $n \geq 1$ bildet dann eine Markov-Kette, die *zyklischer Random Walk* genannt wird; vgl. auch Übungsaufgabe 1.3.

- Beachte

- Dieses Modell einer Markov-Kette kann auf die folgende Weise *experimentell* realisiert werden: Wir werfen zunächst 4-mal eine Münze und registrieren, wie oft dabei das Ereignis „Zahl“ eintritt. Die Anzahl x_0 dieser Ereignisse fassen wir als Realisierung des zufälligen Anfangszustandes X_0 auf; vgl. das *Bernoulli-Schema* in Abschnitt WR-3.2.1.
- Danach werfen wir n -mal einen Würfel und registrieren die jeweiligen Augenzahlen. Die Augenzahl z_i , die beim i -ten Wurf des Würfels eintritt, fassen wir als Realisierung des zufälligen „Zuwachses“ Z_i auf; $i = 1, \dots, n$.
- Der neue „Systemzustand“ x_n ergibt sich dann durch die „Aktualisierung“ des alten Systemzustandes x_{n-1} gemäß (13), und zwar unter Berücksichtigung des „Zuwachses“ z_{n-1} .
- Wenn das Experiment nicht tatsächlich durch das Werfen einer Münze bzw. eines Würfels, sondern durch die Erzeugung entsprechender *Pseudozufallszahlen* $x_0, z_1, z_2, \dots$ mit dem Computer durchgeführt wird, dann spricht man von *Monte-Carlo-Simulation*.
- Methoden zur Konstruktion von *dynamischen Simulationsalgorithmen*, die auf Markov-Ketten beruhen, werden im zweiten Teil der Vorlesung ausführlich behandelt.

2.1.3 Rekursive Darstellung

- In diesem Abschnitt zeigen wir,

- wie sich Markov-Ketten aus Folgen von unabhängigen und identisch verteilten Zufallsvariablen konstruieren lassen,
- dass somit die Rekursionsformeln (9), (10), (11) und (13) Spezialfälle eines allgemeinen Konstruktionsprinzips für Markov-Ketten sind
- und dass auch umgekehrt jede Markov-Kette als Lösung einer rekursiven stochastischen Gleichung aufgefasst werden kann.

- So wie bisher sei $E = \{1, 2, \dots, \ell\}$ eine endliche (oder abzählbar unendliche) Menge.
 - Außerdem sei $(D, \mathcal{D})$ ein beliebiger Maßraum, beispielsweise $D = \mathbb{R}^d$ der d -dimensionale euklidische Raum und $\mathcal{D} = \mathcal{B}(\mathbb{R}^d)$ die σ -Algebra der Borel-Mengen in $\mathbb{R}^d$, oder $D = [0, 1]$ das Einheitsintervall und $\mathcal{D} = \mathcal{B}([0, 1])$ die σ -Algebra der Borel-Mengen in $[0, 1]$.
 - Sei nun $Z_1, Z_2, \dots : \Omega \rightarrow D$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen, die ihre Werte in D annehmen, und die Zufallsvariable $X_0 : \Omega \rightarrow E$ sei unabhängig von $Z_1, Z_2, \dots$.
 - Die Zufallsvariablen $X_1, X_2, \dots : \Omega \rightarrow E$ seien durch die *stochastische Rekursionsgleichung*

$$X_n = \varphi(X_{n-1}, Z_n) \quad (14)$$

gegeben, wobei $\varphi : E \times D \rightarrow E$ eine beliebige (messbare) Funktion ist.

Theorem 2.2

- Die Zufallsvariablen $X_0, X_1, \dots : \Omega \rightarrow E$ seien rekursiv durch (14) gegeben.
- Dann gilt für beliebige $n \geq 1$ und $i_0, i_1, \dots, i_n \in E$,

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

falls $P(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$.

Beweis

- Aus (14) ergibt sich, dass

$$\begin{aligned} P(X_n = i_n \mid X_{n-1} = i_{n-1}, \dots, X_0 = i_0) &= P(\varphi(X_{n-1}, Z_n) = i_n \mid X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(\varphi(i_{n-1}, Z_n) = i_n \mid X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(\varphi(i_{n-1}, Z_n) = i_n), \end{aligned}$$

- wobei sich die letzte Gleichheit aus dem Transformationssatz für unabhängige Zufallsvariablen ergibt (vgl. Theorem WR-3.18),
- denn die Zufallsvariablen $X_0, \dots, X_{n-1}$ sind Funktionen von $Z_1, \dots, Z_{n-1}$ und somit unabhängig von $\varphi(i_{n-1}, Z_n)$.
- Auf die gleiche Weise ergibt sich, dass

$$\begin{aligned} P(\varphi(i_{n-1}, Z_n) = i_n) &= P(\varphi(i_{n-1}, Z_n) = i_n \mid X_{n-1} = i_{n-1}) \\ &= P(\varphi(X_{n-1}, Z_n) = i_n \mid X_{n-1} = i_{n-1}) \\ &= P(X_n = i_n \mid X_{n-1} = i_{n-1}). \end{aligned}$$

□

Beachte

- Aus dem Beweis von Theorem 2.2 ergibt sich, dass die bedingte Wahrscheinlichkeit

$$p_{ij} = P(X_n = j \mid X_{n-1} = i)$$

gegeben ist durch $p_{ij} = P(\varphi(i, Z_n) = j)$.

- Dabei hängt p_{ij} nicht von n ab, weil die „Innovationen“ Z_n identisch verteilt sind.

- Außerdem ist die (gemeinsame) Wahrscheinlichkeit $P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n)$ gegeben durch

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}, \quad (15)$$

wobei $\alpha_{i_0} = P(X_0 = i_0)$.

- Die durch (14) rekursiv gegebene Folge $X_0, X_1, \dots$ von Zufallsvariablen ist also eine Markov-Kette im Sinne der Definitionsgleichung (3).

Wir zeigen nun, dass sich auch umgekehrt jede Markov-Kette als Lösung einer rekursiven stochastischen Gleichung auffassen lässt.

- Sei $X_0, X_1, \dots : \Omega \rightarrow E$ eine beliebige Markov-Kette mit dem Zustandsraum $E = \{1, 2, \dots, \ell\}$, der Anfangsverteilung $\alpha = (\alpha_1, \dots, \alpha_\ell)^\top$ und der Übergangsmatrix $\mathbf{P} = (p_{ij})$.
- Mit Hilfe einer Rekursionsgleichung, die die gleiche Form wie (14) hat, wird eine Markov-Kette $X'_0, X'_1, \dots$ mit (vorgegebener) Anfangsverteilung α und Übergangsmatrix $\mathbf{P}$ rekursiv konstruiert, so dass

$$P(X_0 = i_0, \dots, X_n = i_n) = P(X'_0 = i_0, \dots, X'_n = i_n), \quad \forall i_0, \dots, i_n \in E \quad (16)$$

für jedes $n \geq 0$ gilt:

1. Wir gehen von einer Folge $Z_0, Z_1, \dots$ von unabhängigen und identisch auf $(0, 1]$ -gleichverteilten Zufallsvariablen aus.
2. Zunächst wird die E -wertige Zufallsvariable X'_0 wie folgt definiert:

$$X'_0 = k \quad \text{genau dann, wenn} \quad Z_0 \in \left(\sum_{i=1}^{k-1} \alpha_i, \sum_{i=1}^k \alpha_i \right],$$

für jedes $k = 1, \dots, \ell$, d.h.,

$$X'_0 = \sum_{k=1}^{\ell} k \mathbb{I} \left(\sum_{i=1}^{k-1} \alpha_i < Z_0 \leq \sum_{i=1}^k \alpha_i \right). \quad (17)$$

3. Die Zufallsvariablen $X'_1, X'_2, \dots$ werden dann rekursiv definiert: Und zwar setzen wir

$$X'_n = \varphi(X'_{n-1}, Z_n), \quad (18)$$

wobei die Funktion $\varphi : E \times (0, 1] \rightarrow E$ gegeben ist durch

$$\varphi(i, z) = \sum_{k=1}^{\ell} k \mathbb{I} \left(\sum_{j=1}^{k-1} p_{ij} < z \leq \sum_{j=1}^k p_{ij} \right). \quad (19)$$

- Man kann sich leicht überlegen, dass die Wahrscheinlichkeiten $P(X'_0 = i_0, X'_1 = i_1, \dots, X'_n = i_n)$ für die in (17)–(18) rekursiv definierte Folge $\{X'_n\}$ durch (3) gegeben sind, d.h., $\{X'_n\}$ ist eine Markov-Kette mit der Anfangsverteilung α und der Übergangsmatrix $\mathbf{P}$, vgl. Übungsaufgabe 2.1.

Beachte

- Folgen von Zufallsvariablen $\{X_i\}$ und $\{X'_i\}$, für die (16) gilt, werden *stochastisch äquivalent* genannt.
- Das in (17)–(19) gegebene Konstruktionsprinzip kann zur Monte-Carlo-Simulation von Markov-Ketten mit vorgegebener Anfangsverteilung und Übergangsmatrix genutzt werden.
- Markov-Ketten mit abzählbar unendlichem Zustandsraum können auf die gleiche Weise konstruiert bzw. simuliert werden, wobei dann in (17)–(19) lediglich die Komponenten bzw. Eintragungen von unendlich dimensional Vektoren α bzw. Matrizen $\mathbf{P}$ betrachtet werden müssen.

2.1.4 Matrix der n -stufigen Übergangswahrscheinlichkeiten

- Sei $X_0, X_1, \dots : \Omega \rightarrow E$ eine Markov-Kette mit dem Zustandsraum $E = \{1, 2, \dots, \ell\}$, der Anfangsverteilung $\alpha = (\alpha_1, \dots, \alpha_\ell)^\top$ und der Übergangsmatrix $\mathbf{P} = (p_{ij})$.
- Für beliebige, jedoch fest vorgegebene $n \geq 1$ und $i, j \in E$ kann das Produkt $p_{ii_1} p_{i_1 i_2} \dots p_{i_{n-1} j}$ als die Wahrscheinlichkeit des „Pfades“ $i \rightarrow i_1 \rightarrow \dots \rightarrow i_{n-1} \rightarrow j$ aufgefasst werden.
- Analog hierzu kann die Summe

$$p_{ij}^{(n)} = \sum_{i_1, \dots, i_{n-1} \in E} p_{ii_1} p_{i_1 i_2} \dots p_{i_{n-1} j} \quad (20)$$

als die Wahrscheinlichkeit des Überganges vom Zustand i in den Zustand j in n Schritten aufgefasst werden, wobei

$$p_{ij}^{(n)} = P(X_n = j \mid X_0 = i), \quad (21)$$

falls $P(X_0 = i) > 0$.

Beachte

- Die Matrix $\mathbf{P}^{(n)} = (p_{ij}^{(n)})_{i,j=1,\dots,\ell}$ wird die n -stufige Übergangsmatrix der Markov-Kette $\{X_n\}$ genannt.
- Mit der Schreibweise $\mathbf{P}^{(0)} = \mathbf{I}$, wobei $\mathbf{I}$ die $\ell \times \ell$ -dimensionale Einheitsmatrix bezeichnet, lassen sich die folgenden Darstellungsformeln für $\mathbf{P}^{(n)}$ angeben.

Lemma 2.1 Für beliebige $n, m = 0, 1, \dots$ gilt

$$\mathbf{P}^{(n)} = \mathbf{P}^n \quad (22)$$

und somit

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \mathbf{P}^{(m)}. \quad (23)$$

Beweis Die Gleichung (22) folgt unmittelbar aus (20) und aus der Definition der Matrix-Multiplikation. $\square$

Beispiel (Wettervorhersage)

- Sei $E = \{1, 2\}$, und sei

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ p' & 1-p' \end{pmatrix}$$

eine beliebige Übergangsmatrix mit $0 < p, p' \leq 1$.

- Man kann zeigen (vgl. Übungsaufgabe 2.3), dass dann die n -stufige Übergangsmatrix $\mathbf{P}^{(n)} = \mathbf{P}^n$ gegeben ist durch

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

Beachte

- Die Matrix-Identität (23) wird in der Literatur *Gleichung von Chapman-Kolmogorov* genannt.
- Unmittelbar aus (23) ergeben sich die folgenden nützlichen Abschätzungen.

Korollar 2.2 Für beliebige $n, m, r = 0, 1, \dots$ und $i, j, k \in E$ gilt

$$p_{ii}^{(n+m)} \geq p_{ij}^{(n)} p_{ji}^{(m)} \quad (24)$$

und

$$p_{ij}^{(r+n+m)} \geq p_{ik}^{(r)} p_{kk}^{(n)} p_{kj}^{(m)}. \quad (25)$$

Außerdem ergibt sich aus Lemma 2.1 die folgende Darstellung der Verteilung des zufälligen Zustandes X_n der Markov-Kette zum Zeitpunkt n .

Theorem 2.3

- Sei $X_0, X_1, \dots$ eine Markov-Kette mit dem Zustandsraum $E = \{1, \dots, \ell\}$, der Anfangsverteilung α und der (einstufigen) Übergangsmatrix $\mathbf{P}$.
- Dann ist der Vektor $\alpha_n = (\alpha_{n1}, \dots, \alpha_{n\ell})^\top$ der (Einzel-) Wahrscheinlichkeiten $\alpha_{ni} = P(X_n = i)$ von X_n gegeben durch

$$\alpha_n^\top = \alpha^\top \mathbf{P}^n. \quad (26)$$

Beweis

- Aus der Formel der totalen Wahrscheinlichkeit (vgl. Theorem WR-2.6) und aus (21) folgt, dass

$$P(X_n = j) = \sum_{i \in E} \alpha_i P(X_n = j \mid X_0 = i) = \sum_{i \in E} \alpha_i p_{ij}^{(n)},$$

wobei wir $P(X_n = j \mid X_0 = i) = 0$ setzen, falls $\alpha_i = P(X_0 = i) = 0$.

- Die Behauptung (26) ergibt sich nun aus Lemma 2.1. □

Beachte

- Wegen Theorem 2.3 kann die Berechnung der Wahrscheinlichkeiten $\alpha_{ni} = P(X_n = i)$ auf die Berechnung der n -ten Potenz $\mathbf{P}^n$ der Übergangsmatrix $\mathbf{P}$ zurückgeführt werden.
- Dabei ist in vielen Fällen die sogenannte *Spektraldarstellung* von $\mathbf{P}^n$ nützlich, die auf die folgende Weise mit Hilfe der Eigenwerte bzw. Eigenvektoren der Übergangsmatrix $\mathbf{P}$ bestimmt werden kann.
- Zur Erinnerung

- Sei $\mathbf{A}$ eine (nicht notwendig stochastische) $\ell \times \ell$ Matrix, seien $\phi, \psi \neq \mathbf{0}$ zwei ℓ -dimensionale (Spalten-) Vektoren, so daß jeweils mindestens eine ihrer Komponenten von 0 verschieden ist, und sei θ eine beliebige (reelle oder komplexe) Zahl.

- Falls

$$\mathbf{A}\phi = \theta\phi \quad \text{bzw.} \quad \psi^\top \mathbf{A} = \theta\psi^\top, \quad (27)$$

dann ist θ ein *Eigenwert* von $\mathbf{A}$, und ϕ bzw. ψ ist ein rechter bzw. linker (zu θ gehörender) *Eigenvektor*.

- Weil (27) und

$$(\mathbf{A} - \theta\mathbf{I})\phi = \mathbf{0} \quad \text{bzw.} \quad \psi^\top (\mathbf{A} - \theta\mathbf{I}) = \mathbf{0}^\top,$$

äquivalent sind, ist θ genau dann ein Eigenwert von $\mathbf{A}$, wenn θ eine Lösung der sogenannten *charakteristischen Gleichung* ist:

$$\det(\mathbf{A} - \theta\mathbf{I}) = 0. \quad (28)$$

- Weil (28) eine algebraische Gleichung der Ordnung ℓ ist, besitzt sie somit ℓ Lösungen $\theta_1, \dots, \theta_\ell$, die komplex sein können und die nicht alle voneinander verschieden sein müssen.
- Wir nehmen o.B.d.A. an, dass die Eigenwerte $\theta_1, \dots, \theta_\ell$ so numeriert sind, dass

$$|\theta_1| \geq |\theta_2| \geq \dots \geq |\theta_\ell|.$$

- Für jeden Eigenwert θ_i existieren (jeweils von $\mathbf{0}$ verschiedene) rechte bzw. linke Eigenvektoren ϕ_i bzw. ψ_i .
- Sei $\Phi = (\phi_1, \dots, \phi_\ell)$ die $\ell \times \ell$ Matrix, die aus den rechten Eigenvektoren $\phi_1, \dots, \phi_\ell$ besteht, und sei

$$\Psi = \begin{pmatrix} \psi_1^\top \\ \vdots \\ \psi_\ell^\top \end{pmatrix}$$

die $\ell \times \ell$ Matrix, die aus den linken Eigenvektoren $\psi_1, \dots, \psi_\ell$ gebildet wird.

- Mit dieser Schreibweise ergibt sich, dass

$$\mathbf{A}\Phi = \Phi \operatorname{diag}(\theta), \quad (29)$$

wobei $\theta = (\theta_1, \dots, \theta_\ell)^\top$ und $\operatorname{diag}(\theta)$ die Diagonalmatrix mit den Diagonalelementen $\theta_1, \dots, \theta_\ell$ bezeichnet.

- Falls die Eigenvektoren $\phi_1, \dots, \phi_\ell$ linear unabhängig sind,
 - dann existiert die inverse Matrix Φ^{-1} , und wir können $\Psi = \Phi^{-1}$ setzen.
 - Außerdem ergibt sich in diesem Fall aus (29), dass

$$\mathbf{A} = \Phi \operatorname{diag}(\theta) \Phi^{-1} = \Phi \operatorname{diag}(\theta) \Psi$$

und somit

$$\mathbf{A}^n = \Phi (\operatorname{diag}(\theta))^n \Phi^{-1} = \Phi (\operatorname{diag}(\theta))^n \Psi.$$

- Hieraus ergibt sich die *Spektraldarstellung* von $\mathbf{A}$:

$$\mathbf{A}^n = \sum_{i=1}^{\ell} \theta_i^n \phi_i \psi_i^\top. \quad (30)$$

Beachte

- Die Anwendung von (30) für die Übergangsmatrix $\mathbf{A} = \mathbf{P}$ führt zu einem einfachen Algorithmus zur Berechnung der n -ten Potenz $\mathbf{P}^n$ in (26).
 - Dabei sind lediglich die Eigenwerte und Eigenvektoren von $\mathbf{P}$ zu berechnen, wofür Standard-Software wie MAPLE, MATLAB oder MATHEMATICA verwendet werden kann.
 - Ein wesentlicher Vorteil der Spektraldarstellung (30) besteht darin, dass sich bei der Verwendung von (30) die Komplexität der numerischen Berechnung von $\mathbf{P}^n$ mit wachsendem n *nicht* erhöht.
- Allerdings wird bei der Herleitung von (30) vorausgesetzt, dass die Eigenvektoren $\phi_1, \dots, \phi_\ell$ linear unabhängig sind. Das folgende Lemma liefert eine hinreichende Bedingung hierfür.

Lemma 2.2

- Falls sämtliche Eigenwerte $\theta_1, \dots, \theta_\ell$ von $\mathbf{A}$ voneinander verschieden sind, dann sind die rechten Eigenvektoren $\phi_1, \dots, \phi_\ell$ linear unabhängig.

- Wenn darüber hinaus die linken Eigenvektoren $\psi_1, \dots, \psi_\ell$ durch den Ansatz $\Psi = \Phi^{-1}$ gegeben sind, dann gilt

$$\psi_i^\top \phi_j = \begin{cases} 1, & \text{falls } i = j, \\ 0, & \text{falls } i \neq j. \end{cases} \quad (31)$$

Beweis

- Die Gültigkeit der ersten Teilaussage zeigen wir mit vollständiger Induktion.
 - Weil jeder Eigenvektor ϕ_1 mindestens eine Komponente hat, die von 0 verschieden ist, impliziert $a_1 \phi_1 = \mathbf{0}$, dass $a_1 = 0$.
 - Seien nun sämtliche Eigenwerte $\theta_1, \dots, \theta_\ell$ von $\mathbf{A}$ voneinander verschieden, und für ein gewisses $k \leq \ell$ seien die Eigenvektoren $\phi_1, \dots, \phi_{k-1}$ linear unabhängig.
 - Um zu beweisen, dass dann auch die Eigenvektoren $\phi_1, \dots, \phi_k$ linear unabhängig sind, genügt es zu zeigen, dass

$$\sum_{j=1}^k a_j \phi_j = \mathbf{0} \quad (32)$$

die Gültigkeit von $a_1 = \dots = a_k = 0$ impliziert.

- Die Koeffizienten $a_1, \dots, a_k$ seien nun so gewählt, dass (32) gilt. Dann gilt auch

$$\mathbf{0} = \mathbf{A} \mathbf{0} = \sum_{j=1}^k a_j \mathbf{A} \phi_j = \sum_{j=1}^k a_j \theta_j \phi_j.$$

- Auf die gleiche Weise ergibt sich, dass

$$\mathbf{0} = \theta_k \mathbf{0} = \theta_k \sum_{j=1}^k a_j \phi_j = \sum_{j=1}^k \theta_k a_j \phi_j$$

und somit

$$\mathbf{0} = \sum_{j=1}^{k-1} (\theta_k - \theta_j) a_j \phi_j.$$

- Weil angenommen wird, dass die Eigenvektoren $\phi_1, \dots, \phi_{k-1}$ linear unabhängig sind, folgt hieraus, dass

$$(\theta_k - \theta_1) a_1 = (\theta_k - \theta_2) a_2 = \dots = (\theta_k - \theta_{k-1}) a_{k-1} = 0$$

bzw. $a_1 = a_2 = \dots = a_{k-1} = 0$, weil $\theta_k \neq \theta_j$ für $1 \leq j \leq k-1$.

- Wegen (32) muss dann auch $a_k = 0$ gelten.

- Falls sämtliche Eigenwerte $\theta_1, \dots, \theta_\ell$ von $\mathbf{A}$ voneinander verschieden sind,
 - dann besteht also die $\ell \times \ell$ Matrix Φ aus ℓ linear unabhängigen Spaltenvektoren,
 - und Φ ist somit invertierbar.
 - Die Matrix Ψ der linken Eigenvektoren kann also durch den Ansatz $\Psi = \Phi^{-1}$ gebildet werden, woraus sich unmittelbar die Gültigkeit von (31) ergibt. $\square$

2.2 Ergodizität und Stationarität

2.2.1 Grundlegende Definitionen und quasi-positive Übergangsmatrizen

- Wenn die Anzahl ℓ der potentiell möglichen Zustände der Markov-Kette $X_0, X_1, \dots$ sehr groß ist, dann ist die in Abschnitt 2.1.4 diskutierte Spektraldarstellung (30) der n -stufigen Übergangsmatrix $\mathbf{P}^{(n)} = \mathbf{P}^n$ kein geeignetes Hilfsmittel, um

- die bedingten (Einzel-) Wahrscheinlichkeiten $p_{ij}^{(n)} = P(X_n = j \mid X_0 = i)$ des zufälligen Zustandes X_n
- bzw. die (unbedingten) Wahrscheinlichkeiten $P(X_n = j) = \sum_{i=1}^{\ell} \alpha_i p_{ij}^{(n)}$ von X_n

nach $n \gg 1$ (Zeit-) Schritten zu berechnen.

- Man kann jedoch Bedingungen angeben,
 - unter denen die Grenzwerte $\lim_{n \rightarrow \infty} p_{ij}^{(n)}$ bzw. $\lim_{n \rightarrow \infty} P(X_n = j)$ existieren (gleich sind und nicht von i abhängen),
 - so dass anstelle von $p_{ij}^{(n)}$ bzw. $P(X_n = j)$ der Grenzwert $\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)} = \lim_{n \rightarrow \infty} P(X_n = j)$ als *Näherungslösung* berechnet werden kann, falls $n \gg 1$.

Dies führt zu dem folgenden Begriff der *Ergodizität* von Markov-Ketten.

Definition Die Markov-Kette $X_0, X_1, \dots$ mit der Übergangsmatrix $\mathbf{P} = (p_{ij})$ bzw. den zugehörigen n -stufigen Übergangsmatrizen $\mathbf{P}^{(n)} = (p_{ij}^{(n)})$ ($= \mathbf{P}^n$) heißt *ergodisch*, falls die Grenzwerte

$$\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)} \quad (33)$$

1. für jedes $j \in E$ existieren,
2. positiv sind und nicht von $i \in E$ abhängen,
3. eine Wahrscheinlichkeitsfunktion $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{\ell})^{\top}$ bilden, d.h., $\sum_{j \in E} \pi_j = 1$.

Beispiel (*Wettervorhersage*)

- Um den Begriff der Ergodizität von Markov-Ketten zu verdeutlichen, kehren wir zunächst zu dem einfachen Beispiel der Wettervorhersage zurück, das bereits in den Abschnitten 2.1.2 und 2.1.4 diskutiert wurde.
- Sei also $E = \{1, 2\}$, und sei

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ p' & 1-p' \end{pmatrix}$$

eine beliebige Übergangsmatrix mit $0 < p, p' \leq 1$.

- Die n -stufige Übergangsmatrix $\mathbf{P}^{(n)} = \mathbf{P}^n$ ist gegeben durch

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

- Falls $p + p' < 2$, dann ergibt sich hieraus und aus (26), dass

$$\lim_{n \rightarrow \infty} \mathbf{P}^n = \frac{1}{p+p'} \begin{pmatrix} p' & p \\ p' & p \end{pmatrix}$$

bzw.

$$\boldsymbol{\pi} = \lim_{n \rightarrow \infty} \boldsymbol{\alpha}_n = \left(\frac{p'}{p+p'}, \frac{p}{p+p'} \right)^{\top}, \quad (34)$$

wobei die Grenzverteilung $\boldsymbol{\pi}$ in (34) *nicht* von der Wahl der Anfangsverteilung $\boldsymbol{\alpha}$ ($= \boldsymbol{\alpha}_0$) abhängt.

- Wenn jedoch $p + p' = 2$, dann gilt

$$\mathbf{P}^n = \begin{cases} \mathbf{P}, & \text{falls } n \text{ ungerade,} \\ \mathbf{I}, & \text{falls } n \text{ gerade ist.} \end{cases}$$

Die Ergodizität von Markov-Ketten mit allgemeinem (endlichen) Zustandsraum lässt sich mit Hilfe des folgenden Begriffes aus der Theorie positiver Matrizen charakterisieren.

Definition

- Die $\ell \times \ell$ Matrix $\mathbf{A} = (a_{ij})$ heißt *nichtnegativ*, falls sämtliche Eintragungen a_{ij} von $\mathbf{A}$ nichtnegativ sind.
- Die nichtnegative Matrix $\mathbf{A}$ heißt *quasi-positiv*, falls es eine natürliche Zahl $n_0 \geq 1$ gibt, so dass sämtliche Eintragungen von $\mathbf{A}^{n_0}$ positiv sind.

Beachte Falls $\mathbf{A}$ eine stochastische Matrix ist, für die es eine natürliche Zahl $n_0 \geq 1$ gibt, so dass sämtliche Eintragungen von $\mathbf{A}^{n_0}$ positiv sind, dann sind auch sämtliche Eintragungen von $\mathbf{A}^n$ für *jede* natürliche Zahl $n \geq n_0$ positiv.

Theorem 2.4 Die Markov-Kette $X_0, X_1, \dots$ mit dem Zustandsraum $E = \{1, \dots, \ell\}$ und der Übergangsmatrix $\mathbf{P}$ ist genau dann ergodisch, wenn $\mathbf{P}$ quasi-positiv ist.

Beweis

- Wir zeigen zunächst, dass die Bedingung

$$\min_{i,j \in E} p_{ij}^{(n_0)} > 0 \quad (35)$$

für ein $n_0 \in \mathbb{N}$ hinreichend für die Ergodizität von $\{X_n\}$ ist.

- Sei $m_j^{(n)} = \min_{i \in E} p_{ij}^{(n)}$ und $M_j^{(n)} = \max_{i \in E} p_{ij}^{(n)}$. Aus der Chapman-Kolmogorov-Gleichung (23) ergibt sich dann, dass

$$p_{ij}^{(n+1)} = \sum_{k \in E} p_{ik} p_{kj}^{(n)}$$

und somit

$$m_j^{(n+1)} = \min_i p_{ij}^{(n+1)} = \min_i \sum_k p_{ik} p_{kj}^{(n)} \geq \min_i \sum_k p_{ik} \min_k p_{kj}^{(n)} = m_j^{(n)},$$

d.h., $m_j^{(n)} \leq m_j^{(n+1)}$ für jedes $n \geq 0$, wobei wir $\mathbf{P}^{(0)} = \mathbf{I}$ setzen. Auf die gleiche Weise ergibt sich, dass $M_j^{(n)} \geq M_j^{(n+1)}$ für jedes $n \geq 0$ gilt.

- Um die Existenz der Grenzwerte π_j in (33) zu beweisen, genügt es also zu zeigen, dass für jedes $j \in E$

$$\lim_{n \rightarrow \infty} (M_j^{(n)} - m_j^{(n)}) = 0. \quad (36)$$

- Hierfür betrachten wir für beliebige, jedoch fest vorgegebene Zustände $i_0, j_0 \in E$ die Mengen $E' = \{k \in E : p_{i_0 k}^{(n_0)} \geq p_{j_0 k}^{(n_0)}\}$ und $E'' = E \setminus E'$.

- Sei $a = \min_{i,j \in E} p_{ij}^{(n_0)} > 0$. Dann gilt

$$\sum_{k \in E'} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) = 1 - \sum_{k \in E''} p_{i_0 k}^{(n_0)} - \sum_{k \in E'} p_{j_0 k}^{(n_0)} \leq 1 - \ell a$$

und

$$\sum_{k \in E''} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) = - \sum_{k \in E'} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) .$$

- Durch die erneute Anwendung der Chapman-Kolmogorov-Gleichung (23) ergibt sich hieraus, dass für beliebige $n \geq 0$ und $j \in E$

$$\begin{aligned} p_{i_0 j}^{(n_0+n)} - p_{j_0 j}^{(n_0+n)} &= \sum_{k \in E} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) p_{kj}^{(n)} \\ &= \sum_{k \in E'} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) p_{kj}^{(n)} + \sum_{k \in E''} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) p_{kj}^{(n)} \\ &\leq \sum_{k \in E'} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) M_j^{(n)} + \sum_{k \in E''} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) m_j^{(n)} \\ &= \sum_{k \in E'} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) M_j^{(n)} - \sum_{k \in E'} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) m_j^{(n)} \\ &= \sum_{k \in E'} (p_{i_0 k}^{(n_0)} - p_{j_0 k}^{(n_0)}) (M_j^{(n)} - m_j^{(n)}) \\ &\leq (1 - \ell a) (M_j^{(n)} - m_j^{(n)}) . \end{aligned}$$

- Hieraus folgt, dass $M_j^{(n_0+n)} - m_j^{(n_0+n)} \leq (M_j^{(n)} - m_j^{(n)})(1 - \ell a)$ und (mit vollständiger Induktion) dass für jedes $k \geq 1$

$$M_j^{(kn_0+n)} - m_j^{(kn_0+n)} \leq (M_j^{(n)} - m_j^{(n)})(1 - \ell a)^k . \quad (37)$$

- Es gibt also eine (unbegrenzt wachsende) Folge $n_1, n_2, \dots$ von natürlichen Zahlen, so dass für jedes $j \in E$

$$\lim_{k \rightarrow \infty} (M_j^{(n_k)} - m_j^{(n_k)}) = 0 . \quad (38)$$

- Weil die Differenzen $M_j^{(n)} - m_j^{(n)}$ monoton in n sind, gilt (38) für *jede* (unbegrenzt wachsende) Folge $n_1, n_2, \dots$ von natürlichen Zahlen.
- Damit ist die Gültigkeit von (36) bewiesen.

- Die Grenzwerte π_j sind positiv, weil

$$\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)} \geq \lim_{n \rightarrow \infty} m_j^{(n)} \geq m_j^{(n_0)} \geq a > 0 .$$

- Außerdem gilt $\sum_{j \in E} \pi_j = \sum_{j \in E} \lim_{n \rightarrow \infty} p_{ij}^{(n)} = \lim_{n \rightarrow \infty} \sum_{j \in E} p_{ij}^{(n)} = 1$, weil die Vertauschung von Grenzwert und *endlicher* Summation erlaubt ist.
- Die Notwendigkeit der Bedingung (35) ergibt sich unmittelbar aus $\min_{j \in E} \pi_j > 0$ und (33), wenn dabei berücksichtigt wird, dass der Zustandsraum E endlich ist. $\square$

Beachte

- Weil die Grenzwerte $\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ ergodischer Markov-Ketten nicht von i abhängen, ergibt sich hieraus und aus der Endlichkeit des Zustandsraumes $E = \{1, \dots, \ell\}$, dass

$$\lim_{n \rightarrow \infty} \alpha_n^\top = \alpha^\top \lim_{n \rightarrow \infty} \mathbf{P}^{(n)} = \pi^\top .$$

- Im Beweis von Theorem 2.4 wurde nicht nur die Existenz der Grenzwerte $\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ gezeigt, sondern die folgende *Abschätzung der Konvergenzgeschwindigkeit* hergeleitet: Aus (37) ergibt sich nämlich, dass

$$\sup_{i,j \in E} |p_{ij}^{(n)} - \pi_j| \leq \sup_{j \in E} (M_j^{(n)} - m_j^{(n)}) \leq (1 - \ell a)^{\lfloor n/n_0 \rfloor} \quad (39)$$

bzw.

$$\sup_{j \in E} |\alpha_{nj} - \pi_j| \leq (1 - \ell a)^{\lfloor n/n_0 \rfloor}, \quad (40)$$

wobei $\lfloor n/n_0 \rfloor$ den ganzzahligen Teil von n/n_0 bezeichnet.

- Im Zusammenhang mit (39) und (40) spricht man in der Literatur von *geometrischen Schranken* der Konvergenzgeschwindigkeit.

Wir zeigen nun noch, dass die Grenzwerte $\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ auch als Lösung eines linearen Gleichungssystems aufgefasst werden können.

Theorem 2.5

- Sei $X_0, X_1, \dots$ eine ergodische Markov-Kette mit dem Zustandsraum $E = \{1, \dots, \ell\}$ und der Übergangsmatrix $\mathbf{P} = (p_{ij})$.
- Dann ist der Vektor $\boldsymbol{\pi} = (\pi_1, \dots, \pi_\ell)^\top$ der Grenzwerte $\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ die eindeutig bestimmte (positive) Lösung des linearen Gleichungssystems

$$\pi_j = \sum_{i \in E} \pi_i p_{ij}, \quad j \in E, \quad (41)$$

die gleichzeitig der Normierungsbedingung $\sum_{j \in E} \pi_j = 1$ genügt.

Beweis

- Aus der Definitionsgleichung (33) der Grenzwerte π_j und aus der Chapman-Kolmogorov-Gleichung (23) ergibt sich durch Vertauschen von Grenzwert und Summation, dass

$$\pi_j \stackrel{(33)}{=} \lim_{n \rightarrow \infty} p_{kj}^{(n)} \stackrel{(23)}{=} \lim_{n \rightarrow \infty} \sum_{i \in E} p_{ki}^{(n-1)} p_{ij} = \sum_{i \in E} \lim_{n \rightarrow \infty} p_{ki}^{(n-1)} p_{ij} \stackrel{(33)}{=} \sum_{i \in E} \pi_i p_{ij}.$$

- Wir nehmen nun an, dass es noch eine weitere positive Lösung $\boldsymbol{\pi}' = (\pi'_1, \dots, \pi'_\ell)^\top$ von (41) gibt, so dass $\pi'_j = \sum_{i \in E} \pi'_i p_{ij}$ für jedes $j \in E$ und $\sum_{j \in E} \pi'_j = 1$.
- Mit vollständiger Induktion kann man dann zeigen, dass

$$\pi'_j = \sum_{i \in E} \pi'_i p_{ij}^{(n)}, \quad j \in E, \quad (42)$$

für jedes $n = 1, 2, \dots$

- Insbesondere ergibt sich somit aus (42), dass

$$\pi'_j \stackrel{(42)}{=} \lim_{n \rightarrow \infty} \sum_{i \in E} \pi'_i p_{ij}^{(n)} = \sum_{i \in E} \pi'_i \lim_{n \rightarrow \infty} p_{ij}^{(n)} \stackrel{(33)}{=} \pi_j. \quad \square$$

Beachte

- In Matrix-Schreibweise hat das lineare Gleichungssystem (41) die Form $\boldsymbol{\pi}^\top = \boldsymbol{\pi}^\top \mathbf{P}$.
- Wenn die Anzahl ℓ der Zustände nicht zu groß ist, dann kann dieses Gleichungssystem ein Ausgangspunkt zur numerischen Bestimmung der Wahrscheinlichkeitsfunktion $\boldsymbol{\pi}$ sein; vgl. Abschnitt 2.2.5.
- Falls $\ell \gg 1$, dann ist in vielen Fällen dynamische Monte-Carlo-Simulation besser zur Bestimmung von $\boldsymbol{\pi}$ geeignet; vgl. Abschnitt 3.3.

2.2.2 Abschätzung der Konvergenzgeschwindigkeit; Perron–Frobenius–Theorem

- Zur Erinnerung
 - Falls $\{X_n\}$ eine Markov-Kette ist, so dass sämtliche Eintragungen p_{ij} der (1-stufigen) Übergangsmatrix $\mathbf{P}$ positiv sind,
 - dann hat die in (40) hergeleitete geometrische Schranke der Konvergenzgeschwindigkeit gegen die Grenzverteilung $\boldsymbol{\pi} = (\pi_1, \dots, \pi_\ell)^\top$ die folgende Form:

$$\sup_{j \in E} |\alpha_{nj} - \pi_j| = O((1 - \ell a)^n), \quad (43)$$

wobei $a = \min_{i,j \in E} p_{ij} > 0$.

- Wenn das Minimum a der Eintragungen p_{ij} der Übergangsmatrix $\mathbf{P}$ nahe bei 0 liegt, dann ist die Schranke in (43) nur von geringem Nutzen.
- In manchen Fällen lässt sich jedoch die „Basis“ $1 - \ell a$ der Konvergenzabschätzung in (43) verbessern.

Beispiel (Wettervorhersage)

- Sei $E = \{1, 2\}$ und

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ p' & 1-p' \end{pmatrix}, \quad \text{wobei } 0 < p, p' < 1.$$

- In Abschnitt 2.2.1 hatten wir gezeigt,
 - dass dann die n -stufige Übergangsmatrix $\mathbf{P}^{(n)} = \mathbf{P}^n$ gegeben ist durch

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

- und dass somit

$$(\mathbf{P}^\infty =) \lim_{n \rightarrow \infty} \mathbf{P}^n = \frac{1}{p+p'} \begin{pmatrix} p' & p \\ p' & p \end{pmatrix}.$$

- In diesem Fall ergibt sich also, dass

$$\mathbf{P}^n - \mathbf{P}^\infty = \frac{(1-p-p')^n}{p+p'} \begin{pmatrix} p & -p \\ -p' & p' \end{pmatrix} \quad (= O(|1-p-p'|^n)), \quad (44)$$

wobei $p+p' > 2a = 2 \min_{i,j \in E} p_{ij}$ und somit $|1-p-p'| < 1-2a$, falls $p \neq p'$.

- Beachte
 - Die „Basis“ $|\theta_2| = |1-p-p'|$ der Konvergenzgeschwindigkeit in (44) ist der Betrag des „zweitgrößten“ Eigenwertes θ_2 der Übergangsmatrix $\mathbf{P}$,
 - denn die charakteristische Gleichung

$$(\det(\mathbf{P} - \theta \mathbf{I}) =) (1-p-\theta)(1-p'-\theta) - pp' = 0$$

von $\mathbf{P}$ besitzt die beiden Lösungen $\theta_1 = 1$ und $\theta_2 = 1-p-p'$.

Im allgemeinen lassen sich geometrische Konvergenzabschätzungen der Form (44) mit Hilfe des folgenden *Theorems von Perron–Frobenius* für quasi-positive Matrizen herleiten.

Theorem 2.6

- Sei $\mathbf{A}$ eine quasi-positive $\ell \times \ell$ Matrix mit den Eigenwerten $\theta_1, \dots, \theta_\ell$, so dass $|\theta_1| \geq \dots \geq |\theta_\ell|$.
- Dann gilt:
 - (a) Der Eigenwert θ_1 ist reellwertig und positiv,
 - (b) $\theta_1 > |\theta_i|$ für jedes $i = 2, \dots, \ell$,
 - (c) die rechten bzw. linken (zu θ_1 gehörenden) Eigenvektoren ϕ_1 bzw. ψ_1 sind (bis auf einen konstanten Faktor) eindeutig bestimmt und können so gewählt werden, dass sämtliche Komponenten von ϕ_1 bzw. ψ_1 positiv sind.

Einen Beweis von Theorem 2.6 kann man beispielsweise in Kapitel 1 des Buches E. Seneta (1981) *Non-Negative Matrices and Markov Chains*, Springer-Verlag, New York finden.

Korollar 2.3 Sei $\mathbf{P}$ eine quasi-positive Übergangsmatrix. Dann gilt

- $\theta_1 = 1, \phi_1 = \mathbf{e}$ und $\psi_1 = \boldsymbol{\pi}$, wobei $\mathbf{e} = (1, \dots, 1)^\top$ und $\boldsymbol{\pi} = (\pi_1, \dots, \pi_\ell)^\top$.
- $|\theta_i| < 1$ für jedes $i = 2, \dots, \ell$.

Beweis

- Weil $\mathbf{P}$ eine stochastische Matrix ist,
 - gilt offenbar $\mathbf{P}\mathbf{e} = \mathbf{e}$, und aus (41) ergibt sich, dass $\boldsymbol{\pi}^\top \mathbf{P} = \boldsymbol{\pi}^\top$.
 - Hieraus folgt, dass 1 ein Eigenwert von $\mathbf{P}$ ist und dass $\mathbf{e}$ bzw. $\boldsymbol{\pi}$ ein rechter bzw. linker Eigenvektor dieses Eigenwertes ist.
- Sei nun
 - θ ein beliebiger Eigenwert von $\mathbf{P}$, und sei $\phi = (\phi_1, \dots, \phi_\ell)^\top$ ein zu θ gehörender Eigenvektor.
 - Dann ergibt sich aus der Definitionsgleichung (27) von θ und ϕ , dass

$$|\theta| |\phi_i| \leq \sum_{j=1}^{\ell} p_{ij} |\phi_j| \leq \max_{j \in E} |\phi_j|, \quad \forall i \in E.$$

- Hieraus folgt, dass $|\theta| \leq 1$ und dass somit $\theta_1 = 1$ der größte Eigenwert von $\mathbf{P}$ ist.
- Aus Theorem 2.6 folgt nun, dass $|\theta_i| < 1$ für jedes $i = 2, \dots, \ell$. □

Aus Korollar 2.3 ergibt sich die folgende geometrische Konvergenzabschätzung.

Korollar 2.4 Sei $\mathbf{P}$ eine quasi-positive Übergangsmatrix, so dass sämtliche Eigenwerte $\theta_1, \dots, \theta_\ell$ von $\mathbf{P}$ voneinander verschieden sind. Dann gilt

$$\sup_{j \in E} |\alpha_{nj} - \pi_j| = O(|\theta_2|^n). \quad (45)$$

Beweis

- Aus Korollar 2.3 ergibt sich, dass

$$\lim_{n \rightarrow \infty} \sum_{i=2}^{\ell} \theta_i^n \phi_i \psi_i^\top = \mathbf{0}, \quad (46)$$

weil $|\theta_i| < 1$ für jedes $i = 2, \dots, \ell$.

- Außerdem ergibt sich aus Korollar 2.3, dass $\theta_1 = 1$ und $\phi_1 = (1, \dots, 1)^\top$ bzw. $\psi_1 = \pi$ für den rechten bzw. linken (zu θ_1 gehörenden) Eigenvektor ϕ_1 bzw. ψ_1 .
- Aus der Spektraldarstellung (30) von $\mathbf{P}^n$, d.h.

$$\mathbf{P}^n = \sum_{i=1}^{\ell} \theta_i^n \phi_i \psi_i^\top,$$

folgt nun, dass

$$\mathbf{P}^n - \begin{pmatrix} \pi^\top \\ \vdots \\ \pi^\top \end{pmatrix} = \mathbf{P}^n - \theta_1^n \phi_1 \psi_1^\top = \sum_{i=2}^{\ell} \theta_i^n \phi_i \psi_i^\top.$$

- Weil $\alpha_n^\top = \alpha^\top \mathbf{P}^n$ (vgl. Theorem 2.3), ergibt sich hieraus und aus (46) die Gültigkeit von (45). $\square$

Beispiel (Konsensbildung)

vgl. C. Hesse (2003) *Angewandte Wahrscheinlichkeitstheorie*. Vieweg-Verlag, Braunschweig, S. 349

- Eine Experten-Kommission, die aus ℓ Mitgliedern besteht, soll eine Prognose für einen (Wirtschafts-) Parameter $\mu \in \mathbb{R}$ erstellen, beispielsweise der „Rat der Fünf Weisen“ eine Prognose für das Wirtschaftswachstum im kommenden Jahr.
 - Dabei entwickelt zunächst jeder der ℓ Experten eine eigene Prognose für μ , wobei die einzelnen Prognosewerte mit $\hat{\mu}_1^{(0)}, \dots, \hat{\mu}_\ell^{(0)}$ bezeichnet werden und typischerweise voneinander verschieden sind.
 - Wie können nun die Experten hieraus eine *gemeinsame* Prognose der gesamten Kommission, d.h., einen *Konsens* bilden?
 - Eine einfache Vorgehensweise wäre die Bildung des arithmetischen Mittels $(\hat{\mu}_1^{(0)} + \dots + \hat{\mu}_\ell^{(0)})/\ell$, wobei jedoch die Kompetenzunterschiede zwischen den einzelnen Experten unberücksichtigt bleiben würden.
- Eine alternative Vorgehensweise besteht darin, dass jeder Experte seine eigene Prognose nach Kenntnisnahme der Prognosen der anderen $\ell - 1$ Experten und unter Berücksichtigung seiner Einschätzung der prognostischen Fähigkeiten der Kommissionsmitglieder modifiziert.
 - Für beliebige $i, j \in \{1, \dots, \ell\}$ ordnet dabei der i -te Experte dem j -ten Experten die „Vertrauenswahrscheinlichkeit“ p_{ij} zu, so dass

$$p_{ij} > 0 \quad \text{und} \quad \sum_{j=1}^{\ell} p_{ij} = 1, \quad \forall i, j \in \{1, \dots, \ell\}$$

- und der i -te Experte modifiziert seine ursprüngliche Prognose $\hat{\mu}_i^{(0)}$ zu

$$\hat{\mu}_i^{(1)} = \sum_{j=1}^{\ell} p_{ij} \hat{\mu}_j^{(0)}.$$

- In den meisten Fällen werden die modifizierten Prognosen $\hat{\mu}_1^{(1)}, \dots, \hat{\mu}_\ell^{(1)}$ immer noch voneinander verschieden sein, weshalb das Verfahren solange wiederholt wird, bis die Unterschiede hinreichend klein geworden sind.
- Aus Theorem 2.4 folgt,
 - dass dieses Ziel erreichbar ist, wenn das Modifikationsverfahren oft genug wiederholt wird,

– denn gemäß Theorem 2.4 existieren die Grenzwerte

$$\lim_{n \rightarrow \infty} \hat{\mu}_i^{(n)} = \sum_{j=1}^{\ell} \pi_j \hat{\mu}_j^{(0)} \quad (47)$$

und hängen nicht von i ab,

– wobei der Vektor $\boldsymbol{\pi} = (\pi_1, \dots, \pi_\ell)^\top$ der Grenzwerte $\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ die (eindeutig bestimmte) Lösung des linearen Gleichungssystems (41) ist, d.h., es gilt $\boldsymbol{\pi}^\top = \boldsymbol{\pi}^\top \mathbf{P}$ mit $\mathbf{P} = (p_{ij})$.

- Die Gültigkeit von (47) lässt sich dabei wie folgt begründen. Es gilt nämlich

$$\begin{aligned} \lim_{n \rightarrow \infty} \hat{\mu}_i^{(n)} &= \lim_{n \rightarrow \infty} \sum_{j=1}^{\ell} p_{ij} \hat{\mu}_j^{(n-1)} = \lim_{n \rightarrow \infty} \sum_{j=1}^{\ell} p_{ij}^{(n)} \hat{\mu}_j^{(0)} \\ &= \sum_{j=1}^{\ell} \lim_{n \rightarrow \infty} p_{ij}^{(n)} \hat{\mu}_j^{(0)} = \sum_{j=1}^{\ell} \pi_j \hat{\mu}_j^{(0)}. \end{aligned}$$

- Der Konsens der Experten-Kommission, d.h. die gemeinsame Prognose des (unbekannten) Parameters μ ist dann gegeben durch

$$\hat{\mu} = \sum_{j=1}^{\ell} \pi_j \hat{\mu}_j^{(0)}. \quad (48)$$

Beachte

- Für große ℓ kann die direkte (algebraische) Lösung des linearen Gleichungssystems (41) schwierig sein.
- Dann spielt die Abschätzung der Konvergenzgeschwindigkeit in (47) eine wichtige Rolle bei der praktischen Durchführung der in (47) beschriebenen Konsensbildung.
- Wir betrachten hierfür das folgende Zahlenbeispiel.
 - Sei $\ell = 3$, und sei

$$\mathbf{P} = \begin{pmatrix} \frac{2}{6} & \frac{1}{6} & \frac{3}{6} \\ \frac{1}{4} & \frac{1}{4} & \frac{2}{4} \\ \frac{2}{8} & \frac{1}{8} & \frac{5}{8} \end{pmatrix}. \quad (49)$$

- Die Eintragungen dieser stochastischen Matrix bedeuten, dass die Kompetenz des dritten Experten von allen Kommissionsmitgliedern besonders hoch eingeschätzt wird.
- Die Lösung $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)^\top$ des entsprechenden linearen Gleichungssystems (41) ist dann gegeben durch

$$\pi_1 = \frac{21}{77}, \quad \pi_2 = \frac{12}{77}, \quad \pi_3 = \frac{44}{77},$$

d.h., die Prognose $\hat{\theta}_3^{(0)}$ des dritten (als besonders kompetent eingeschätzten) Experten wird am stärksten gewichtet.

- Die Eigenwerte der in (49) gegebenen Übergangsmatrix sind $\theta_1 = 1$, $\theta_2 = 1/8$ und $\theta_3 = 1/12$.
- Gemäß (43) ergibt sich dabei die „Basis“

$$1 - 3a = 1 - 3 \min_{i,j=1,2,3} p_{ij} = 1 - \frac{3}{8} = \frac{5}{8}$$

der Konvergenzabschätzung, während Korollar 2.4 die folgende (deutlich verbesserte) geometrische Konvergenzabschätzung

$$\max_{i \in \{1, \dots, \ell\}} |\hat{\mu} - \hat{\mu}_i^{(n)}| = O(|\theta_2|^n)$$

liefert, wobei $\theta_2 = 1/8$ der zweitgrößte Eigenwert der in (49) gegebenen stochastischen Matrix $\mathbf{P}$ ist, vgl. auch Übungsaufgabe 3.3.

2.2.3 Irreduzibilität und Aperiodizität

- Zur Erinnerung
 - In Theorem 2.4 hatten wir die Ergodizität der Markov-Kette $X_0, X_1, \dots$ durch die Quasi-Positivität ihrer Übergangsmatrix $\mathbf{P}$ charakterisiert.
 - Es kann sich jedoch als schwierig erweisen, diese Eigenschaft von $\mathbf{P}$ direkt nachzuweisen (insbesondere dann, wenn $\ell \gg 1$).
- Wir leiten deshalb noch eine andere (probabilistische) Charakterisierung der Ergodizität von Markov-Ketten mit endlichem Zustandsraum her. Dabei benötigen wir den folgenden Begriff.
 - Für beliebige, jedoch fest vorgegebene Zustände $i, j \in E$ sagen wir, dass der Zustand j vom Zustand i *erreichbar* ist, falls $p_{ij}^{(n)} > 0$ für ein $n \geq 0$ gilt, wobei $\mathbf{P}^{(0)} = \mathbf{I}$. (Schreibweise: $i \rightarrow j$)
 - Eine andere (äquivalente) Definitionsmöglichkeit der Erreichbarkeit von Zuständen ist folgendermaßen gegeben.
 - Hierfür sei $\tau_j = \min\{n \geq 0 : X_n = j\}$ die Anzahl der Schritte bis die Markov-Kette $\{X_n\}$ zum ersten Mal den Zustand $j \in E$ erreicht. Dabei setzen wir $\tau_j = \infty$, falls $X_n \neq j$ für jedes $n \geq 0$.

Theorem 2.7 *Sei $i \in E$ so gewählt, dass $P(X_0 = i) > 0$. In diesem Fall ist der Zustand $j \in E$ genau dann von i erreichbar, wenn $P(\tau_j < \infty \mid X_0 = i) > 0$.*

Beweis

- Die Notwendigkeit der Bedingung ist offensichtlich, denn es gilt

$$\{X_n = j\} \subset \{\tau_j \leq n\} \subset \{\tau_j < \infty\} \quad \text{und somit} \quad 0 < p_{ij}^{(n)} \leq P(\tau_j < \infty \mid X_0 = i)$$

für ein $n \geq 0$, falls j von i erreichbar ist.

- Falls umgekehrt $i \neq j$ und $p_{ij}^{(n)} = 0$ für jedes $n \geq 0$ gilt, dann folgt hieraus, dass

$$\begin{aligned} P(\tau_j < \infty \mid X_0 = i) &= \lim_{n \rightarrow \infty} P(\tau_j < n \mid X_0 = i) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{k=0}^{n-1} \{X_k = j\} \mid X_0 = i\right) \\ &\leq \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} P(X_k = j \mid X_0 = i) = \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} p_{ij}^{(k)} = 0. \end{aligned}$$

□

Beachte

- Die Eigenschaft der Erreichbarkeit ist
 - *transitiv*, d.h., aus $i \rightarrow k$ und $k \rightarrow j$ folgt $i \rightarrow j$.

- Dies ergibt sich unmittelbar aus der Ungleichung $p_{ij}^{(r+m)} \geq p_{ik}^{(r)} p_{kj}^{(m)}$ (vgl. Korollar 2.2) und aus der Definition der Erreichbarkeit.
- Falls darüber hinaus $i \rightarrow j$ und $j \rightarrow i$, dann sagen wir, dass die Zustände i und j *kommunizieren*. (Schreibweise: $i \leftrightarrow j$)
- Die Eigenschaft des Kommunizierens ist eine *Äquivalenzrelation*, denn es gilt
 - (a) $i \leftrightarrow i$ (Reflexivität),
 - (b) $i \leftrightarrow j$ genau dann, wenn $j \leftrightarrow i$ (Symmetrie),
 - (c) $i \leftrightarrow k$ und $k \leftrightarrow j$ implizieren, dass $i \leftrightarrow j$ (Transitivität).
- Dies bedeutet insbesondere,
 - dass der Zustandsraum E in (disjunkte und ausschöpfende) Äquivalenzklassen bezüglich der Äquivalenzrelation $\leftrightarrow$ zerlegt werden kann.
 - Dabei heißt die Markov-Kette $\{X_n\}$ bzw. ihre Übergangsmatrix $\mathbf{P} = (p_{ij})$ *irreduzibel*, wenn der Zustandsraum E nur aus einer solchen Äquivalenzklasse besteht, d.h., wenn $i \leftrightarrow j$ für sämtliche $i, j \in E$ gilt.

Beispiele

- Unmittelbar aus der Definition der Irreduzibilität ergibt sich, dass die 2×2 Matrizen

$$\mathbf{P}_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \quad \text{und} \quad \mathbf{P}_2 = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$$

irreduzibel sind.

- Die aus $\mathbf{P}_1$ und $\mathbf{P}_2$ gebildete 4×4 Matrix $\mathbf{P}$ mit der folgenden *Blockstruktur*

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 \end{pmatrix}$$

ist jedoch nicht irreduzibel.

Neben der Irreduzibilität benötigen wir noch eine weitere Eigenschaft der Zustände, nämlich die sogenannte *Aperiodizität*, um die Ergodizität von Markov-Ketten auf einfache Weise charakterisieren zu können.

Definition

- Die *Periode* d_i des Zustandes $i \in E$ ist gegeben durch $d_i = \text{ggT}\{n \geq 1 : p_{ii}^{(n)} > 0\}$, wobei „ggT“ den größten gemeinsamen Teiler bezeichnet. Dabei setzen wir $d_i = \infty$, falls $p_{ii}^{(n)} = 0$ für jedes $n \geq 1$.
- Der Zustand $i \in E$ heißt *aperiodisch*, falls $d_i = 1$.
- Die Markov-Kette $\{X_n\}$ bzw. ihre Übergangsmatrix $\mathbf{P} = (p_{ij})$ heißt *aperiodisch*, wenn sämtliche Zustände von $\{X_n\}$ aperiodisch sind.

Wir zeigen nun, dass die Perioden d_i und d_j übereinstimmen, wenn die Zustände i, j zu einundderselben Äquivalenzklasse von kommunizierenden Zuständen gehören. Dabei verwenden wir die Schreibweise $i \rightarrow j[n]$, falls $p_{ij}^{(n)} > 0$.

Theorem 2.8 Falls die Zustände $i, j \in E$ kommunizieren, dann gilt $d_i = d_j$.

Beweis

- Falls $j \rightarrow j[n]$, $i \rightarrow j[k]$ und $j \rightarrow i[m]$ für gewisse $k, m, n \geq 1$ gilt, dann ergibt sich aus den Ungleichungen in Korollar 2.2, dass $i \rightarrow i[k+m]$ und $i \rightarrow i[k+m+n]$.
- Dies bedeutet, dass die natürlichen Zahlen $k+m$ und $k+m+n$ durch d_i teilbar sind.
- Dann ist auch $n = (k+m+n) - (k+m)$ durch d_i teilbar.
- Somit ist d_i ein gemeinsamer Teiler für diejenigen natürlichen Zahlen n , für die $p_{jj}^{(n)} > 0$ gilt, d.h. $d_i \leq d_j$.
- Auf die gleiche Weise ergibt sich (aus Symmetriegründen), dass auch $d_j \leq d_i$ gilt. $\square$

Korollar 2.5 Die Markov-Kette $\{X_n\}$ sei irreduzibel. Dann besitzen sämtliche Zustände von $\{X_n\}$ die gleiche Periode.

Um zeigen zu können,

- dass die in Abschnitt 2.2.1 betrachtete Charakterisierung (vgl. Theorem 2.4) der Ergodizität von Markov-Ketten gleichbedeutend mit ihrer Irreduzibilität und Aperiodizität ist,
- benötigen wir noch den folgenden elementaren Hilfssatz aus der *Zahlentheorie*.

Lemma 2.3 Sei $k = 1, 2, \dots$ eine beliebige, jedoch fest vorgegebene natürliche Zahl. Dann gibt es eine natürliche Zahl $n_0 \geq 1$, so dass

$$\{n_0, n_0 + 1, n_0 + 2, \dots\} \subset \{n_1 k + n_2(k+1); n_1, n_2 \geq 0\}.$$

Beweis

- Falls $n \geq k^2$, dann gibt es ganze Zahlen $m, d \geq 0$, so dass $n - k^2 = mk + d$ und $d < k$.
- Hieraus folgt, dass $n = (k - d + m)k + d(k+1)$ und somit

$$n \in \{n_1 k + n_2(k+1); n_1, n_2 \geq 0\},$$

d.h., wir können $n_0 = k^2$ setzen. $\square$

Theorem 2.9 Die Übergangsmatrix $\mathbf{P}$ ist genau dann quasi-positiv, wenn $\mathbf{P}$ irreduzibel und aperiodisch ist.

Beweis

- Wir nehmen zunächst an, dass die Übergangsmatrix $\mathbf{P}$ irreduzibel und aperiodisch ist.
 - Für jedes $i \in E$ betrachten wir die Menge $J(i) = \{n \geq 1 : p_{ii}^{(n)} > 0\}$, deren größter gemeinsamer Teiler wegen der Aperiodizität von $\mathbf{P}$ gleich 1 ist.
 - Aus den Ungleichungen in Korollar 2.2 ergibt sich, dass

$$p_{ii}^{(n+m)} \geq p_{ii}^{(n)} p_{ii}^{(m)}$$

und somit

$$n + m \in J(i), \quad \text{falls } n, m \in J(i). \quad (50)$$

- Wir zeigen, dass die Menge $J(i)$ zwei aufeinanderfolgende Zahlen enthält.

- Falls umgekehrt $J(i)$ nicht zwei aufeinanderfolgende Zahlen enthalten würde, dann würde es einen Minimalabstand $k \geq 2$ zwischen den Zahlen in $J(i)$ geben.
- Hieraus würde die Gültigkeit von $mk + d \in J(i)$ für gewisse $m = 0, 1, \dots$ und $d = 1, \dots, k - 1$ folgen, weil ansonsten $n = mk$ für alle $n \in J(i)$ gelten würde.
- Dies würde jedoch ein Widerspruch zu unserer Annahme sein, dass $\text{ggT}(J(i)) = 1$.
- Sei nun $n_1, n_1 + k \in J(i)$. Wegen (50) gilt dann auch $a(n_1 + k) \in J(i)$ und $n + bn_1 \in J(i)$ für beliebige $a, b \in \mathbb{N}$, wobei

$$n = mk + d \in J(i). \quad (51)$$

- Wir zeigen jedoch,
 - dass es dann natürliche Zahlen $a, b \in \{1, 2, \dots\}$ gibt, so dass die Differenz zwischen $a(n_1 + k) \in J(i)$ und $n + bn_1 \in J(i)$ kleiner als k ist.
 - Und zwar ergibt sich aus (51), dass

$$a(n_1 + k) - n - bn_1 = (a - b)n_1 + (a - m)k - d$$

und somit für $a = b = m + 1$

$$a(n_1 + k) - n - bn_1 = k - d < k.$$

- Deshalb enthält die Menge $J(i)$ zwei aufeinanderfolgende Zahlen.
- Aus (50) und aus Lemma 2.3 ergibt sich nun, dass es für jedes $i \in E$ ein $n(i) \geq 1$ gibt, so dass

$$J(i) \supset \{n(i), n(i) + 1, \dots\}. \quad (52)$$

- Hieraus, aus der Irreduzibilität von $\mathbf{P}$ und aus der Ungleichung (25) in Korollar 2.2, d.h.,

$$p_{ij}^{(r+n+m)} \geq p_{ik}^{(r)} p_{kk}^{(n)} p_{kj}^{(m)},$$

folgt, dass es für jedes Paar $i, j \in E$ von Zuständen eine natürliche Zahl $n(ij) \geq 1$ gibt, so dass

$$J(ij) = \{n \geq 0 : p_{ij}^{(n)} > 0\} \supset \{n(ij), n(ij) + 1, \dots\},$$

d.h., $\mathbf{P}$ ist quasi-positiv.

- Umgekehrt ergibt sich die Irreduzibilität und Aperiodizität von quasi-positiven Übergangsmatrizen unmittelbar aus den Definitionen dieser Begriffe. $\square$

Beachte

- Ein einfaches Beispiel einer Markov-Kette, die *nicht* irreduzibel ist,
 - kann durch das bereits mehrfach betrachtete Modell der *Wettervorhersage* gegeben werden, wobei $E = \{1, 2\}$ und

$$\mathbf{P} = \begin{pmatrix} 1 - p & p \\ p' & 1 - p' \end{pmatrix}.$$

- Falls $p = 0$ oder $p' = 0$, dann ist die zugehörige Markov-Kette offenbar nicht irreduzibel (und damit wegen Theorem 2.9 auch nicht ergodisch).
- Es ist dennoch möglich, dass das lineare Gleichungssystem

$$\boldsymbol{\alpha}^\top = \boldsymbol{\alpha}^\top \mathbf{P} \quad (53)$$

eine (oder unendlich viele) Wahrscheinlichkeitslösungen $\boldsymbol{\alpha}^\top = (\alpha_1, \alpha_2)$ besitzt.

- Wenn beispielsweise $p = 0$ und $p' > 0$ gilt, dann ist $i = 1$ ein sogenannter *absorbierender Zustand*, und $\alpha^\top = (1, 0)$ ist die (eindeutig bestimmte) Lösung des linearen Gleichungssystems (53).
- Wenn $p = 0$ und $p' = 0$, dann ist *jede* Wahrscheinlichkeitsfunktion $\alpha^\top = (\alpha_1, \alpha_2)$ Lösung des linearen Gleichungssystems (53).
- Wir geben nun noch Beispiele von Markov-Ketten $X_0, X_1, \dots : \Omega \rightarrow E$ an, die *nicht* aperiodisch sind.
 - Dabei sind die Zufallsvariablen $X_0, X_1, \dots : \Omega \rightarrow E$ *nicht* durch eine stochastische Rekursionsgleichung $X_n = \varphi(X_{n-1}, Z_n)$ der Form (14) gegeben, so dass die „Zuwächse“ $Z_1, Z_2, \dots : \Omega \rightarrow D$ unabhängige und identisch verteilte Zufallsvariablen sind.
 - Wir setzen nämlich lediglich voraus, dass die Zufallsvariablen $Z_1, Z_2, \dots : \Omega \rightarrow D$ in dem folgenden Sinne „bedingt unabhängig“ sind.
 - Wie in Abschnitt 2.1.3 gezeigt wurde, kann man jedoch stets eine zu $X_0, X_1, \dots$ stochastisch äquivalente Markov-Kette konstruieren, deren Zuwächse unabhängig sind, vgl. das in (17)–(19) betrachtete Konstruktionsprinzip.
- Seien also E und D beliebige endliche (oder abzählbar unendliche) Mengen, sei $\varphi : E \times D \rightarrow E$ eine beliebige Funktion, und seien $X_0, X_1, \dots : \Omega \rightarrow E$ bzw. $Z_1, Z_2, \dots : \Omega \rightarrow D$ Zufallsvariablen,
 - so dass

$$X_n = \varphi(X_{n-1}, Z_n) \quad (54)$$

- und dass für jedes $n \in \mathbb{N}$ die Zufallsvariable Z_n bei gegebenem X_{n-1} *bedingt unabhängig* von den Zufallsvariablen $Z_1, \dots, Z_{n-1}, X_0, \dots, X_{n-2}$ ist.
- D.h., für beliebige $n \in \mathbb{N}$, $i_0, i_1, \dots, i_{n-1} \in E$ und $k_1, \dots, k_n \in D$ gelte

$$\begin{aligned} P(Z_n = k_n, Z_{n-1} = k_{n-1}, \dots, Z_1 = k_1, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ = P(Z_n = k_n \mid X_{n-1} = i_{n-1}) P(Z_{n-1} = k_{n-1}, \dots, Z_1 = k_1, X_{n-1} = i_{n-1}, \dots, X_0 = i_0), \end{aligned}$$
 wobei wir $P(Z_n = k_n \mid X_{n-1} = i_{n-1}) = 0$ setzen, falls $P(X_{n-1} = i_{n-1}) = 0$.
- Außerdem wird vorausgesetzt, dass für beliebige $i \in E$ und $k \in D$ die Wahrscheinlichkeiten $P(Z_n = k \mid X_{n-1} = i)$ *nicht* von $n \in \mathbb{N}$ abhängen.
- Man kann zeigen (vgl. Übungsaufgabe 4.2), dass die durch (54) rekursiv gegebene Folge $X_0, X_1, \dots : \Omega \rightarrow E$ eine Markov-Kette ist, deren Übergangsmatrix $\mathbf{P} = (p_{ij})$ gegeben ist durch

$$p_{ij} = P(\varphi(i, Z_1) = j \mid X_0 = i),$$

falls $P(X_0 = i) > 0$ für jedes $i \in E$.

Beispiel (Diffusionsmodell)

vgl. P. Brémaud (1999) *Markov Chains*. Springer-Verlag, New York, S.76

- Das folgende einfache Modell zur Beschreibung von Diffusionsvorgängen durch eine Membran wurde im Jahre 1907 von den Physikern Tatiana und Paul Ehrenfest vorgeschlagen. Dabei geht es um die Modellierung des Wärmeaustausches zwischen zwei Systemen mit unterschiedlichen Temperaturen.
 - Wir betrachten ℓ Partikel, die auf zwei miteinander durchlässig verbundene, aber nach außen isolierte Behälter A und B verteilt sind.
 - Angenommen zum Zeitpunkt $n - 1$ befinden sich $X_{n-1} = i$ Partikel in A . Dann wird eine der ℓ insgesamt vorhandenen Partikel zufällig ausgewählt und in den jeweils anderen Behälter überführt.
 - Für den Zustand X_n des Systems zum Zeitpunkt n gilt somit entweder $X_n = i - 1$ mit Wahrscheinlichkeit i/ℓ (falls die ausgewählte Partikel im Behälter A war) oder $X_n = i + 1$ mit Wahrscheinlichkeit $(\ell - i)/\ell$ (falls die ausgewählte Partikel im Behälter B war).
- Die Zufallsvariablen $X_0, X_1, \dots : \Omega \rightarrow \{0, 1, \dots, \ell\}$ können also rekursiv definiert werden

- durch die stochastische Rekursionsgleichung

$$X_n = X_{n-1} + Z_n, \quad (55)$$

- wobei die Zufallsvariable Z_n bei gegebenem X_{n-1} *bedingt unabhängig* von den Zufallsvariablen $Z_1, \dots, Z_{n-1}, X_0, \dots, X_{n-1}$ ist mit $P(Z_n = -1) + P(Z_n = 1) = 1$ und

$$P(Z_n = -1 \mid X_{n-1} = i) = \frac{i}{\ell}, \quad \text{falls } P(X_{n-1} = i) > 0.$$

- Für die Eintragungen der Übergangsmatrix $\mathbf{P} = (p_{ij})$ gilt somit

$$p_{ij} = \begin{cases} \frac{\ell-i}{\ell}, & \text{falls } i < \ell \text{ und } j = i+1, \\ \frac{i}{\ell}, & \text{falls } i > 0 \text{ und } j = i-1, \\ 0, & \text{sonst.} \end{cases}$$

- Dies bedeutet insbesondere, dass $d_i = \text{ggT}\{n \geq 1 : p_{ii}^{(n)} > 0\} = 2$ für jedes $i \in \{0, 1, \dots, \ell\}$, d.h., die durch (55) gegebene Markov-Kette ist nicht aperiodisch (und damit wegen Theorem 2.9 auch nicht ergodisch).

- Das lineare Gleichungssystem

$$\boldsymbol{\alpha}^\top = \boldsymbol{\alpha}^\top \mathbf{P} \quad (56)$$

besitzt *dennoch* eine (eindeutig bestimmte) Wahrscheinlichkeitslösung $\boldsymbol{\alpha}^\top = (\alpha_0, \dots, \alpha_\ell)$ mit

$$\alpha_i = \frac{1}{2^\ell} \binom{\ell}{i}, \quad \forall i \in \{0, 1, \dots, \ell\}. \quad (57)$$

Beachte

- Das Diffusionsmodell von Ehrenfest ist ein Spezialfall der folgenden Klasse von Markov-Ketten, die in der Literatur *Geburts- und Todesprozesse mit zwei reflektierenden Schranken* genannt werden.
- Dabei wird erneut der Zustandsraum $E = \{0, 1, \dots, \ell\}$ betrachtet, während die Übergangsmatrix $\mathbf{P} = (p_{ij})$ gegeben ist durch

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & & & & \\ q_1 & r_1 & p_1 & & & \\ & q_2 & r_2 & p_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & q_i & r_i & p_i \\ & & & & \ddots & \ddots & \ddots \\ & & & & & q_{\ell-1} & r_{\ell-1} & p_{\ell-1} \\ & & & & & & 1 & 0 \end{pmatrix}, \quad (58)$$

wobei $p_i > 0$, $q_i > 0$ und $p_i + q_i + r_i = 1$ für jedes $i \in \{1, \dots, \ell-1\}$.

- Das lineare Gleichungssystem $\boldsymbol{\alpha}^\top = \boldsymbol{\alpha}^\top \mathbf{P}$ hat dann die Form

$$\alpha_i = \begin{cases} p_{i-1}\alpha_{i-1} + r_i\alpha_i + q_{i+1}\alpha_{i+1}, & \text{falls } 0 < i < \ell, \\ q_1\alpha_1, & \text{falls } i = 0, \\ p_{\ell-1}\alpha_{\ell-1}, & \text{falls } i = \ell. \end{cases}$$

- Man kann zeigen (vgl. Übungsaufgabe 4.3), dass

$$\alpha_i = \alpha_0 \frac{p_1 p_2 \cdots p_{i-1}}{q_1 q_2 \cdots q_i},$$

- wobei $\alpha_0 > 0$ durch die Normierungsbedingung $\sum_{i=0}^{\ell} \alpha_i = 1$ gegeben ist, d.h.

$$\alpha_0 \left(1 + \frac{1}{q_1} + \frac{p_1}{q_1 q_2} + \cdots + \frac{p_1 p_2 \cdots p_{\ell-1}}{q_1 q_2 \cdots q_{\ell}} \right) = 1$$

bzw.

$$\alpha_0 = \left(1 + \frac{1}{q_1} + \frac{p_1}{q_1 q_2} + \cdots + \frac{p_1 p_2 \cdots p_{\ell-1}}{q_1 q_2 \cdots q_{\ell}} \right)^{-1}.$$

- Weil wir voraussetzen, dass $p_i > 0$ und $q_i > 0$ für jedes $i \in \{1, \dots, \ell - 1\}$ gilt, sind Geburts- und Todesprozesse mit zwei reflektierenden Schranken offenbar irreduzibel.
- Wenn zusätzlich $r_i > 0$ für ein $i \in \{1, \dots, \ell - 1\}$ gilt, dann sind Geburts- und Todesprozesse mit zwei reflektierenden Schranken auch aperiodisch (und somit ergodisch wegen Theorem 2.9), vgl. Übungsaufgabe 4.3.

2.2.4 Stationäre Anfangsverteilungen

- *Zur Erinnerung*

- Falls $\{X_n\}$ eine irreduzible und aperiodische Markov-Kette mit dem (endlichen) Zustandsraum $E = \{1, \dots, \ell\}$ und der (quasi-positiven) Übergangsmatrix $\mathbf{P} = (p_{ij})$ ist,
- dann ist die Grenzverteilung $\boldsymbol{\pi} = \lim_{n \rightarrow \infty} \boldsymbol{\alpha}_n$ die eindeutig bestimmte Lösung der folgenden Matrix-Gleichung (vgl. Theorem 2.5):

$$\boldsymbol{\alpha}^\top = \boldsymbol{\alpha}^\top \mathbf{P}. \quad (59)$$

- Wenn nicht vorausgesetzt wird, dass die Markov-Kette $\{X_n\}$ irreduzibel ist, dann kann (59) mehr als eine (Wahrscheinlichkeits-) Lösung besitzen.

- Falls die Anfangsverteilung $\boldsymbol{\alpha}_0$ von $\{X_n\}$ eine solche Lösung von (59) ist, dann implizieren Theorem 2.3 und (59) darüber hinaus, dass

$$\boldsymbol{\alpha}_1^\top = \boldsymbol{\alpha}_0^\top \mathbf{P} = \boldsymbol{\alpha}_0^\top$$

und somit $\boldsymbol{\alpha}_n = \boldsymbol{\alpha}_0$ für jedes $n \geq 0$.

- Wegen dieser Invarianzeigenschaft wird jede (Wahrscheinlichkeits-) Lösung $\boldsymbol{\alpha}$ von (59) eine *stationäre Anfangsverteilung* von $\{X_n\}$ genannt.

- Umgekehrt kann man zeigen, dass,

- die Matrix-Gleichung (59) *genau eine* (Wahrscheinlichkeits-) Lösung $\boldsymbol{\alpha}$ besitzt, falls $\mathbf{P}$ irreduzibel ist,
- wobei diese Lösung $\boldsymbol{\alpha}$ von (59) jedoch *nicht* notwendig die Grenzverteilung $\boldsymbol{\pi} = \lim_{n \rightarrow \infty} \boldsymbol{\alpha}_n$ sein muss, die nämlich dann nicht existiert, falls $\mathbf{P}$ nicht aperiodisch ist.

Theorem 2.10

- Sei $\mathbf{P} = (p_{ij})_{i,j \in E}$ eine irreduzible Übergangsmatrix, wobei $E = \{1, \dots, \ell\}$.

- Für beliebige, aber fest vorgegebene $i, j \in E$ konvergieren dann die Eintragungen $q_{ij}^{(n)}$ der stochastischen $(\ell \times \ell)$ -dimensionalen Matrizen $\mathbf{Q}_n = (q_{ij}^{(n)})$ mit

$$\mathbf{Q}_n = \frac{1}{n} (\mathbf{P} + \mathbf{P}^2 + \dots + \mathbf{P}^n) \quad (60)$$

gegen einen Grenzwert

$$\alpha_j = \lim_{n \rightarrow \infty} q_{ij}^{(n)} > 0, \quad (61)$$

der nicht von i abhängt, so dass $\sum_{j=1}^{\ell} \alpha_j = 1$ und dass der Vektor $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{\ell})^{\top}$ eine Lösung der Matrix-Gleichung (59) ist.

- Die in (60)–(61) gegebene Verteilung $\boldsymbol{\alpha}$ ist die einzige Wahrscheinlichkeitslösung von (59).

Einen Beweis von Theorem 2.10 kann man beispielsweise in Kapitel 7 des Buches E. Behrends (2000) *Introduction to Markov Chains*, Vieweg, Braunschweig finden.

Beachte

- Außer der Invarianzeigenschaft $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}_1 = \dots$ besitzt die Markov-Kette $\{X_n\}$ mit der stationären Anfangsverteilung $\boldsymbol{\alpha}_0$ noch eine wesentlich stärkere Invarianzeigenschaft (und zwar für sämtliche endlich-dimensionalen Verteilungen).
- In diesem Zusammenhang betrachten wir den folgenden Begriff der (streng) stationären Folge von Zufallsvariablen.

Definition

- Sei $X_0, X_1, \dots : \Omega \rightarrow E$ eine beliebige Folge von Zufallsvariablen mit Werten in $E = \{1, \dots, \ell\}$ (die nicht unbedingt eine Markov-Kette sein müssen).
- Die Folge $\{X_n\}$ von E -wertigen Zufallsvariablen heißt *stationär*, falls für beliebige $k, n \in \{0, 1, \dots\}$ und $i_0, \dots, i_n \in E$ gilt, dass

$$P(X_k = i_0, X_{k+1} = i_1, \dots, X_{k+n} = i_n) = P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n). \quad (62)$$

Theorem 2.11

- Sei $X_0, X_1, \dots : \Omega \rightarrow E$ eine Markov-Kette mit dem Zustandsraum $E = \{1, \dots, \ell\}$.
- Die Markov-Kette $\{X_n\}$ ist genau dann eine stationäre Folge von Zufallsvariablen, wenn sie eine stationäre Anfangsverteilung besitzt.

Beweis

- Die Notwendigkeit der Bedingung ergibt sich unmittelbar
 - aus Theorem 2.3 sowie aus den Definitionen der Begriffe „stationäre Anfangsverteilung“ bzw. „stationäre Folge von Zufallsvariablen“,
 - denn die Gültigkeit von (62) impliziert insbesondere, dass $P(X_1 = i) = P(X_0 = i)$ für jedes $i \in E$
 - und aus Theorem 2.3 ergibt sich somit, dass $\boldsymbol{\alpha}_0^{\top} = \boldsymbol{\alpha}_1^{\top} = \boldsymbol{\alpha}_0^{\top} \mathbf{P}$, d.h. $\boldsymbol{\alpha}_0$ ist eine stationäre Anfangsverteilung.
- Sei nun umgekehrt $\boldsymbol{\alpha}_0$ eine stationäre Anfangsverteilung der Markov-Kette $\{X_n\}$.

- Dann ergibt sich aus der Definitionsgleichung (3) der Markov-Kette $\{X_n\}$, dass

$$\begin{aligned}
& P(X_k = i_0, X_{k+1} = i_1, \dots, X_{k+n} = i_n) \\
&= \sum_{i'_0, \dots, i'_{k-1} \in E} P(X_0 = i'_0, \dots, X_{k-1} = i'_{k-1}, X_k = i_0, X_{k+1} = i_1, \dots, X_{k+n} = i_n) \\
&= \sum_{i'_0, \dots, i'_{k-1} \in E} \alpha_{0, i'_0} p_{i'_0 i'_1} \cdots p_{i'_{k-2} i'_{k-1}} p_{i'_{k-1} i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \\
&= \left(\alpha_0^\top \mathbf{P}^k \right)_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} = \alpha_{0, i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \\
&= P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n),
\end{aligned}$$

- wobei in der vorletzten Gleichheit die Stationarität der Anfangsverteilung α_0 und in der letzten Gleichheit erneut die Definitionsgleichung (3) der Markov-Kette $\{X_n\}$ verwendet wurde. $\square$

Beachte

- Für einige Beispiele von Markov-Ketten mit speziell strukturierten Übergangsmatrizen hatten wir bereits in den Abschnitten 2.2.2 und 2.2.3 die zugehörigen stationären Anfangsverteilungen bestimmt.
- Wir diskutieren nun noch zwei weitere Beispiele dieses Typs.
 - Der Zustandsraum ist dabei jeweils *unendlich*, so dass außer der Quasi-Positivität (bzw. der Irreduzibilität und Aperiodizität der Übergangsmatrix) noch eine weitere Bedingung erforderlich ist, um die Ergodizität der betrachteten Markov-Ketten zu garantieren.
 - Es ist dies eine sogenannte *Kontraktionsbedingung*, die verhindert, dass die „Wahrscheinlichkeitsmasse“ ins Unendliche abdriftet.

Beispiele

1. Warteschlangen

vgl. T. Rolski, H. Schmidli, V. Schmidt, J. Teugels (2002) *Stochastic Processes for Insurance and Finance*. J. Wiley & Sons, Chichester, S. 147 ff.

- Wir betrachten das bereits in Abschnitt 2.1.2 diskutierte Beispiel
 - der rekursiv definierten Markov-Kette $X_0, X_1, \dots, \Omega \rightarrow \{0, 1, \dots\}$ mit $X_0 = 0$ und

$$X_n = \max\{0, X_{n-1} + Z_n - 1\}, \quad \forall n \geq 1, \quad (63)$$

- wobei die Zufallsvariablen $Z, Z_1, Z_2, \dots: \Omega \rightarrow \{0, 1, \dots\}$ unabhängig und identisch verteilt sind und die Übergangsmatrix $\mathbf{P} = (p_{ij})$ gegeben ist durch

$$p_{ij} = \begin{cases} P(Z = j + 1 - i), & \text{falls } j + 1 \geq i > 0 \text{ oder } j > i = 0, \\ P(Z = 0) + P(Z = 1), & \text{falls } j = i = 0, \\ 0, & \text{sonst.} \end{cases} \quad (64)$$

- Man kann zeigen, dass
 - die in (63) rekursiv definierte Markov-Kette $\{X_n\}$ bzw. die zugehörige (in (64) gegebene) Übergangsmatrix irreduzibel und aperiodisch ist, falls

$$P(Z = 0) > 0, \quad P(Z = 1) > 0 \quad \text{und} \quad P(Z = 2) > 0, \quad (65)$$

- die Lösung der Rekursionsgleichung (63) sich für jedes $n \geq 1$ darstellen lässt in der Form

$$X_n = \max\left\{0, \max_{k \in \{1, \dots, n\}} \sum_{r=k}^n (Z_r - 1)\right\} \stackrel{d}{=} \max\left\{0, \max_{k \in \{1, \dots, n\}} \sum_{r=1}^k (Z_r - 1)\right\}, \quad (66)$$

- die (Grenz-) Wahrscheinlichkeiten π_i für jedes $i \in \{0, 1, \dots\}$ existieren, wobei

$$\begin{aligned}\pi_i &= \lim_{n \rightarrow \infty} P\left(\max\left\{0, \max_{k \in \{1, \dots, n\}} \sum_{r=1}^k (Z_r - 1) = i\right\}\right) \\ &= \begin{cases} P\left(\sup_{k \in \{1, 2, \dots\}} \sum_{r=1}^k (Z_r - 1) = i\right) & \text{für } i > 0, \\ P\left(\sup_{k \in \{1, 2, \dots\}} \sum_{r=1}^k (Z_r - 1) \leq 0\right) & \text{für } i = 0. \end{cases}\end{aligned}$$

Dabei gilt

$$\begin{aligned}\pi_i &= 0 \quad \text{für jedes } i \in \{0, 1, \dots\}, & \text{falls } \mathbb{E} Z \geq 1, \\ \pi_i &> 0 \quad \text{für jedes } i \in \{0, 1, \dots\} \text{ und } \sum_{i \geq 0} \pi_i = 1, & \text{falls (65) gilt und } \mathbb{E} Z < 1.\end{aligned}$$

- Für Markov-Ketten mit (abzählbar) unendlichem Zustandsraum
 - folgt also die Ergodizität im allgemeinen *nicht* aus der Irreduzibilität und Aperiodizität,
 - sondern es muss noch zusätzlich eine gewisse *Kontraktionsbedingung* erfüllt sein,
 - wobei dies im Fall des hier betrachteten Beispiels die Bedingung ist, dass die Markov-Kette $\{X_n\}$ eine *negative Drift* besitzt, d.h., dass $\mathbb{E}(Z - 1) < 0$ gilt.
- Falls die Bedingungen (65) erfüllt sind und $\mathbb{E} Z < 1$ gilt,
 - dann besitzt die Matrixgleichung $\boldsymbol{\alpha}^\top = \boldsymbol{\alpha}^\top \mathbf{P}$ zwar eine (eindeutig bestimmte) Wahrscheinlichkeitslösung $\boldsymbol{\alpha}^\top = (\alpha_0, \alpha_1, \dots)$,
 - die mit $\boldsymbol{\pi}^\top = (\pi_0, \pi_1, \dots)$ ($= \lim_{n \rightarrow \infty} \boldsymbol{\alpha}_n^\top$) übereinstimmt, sich jedoch im allgemeinen *nicht* explizit bestimmen lässt.
 - Man kann aber eine einfache Formel für die *erzeugende Funktion* $g_\pi : (-1, 1) \rightarrow [0, 1]$ von $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots)^\top$ angeben, wobei

$$g_\pi(s) = \sum_{i=0}^{\infty} s^i \pi_i \quad \left(= \mathbb{E} s^{X_\infty} \right)$$

und

$$X_\infty = \max\left\{0, \sup_{k \in \{1, 2, \dots\}} \sum_{r=1}^k (Z_r - 1)\right\}. \quad (67)$$

- Und zwar gilt

$$g_\pi(s) = \frac{(1 - \rho)(1 - s)}{g_Z(s) - s}, \quad \forall s \in (-1, 1), \quad (68)$$

wobei $\rho = \mathbb{E} Z$ und $g_Z(s) = \mathbb{E} s^Z$ die erzeugende Funktion von Z ist.

- *Beweis* von (68)
 - Aus (67) ergibt sich, dass $X_\infty \stackrel{d}{=} \max\{0, X_\infty + (Z - 1)\}$.
 - Mit der Schreibweise $x_+ = \max\{0, x\}$ gilt somit

$$\begin{aligned}g_\pi(s) &= \mathbb{E} s^{X_\infty} = \mathbb{E} s^{(X_\infty + Z - 1)_+} \\ &= \mathbb{E} \left(s^{(X_\infty + Z - 1)_+} \mathbf{1}(X_\infty + Z - 1 \geq 0) \right) + \mathbb{E} \left(s^{(X_\infty + Z - 1)_+} \mathbf{1}(X_\infty + Z - 1 = -1) \right) \\ &= \frac{1}{s} \sum_{k=1}^{\infty} s^k P(X_\infty + Z = k) + P(X_\infty + Z = 0) \\ &= s^{-1} g_\pi(s) g_Z(s) + (1 - s^{-1}) P(X_\infty + Z = 0),\end{aligned}$$

d.h.,

$$g_\pi(s) = \frac{(s - 1)P(X_\infty + Z = 0)}{s - g_Z(s)}. \quad (69)$$

– Weil

$$\lim_{s \uparrow 1} g_\pi(s) = 1 \quad \text{und} \quad \lim_{s \uparrow 1} \frac{d}{ds} g_Z(s) = \mathbb{E} Z$$

ergibt sich aus der Regel von L'Hospital, dass

$$1 = \frac{P(X_\infty + Z = 0)}{1 - \rho}.$$

– Somit ergibt sich (68) aus (69).

2. Geburts- und Todesprozesse mit einer reflektierenden Schranke

- Wir modifizieren das in Abschnitt 2.2.3 diskutierte Beispiel des Geburts- und Todesprozesses dahingehend, dass wir nun den (unendlichen) Zustandsraum $E = \{0, 1, \dots\}$ und die Übergangsmatrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & & & \\ q_1 & r_1 & p_1 & & \\ & q_2 & r_2 & p_2 & \\ & & \ddots & \ddots & \ddots \\ & & & q_i & r_i & p_i \\ & & & & \ddots & \ddots & \ddots \end{pmatrix} \quad (70)$$

betrachten, wobei $p_i > 0$, $q_i > 0$ und $p_i + q_i + r_i = 1$ für jedes $i \in \{1, 2, \dots\}$ vorausgesetzt wird.

- Das lineare Gleichungssystem $\boldsymbol{\alpha}^\top = \boldsymbol{\alpha}^\top \mathbf{P}$ hat dann die Form

$$\alpha_i = \begin{cases} p_{i-1}\alpha_{i-1} + r_i\alpha_i + q_{i+1}\alpha_{i+1}, & \text{falls } i > 0, \\ q_1\alpha_1, & \text{falls } i = 0. \end{cases} \quad (71)$$

- Ähnlich wie bei Geburts- und Todesprozessen mit zwei reflektierenden Schranken kann man zeigen, dass
 - das Gleichungssystem (71) eine (eindeutig bestimmte) Wahrscheinlichkeitslösung $\boldsymbol{\alpha}^\top$ besitzt, falls

$$\sum_{j=1}^{\infty} \frac{p_1 p_2 \cdots p_j}{q_1 q_2 \cdots q_{j+1}} < \infty, \quad (72)$$

- die Lösung $\boldsymbol{\alpha}^\top = (\alpha_0, \alpha_1, \dots)$ von (71) in diesem Fall gegeben ist durch

$$\alpha_i = \alpha_0 \frac{p_1 p_2 \cdots p_{i-1}}{q_1 q_2 \cdots q_i}, \quad \forall i > 0,$$

- wobei $\alpha_0 > 0$ durch die Normierungsbedingung $\sum_{i=0}^{\infty} \alpha_i = 1$ gegeben ist, d.h.

$$\alpha_0 \left(1 + \frac{1}{q_1} + \sum_{j=1}^{\infty} \frac{p_1 p_2 \cdots p_j}{q_1 q_2 \cdots q_{j+1}} \right) = 1$$

bzw.

$$\alpha_0 = \left(1 + \frac{1}{q_1} + \sum_{j=1}^{\infty} \frac{p_1 p_2 \cdots p_j}{q_1 q_2 \cdots q_{j+1}} \right)^{-1}.$$

- Weil wir voraussetzen, dass $p_i > 0$ und $q_i > 0$ für jedes $i \in \{1, 2, \dots\}$ gilt, sind Geburts- und Todesprozesse mit einer reflektierenden Schranke offenbar irreduzibel.
- Wenn zusätzlich $r_i > 0$ für ein $i \in \{1, 2, \dots\}$ gilt, dann sind Geburts- und Todesprozesse mit einer reflektierenden Schranke auch aperiodisch (sowie ergodisch, falls die Kontraktionsbedingung (72) erfüllt ist).

2.2.5 Direkte und iterative Berechnungsmethoden

Wir zeigen nun, wie die stationäre Anfangsverteilung α_0 ($= \pi = \lim_{n \rightarrow \infty} \alpha_n$) der Markov-Kette $\{X_n\}$ mit Methoden der linearen Algebra berechnet werden kann, falls die Übergangsmatrix $\mathbf{P}$ nicht speziell strukturiert (jedoch quasi-positiv) ist und falls die Anzahl ℓ der Zustände nicht zu groß ist.

Theorem 2.12

- Die Übergangsmatrix $\mathbf{P}$ der Markov-Kette $\{X_n\}$ sei quasi-positiv.
- Dann ist die Matrix $\mathbf{I} - \mathbf{P} + \mathbf{E}$ invertierbar, und die (eindeutig bestimmte) Wahrscheinlichkeitslösung $\pi = \lim_{n \rightarrow \infty} \alpha_n$ der Matrix-Gleichung $\pi^\top = \pi^\top \mathbf{P}$ ist gegeben durch

$$\pi^\top = \mathbf{e}^\top (\mathbf{I} - \mathbf{P} + \mathbf{E})^{-1}, \quad (73)$$

wobei $\mathbf{e} = (1, \dots, 1)^\top$ und sämtliche Eintragungen der $\ell \times \ell$ Matrix $\mathbf{E}$ gleich 1 sind.

Beweis

- Um zu beweisen, dass die Matrix $\mathbf{I} - \mathbf{P} + \mathbf{E}$ invertierbar ist, zeigen wir, dass die einzige Lösung der Gleichung

$$(\mathbf{I} - \mathbf{P} + \mathbf{E}) \mathbf{x} = \mathbf{0} \quad (74)$$

durch $\mathbf{x} = \mathbf{0}$ gegeben ist.

- Weil π der Gleichung $\pi^\top = \pi^\top \mathbf{P}$ genügt, gilt auch

$$\pi^\top (\mathbf{I} - \mathbf{P}) = \mathbf{0}. \quad (75)$$

- Somit folgt aus (74), dass

$$0 = \pi^\top (\mathbf{I} - \mathbf{P} + \mathbf{E}) \mathbf{x} = 0 + \pi^\top \mathbf{E} \mathbf{x},$$

d.h.

$$\pi^\top \mathbf{E} \mathbf{x} = 0. \quad (76)$$

- Andererseits gilt offenbar $\pi^\top \mathbf{E} = \mathbf{e}^\top$ und somit wegen (76) auch

$$\mathbf{e}^\top \mathbf{x} = 0 \quad \text{bzw.} \quad \mathbf{E} \mathbf{x} = \mathbf{0}. \quad (77)$$

- Hieraus und aus (74) folgt, dass $(\mathbf{I} - \mathbf{P}) \mathbf{x} = \mathbf{0}$ bzw. $\mathbf{P} \mathbf{x} = \mathbf{x}$.

- Damit gilt auch $\mathbf{x} = \mathbf{P}^n \mathbf{x}$ für jedes $n \geq 1$.

- Aus Theorem 2.4 ergibt sich nun, dass $\mathbf{P}^n \rightarrow \mathbf{\Pi}$,

- wobei $\mathbf{\Pi}$ die $\ell \times \ell$ Matrix bezeichnet, die aus den ℓ identischen (Zeilen-) Vektoren $\pi^\top$ besteht.

- Mit anderen Worten: Für $n \rightarrow \infty$ gilt

$$\mathbf{x} = \mathbf{P}^n \mathbf{x} \rightarrow \mathbf{\Pi} \mathbf{x},$$

d.h. $x_i = \sum_{j=1}^{\ell} \pi_j x_j$ für jedes $i = 1, \dots, \ell$.

- Weil die rechten Seiten dieser Gleichungen nicht von i abhängen, ergibt sich, dass $\mathbf{x} = c \mathbf{e}$ für eine Konstante $c \in \mathbb{R}$.

- Wegen (77) gilt außerdem, dass

$$0 = \mathbf{e}^\top \mathbf{x} = c \mathbf{e}^\top \mathbf{e} = c \ell$$

und somit $c = 0$ bzw. $\mathbf{x} = \mathbf{0}$.

- Die Matrix $\mathbf{I} - \mathbf{P} + \mathbf{E}$ ist also invertierbar.

- Wegen (75) ergibt sich schließlich, dass

$$\boldsymbol{\pi}^\top (\mathbf{I} - \mathbf{P} + \mathbf{E}) = \boldsymbol{\pi}^\top \mathbf{E} = \mathbf{e}^\top$$

bzw.

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

□

Beachte

- Bei einer größeren Anzahl ℓ von Zuständen kann die numerische Berechnung der inversen Matrix $(\mathbf{I} - \mathbf{P} + \mathbf{E})^{-1}$ in (73) schwierig sein.
- In diesem Fall ist es oft besser, die Matrix-Gleichung $\boldsymbol{\pi}^\top = \boldsymbol{\pi}^\top \mathbf{P}$ iterativ zu lösen.
- Wenn die Übergangsmatrix $\mathbf{P}$ quasi-positiv ist und somit $\pi_\ell > 0$ gilt, kann man dabei zunächst $\hat{\pi}_\ell = 1$ setzen und die modifizierte Gleichung

$$\hat{\boldsymbol{\pi}}^\top (\mathbf{I} - \hat{\mathbf{P}}) = \mathbf{b}^\top \quad (78)$$

lösen, wobei $\hat{\mathbf{P}} = (p_{ij})_{i,j=1,\dots,\ell-1}$ and $\hat{\boldsymbol{\pi}}^\top = (\hat{\pi}_1, \dots, \hat{\pi}_{\ell-1})$, $\mathbf{b}^\top = (p_{\ell 1}, \dots, p_{\ell, \ell-1})$.

- Die ursprünglich gesuchte Wahrscheinlichkeitsfunktion $\boldsymbol{\pi}^\top = (\pi_1, \dots, \pi_\ell)$ ist dann gegeben durch

$$\pi_i = \hat{\pi}_i / c, \quad \forall i = 1, \dots, \ell,$$

wobei $c = \hat{\pi}_1 + \dots + \hat{\pi}_\ell$.

- Bei der iterativen Lösung der modifizierten Matrixgleichung (78) wird die Tatsache genutzt, dass die Matrix $\mathbf{I} - \hat{\mathbf{P}}$ in (78) invertierbar ist und dass $(\mathbf{I} - \hat{\mathbf{P}})^{-1}$ in eine Potenzreihe entwickelt werden kann.
- Dies ergibt sich aus den folgenden zwei Hilfssätzen.

Lemma 2.4

- Sei $\mathbf{A}$ eine $\ell \times \ell$ Matrix, so dass $\mathbf{A}^n \rightarrow \mathbf{0}$ für $n \rightarrow \infty$.
- Dann ist die Matrix $\mathbf{I} - \mathbf{A}$ invertierbar, und für jedes $n = 1, 2, \dots$ gilt

$$\mathbf{I} + \mathbf{A} + \dots + \mathbf{A}^{n-1} = (\mathbf{I} - \mathbf{A})^{-1} (\mathbf{I} - \mathbf{A}^n). \quad (79)$$

Beweis

- Offenbar gilt für jedes $n = 1, 2, \dots$

$$\begin{aligned} (\mathbf{I} - \mathbf{A})(\mathbf{I} + \mathbf{A} + \dots + \mathbf{A}^{n-1}) &= \mathbf{I} + \mathbf{A} + \dots + \mathbf{A}^{n-1} - \mathbf{A} - \dots - \mathbf{A}^n \\ &= \mathbf{I} - \mathbf{A}^n. \end{aligned} \quad (80)$$

- Außerdem ist die Matrix $\mathbf{I} - \mathbf{A}^n$ für jedes hinreichend große n invertierbar, weil $\mathbf{A}^n \rightarrow \mathbf{0}$ vorausgesetzt wird.
- Somit gilt für jedes hinreichend große n

$$\begin{aligned} 0 &\neq \det(\mathbf{I} - \mathbf{A}^n) \\ &= \det\left((\mathbf{I} - \mathbf{A})(\mathbf{I} + \mathbf{A} + \dots + \mathbf{A}^{n-1})\right) \\ &= \det(\mathbf{I} - \mathbf{A}) \det(\mathbf{I} + \mathbf{A} + \dots + \mathbf{A}^{n-1}). \end{aligned}$$

- Hieraus folgt, dass $\det(\mathbf{I} - \mathbf{A}) \neq 0$ und dass somit die Matrix $\mathbf{I} - \mathbf{A}$ invertierbar ist.

- Die Behauptung (79) ergibt sich nun aus (80). □

Lemma 2.5

- Die stochastische Matrix $\mathbf{P}$ sei quasi-positiv, und $\hat{\mathbf{P}}$ sei die in (78) eingeführte $(\ell - 1) \times (\ell - 1)$ Matrix.
- Dann gilt $\hat{\mathbf{P}}^n \rightarrow \mathbf{0}$ für $n \rightarrow \infty$, die Matrix $\mathbf{I} - \hat{\mathbf{P}}$ ist invertierbar, und es gilt

$$(\mathbf{I} - \hat{\mathbf{P}})^{-1} = \sum_{n=0}^{\infty} \hat{\mathbf{P}}^n. \quad (81)$$

Beweis

- Wegen Lemma 2.4 genügt es zu zeigen, dass $\hat{\mathbf{P}}^n \rightarrow \mathbf{0}$.
- Weil vorausgesetzt wird, dass $\mathbf{P}$ quasi-positiv ist, gibt es eine natürliche Zahl $n_0 \geq 1$, so dass

$$\delta = \max_{i \in \hat{E}} \sum_{j \in \hat{E}} p_{ij}^{(n_0)} < 1, \quad \text{wobei } \hat{E} = \{1, \dots, \ell - 1\}.$$

- Außerdem gilt

$$(\hat{\mathbf{P}}^n)_{ij} = \sum_{i_1, \dots, i_{n-1} \in \hat{E}} p_{ii_1} p_{i_1 i_2} \dots p_{i_{n-1} j} \leq \sum_{i_1, \dots, i_{n-1} \in E} p_{ii_1} p_{i_1 i_2} \dots p_{i_{n-1} j} = (\mathbf{P}^n)_{ij}$$

und somit $0 \leq (\hat{\mathbf{P}}^n)_{ij} \leq (\mathbf{P}^n)_{ij} = p_{ij}^{(n)} < 1$ für jedes $n \geq n_0$; $i, j \in \hat{E}$.

- Mit der Darstellung $n = kn_0 + m$ für gewisse $k \geq 1$ und $m \in \{0, \dots, n_0 - 1\}$ gilt somit

$$\begin{aligned} (\hat{\mathbf{P}}^n)_{ij} &= \sum_{i_1, \dots, i_k \in \hat{E}} (\hat{\mathbf{P}}^{n_0})_{ii_1} (\hat{\mathbf{P}}^{n_0})_{i_1 i_2} \dots (\hat{\mathbf{P}}^{n_0})_{i_{k-1} i_k} (\hat{\mathbf{P}}^m)_{i_k j} \\ &\leq \sum_{i_1, \dots, i_k \in \hat{E}} p_{ii_1}^{(n_0)} p_{i_1 i_2}^{(n_0)} \dots p_{i_{k-1} i_k}^{(n_0)} \\ &= \sum_{i_1, \dots, i_{k-1} \in \hat{E}} p_{ii_1}^{(n_0)} p_{i_1 i_2}^{(n_0)} \dots p_{i_{k-2} i_{k-1}}^{(n_0)} \left(\sum_{i_k \in \hat{E}} p_{i_{k-1} i_k}^{(n_0)} \right) \\ &\leq \delta \sum_{i_1, \dots, i_{k-1} \in \hat{E}} p_{ii_1}^{(n_0)} p_{i_1 i_2}^{(n_0)} \dots p_{i_{k-2} i_{k-1}}^{(n_0)} \\ &\vdots \\ &\leq \delta^k. \end{aligned}$$

- Dies ergibt, dass $\lim_{n \rightarrow \infty} (\hat{\mathbf{P}}^n)_{ij} \leq \lim_{k \rightarrow \infty} \delta^k = 0$. □

Beachte

- Wegen Lemma 2.5 ist die Lösung $\hat{\boldsymbol{\pi}}^\top$ der Gleichung (78), d.h. $\hat{\boldsymbol{\pi}}^\top (\mathbf{I} - \hat{\mathbf{P}}) = \mathbf{b}^\top$, gegeben durch

$$\hat{\boldsymbol{\pi}}^\top = \mathbf{b}^\top \sum_{n=0}^{\infty} \hat{\mathbf{P}}^n, \quad (82)$$

wodurch eine iterative Berechnung von $\hat{\boldsymbol{\pi}}^\top = (\hat{\pi}_1, \dots, \hat{\pi}_{\ell-1})$ möglich ist. Dabei beginnen wir die Iteration mit $\mathbf{b}_0^\top = \mathbf{b}^\top$ und setzen dann $\mathbf{b}_{n+1}^\top = \mathbf{b}_n^\top \hat{\mathbf{P}}$ für jedes $n \geq 0$.

- Somit lässt sich (82) umformulieren zu

$$\hat{\boldsymbol{\pi}}^\top = \sum_{n=0}^{\infty} \mathbf{b}_n^\top, \quad (83)$$

und die Größe $\sum_{n=0}^{n_0} \mathbf{b}_n^\top$ kann für hinreichend großes $n_0 \geq 1$ als Näherung von $\hat{\boldsymbol{\pi}}^\top$ dienen.

2.3 Abschätzung der Konvergenzgeschwindigkeit; Reversibilität

2.3.1 Definition und Beispiele

- Eine stationäre Markov-Kette $X_0, X_1, \dots : \Omega \rightarrow E$ bzw. das zugehörige Paar $(\mathbf{P}, \alpha)$ bestehend aus der Übergangsmatrix $\mathbf{P}$ und der (stationären) Anfangsverteilung α heißt *reversibel*, wenn ihre endlich-dimensionalen Verteilungen nicht von der Orientierung der Zeitachse abhängen, d.h., wenn

$$P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i_n) = P(X_n = i_0, X_{n-1} = i_1, \dots, X_1 = i_{n-1}, X_0 = i_n) \quad (84)$$

für beliebige $n \geq 0$ und $i_0, \dots, i_n \in E$.

- Die Reversibilität von Markov-Ketten ist insbesondere bei der Konstruktion von dynamischen Simulationsalgorithmen eine nützliche Modelleigenschaft, vgl. die Abschnitte 3.3 bis 3.5.

Wir leiten zunächst eine einfache Charakterisierung der Reversibilität von stationären (jedoch nicht unbedingt ergodischen) Markov-Ketten her.

Theorem 2.13

- Sei $X_0, X_1, \dots : \Omega \rightarrow E$ eine Markov-Kette mit dem Zustandsraum E , mit der Übergangsmatrix $\mathbf{P} = (p_{ij})$ und der stationären Anfangsverteilung $\alpha = (\alpha_1, \alpha_2, \dots)^\top$, so dass $\alpha_i > 0$ für jedes $i \in E$.
- Die Markov-Kette ist genau dann reversibel, wenn

$$\alpha_i p_{ij} = \alpha_j p_{ji} \quad \text{für beliebige } i, j \in E. \quad (85)$$

Beweis

- Die Notwendigkeit der Bedingung (85) ergibt sich unmittelbar aus der Definitionsgleichung (84). Aus (84) folgt nämlich insbesondere, dass

$$P(X_0 = i, X_1 = j) = P(X_1 = i, X_0 = j) \quad \text{für beliebige } i, j \in E.$$

- Somit gilt auch

$$\begin{aligned} \alpha_i p_{ij} &= P(X_0 = i, X_1 = j) \\ &= P(X_1 = i, X_0 = j) \\ &= \alpha_j p_{ji}. \end{aligned}$$

- Wenn umgekehrt (85) gilt, dann ergibt sich aus der Definitionsgleichung (3) von Markov-Ketten, dass

$$\begin{aligned} &P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i_n) \\ &\stackrel{(3)}{=} \alpha_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \\ &\stackrel{(85)}{=} p_{i_1 i_0} \alpha_{i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \\ &\quad \vdots \\ &\stackrel{(85)}{=} p_{i_1 i_0} p_{i_2 i_1} \cdots p_{i_n i_{n-1}} \alpha_{i_n} \\ &= \alpha_{i_n} p_{i_n i_{n-1}} \cdots p_{i_2 i_1} p_{i_1 i_0} \\ &\stackrel{(3)}{=} P(X_0 = i_n, X_1 = i_{n-1}, \dots, X_{n-1} = i_1, X_n = i_0) \\ &= P(X_n = i_0, X_{n-1} = i_1, \dots, X_1 = i_{n-1}, X_0 = i_n). \end{aligned}$$

- Damit ist die Gültigkeit von (84) bewiesen. □

Beachte

- Im Beweis von Theorem 2.13 wird nicht benötigt,
 - dass die stationäre Markov-Kette $X_0, X_1, \dots$ ergodisch ist,
 - sondern lediglich, dass sämtliche Komponenten α_{0i} der stationären Anfangsverteilung $\alpha = (\alpha_1, \alpha_2, \dots)^\top$ positiv sind.
- Mit anderen Worten
 - Falls die Übergangsmatrix $\mathbf{P}$ nicht irreduzibel oder nicht aperiodisch ist und somit die Grenzverteilung π nicht existiert bzw. nicht eindeutig bestimmt ist,
 - dann bleibt die Gültigkeit von Theorem 2.13 dennoch erhalten, wenn α eine beliebige stationäre Anfangsverteilung ist mit $\alpha_i > 0$ für jedes $i \in E$.
- Weil $\mathbf{P} = (p_{ij})$ eine stochastische Matrix ist, ergibt sich aus (85), dass für beliebige $i \in E$

$$\alpha_i = \alpha_i \sum_{j \in E} p_{ij} = \sum_{j \in E} \alpha_i p_{ij} \stackrel{(85)}{=} \sum_{j \in E} \alpha_j p_{ji}.$$

- Mit anderen Worten: Jede Anfangsverteilung α , die der sogenannten *Detailed-Balance-Bedingung* (85) genügt, ist notwendigerweise eine stationäre Anfangsverteilung, d.h., sie genügt dann auch der „globalen“ Balance-Bedingung $\alpha^\top = \alpha^\top \mathbf{P}$.

Beispiele

1. Diffusionsmodell

- Wir kehren zu dem bereits in Abschnitt 2.2.3 betrachteten Diffusionsmodell mit dem endlichen Zustandsraum $E = \{0, 1, \dots, \ell\}$, der irreduziblen (jedoch nicht aperiodischen) Übergangsmatrix $\mathbf{P} = (p_{ij})$, wobei

$$p_{ij} = \begin{cases} \frac{\ell - i}{\ell}, & \text{falls } i < \ell \text{ und } j = i + 1, \\ \frac{i}{\ell}, & \text{falls } i > 0 \text{ und } j = i - 1, \\ 0, & \text{sonst,} \end{cases} \quad (86)$$

und der (gemäß Theorem 2.10 eindeutig bestimmten, jedoch nicht ergodischen) stationären Anfangsverteilung

$$\alpha^\top = (\alpha_0, \dots, \alpha_\ell), \quad \text{wobei } \alpha_i = \frac{1}{2^\ell} \binom{\ell}{i}, \quad \forall i \in \{0, 1, \dots, \ell\}. \quad (87)$$

- Man kann zeigen, dass dann

$$\alpha_i p_{ij} = \alpha_j p_{ji}$$

für beliebige $i, j \in E$ gilt, d.h., das in (86) bzw. (87) gegebene Paar $(\mathbf{P}, \alpha)$ ist reversibel.

2. Geburts- und Todesprozesse

- Für die ebenfalls in Abschnitt 2.2.3 betrachteten Geburts- und Todesprozesse mit einer oder zwei reflektierenden Schranken sei die Übergangsmatrix $\mathbf{P} = (p_{ij})$ so beschaffen, dass die Gleichung $\alpha^\top = \alpha^\top \mathbf{P}$ eine (eindeutig bestimmte) Wahrscheinlichkeitslösung $\alpha^\top = (\alpha_1, \alpha_2, \dots)$ besitzt.
- Dann kann man zeigen, dass

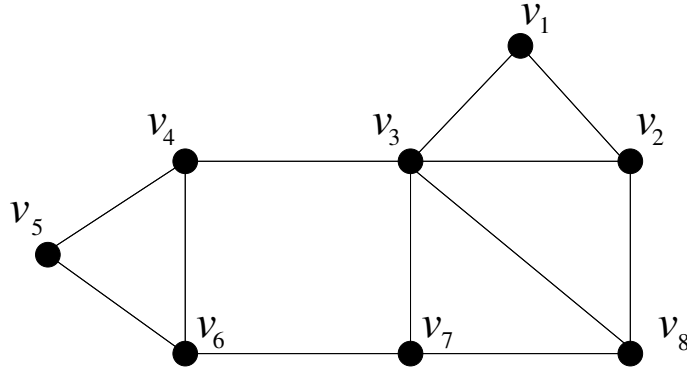
$$\alpha_i p_{ij} = \alpha_j p_{ji}$$

für beliebige $i, j \in E$ gilt.

3. Zufällige Irrfahrten auf Graphen

- Wir betrachten einen verbundenen Graph $G = (V, K)$
 - mit der Menge $V = \{v_1, \dots, v_\ell\}$ von ℓ Eckpunkten und der Menge K von Kanten, die jeweils zwei Eckpunkte miteinander verbinden,
 - so dass es für jedes Paar $v_i, v_j \in V$ von Eckpunkten einen „Pfad“ von Kanten aus K gibt, der die Eckpunkte v_i und v_j miteinander verbindet.
- In der Abbildung ist ein solcher Graph $G = (V, K)$ dargestellt, und zwar mit der Menge $V = \{v_1, \dots, v_8\}$ von 8 Eckpunkten und der Menge K von 12 Kanten, wobei

$$K = \{(v_1, v_2), (v_1, v_3), (v_2, v_3), (v_2, v_8), (v_3, v_4), (v_3, v_7), (v_3, v_8), (v_4, v_5), (v_4, v_6), (v_5, v_6), (v_6, v_7), (v_7, v_8)\}.$$



- Eine zufällige Irrfahrt auf dem Graph $G = (V, K)$ ist eine Markov-Kette $X_0, X_1, \dots : \Omega \rightarrow E$
 - mit dem Zustandsraum $E = \{1, \dots, \ell\}$ und der Übergangsmatrix $\mathbf{P} = (p_{ij})$, wobei

$$p_{ij} = \begin{cases} \frac{1}{d_i}, & \text{falls die Eckpunkte } v_i \text{ und } v_j \text{ Nachbarn sind,} \\ 0, & \text{sonst.} \end{cases} \quad (88)$$

- Dabei werden zwei Eckpunkte v_i und v_j *Nachbarn* genannt, falls sie die Endpunkte einundderselben Kante sind, und d_i bezeichnet die Anzahl der Nachbarn von v_i .
- Man kann zeigen (vgl. Übungsaufgabe 5.4), dass
 - die in (88) gegebene Übergangsmatrix irreduzibel ist,
 - die (gemäß Theorem 2.10 eindeutig bestimmte) stationäre Anfangsverteilung α gegeben ist durch

$$\alpha = \left(\frac{d_1}{d}, \dots, \frac{d_\ell}{d} \right)^\top, \quad \text{wobei } d = \sum_{i=1}^{\ell} d_i, \quad (89)$$

- das in (88)–(89) gegebene Paar $(\mathbf{P}, \alpha)$ reversibel ist, denn für beliebige $i, j \in \{1, \dots, \ell\}$ gilt

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

- Für das in der Abbildung betrachtete Zahlenbeispiel ist die in (88) gegebene Übergangsmatrix $\mathbf{P}$ nicht nur irreduzibel, sondern auch aperiodisch, und die stationäre Anfangsverteilung α ($= \pi = \lim_{n \rightarrow \infty} \alpha_n$) ist in diesem Fall gegeben durch

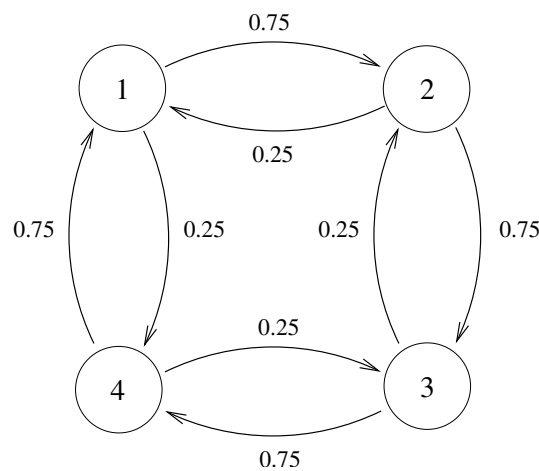
$$\alpha = \left(\frac{2}{24}, \frac{3}{24}, \frac{5}{24}, \frac{3}{24}, \frac{2}{24}, \frac{3}{24}, \frac{3}{24}, \frac{3}{24} \right)^\top.$$

4. Zyklische zufällige Irrfahrten

- Das folgende Beispiel einer zyklischen zufälligen Irrfahrt ist *nicht* reversibel.
 - Dabei sei $E = \{1, 2, 3, 4\}$ und

$$\mathbf{P} = \begin{pmatrix} 0 & 0.75 & 0 & 0.25 \\ 0.25 & 0 & 0.75 & 0 \\ 0 & 0.25 & 0 & 0.75 \\ 0.75 & 0 & 0.25 & 0 \end{pmatrix} \quad (90)$$

- d.h., der Übergangsgraph ist gegeben durch



- Die in (90) gegebene Übergangsmatrix ist offenbar irreduzibel, jedoch nicht aperiodisch, und die (wegen Theorem 2.10 eindeutig bestimmte) Anfangsverteilung α ist gegeben durch $\alpha = (1/4, 1/4, 1/4, 1/4)^\top$.
- Andererseits gilt jedoch

$$\alpha_1 p_{12} = \frac{1}{4} \frac{3}{4} = \frac{3}{16} > \frac{1}{16} = \frac{1}{4} \frac{1}{4} = \alpha_2 p_{21}.$$

- Es ist intuitiv klar, warum diese zyklische zufällige Irrfahrt nicht reversibel ist, denn die „Fortbewegung“ im Uhrzeigersinn ist viel wahrscheinlicher als die „Fortbewegung“ gegen den Uhrzeigersinn.

5. Doppelt-stochastische Übergangsmatrix

- Wir betrachten nun noch das folgende Beispiel einer Übergangsmatrix $\mathbf{P} = (p_{ij})$ und einer stationären Anfangsverteilung $\alpha = (\alpha_1, \dots, \alpha_\ell)^\top$, die *nicht* reversibel sind: Für $a, b > 0$ mit $b < a$ und $2a + b = 1$ sei

$$\mathbf{P} = \begin{pmatrix} a & a-b & 2b \\ a+b & b & a-b \\ 0 & a+b & a \end{pmatrix}. \quad (91)$$

- Diese Übergangsmatrix $\mathbf{P}$ ist *doppelt-stochastisch*, d.h., die transponierte Matrix $\mathbf{P}^\top$ ist ebenfalls eine stochastische Matrix, und $\mathbf{P}$ ist offenbar auch quasi-positiv.
- Für die (eindeutig bestimmte) stationäre Anfangsverteilung $\boldsymbol{\pi} = \lim_{n \rightarrow \infty} \boldsymbol{\alpha}_n$ gilt dann $\boldsymbol{\pi} = (1/3, 1/3, 1/3)^\top$.
- Weil jedoch die in (91) gegebene Übergangsmatrix $\mathbf{P}$ nicht symmetrisch ist, ist somit $(\mathbf{P}, \boldsymbol{\alpha})$ nicht reversibel.

2.3.2 Rekursive Konstruktion der „Vergangenheit“

- *Zur Erinnerung*

- In Abschnitt 2.1.3 hatten wir gezeigt, dass sich eine stationäre Markov-Kette $X_0, X_1, \dots$ mit der Übergangsmatrix $\mathbf{P} = (p_{ij})$ und der stationären Anfangsverteilung $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_\ell)^\top$ wie folgt konstruieren lässt.
- Dabei sind wir von einer Folge $Z_0, Z_1, \dots$ von unabhängigen und identisch auf $[0, 1]$ -gleichverteilten Zufallsvariablen ausgegangen und definierten zunächst

$$X_0 = k \quad \text{genau dann, wenn} \quad Z_0 \in \left(\sum_{i=1}^{k-1} \alpha_i, \sum_{i=1}^k \alpha_i \right],$$

für jedes $k = 1, \dots, \ell$, d.h.,

$$X_0 = \sum_{k=1}^{\ell} k \mathbb{I} \left(\sum_{i=1}^{k-1} \alpha_i < Z_0 \leq \sum_{i=1}^k \alpha_i \right). \quad (92)$$

- Die Zufallsvariablen $X_1, X_2, \dots$ wurden dann rekursiv definiert: Und zwar setzten wir

$$X_n = \varphi(X_{n-1}, Z_n) \quad \text{für } n = 1, 2, \dots, \quad (93)$$

wobei die Funktion $\varphi : E \times [0, 1] \rightarrow E$ gegeben war durch

$$\varphi(i, z) = \sum_{k=1}^{\ell} k \mathbb{I} \left(\sum_{j=1}^{k-1} p_{ij} < z \leq \sum_{j=1}^k p_{ij} \right). \quad (94)$$

- Wenn das Paar $(\mathbf{P}, \boldsymbol{\alpha})$ reversibel ist, dann lässt sich die in (92)–(94) konstruierte stationäre Markov-Kette $X_0, X_1, \dots$ wie folgt in die „Vergangenheit“ fortsetzen.
 - Und zwar setzen wir zunächst die Folge $Z_0, Z_1, \dots$ von unabhängigen und identisch auf $[0, 1]$ -gleichverteilten Zufallsvariablen zu einer beidseitig unendlichen Folge $\dots, Z_{-1}, Z_0, Z_1, \dots$ von unabhängigen und identisch auf $[0, 1]$ -gleichverteilten Zufallsvariablen fort.
 - Wegen der angenommenen Unabhängigkeit dieser Zufallsvariablen ist dies unproblematisch, denn der Wahrscheinlichkeitsraum, über dem die Zufallsvariablen $\dots, Z_{-1}, Z_0, Z_1, \dots$ definiert sind, kann durch die Bildung eines entsprechenden Produktraumes bzw. Produktmaßes konstruiert werden.
 - Die Zufallsvariablen $X_{-1}, X_{-2}, \dots$ werden nun rekursiv definiert: Und zwar setzen wir

$$X_{n-1} = \varphi(X_n, Z_{n-1}) \quad \text{für } n = 0, -1, \dots, \quad (95)$$

wobei die Funktion $\varphi : E \times [0, 1] \rightarrow E$ in (94) gegeben ist.

Theorem 2.14

- Sei $X_0, X_1, \dots : \Omega \rightarrow E$ eine reversible Markov-Kette mit dem Zustandsraum E , mit der Übergangsmatrix $\mathbf{P} = (p_{ij})$ und der stationären Anfangsverteilung $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_\ell)^\top$, so dass $\alpha_i > 0$ für jedes $i \in E$.

- Dann ist die durch (92)–(95) gegebene, beidseitig unendliche Folge $\dots, X_{-1}, X_0, X_1, \dots : \Omega \rightarrow E$ eine stationäre Markov-Kette mit der Übergangsmatrix $\mathbf{P} = (p_{ij})$ und der eindimensionalen Randverteilung α .
- D.h., für beliebige $k \in \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ bzw. $i_k, i_{k+1}, \dots, i_n \in E$ und $m \geq 1$ gilt

$$\begin{aligned} P(X_k = i_k, X_{k+1} = i_{k+1}, \dots, X_{n-1} = i_{n-1}, X_n = i_n) \\ &= P(X_{k+m} = i_k, X_{k+m+1} = i_{k+1}, \dots, X_{n+m-1} = i_{n-1}, X_{n+m} = i_n) \\ &= \alpha_{i_k} p_{i_k i_{k+1}} \dots p_{i_{n-1} i_n}. \end{aligned}$$

Der Beweis von Theorem 2.14 verläuft ähnlich wie der Beweis der Theoreme 2.11 bzw. 2.13 und wird deshalb weggelassen.

2.3.3 Bestimmung der Konvergenzgeschwindigkeit bei Reversibilität

- Sei $E = \{1, \dots, \ell\}$, und sei $\mathbf{P}$ eine quasi-positive (d.h. irreduzible und aperiodische) Übergangsmatrix.
 - Für den Fall, dass sämtliche Eigenwerte $\theta_1, \dots, \theta_\ell$ von $\mathbf{P}$ voneinander verschieden sind, hatten wir mit Hilfe des Theorems von Perron-Frobenius gezeigt (vgl. Korollar 2.4), dass

$$\max_{j \in E} |\alpha_{nj} - \pi_j| = O(|\theta_2|^n), \quad (96)$$

wobei $\pi = (\pi_1, \dots, \pi_\ell)^\top$ die (eindeutig bestimmte) Lösung der Gleichung $\pi^\top = \pi^\top \mathbf{P}$ ist.

- Falls $(\mathbf{P}, \pi)$ zusätzlich reversibel ist, dann kann man zeigen, dass die in (96) betrachtete Basis $|\theta_2|$ der Konvergenzgeschwindigkeit *nicht* verbessert werden kann.
- Sei also $(\mathbf{P}, \pi)$ reversibel, wobei $\mathbf{P}$ eine irreduzible und aperiodische Übergangsmatrix ist.
 - Dann ergibt sich aus der Detailed-Balance-Bedingung (85), dass die Matrix $\mathbf{D}\mathbf{P}\mathbf{D}^{-1}$ symmetrisch ist, wobei $\mathbf{D} = \text{diag}(\sqrt{\pi_i})$.
 - Weil die Eigenwerte $\theta_1, \dots, \theta_\ell$ von $\mathbf{P}$ mit den Eigenwerten von $\mathbf{D}\mathbf{P}\mathbf{D}^{-1}$ übereinstimmen, gilt somit $\theta_i \in \mathbb{R}$ für jedes $i \in E$,
 - und die rechten Eigenvektoren $\phi_1^*, \dots, \phi_\ell^*$ von $\mathbf{D}\mathbf{P}\mathbf{D}^{-1}$ können so gewählt werden, dass ihre Komponenten reellwertig sind,
 - dass $\phi_1^*, \dots, \phi_\ell^*$ gleichzeitig linke Eigenvektoren von $\mathbf{D}\mathbf{P}\mathbf{D}^{-1}$ sind und dass sowohl die Zeilen als auch die Spalten der $\ell \times \ell$ Matrix $(\phi_1^*, \dots, \phi_\ell^*)$ orthonormale Vektoren sind (vgl. beispielsweise Lemma 3.8 des Vorlesungsskriptes „Statistik II“).
- Aus der Spektraldarstellung (30) von $\mathbf{A} = \mathbf{D}\mathbf{P}\mathbf{D}^{-1}$ ergibt sich nun, dass für jedes $n \geq 1$

$$\mathbf{P}^n = (\mathbf{D}^{-1}\mathbf{A}\mathbf{D})^n = \mathbf{D}^{-1}\mathbf{A}^n\mathbf{D} = \sum_{k=1}^{\ell} \theta_k^n \mathbf{D}^{-1} \phi_k^* (\phi_k^*)^\top \mathbf{D}.$$

- Mit $\theta_1 = 1$ und $\phi_1^* = (\sqrt{\pi_1}, \dots, \sqrt{\pi_\ell})^\top$ ergibt sich hieraus, dass für beliebige $i, j \in E$

$$p_{ij}^{(n)} = \pi_j + \sqrt{\frac{\pi_j}{\pi_i}} \sum_{k=2}^{\ell} \theta_k^n \phi_{ki}^* \phi_{kj}^*, \quad \text{wobei } \phi_k^* = (\phi_{k1}^*, \dots, \phi_{k\ell}^*)^\top. \quad (97)$$

- Falls n eine gerade Zahl ist bzw. sämtliche Eigenwerte $\theta_2, \dots, \theta_\ell$ nichtnegativ sind, dann gilt somit

$$\sup_{\alpha_0} \max_{j \in E} |\alpha_{nj} - \pi_j| \geq \max_{j \in E} |p_{jj}^{(n)} - \pi_j| = \max_{j \in E} \left| \sum_{k=2}^{\ell} \theta_k^n (\phi_{kj}^*)^2 \right| \geq \theta_2^n \max_{j \in E} (\phi_{2j}^*)^2.$$

- Hieraus folgt, dass $|\theta_2|$ die kleinste positive Zahl ist, für die die in (96) betrachtete Abschätzung der Konvergenzgeschwindigkeit gleichmäßig für alle Anfangsverteilungen α_0 gilt.

Beachte

- Aus (97) ergibt sich, dass die Konvergenzabschätzung (96) wie folgt präzisiert werden kann. Und zwar gilt

$$|p_{ij}^{(n)} - \pi_j| \leq \frac{1}{\sqrt{\min_{i \in E} \pi_i}} \sum_{k=2}^{\ell} |\theta_k|^n |\phi_{ki}^*| |\phi_{kj}^*| \leq \frac{\sum_{k=2}^{\ell} |\phi_{ki}^*| |\phi_{kj}^*|}{\sqrt{\min_{i \in E} \pi_i}} |\theta_2|^n \leq \frac{1}{\sqrt{\min_{i \in E} \pi_i}} |\theta_2|^n,$$

weil die Spaltenvektoren $\phi_1^*, \dots, \phi_\ell^*$ und damit auch die Zeilenvektoren $(\phi_{1,j}, \dots, \phi_{\ell,j})$ für $j = 1, \dots, \ell$ eine orthonormale Basis in $\mathbb{R}^\ell$ bilden und sich deshalb aus der Ungleichung von Cauchy-Schwarz ergibt, dass

$$\sum_{k=2}^{\ell} |\phi_{ki}^*| |\phi_{kj}^*| \leq \left(\sum_{k=1}^{\ell} (\phi_{ki}^*)^2 \right)^{1/2} \left(\sum_{k=1}^{\ell} (\phi_{kj}^*)^2 \right)^{1/2} = 1.$$

- Somit gilt

$$\max_{j \in E} |\alpha_{nj} - \pi_j| \leq \frac{1}{\sqrt{\min_{i \in E} \pi_i}} |\theta_2|^n. \quad (98)$$

- Die praktische Nutzbarkeit der Konvergenzabschätzung (98) kann jedoch aus mehreren Gründen problematisch sein, weil
 - der Vorfaktor in (98) nicht von der Wahl der Anfangsverteilung α_0 abhängt,
 - bei der Herleitung der Konvergenzabschätzung (98) vorausgesetzt wird, dass die Markov-Kette reversibel ist,
 - die Bestimmung des Eigenwertes θ_2 schwierig sein kann, falls die Anzahl der Zustände ℓ groß ist.
- In Abschnitt 2.3.5 betrachten wir deshalb eine *alternative Konvergenzabschätzung*,
 - deren Vorfaktor von der gewählten Anfangsverteilung abhängt und
 - für die nicht vorausgesetzt wird, dass die betrachtete Markov-Kette reversibel ist,
 - und leiten in Abschnitt 2.3.7 eine Schranke für den zweitgrößten Betrag $|\theta_2|$ der Eigenwerte von reversiblen Übergangsmatrizen her.

2.3.4 Multiplikativ reversible Version der Übergangsmatrix; Spektraldarstellung

Wir diskutieren zunächst eine Methode, mit deren Hilfe (ergodische) Übergangsmatrizen so transformiert werden können, dass sich die Eigenschaft der Reversibilität einstellt.

- Sei $\mathbf{P} = (p_{ij})$ eine irreduzible und aperiodische (jedoch nicht notwendig reversible) Übergangsmatrix, und sei $\boldsymbol{\pi} = (\pi_1, \dots, \pi_\ell)^\top$ die zugehörige stationäre Anfangsverteilung, so dass $\pi_i > 0$ für jedes $i \in E$.
- Außerdem betrachten wir die stochastische Matrix $\tilde{\mathbf{P}} = (\tilde{p}_{ij})$ mit

$$\tilde{p}_{ij} = \frac{\pi_j p_{ji}}{\pi_i}, \quad (99)$$

d.h., $\tilde{\mathbf{P}} = \mathbf{D}^{-2} \mathbf{P}^\top \mathbf{D}^2$ mit $\mathbf{D} = \text{diag}(\sqrt{\pi_i})$ ist ebenfalls eine irreduzible und aperiodische Übergangsmatrix mit der gleichen stationären Anfangsverteilung $\boldsymbol{\pi} = (\pi_1, \dots, \pi_\ell)^\top$.

- Das Paar $(\mathbf{M}, \pi)$ mit $\mathbf{M} = \mathbf{P}\tilde{\mathbf{P}}$ ist reversibel, denn für die stochastische Matrix $\mathbf{M} = (m_{ij})$ gilt

$$\pi_i m_{ij} = \pi_i \sum_{k=1}^{\ell} p_{ik} \frac{\pi_j p_{jk}}{\pi_k} = \pi_j \sum_{k=1}^{\ell} p_{jk} \frac{\pi_i p_{ik}}{\pi_k} = \pi_j m_{ji}.$$

Definition Die Matrix $\mathbf{M} = \mathbf{P}\tilde{\mathbf{P}}$ heißt die *multiplikativ reversible* Version der Übergangsmatrix $\mathbf{P}$.

Beachte

- Sämtliche Eigenwerte $\theta_{\mathbf{M},1}, \dots, \theta_{\mathbf{M},\ell}$ von $\mathbf{M}$ sind reell und liegen im Intervall $[0, 1]$, denn $\mathbf{M}$ hat die gleichen Eigenwerte wie die symmetrische und nichtnegativ definite Matrix $\mathbf{M}^* = \mathbf{D}\mathbf{M}\mathbf{D}^{-1}$, wobei

$$m_{ij}^* = \frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} m_{ij} = \frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} \sum_{k=1}^{\ell} p_{ik} \frac{\pi_j p_{jk}}{\pi_k} = \sum_{k=1}^{\ell} \left(\frac{\sqrt{\pi_i}}{\sqrt{\pi_k}} p_{ik} \right) \left(\frac{\sqrt{\pi_j}}{\sqrt{\pi_k}} p_{jk} \right)$$

und somit

$$\mathbf{M}^* = \mathbf{D}\mathbf{M}\mathbf{D}^{-1} = (\mathbf{D}\mathbf{P}\mathbf{D}^{-1})(\mathbf{D}\tilde{\mathbf{P}}\mathbf{D}^{-1})^{\top}.$$

- Die symmetrische Matrix $\mathbf{M}^*$ ist also diagonalisierbar, und die rechten bzw. linken (zu $\theta_i^* = \theta_{\mathbf{M},i}$ gehörenden) Eigenvektoren ϕ_i^* bzw. ψ_i^* können so gewählt werden, dass
 - $\phi_i^* = \psi_i^*$ für jedes $i \in E$ gilt und
 - die Vektoren $\phi_1^*, \dots, \phi_{\ell}^*$ eine orthonormale Basis in $\mathbb{R}^{\ell}$ bilden.
- Dann sind $\phi_1, \dots, \phi_{\ell}$ bzw. $\psi_1, \dots, \psi_{\ell}$ mit

$$\phi_i = \mathbf{D}^{-1}\phi_i^* \quad \text{bzw.} \quad \psi_i = \mathbf{D}\psi_i^*, \quad \forall i \in E, \quad (100)$$

rechte bzw. linke Eigenvektoren von $\mathbf{M}$, denn für jedes $i \in E$ gilt

$$\mathbf{M}\phi_i = \mathbf{M}\mathbf{D}^{-1}\phi_i^* = \mathbf{D}^{-1}\mathbf{D}\mathbf{M}\mathbf{D}^{-1}\phi_i^* = \mathbf{D}^{-1}\theta_{\mathbf{M},i}\phi_i^* = \theta_{\mathbf{M},i}\phi_i$$

bzw.

$$\psi_i^{\top}\mathbf{M} = (\mathbf{D}\psi_i^*)^{\top}\mathbf{M} = (\psi_i^*)^{\top}\mathbf{D}\mathbf{M}\mathbf{D}^{-1}\mathbf{D} = \theta_{\mathbf{M},i}(\psi_i^*)^{\top}\mathbf{D} = \theta_{\mathbf{M},i}\psi_i^{\top}.$$

Hieraus ergibt sich insbesondere die folgende *Spektraldarstellung* der multiplikativ reversiblen Version $\mathbf{M}$ der Übergangsmatrix $\mathbf{P}$; vgl. auch die in Formel (30) gegebene Spektraldarstellung.

Theorem 2.15 Für beliebige $n \in \mathbb{N}$ und $\mathbf{x} \in \mathbb{R}^{\ell}$ gilt

$$\mathbf{M}^n \mathbf{x} = \sum_{i=1}^{\ell} \theta_{\mathbf{M},i}^n \phi_i \psi_i^{\top} \mathbf{x}. \quad (101)$$

wobei ϕ_i und ψ_i die in (100) definierten (rechten bzw. linken) Eigenvektoren von $\mathbf{M}$ sind.

Beweis

- Weil die in (100) definierten (rechten) Eigenvektoren $\phi_1, \dots, \phi_{\ell}$ von $\mathbf{M}$ ebenfalls eine Basis in $\mathbb{R}^{\ell}$ bilden, gibt es für jedes $\mathbf{x} \in \mathbb{R}^{\ell}$ einen (eindeutig bestimmten) Vektor $(x_1^{(r)}, \dots, x_{\ell}^{(r)})^{\top} \in \mathbb{R}^{\ell}$, so dass

$$\mathbf{x} = \sum_{i=1}^{\ell} x_i^{(r)} \phi_i.$$

- Außerdem gilt $\mathbf{M}\phi_i = \theta_{\mathbf{M},i}\phi_i$ und somit auch $\mathbf{M}^n \phi_i = \theta_{\mathbf{M},i}^n \phi_i$ für beliebige $i \in E$ und $n \in \mathbb{N}$.

- Hieraus folgt, dass

$$\mathbf{M}^n \mathbf{x} = \sum_{i=1}^{\ell} x_i^{(r)} \mathbf{M}^n \phi_i = \sum_{i=1}^{\ell} x_i^{(r)} \theta_{\mathbf{M},i}^n \phi_i.$$

- Andererseits ergibt sich aus (100), dass für beliebige $i \in E$ und $\mathbf{x} \in \mathbb{R}^{\ell}$

$$\psi_i^{\top} \mathbf{x} = (\mathbf{D} \psi_i^*)^{\top} \mathbf{x} = (\psi_i^*)^{\top} \sum_{j=1}^{\ell} x_j^{(r)} \mathbf{D} \phi_j = (\psi_i^*)^{\top} \sum_{j=1}^{\ell} x_j^{(r)} \phi_j^* = \sum_{j=1}^{\ell} x_j^{(r)} (\psi_i^*)^{\top} \phi_j^* = x_i^{(r)}, \quad (102)$$

wobei in der letzten Gleichheit berücksichtigt wurde, dass $\psi_i^* = \phi_i^*$ für jedes $i \in E$ gilt und dass die Eigenvektoren $\phi_1^*, \dots, \phi_{\ell}^*$ von $\mathbf{M}^*$ eine orthonormale Basis in $\mathbb{R}^{\ell}$ bilden.

- Damit ist die Spektraldarstellung (101) bewiesen. $\square$

2.3.5 Alternative Abschätzung der Konvergenzgeschwindigkeit; χ^2 -Kontrast

Mit Hilfe der multiplikativ reversiblen Version $\mathbf{M} = \mathbf{P} \tilde{\mathbf{P}}$ der (ergodischen, jedoch nicht notwendig reversiblen) Übergangsmatrix $\mathbf{P}$ leiten wir nun eine alternative Abschätzung für die Geschwindigkeit der Konvergenz $\alpha^{\top} \mathbf{P}^n \rightarrow \pi^{\top}$ her, wenn $n \rightarrow \infty$; vgl. Theorem 2.16.

Dabei erweisen sich die folgenden (abkürzenden) Bezeichnungen und Hilfssätze im Beweis von Theorem 2.16 als nützlich.

- Sei $\mathcal{L}(E)$ die Familie aller Funktionen,
 - die auf E definiert sind und die in die reelle Zahlengerade $\mathbb{R}$ abbilden, und
 - sei $\pi = (\pi_1, \dots, \pi_{\ell})^{\top}$ eine beliebige positive Wahrscheinlichkeitsfunktion aus $\mathcal{L}(E)$, d.h., es gelte $\pi_i > 0$ für jedes $i \in E$ und $\sum_{i=1}^{\ell} \pi_i = 1$.
- Für beliebige Vektoren $\mathbf{x} = (x_1, \dots, x_{\ell})^{\top} \in \mathcal{L}(E)$ und $\mathbf{y} = (y_1, \dots, y_{\ell})^{\top} \in \mathcal{L}(E)$ bezeichnen wir mit $(\mathbf{x}, \mathbf{y})_{\pi}$ das *Skalarprodukt*

$$(\mathbf{x}, \mathbf{y})_{\pi} = \sum_{i=1}^{\ell} x_i y_i \pi_i \quad (103)$$

und mit $\|\mathbf{x}\|_{\pi}$ die zugehörige Norm, d.h.

$$\|\mathbf{x}\|_{\pi} = \sqrt{\sum_{i=1}^{\ell} x_i^2 \pi_i}$$

- Unter dem (bezüglich π gewichteten) *Mittel* $(\mathbf{x})_{\pi}$ bzw. der *Varianz* $\text{Var}_{\pi}(\mathbf{x})$ von $\mathbf{x} \in \mathcal{L}(E)$ verstehen wir die Größen

$$(\mathbf{x})_{\pi} = \sum_{i=1}^{\ell} x_i \pi_i \quad \left(= (\mathbf{x}, \mathbf{e})_{\pi} \right) \quad (104)$$

und

$$\text{Var}_{\pi}(\mathbf{x}) = \|\mathbf{x}\|_{\pi}^2 - (\mathbf{x})_{\pi}^2. \quad (105)$$

Lemma 2.6 Für jedes $\mathbf{x} \in \mathcal{L}(E)$ gilt die Identität

$$\text{Var}_{\pi}(\mathbf{x}) = \text{Var}_{\pi}(\tilde{\mathbf{P}}\mathbf{x}) + ((\mathbf{I} - \mathbf{M})\mathbf{x}, \mathbf{x})_{\pi}. \quad (106)$$

Beweis

- Mit der Schreibweise $\hat{\mathbf{x}} = \mathbf{x} - (\mathbf{x})_{\pi} \mathbf{e}$ gilt, dass $(\hat{\mathbf{x}})_{\pi} = 0$ und

$$(\tilde{\mathbf{P}}\hat{\mathbf{x}})_{\pi} = \sum_{i=1}^{\ell} \left(\sum_{j=1}^{\ell} \tilde{p}_{ij} (x_j - (\mathbf{x})_{\pi}) \right) \pi_i = \sum_{i,j=1}^{\ell} \tilde{p}_{ij} x_j \pi_i - (\mathbf{x})_{\pi} = \sum_{i,j=1}^{\ell} \pi_j p_{ji} x_j - (\mathbf{x})_{\pi} = 0,$$

wobei sich die vorletzte Gleichheit aus der Definitionsgleichung (99) der Matrix $\tilde{\mathbf{P}}$ ergibt.

- Hieraus folgt, dass

$$\|\hat{\mathbf{x}}\|_{\pi}^2 = \text{Var}_{\pi}(\hat{\mathbf{x}}) = \text{Var}_{\pi}(\mathbf{x}) \quad \text{und} \quad \|\tilde{\mathbf{P}}\hat{\mathbf{x}}\|_{\pi}^2 = \text{Var}_{\pi}(\tilde{\mathbf{P}}\hat{\mathbf{x}}) = \text{Var}_{\pi}(\tilde{\mathbf{P}}\mathbf{x}). \quad (107)$$

- Andererseits gilt

$$\begin{aligned} \|\tilde{\mathbf{P}}\hat{\mathbf{x}}\|_{\pi}^2 &= (\tilde{\mathbf{P}}\hat{\mathbf{x}}, \tilde{\mathbf{P}}\hat{\mathbf{x}})_{\pi} = \sum_{i=1}^{\ell} \left(\sum_{k=1}^{\ell} \tilde{p}_{ik} \hat{x}_k \right)^2 \pi_i \\ &= \sum_{i=1}^{\ell} \sum_{j,k=1}^{\ell} \tilde{p}_{ij} \underbrace{\tilde{p}_{ik}}_{=\frac{\pi_k p_{ki}}{\pi_i}} \hat{x}_j \hat{x}_k \pi_i \\ &= \sum_{k=1}^{\ell} \sum_{j=1}^{\ell} \underbrace{(\mathbf{P}\tilde{\mathbf{P}})_{kj}}_{=(\mathbf{P}\tilde{\mathbf{P}}\mathbf{x})_k} \hat{x}_j \hat{x}_k \pi_k \\ &= (\mathbf{P}\tilde{\mathbf{P}}\hat{\mathbf{x}}, \hat{\mathbf{x}})_{\pi} = (\mathbf{M}\hat{\mathbf{x}}, \hat{\mathbf{x}})_{\pi} \end{aligned}$$

und somit

$$\|\hat{\mathbf{x}}\|_{\pi}^2 - \|\tilde{\mathbf{P}}\hat{\mathbf{x}}\|_{\pi}^2 = \|\hat{\mathbf{x}}\|_{\pi}^2 - (\mathbf{M}\hat{\mathbf{x}}, \hat{\mathbf{x}})_{\pi} = ((\mathbf{I} - \mathbf{M})\hat{\mathbf{x}}, \hat{\mathbf{x}})_{\pi} = ((\mathbf{I} - \mathbf{M})\mathbf{x}, \mathbf{x})_{\pi},$$

weil $\mathbf{M}$ eine stochastische Matrix ist mit $\pi^{\top} \mathbf{M} = \pi^{\top}$ und weil deshalb $(\mathbf{I} - \mathbf{M})\mathbf{e} = \mathbf{0}$ bzw.

$$((\mathbf{I} - \mathbf{M})\mathbf{x}, \mathbf{e})_{\pi} = \sum_{i,j=1}^{\ell} (\delta_j(i) - m_{ij}) x_j \pi_i = \sum_{i=1}^{\ell} x_i \pi_i - \sum_{j=1}^{\ell} x_j \underbrace{\sum_{i=1}^{\ell} \pi_i m_{ij}}_{=\pi_j} = 0.$$

- Hieraus und aus (107) ergibt sich die Behauptung. □

Wir führen nun noch die folgenden Begriffe ein.

- Sei $E = \{1, \dots, \ell\}$, seien $\alpha = (\alpha_1, \dots, \alpha_{\ell})^{\top}$ und $\beta = (\beta_1, \dots, \beta_{\ell})^{\top}$ zwei beliebige Wahrscheinlichkeitsverteilungen auf E , und sei

$$d_{\text{TV}}(\alpha, \beta) = \frac{1}{2} \sum_{i \in E} |\alpha_i - \beta_i|, \quad (108)$$

d.h., der Abstand $d_{\text{TV}}(\alpha, \beta)$ zwischen α und β wird durch die *Totalvariation*

$$|\alpha - \beta| = \sum_{i \in E} |\alpha_i - \beta_i| \quad (109)$$

des „signierten Maßes“ $\alpha - \beta$ ausgedrückt.

- Falls $\beta_i > 0$ für jedes $i \in E$, dann betrachten wir auch die Größe

$$\chi^2(\alpha; \beta) = \sum_{i \in E} \frac{(\alpha_i - \beta_i)^2}{\beta_i} \quad (110)$$

die der χ^2 -Kontrast von α bezüglich β genannt wird.

Der Abstand $d_{TV}(\alpha, \beta)$ zwischen α und β kann wie folgt durch den χ^2 -Kontrast $\chi^2(\alpha; \beta)$ von α bezüglich β abgeschätzt werden.

Lemma 2.7 Falls $\beta_i > 0$ für jedes $i \in E$, dann gilt

$$d_{TV}^2(\alpha, \beta) \leq \frac{1}{4} \chi^2(\alpha; \beta). \quad (111)$$

Beweis

- Aus der Ungleichung von Cauchy-Schwarz ergibt sich unter Berücksichtigung von $\sum_{i \in E} \beta_i = 1$, dass

$$\left(\sum_{i \in E} |\alpha_i - \beta_i| \right)^2 = \left(\sum_{i \in E} \frac{1}{\sqrt{\beta_i}} |\alpha_i - \beta_i| \sqrt{\beta_i} \right)^2 \leq \sum_{i \in E} \frac{1}{\beta_i} (\alpha_i - \beta_i)^2.$$

- Hieraus folgt die Behauptung. □

Die Geschwindigkeit der Konvergenz $\alpha^\top \mathbf{P}^n \rightarrow \pi^\top$, wenn $n \rightarrow \infty$, lässt sich nun wie folgt

- mit Hilfe des zweitgrößten Eigenwertes $\theta_{\mathbf{M},2}$ der multiplikativ reversiblen Version $\mathbf{M} = \mathbf{P}\tilde{\mathbf{P}}$ der (ergodischen) Übergangsmatrix $\mathbf{P}$ sowie
- des χ^2 -Kontrastes $\chi^2(\alpha; \pi)$ der Anfangsverteilung α bezüglich der stationären Grenzverteilung π abschätzen.

Theorem 2.16 Für jede beliebige Anfangsverteilung α und für jedes $n \in \mathbb{N}$ gilt

$$d_{TV}^2((\alpha^\top \mathbf{P}^n)^\top, \pi) \leq \frac{\chi^2(\alpha; \pi)}{4} \theta_{\mathbf{M},2}^n. \quad (112)$$

Beweis

- Sei $\rho_n = (\rho_{n1}, \dots, \rho_{n\ell})^\top$ mit $\rho_{ni} = (\alpha^\top \mathbf{P}^n)_i / \pi_i$.
– Dann gilt für jedes $i \in E$

$$\sum_{k=1}^{\ell} \frac{\pi_k \rho_{ki}}{\pi_i} \frac{(\alpha^\top \mathbf{P}^n)_k}{\pi_k} = \frac{(\alpha^\top \mathbf{P}^{n+1})_i}{\pi_i}$$

und somit

$$\tilde{\mathbf{P}} \rho_n = \rho_{n+1}.$$

- Außerdem ergibt sich aus der Definitionsgleichung (110) des χ^2 -Kontrastes $\chi_n^2 = \chi^2((\boldsymbol{\alpha}^\top \mathbf{P}^n)^\top; \boldsymbol{\pi})$ von $(\boldsymbol{\alpha}^\top \mathbf{P}^n)^\top$ bezüglich $\boldsymbol{\pi}$, dass

$$\begin{aligned}\chi_n^2 &= \sum_{i=1}^{\ell} \frac{((\boldsymbol{\alpha}^\top \mathbf{P}^n)_i - \pi_i)^2}{\pi_i} = \sum_{i=1}^{\ell} \left(\frac{(\boldsymbol{\alpha}^\top \mathbf{P}^n)_i}{\pi_i} - 1 \right)^2 \pi_i \\ &= \sum_{i=1}^{\ell} (\rho_{ni} - (\rho_n)_\pi)^2 \pi_i = \text{Var}_{\boldsymbol{\pi}}(\rho_n),\end{aligned}$$

d.h.,

$$\chi_n^2 = \text{Var}_{\boldsymbol{\pi}}(\rho_n). \quad (113)$$

- Wegen der in Lemma 2.6 hergeleiteten Identität (106) gilt somit, dass

$$\chi_n^2 = \chi_{n+1}^2 + ((\mathbf{I} - \mathbf{M})\rho_n, \rho_n)_\pi. \quad (114)$$

- Andererseits ergibt sich aus der Spektraldarstellung (101) von $\mathbf{M}$, die in Theorem 2.15 hergeleitet wurde, dass

$$\begin{aligned}((\mathbf{I} - \mathbf{M})\rho_n, \rho_n)_\pi &= (\rho_n, \rho_n)_\pi - (\mathbf{M}\rho_n, \rho_n)_\pi \\ &= (\rho_n, \rho_n)_\pi - \sum_{i=1}^{\ell} \theta_{\mathbf{M},i} (\phi_i \psi_i^\top \rho_n, \rho_n)_\pi \\ &= (\rho_n, \rho_n)_\pi - 1 - \sum_{i=2}^{\ell} \theta_{\mathbf{M},i} (\phi_i \psi_i^\top \rho_n, \rho_n)_\pi,\end{aligned}$$

weil $\theta_{\mathbf{M},1} = 1$, $\phi_1 = \mathbf{e}$ bzw. $\psi_1^\top = \boldsymbol{\pi}^\top$ und somit

$$(\phi_1 \psi_1^\top \rho_n, \rho_n)_\pi = (\mathbf{I} \rho_n, \rho_n)_\pi = (\mathbf{e}, \rho_n)_\pi = (\rho_n)_\pi = 1 \quad \left(= (\rho_n)_\pi^2 \right).$$

- Weil die in (100) definierten (rechten) Eigenvektoren $\phi_1, \dots, \phi_\ell$ von $\mathbf{M}$ eine Basis in $\mathbb{R}^\ell$ bilden, gibt es also einen (eindeutig bestimmten) Vektor $(\rho_{n1}^{(r)}, \dots, \rho_{n\ell}^{(r)})^\top \in \mathbb{R}^\ell$, so dass $\rho_n = \sum_{i=1}^{\ell} \rho_{ni}^{(r)} \phi_i$.
- In (102) hatten wir darüber hinaus gezeigt, dass $\psi_i^\top \rho_n = \rho_{ni}^{(r)}$ gilt. Wegen $\psi_1^\top = \boldsymbol{\pi}^\top$ gilt somit insbesondere, dass $\psi_1^\top \rho_n = 1$.
- Weil außerdem $\phi_i = \mathbf{D}^{-1} \phi_i^*$ bzw. $\psi_i = \mathbf{D} \phi_i^*$ für jedes $i \in E$ gilt, ergibt sich hieraus, dass

$$\begin{aligned}\sum_{i=2}^{\ell} \theta_{\mathbf{M},i} (\phi_i \psi_i^\top \rho_n, \rho_n)_\pi &= \sum_{i=2}^{\ell} \sum_{j=1}^{\ell} \theta_{\mathbf{M},i} \rho_{ni}^{(r)} \rho_{nj}^{(r)} (\phi_i, \phi_j)_\pi \\ &= \sum_{i=2}^{\ell} \sum_{j=1}^{\ell} \theta_{\mathbf{M},i} \rho_{ni}^{(r)} \rho_{nj}^{(r)} \underbrace{(\mathbf{D}^{-1} \phi_i^*, \mathbf{D}^{-1} \phi_j^*)_\pi}_{=\delta_i(j)} \\ &= \sum_{i=2}^{\ell} \theta_{\mathbf{M},i} (\rho_{ni}^{(r)})^2 \\ &\leq \theta_{\mathbf{M},2} \sum_{i=2}^{\ell} (\rho_{ni}^{(r)})^2 = \theta_{\mathbf{M},2} \left(\sum_{i=1}^{\ell} (\rho_{ni}^{(r)})^2 - \underbrace{\left(\rho_{n1}^{(r)} \right)^2}_{=\psi_1^\top \rho_n = 1} \right) \\ &= \theta_{\mathbf{M},2} \left(\sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \rho_{ni}^{(r)} \rho_{nj}^{(r)} (\phi_i, \phi_j)_\pi - 1 \right) = \theta_{\mathbf{M},2} \underbrace{\left((\rho_n, \rho_n)_\pi - 1 \right)}_{=\text{Var}_{\boldsymbol{\pi}}(\rho_n)}.\end{aligned}$$

- Insgesamt haben wir also gezeigt, dass

$$((\mathbf{I} - \mathbf{M})\boldsymbol{\rho}_n, \boldsymbol{\rho}_n)_\pi \geq (1 - \theta_{\mathbf{M},2}) \text{Var}_\pi(\boldsymbol{\rho}_n).$$

- Wegen (113) und (114) folgt hieraus, dass

$$\chi_n^2 \geq \chi_{n+1}^2 + (1 - \theta_{\mathbf{M},2}) \chi_n^2 \quad \text{bzw.} \quad \chi_{n+1}^2 \leq \theta_{\mathbf{M},2} \chi_n^2.$$

- Somit gilt $\chi_n^2 \leq \theta_{\mathbf{M},2}^n \chi_0^2$ für jedes $n \geq 1$, und die Behauptung ergibt sich nun aus Lemma 2.7. $\square$

2.3.6 Dirichlet-Formen und Rayleigh-Theorem

- Sei $E = \{1, \dots, \ell\}$ eine (beliebige) endliche Menge, und sei $\mathbf{P}$ eine $(\ell \times \ell)$ -dimensionale Übergangsmatrix, die irreduzibel und aperiodisch (d.h. quasi-positiv) sowie reversibel ist.
- *Zur Erinnerung*
 - Sämtliche Eigenwerte von $\mathbf{P}$ sind reell (vgl. Abschnitt 2.3.3), und
 - aus dem Theorem von Perron-Frobenius (vgl. Theorem 2.6 bzw. Korollar 2.3) folgt darüber hinaus, dass die Eigenwerte von $\mathbf{P}$ im Intervall $(-1, 1]$ liegen,
 - wobei 1 der größte Eigenwert ist, und die Beträge der anderen Eigenwerte (strikt) kleiner als 1 sind.

Beachte

- In Abweichung von der bisher verwendeten Schreibweise (bei der die Beträge der Eigenwerte betrachtet wurden) ordnen wir nun die Eigenwerte *selbst* der Größe nach und bezeichnen sie mit $\lambda_1, \dots, \lambda_\ell$, so dass

$$1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_\ell > -1.$$

- Für die in Abschnitt 2.3.4 eingeführte multiplikativ reversible Version $\mathbf{M} = \mathbf{P}\tilde{\mathbf{P}}$ der Übergangsmatrix $\mathbf{P}$ gilt sogar

$$1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_\ell > 0,$$

d.h., für die Eigenwerte der Matrix $\mathbf{M}$ stimmen die beiden Notationen $\theta_1, \dots, \theta_\ell$ und $\lambda_1, \dots, \lambda_\ell$ überein.

- Wenn ℓ eine große Zahl ist,
 - dann kann die Berechnung des zweitgrößten Betrages $|\theta_2| = \max\{\lambda_2, |\lambda_\ell|\}$ der Eigenwerte schwierig sein,
 - weshalb wir in Abschnitt 2.3.7 Schranken für λ_2 und λ_ℓ herleiten, deren Berechnung einfacher ist.
- Diese Schranken sind insbesondere dann nützlich, wenn
 - die stationäre (Grenz-) Verteilung $\boldsymbol{\pi}$ zwar (zumindest prinzipiell) bekannt ist,
 - die zugehörige Markov-Kette dennoch mit einer instationären Anfangsverteilung $\boldsymbol{\alpha}$ „gestartet“ wird, beispielsweise in einem vorgegebenen Zustand $i \in E$, wobei dann $\alpha_i = 1$ und $\alpha_j = 0$ für $j \neq i$ gilt.

Um eine obere Schranke für λ_2 herzuleiten, benötigen wir eine Darstellungsformel des Eigenwertes λ_2 ,

- die in der Literatur das *Theorem von Rayleigh* genannt wird und
- die mit Hilfe der sogenannten *Dirichlet-Form*

$$D_{(\mathbf{P}, \boldsymbol{\pi})}(\mathbf{x}, \mathbf{x}) = ((\mathbf{I} - \mathbf{P})\mathbf{x}, \mathbf{x})_\pi \quad (115)$$

des reversiblen Paares $(\mathbf{P}, \boldsymbol{\pi})$ ausgedrückt wird, wobei $(\mathbf{y}, \mathbf{x})_\pi$ das in (103) eingeführte Skalarprodukt von $\mathbf{y}$ und $\mathbf{x}$ bezüglich $\boldsymbol{\pi}$ bezeichnet.

Dabei zeigen wir zunächst den folgenden Hilfssatz.

Lemma 2.8 Für jedes $\mathbf{x} = (x_1, \dots, x_\ell)^\top \in \mathbb{R}^\ell$ gilt

$$D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x}) = \frac{1}{2} \sum_{i,j \in E} \pi_i p_{ij} (x_j - x_i)^2. \quad (116)$$

Beweis Aus der Definitionsgleichung (103) des Skalarproduktes und aus der Reversibilität des Paares $(\mathbf{P}, \pi)$ ergibt sich, dass

$$\begin{aligned} 2((\mathbf{I} - \mathbf{P})\mathbf{x}, \mathbf{x})_\pi &= 2 \sum_{i,j \in E} \pi_i p_{ij} x_i (x_i - x_j) \\ &\stackrel{i \rightarrow j}{=} \sum_{i,j \in E} \pi_i p_{ij} x_i (x_i - x_j) + \sum_{i,j \in E} \pi_j p_{ji} x_j (x_j - x_i) \\ &\stackrel{(85)}{=} \sum_{i,j \in E} \pi_i p_{ij} x_i (x_i - x_j) + \sum_{i,j \in E} \pi_i p_{ij} x_j (x_j - x_i) \\ &= \sum_{i,j \in E} \pi_i p_{ij} (x_j - x_i)^2. \end{aligned} \quad \square$$

Wir beweisen nun das folgende *Theorem von Rayleigh*, das eine Darstellungsformel für den zweitgrößten Eigenwert λ_2 des reversiblen Paares $(\mathbf{P}, \pi)$ liefert.

Theorem 2.17

- Sei $\mathbb{R}_\neq^\ell = \{\mathbf{x} = (x_1, \dots, x_\ell)^\top \in \mathbb{R}^\ell : x_i \neq x_j \text{ für ein Paar } i, j \in E\}$ die Menge derjenigen Vektoren aus $\mathbb{R}^\ell$, für die nicht alle Komponenten gleich sind.
- Für den Eigenwert λ_2 des reversiblen Paares $(\mathbf{P}, \pi)$ gilt dann

$$\lambda_2 = 1 - \inf_{\mathbf{x} \in \mathbb{R}_\neq^\ell} \frac{D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x})}{\text{Var}_\pi(\mathbf{x})}, \quad (117)$$

wobei $\text{Var}_\pi(\mathbf{x})$ die in (105) definierte Varianz der Komponenten von $\mathbf{x}$ bezüglich π bezeichnet.

Beweis

- Aus Lemma 2.8 ergibt sich, dass für beliebige $c \in \mathbb{R}$ und $\mathbf{x} \in \mathbb{R}^\ell$

$$D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x}) = D_{(\mathbf{P}, \pi)}(\mathbf{x} - c\mathbf{e}, \mathbf{x} - c\mathbf{e}).$$

– Somit ist die Behauptung (117) äquivalent mit

$$1 - \lambda_2 = \inf_{\mathbf{x} \in \mathbb{R}_0^\ell} \frac{D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x})}{\text{Var}_\pi(\mathbf{x})}, \quad (118)$$

wobei $\mathbb{R}_0^\ell = \{\mathbf{x} = (x_1, \dots, x_\ell)^\top \in \mathbb{R}^\ell : (\mathbf{x})_\pi = 0, \mathbf{x} \neq \mathbf{0}\}$.

- Die rechten Eigenvektoren $\phi_1, \dots, \phi_\ell$ von $\mathbf{P}$ seien nun so gewählt, dass sie eine orthonormale Basis in $\mathbb{R}^\ell$ bezüglich des Skalarproduktes $(\cdot, \cdot)_\pi$ bilden, d.h., es gilt $(\phi_i, \phi_j)_\pi = 1$, falls $i = j$, und $(\phi_i, \phi_j)_\pi = 0$, falls $i \neq j$, wobei $\phi_1 = \mathbf{e}$.
- Dabei werden zunächst die Eigenvektoren $\phi_1^*, \dots, \phi_\ell^*$ der symmetrischen Matrix $\mathbf{D}\mathbf{P}\mathbf{D}^{-1}$ so gewählt, dass sie orthonormal bezüglich des gewöhnlichen euklidischen Skalarproduktes sind, wobei dann $\phi_i = \mathbf{D}^{-1}\phi_i^*$ für jedes $i \in E$ gesetzt wird (vgl. auch Abschnitt 2.3.3).

– Es gibt somit für jedes $\mathbf{x} \in \mathbb{R}^\ell$ einen (eindeutig bestimmten) Vektor $(x_1^{(r)}, \dots, x_\ell^{(r)})^\top \in \mathbb{R}^\ell$, so dass

$$\mathbf{x} = \sum_{i=1}^{\ell} x_i^{(r)} \phi_i.$$

– Wegen $\lambda_1 = 1$ folgt hieraus, dass

$$(\mathbf{I} - \mathbf{P})\mathbf{x} = \sum_{i=2}^{\ell} (1 - \lambda_i) x_i^{(r)} \phi_i \quad \text{bzw.} \quad D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x}) = \sum_{i=2}^{\ell} (1 - \lambda_i) (x_i^{(r)})^2.$$

- Andererseits ergibt sich aus $\phi_1 = \mathbf{e}$ und aus der Orthonormalität der Eigenvektoren $\phi_1, \dots, \phi_\ell$ bezüglich des Skalarproduktes $(\cdot, \cdot)_\pi$, dass

$$(\mathbf{x})_\pi = (\mathbf{x}, \mathbf{e})_\pi = x_1^{(r)} \quad \text{bzw.} \quad \text{Var}_\pi(\mathbf{x}) = \sum_{i=2}^{\ell} (x_i^{(r)})^2, \quad \text{falls } (\mathbf{x})_\pi = 0.$$

– Somit gilt für jedes $\mathbf{x} \in \mathbb{R}_0^\ell$, dass

$$\begin{aligned} \frac{D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x})}{\text{Var}_\pi(\mathbf{x})} &= \frac{\sum_{i=2}^{\ell} (1 - \lambda_i) (x_i^{(r)})^2}{\sum_{i=2}^{\ell} (x_i^{(r)})^2} \\ &= (1 - \lambda_2) + \frac{\sum_{i=2}^{\ell} (1 - \lambda_i) (x_i^{(r)})^2 - (1 - \lambda_2) \sum_{i=2}^{\ell} (x_i^{(r)})^2}{\sum_{i=2}^{\ell} (x_i^{(r)})^2} \\ &= (1 - \lambda_2) + \frac{\sum_{i=3}^{\ell} (\lambda_2 - \lambda_i) (x_i^{(r)})^2}{\sum_{i=2}^{\ell} (x_i^{(r)})^2} \geq 1 - \lambda_2. \end{aligned}$$

– Damit ist die (äquivalente) Behauptung (118) bewiesen, denn aus dem letzten Ausdruck für den Quotienten $D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x})/\text{Var}_\pi(\mathbf{x})$ ergibt sich außerdem, dass

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

für $\mathbf{x} = \phi_2$, wobei $\phi_2 \in \mathbb{R}_0^\ell$ gilt, weil $\phi_1 = \mathbf{e}$ und ϕ_2 linear unabhängig sind. □

2.3.7 Schranken für die Eigenwerte λ_2 und λ_ℓ

Um Schranken für die Eigenwerte λ_2 und λ_ℓ herzuleiten, benötigen wir nun noch die folgenden Begriffe und Bezeichnungen.

- Für jedes Paar $i, j \in E$ mit $i \neq j$ und $p_{ij} > 0$ bezeichnen wir
 - mit $e = e_{ij}$ die entsprechende (gerichtete) *Kante* des Übergangsgraphen, sowie
 - mit $e^- = i$ bzw. $e^+ = j$ die *Anfangsecke* bzw. *Endecke* von e .
 - Sei $\mathcal{E}$ die Menge aller gerichteten Kanten $e = e_{ij}$ mit $i \neq j$ und $p_{ij} > 0$.

- Außerdem betrachten wir für jedes Paar $i, j \in E$ mit $i \neq j$ genau einen *Pfad* γ_{ij} von i nach j ,
 - der ein Vektor $\gamma_{ij} = (i_0, i_1, \dots, i_{m-1}, i_m)$ von Zuständen ist mit $i = i_0, j = i_m$ und

$$p_{i i_1} p_{i_1 i_2} \cdots p_{i_{m-1} j} > 0,$$

so dass keine der Kanten $e_{i_{k-1} i_k}$ mehrfach auftritt.

- Sei Γ die Menge aller dieser Pfade, und für jeden Pfad $\gamma_{ij} \in \Gamma$ sei

$$|\gamma_{ij}| = \sum_{e \in \gamma_{ij}} \frac{1}{Q(e)} = \frac{1}{\pi_i p_{i i_1}} + \frac{1}{\pi_{i_1} p_{i_1 i_2}} + \dots + \frac{1}{\pi_{i_{m-1}} p_{i_{m-1} j}}, \quad (119)$$

wobei $Q(e_{i_{k-1} i_k}) = \pi_{i_{k-1}} p_{i_{k-1} i_k}$.

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

$$\kappa = \kappa(\Gamma) = \max_{e \in \mathcal{E}} \sum_{\gamma_{ij} \ni e} |\gamma_{ij}| \pi_i \pi_j. \quad (120)$$

- Schließlich betrachten wir

- die erweiterte Kanten-Menge $\mathcal{E}' \supset \mathcal{E}$, die auch die „Kanten“ $i \rightarrow i$ mit $p_{ii} > 0$ enthält.
- für jedes $i \in E$ genau einen Pfad γ_i von i nach i , der eine ungerade Anzahl von Kanten aus $\mathcal{E}'$ enthält, so dass keine der Kanten mehrfach auftritt.
- Sei Γ' die Menge aller dieser Pfade, und für jeden Pfad $\gamma_i \in \Gamma'$ sei

$$|\gamma_i| = \sum_{e \in \gamma_i} \frac{1}{Q(e)}. \quad (121)$$

- Der Koeffizient ζ der Pfadmenge Γ' ist dann gegeben durch

$$\zeta = \zeta(\Gamma') = \max_{e \in \mathcal{E}'} \sum_{\gamma_i \ni e} |\gamma_i| \pi_i. \quad (122)$$

Theorem 2.18 Für die Eigenwerte λ_2 und λ_ℓ von $\mathbf{P}$ gilt

$$1 - \frac{1}{\kappa} \geq \lambda_2 \geq \lambda_\ell \geq -1 + \frac{2}{\zeta} \quad (123)$$

und somit

$$\max\{\lambda_2, |\lambda_\ell|\} \leq 1 - \min\left\{\frac{1}{\kappa}, \frac{2}{\zeta}\right\}. \quad (124)$$

Beweis

- Wir zeigen zunächst, dass $\lambda_2 \leq 1 - \kappa^{-1}$.
 - Wegen Theorem 2.17 genügt es zu zeigen, dass

$$\text{Var}_\pi(\mathbf{x}) \leq \kappa D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^\ell. \quad (125)$$

- Mit den in (119) eingeführten Bezeichnungen gilt

$$\begin{aligned} \text{Var}_\pi(\mathbf{x}) &= \frac{1}{2} \left(2\|\mathbf{x}\|_\pi^2 - 2(\mathbf{x})_\pi^2 \right) \\ &= \frac{1}{2} \left(\sum_{i \in E} x_i^2 \pi_i + \sum_{j \in E} x_j^2 \pi_j - 2 \sum_{i, j \in E} x_i x_j \pi_i \pi_j \right) \\ &= \frac{1}{2} \sum_{i, j \in E} (x_i - x_j)^2 \pi_i \pi_j \\ &= \frac{1}{2} \sum_{i, j \in E} \left(\sum_{e \in \gamma_{ij}} \frac{1}{\sqrt{Q(e)}} \sqrt{Q(e)} (x_{e^-} - x_{e^+}) \right)^2 \pi_i \pi_j. \end{aligned}$$

– Aus der Ungleichung von Cauchy-Schwarz ergibt sich nun, dass

$$\begin{aligned} \text{Var}_{\pi}(\mathbf{x}) &\leq \frac{1}{2} \sum_{i,j \in E} \left(|\gamma_{ij}| \sum_{e \in \gamma_{ij}} Q(e) (x_{e^-} - x_{e^+})^2 \right) \pi_i \pi_j \\ &= \frac{1}{2} \sum_{e \in \mathcal{E}} \left(Q(e) (x_{e^-} - x_{e^+})^2 \left(\sum_{\gamma_{ij} \ni e} |\gamma_{ij}| \pi_i \pi_j \right) \right) \\ &\leq \kappa D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x}), \end{aligned}$$

wobei sich die letzte Ungleichung aus Lemma 2.8 und aus der Definitionsgleichung (120) des Poincaré-Koeffizienten ergibt. Damit ist (125) bewiesen.

- Um den Beweis zu beenden, ist noch zu zeigen, dass $\lambda_\ell \geq -1 + 2\zeta^{-1}$.
– Hierfür nutzen wir die folgende Identität: Für jedes $\mathbf{x} = (x_1, \dots, x_\ell)^\top \in \mathbb{R}^\ell$ gilt

$$\frac{1}{2} \sum_{i,j \in E} (x_i + x_j)^2 \pi_i p_{ij} = (\mathbf{P}\mathbf{x}, \mathbf{x})_\pi + \|\mathbf{x}\|_\pi^2, \quad (126)$$

denn aus der Reversibilität von $(\mathbf{P}, \pi)$ ergibt sich, dass

$$\begin{aligned} \frac{1}{2} \sum_{i,j \in E} (x_i + x_j)^2 \pi_i p_{ij} &= \frac{1}{2} \sum_{i,j \in E} \underbrace{x_i^2 \pi_i p_{ij}}_{= \sum_{i \in E} x_i^2 \pi_i} + \sum_{i,j \in E} x_i x_j \pi_i p_{ij} + \frac{1}{2} \sum_{i,j \in E} \underbrace{x_j^2 \pi_j p_{ji}}_{= \sum_{j \in E} x_j^2 \pi_j} \\ &= \|\mathbf{x}\|_\pi^2 + (\mathbf{P}\mathbf{x}, \mathbf{x})_\pi. \end{aligned}$$

- Sei nun $\gamma_i = (i_0, i_1, \dots, i_{2m}, i_{2m+1})$ mit $i = i_0 = i_{2m+1}$ ein Pfad von i nach i , der eine ungerade Anzahl von Kanten enthält, so dass keine der Kanten mehrfach auftritt.
- Dann gilt

$$\begin{aligned} x_i &= \frac{1}{2} \left((x_i + x_{i_1}) - (x_{i_1} + x_{i_2}) + \dots + (x_{i_{2m}} + x_i) \right) \\ &= \frac{1}{2} \sum_{e \in \gamma_i} (-1)^{n(e)} (x_{e^+} + x_{e^-}), \end{aligned}$$

wobei $n(e) = k$, falls $e = (i_k, i_{k+1}) \in \gamma_i$.

- Deshalb ergibt sich (ähnlich wie im ersten Teil des Beweises) aus der Ungleichung von Cauchy-Schwarz, dass für jedes $\mathbf{x} = (x_1, \dots, x_\ell)^\top \in \mathbb{R}^\ell$

$$\begin{aligned} \|\mathbf{x}\|_\pi^2 &= \sum_{i \in E} \frac{\pi_i}{4} \left(\sum_{e \in \gamma_i} \frac{1}{\sqrt{Q(e)}} \sqrt{Q(e)} (-1)^{n(e)} (x_{e^+} + x_{e^-}) \right)^2 \\ &\leq \sum_{i \in E} \left(\frac{\pi_i}{4} |\gamma_i| \sum_{e \in \gamma_i} (x_{e^+} + x_{e^-})^2 Q(e) \right) \\ &= \frac{1}{4} \sum_{e \in \mathcal{E}'} \left((x_{e^+} + x_{e^-})^2 Q(e) \sum_{\gamma_i \ni e} |\gamma_i| \pi_i \right) \\ &\leq \frac{\zeta}{4} \sum_{e \in \mathcal{E}'} (x_{e^+} + x_{e^-})^2 Q(e). \end{aligned}$$

- Hieraus und aus (126) ergibt sich nun, dass

$$\|\mathbf{x}\|_\pi^2 \leq \frac{\zeta}{2} \left((\mathbf{P}\mathbf{x}, \mathbf{x})_\pi + \|\mathbf{x}\|_\pi^2 \right).$$

- Für $\mathbf{x} = \phi_\ell$ gilt also insbesondere

$$1 \leq \frac{\zeta}{2} (\lambda_\ell + 1) \quad \text{bzw.} \quad \lambda_\ell \geq -1 + \frac{2}{\zeta} .$$

□

Beispiel Zufällige Irrfahrten auf Graphen

- Wir kehren nun zu dem bereits in Abschnitt 2.3.1 diskutierten Beispiel von zufälligen Irrfahrten auf Graphen zurück.
 - Sei also $G = (V, K)$ ein verbundener Graph mit der Menge $V = \{v_1, \dots, v_\ell\}$ von ℓ Eckpunkten und der Menge K von Kanten, die jeweils zwei Eckpunkte miteinander verbinden,
 - so dass es für jedes Paar $v_i, v_j \in V$ von Eckpunkten einen „Pfad“ von Kanten aus K gibt, der die Kanten v_i und v_j miteinander verbindet.
- Eine zufällige Irrfahrt auf dem Graph $G = (V, K)$ ist eine Markov-Kette $X_0, X_1, \dots : \Omega \rightarrow E$
 - mit dem Zustandsraum $E = \{1, \dots, \ell\}$ und der Übergangsmatrix $\mathbf{P} = (p_{ij})$, wobei

$$p_{ij} = \begin{cases} \frac{1}{d_i}, & \text{falls die Eckpunkte } v_i \text{ und } v_j \text{ Nachbarn sind,} \\ 0, & \text{sonst.} \end{cases} \quad (127)$$

- Dabei werden zwei Eckpunkte v_i und v_j *Nachbarn* genannt, falls sie die Endpunkte einundderselben Kante sind, und d_i bezeichnet die Anzahl der Nachbarn von v_i .
- Wir hatten gezeigt, dass
 - die in (127) gegebene Übergangsmatrix $\mathbf{P}$ stets irreduzibel ist (wobei wir nun zusätzlich noch voraussetzen, dass $\mathbf{P}$ auch aperiodisch ist),
 - die (eindeutig bestimmte) stationäre Anfangsverteilung $\boldsymbol{\pi}$ gegeben ist durch

$$\boldsymbol{\pi} = \left(\frac{d_1}{d}, \dots, \frac{d_\ell}{d} \right)^\top, \quad \text{wobei } d = \sum_{i=1}^{\ell} d_i, \quad (128)$$

- und das in (127)–(128) gegebene Paar $(\mathbf{P}, \boldsymbol{\pi})$ reversibel ist.
- Für den in (120) eingeführten Poincaré-Koeffizienten κ gilt dann

$$\kappa = \kappa(\Gamma) = \max_{e \in \mathcal{E}} \sum_{\gamma_{ij} \ni e} |\gamma_{ij}| \pi_i \pi_j ,$$

wobei

$$|\gamma_{ij}| = \sum_{e \in \gamma_{ij}} \frac{1}{Q(e)} = d \Delta(\gamma_{ij})$$

und $\Delta(\gamma_{ij}) = \#\{e : e \in \gamma_{ij}\}$ die Anzahl der Kanten (d.h. die Länge) des Pfades γ_{ij} bezeichnet.

- Hieraus und aus (127)–(128) ergibt sich, dass

$$\kappa(\Gamma) \leq \frac{\delta^2 \Delta \beta}{d}, \quad (129)$$

wobei $d/2$ die Gesamtanzahl aller Kanten,

- $\delta = \max_{i \in E} d_i$ die maximale Anzahl von Kanten ist, die von einem Eckpunkt ausgehen,
- $\Delta = \max_{\gamma \in \Gamma} \Delta(\gamma)$ die maximale Pfadlänge und
- $\beta = \max_{e \in \mathcal{E}} \#\{\gamma \in \Gamma : \gamma \ni e\}$ der sogenannte *Bottleneck-Koeffizient*, d.h., die maximale Anzahl der Pfade ist, die jeweils durch eine einzelne Kante laufen.

- Aus (123) und (129) ergibt sich nun die folgende Abschätzung

$$\lambda_2 \leq 1 - \frac{1}{\kappa} \leq 1 - \frac{d}{\delta^2 \Delta \beta} \quad (130)$$

für den zweitgrößten Eigenwert λ_2 von $\mathbf{P}$.

- Auf ähnliche Weise ergibt sich die Abschätzung

$$\zeta = \zeta(\Gamma') = \max_{e \in \mathcal{E}'} \sum_{\gamma_i \ni e} |\gamma_i| \pi_i \leq \delta \Delta' \beta',$$

wobei

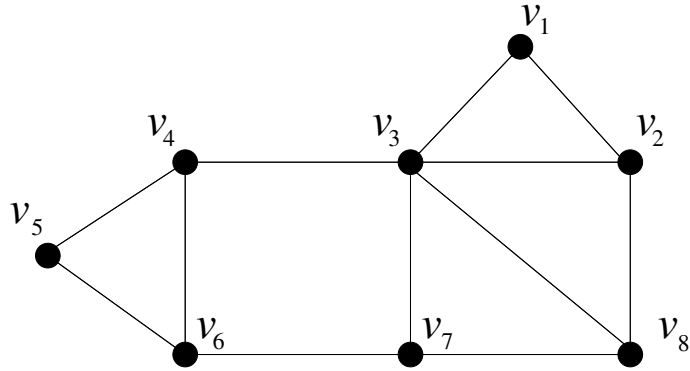
$$|\gamma_i| = \sum_{e \in \gamma_i} \frac{1}{Q(e)} = d \Delta(\gamma_i), \quad \Delta' = \max_{\gamma \in \Gamma'} \Delta(\gamma), \quad \beta' = \max_{e \in \mathcal{E}'} \#\{\gamma \in \Gamma' : \gamma \ni e\}$$

und somit

$$\lambda_\ell \geq -1 + \frac{2}{\zeta} \geq -1 + \frac{2}{\delta \Delta' \beta'}. \quad (131)$$

- Beachte

- Für das bereits in Abschnitt 2.3.1 betrachtete Zahlenbeispiel



gilt insbesondere, dass

$$d = 24, \delta = 5, \Delta = 3, \beta = 7 \quad \text{und} \quad \Delta' = 3, \beta' = 3$$

- Aus (130) und (131) ergibt sich somit

$$\lambda_2 \leq 1 - \frac{24}{25 \cdot 3 \cdot 7} < \frac{24}{25} \quad \text{und} \quad \lambda_8 > -1 + \frac{2}{5 \cdot 3 \cdot 3} = -\frac{43}{45}.$$

bzw.

$$\max\{\lambda_2, |\lambda_8|\} < \frac{24}{25}.$$

3 Monte-Carlo-Simulation

- Neben der Gewinnung von experimentellen Daten im Rahmen von realen Labor- bzw. Feldversuchen erlangte die Erzeugung sogenannter *synthetischer Daten* durch Computer-Simulation in den letzten Jahren eine immer größere Bedeutung.
- Die Gründe für die zunehmende Nützlichkeit von Computer-Simulationen bei der Untersuchung von interessierenden Sachverhalten/Objekten/Vorgängen sind vielfältig:
 - An erster Stelle ist hier natürlich das rasant wachsende Leistungsvermögen moderner Computer-Systeme zu nennen, das sich in den letzten Jahren in einem ungeahnten Ausmaß weiterentwickelt hat und dabei Möglichkeiten eröffnet, die noch vor kurzem völlig unvorstellbar waren.
 - Im Zusammenhang damit ist die Datengewinnung durch Computer-Simulation oft viel *kostengünstiger* und mit *weniger Zeitaufwand* verbunden als die herkömmliche Gewinnung von experimentellen Daten im Rahmen von realen Labor- bzw. Feldversuchen.
 - Außerdem lassen sich (virtuelle) Computer-Experimente unter gleichbleibenden Versuchsbedingungen beliebig oft wiederholen, wogegen beispielsweise bei naturwissenschaftlichen Experimenten das untersuchte Objekt während der Versuche oft beschädigt bzw. zerstört wird.
- Ein weiterer Grund für die Nützlichkeit von Computer-Simulationen besteht darin, dass
 - der Umfang und die Struktur der zu analysierenden Datensätze oft sehr komplex ist,
 - die Verarbeitung und Bewertung der Daten dann typischerweise auf mathematischen Modellen beruht, deren charakteristische Eigenschaften nicht (oder nur teilweise) mit geschlossenen analytischen Formeln beschrieben werden können,
 - die Computer-Simulation der betrachteten Modelle in diesem Fall ein nützliches (alternatives) Analyse-Tool ist.
- Computer-Experimente zur Untersuchung von interessierenden Sachverhalten/Objekten/Vorgängen beruhen auf *stochastischen Simulationsalgorithmen*. Man spricht deshalb auch von *Monte-Carlo-Simulation*. Dabei gibt es unterschiedliche Arten von Simulationsalgorithmen.
 1. Die Grundlage zur Monte-Carlo-Simulation von (einzelnen) Merkmalen/Kenngrößen/Variablen bilden *Zufallszahlen-Generatoren*.
 - Dies sind Algorithmen, durch die Realisierungen von Zufallsvariablen per Computer erzeugt werden können, die *Pseudozufallszahlen* genannt werden.
 - Den Ausgangspunkt bilden dabei sogenannte *Standard-Zufallszahlen-Generatoren*, durch die Realisierungen von Zufallsvariablen erzeugt werden können, die auf dem Einheitsintervall $(0, 1]$ gleichverteilt sind, sogenannte *Standard-Pseudozufallszahlen*.
 - Hiervon ausgehend lassen sich dann durch gewisse *Transformations- bzw. Verwerfungsmethoden* auch Pseudozufallszahlen für Zufallsvariablen mit anderen Verteilungen erzeugen, zum Beispiel für binomialverteilte, Poisson-verteilte oder normalverteilte Zufallsvariablen.
 2. Computer-Experimente zur Untersuchung von *hochdimensionalen Zufallsvektoren* oder der *zeitlichen Entwicklung* von Sachverhalten/Objekten/Vorgängen beruhen auf anspruchsvolleren Algorithmen der *dynamischen Monte-Carlo-Simulation*.
 - Eine zentrale Rolle spielen dabei die Algorithmen der *Markov-Chain-Monte-Carlo-Simulation* (MCMC-Simulation), durch die insbesondere *zeitstationäre Gleichgewichtszustände* von Sachverhalten/Objekten/Vorgängen näherungsweise simuliert werden können.
 - Ein anderes Beispiel, bei denen Algorithmen der MCMC-Simulation angewendet werden, ist die *statistische Analyse von Bilddaten*.
 - Eine aktuelle Forschungsthematik, zu der in den letzten Jahren zahlreiche Publikationen veröffentlicht wurden, sind sogenannte *Kopplungsalgorithmen der perfekten MCMC-Simulation*.
 - Durch solche Kopplungsalgorithmen können beispielsweise zeitstationäre Gleichgewichtszustände von Sachverhalten/Objekten/Vorgängen nicht nur näherungsweise, sondern in einem gewissen Sinne „perfekt“ simuliert werden können.

3.1 Erzeugung von Pseudozufallszahlen

3.1.1 Einfache Anwendungsbeispiele; Monte-Carlo-Schätzer

Wir erinnern zunächst an zwei einfache Beispiele von Fragestellungen, die mit Monte-Carlo-Simulation gelöst werden können und die bereits in der Vorlesung „Wahrscheinlichkeitsrechnung“ erwähnt wurden.

1. Algorithmus zur Bestimmung der Zahl π

- Ein einfacher Computer-Algorithmus zur Monte-Carlo-Simulation der Zahl π ist die folgende (verbesserte) Variante des Buffonschen Nadelexperimentes, vgl. auch die Abschnitte 2.5 bzw. 5.2.3 der Vorlesung „Wahrscheinlichkeitsrechnung“.

- Dieser Algorithmus hängt mit dem folgenden geometrischen Sachverhalt zusammen.

- Wir betrachten das Quadrat

$$B = (-1, 1] \times (-1, 1] \subset \mathbb{R}^2$$

- bzw. den in B eingeschriebenen Kreis

$$C = \{(x, y) : (x, y) \in B, x^2 + y^2 < 1\}$$

- und werfen einen Punkt *willkürlich* in die Menge B .

- In der Sprechweise der Stochastik bedeutet dies:

- Wir betrachten zwei unabhängige Zufallsvariablen S und T , die jeweils gleichverteilt im Intervall $(-1, 1]$ sind, und
- bestimmen die Wahrscheinlichkeit des Ereignisses

$$A = \{(S, T) \in C\} = \{S^2 + T^2 < 1\},$$

dass der „zufällige Punkt“ (S, T) in $C \subset B$ liegt.

- Dann gilt

$$P(A) = P(S^2 + T^2 < 1) = \frac{|C|}{|B|} = \frac{\pi}{4},$$

wobei $|B|$, $|C|$ den Flächeninhalt von B bzw. C bezeichnet.

- Ähnlich wie beim Buffonschen Nadelexperiment ergibt sich nun aus der Gleichung $P(A) = \pi/4$ eine
 - Methode zur *statistischen Schätzung* der Zahl π ,
 - die auf dem *Gesetz der großen Zahlen* beruht und die sich leicht implementieren lässt.
- Und zwar seien $(S_1, T_1), \dots, (S_n, T_n)$ unabhängige und identisch verteilte Zufallsvektoren,
 - deren Verteilung mit der Verteilung von (S, T) übereinstimmt und
 - die wir als ein stochastisches Modell für n (unabhängig durchgeführte) Experimente auffassen.
 - Dann sind $X_1, X_2, \dots, X_n$ mit

$$X_i = \begin{cases} 1, & \text{falls } S_i^2 + T_i^2 < 1 \\ 0 & \text{sonst} \end{cases}$$

unabhängige und identisch verteilte Zufallsvariablen mit dem Erwartungswert $\mathbb{E} X_i = \pi/4$.

- Aus dem starken Gesetz der großen Zahlen (vgl. Theorem WR-5.15) ergibt sich darüber hinaus, dass
 - das arithmetische Mittel

$$Y_n = n^{-1} \sum_{i=1}^n X_i$$

fast sicher gegen die Zahl $\pi/4$ strebt.

- Somit ist Y_n ein *erwartungstreuer* und (stark) *konsistenter Schätzer* für $\pi/4$,
- d.h., für große n ist $4Y_n$ mit hoher Wahrscheinlichkeit eine gute Näherung der Zahl π .
- Bei der Implementierung dieser Simulationsalgorithmus kann man wie folgt vorgehen.
 - Erzeuge $2n$ in $(0, 1]$ gleichverteilte Pseudozufallszahlen $u_1, \dots, u_{2n}$ mit einem Zufallszahlengenerator.
 - Setze $s_i = 2u_i - 1$ und $t_i = 2u_{n+i} - 1$ für $i = 1, \dots, n$.
 - Setze

$$x_i = \begin{cases} 1, & \text{falls } s_i^2 + t_i^2 < 1 \\ 0 & \text{sonst} \end{cases}$$
 - Berechne $4(x_1 + \dots + x_n)/n$.

2. Monte-Carlo-Integration

- Sei $\varphi : [0, 1] \rightarrow [0, 1]$ eine stetige Funktion.
 - Gesucht ist ein Schätzer für den Wert des Integrals $\int_0^1 \varphi(x) dx$, der mit Monte-Carlo-Simulation bestimmt werden soll.
 - Dabei betrachten wir das folgende stochastische Modell.
- Seien $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ unabhängige und identisch verteilte Zufallsvariable, die im Intervall $(0, 1]$ gleichverteilt sind.
 - Außerdem sei $Z_k = \varphi(X_k)$ für jedes $k = 1, 2, \dots$
 - Wegen des Transformationssatzes für unabhängige Zufallsvariablen (vgl. Theorem WR-3.18) sind dann auch $Z_1, Z_2, \dots$ unabhängige und identisch verteilte Zufallsvariable,
 - und es gilt

$$\mathbb{E} Z_1 = \int_0^1 \varphi(x) f_{X_1}(x) dx = \int_0^1 \varphi(x) dx.$$

- Aus dem starken Gesetz der großen Zahlen (vgl. Theorem WR-5.15) ergibt sich darüber hinaus, dass für $n \rightarrow \infty$

$$\frac{1}{n} \sum_{k=1}^n Z_k \xrightarrow{\text{f.s.}} \int_0^1 \varphi(x) dx.$$

- Somit ist $\frac{1}{n} \sum_{k=1}^n Z_k$ ein *erwartungstreuer* und (stark) *konsistenter Schätzer* für $\int_0^1 \varphi(x) dx$,
- d.h., für große n ist $\frac{1}{n} \sum_{k=1}^n Z_k$ mit hoher Wahrscheinlichkeit eine gute Näherung des Integrals $\int_0^1 \varphi(x) dx$.
- Bei der Implementierung dieses Simulationsalgorithmus kann man ähnlich wie in Beispiel 1 vorgehen:
 - Erzeuge n in $(0, 1]$ gleichverteilte Pseudozufallszahlen $x_1, \dots, x_n$ mit einem Zufallszahlengenerator.
 - Setze $z_k = \varphi(x_k)$ für $k = 1, \dots, n$.
 - Berechne $\frac{1}{n} \sum_{k=1}^n z_k$.

3.1.2 Lineare Kongruenzgeneratoren

- Den Ausgangspunkt der meisten Simulationsalgorithmen bilden *Standard-Zufallszahlen-Generatoren*,
 - deren Ziel darin besteht, Folgen $u_1, \dots, u_n$ von Zahlen aus dem Einheitsintervall $(0, 1]$ zu erzeugen, sogenannte *Standard-Pseudozufallszahlen*,
 - die als Realisierungen von unabhängigen und (identisch in $(0, 1]$) gleichverteilten Zufallsvariablen $U_1, \dots, U_n$ aufgefasst werden können.
- Ein weit verbreitetes Verfahren zur Erzeugung von Standard-Pseudozufallszahlen ist die folgende *lineare Kongruenzmethode*,
 - wobei zunächst die Zahlen $z_1, \dots, z_n$ gemäß einer Rekursion vom Typ

$$z_k = (az_{k-1} + c) \mod (m), \quad \forall k = 1, \dots, n \quad (1)$$

erzeugt werden.

- Dabei wird von einem gewissen Startwert $z_0 \in \{0, 1, \dots, m-1\}$ ausgegangen, der auch *Keim* des linearen Kongruenzgenerators genannt wird, und
- $m \in \mathbb{N}$, $a \in \{0, 1, \dots, m-1\}$ bzw. $c \in \{0, 1, \dots, m-1\}$ sind weitere Parameter, die *Modul*, *Faktor* bzw. *Inkrement* des Kongruenzgenerators genannt werden.
- Hieraus ergeben sich durch die Normierung

$$u_k = \frac{z_k}{m} \quad (2)$$

die Standard-Pseudozufallszahlen $u_1, \dots, u_n$.

Wir lösen nun die Rekursionsgleichung (1), d.h., wir zeigen, wie sich die in (1) rekursiv definierte Zahl z_k *direkt* durch den Startwert z_0 sowie durch die Parameter m , a bzw. c ausdrücken lässt.

Theorem 3.1 Für jedes $k \in \{1, \dots, n\}$ gilt

$$z_k = \left(a^k z_0 + c \frac{a^k - 1}{a - 1} \right) \mod (m). \quad (3)$$

Beweis

- Wir führen den Beweis mit vollständiger Induktion, wobei für $k = 1$ die Behauptung (3) identisch ist mit der Rekursionsgleichung (1).
- Es gelte (3) nun für eine gewisse natürliche Zahl $k \geq 1$, d.h., es gibt eine ganze Zahl $j \geq 0$, so dass

$$z_k = a^k z_0 + c \frac{a^k - 1}{a - 1} - jm. \quad (4)$$

- Wir zeigen, dass dann (3) auch für $k + 1$ gilt.
- Aus der Rekursionsgleichung (1) und aus der Induktionsannahme (4) ergibt sich, dass

$$\begin{aligned} z_{k+1} &= (az_k + c) \mod (m) \\ &= \left(a \left(a^k z_0 + c \frac{a^k - 1}{a - 1} - jm \right) + c \right) \mod (m) \\ &= \left(a^{k+1} z_0 + c \frac{a(a^k - 1) + a - 1}{a - 1} - a jm \right) \mod (m) \\ &= \left(a^{k+1} z_0 + c \frac{a^{k+1} - 1}{a - 1} \right) \mod (m). \end{aligned}$$

- Damit ist die Gültigkeit von (3) für $k + 1$ bewiesen. □

Beachte

- Es ist klar, dass mit dem in (1) definierten linearen Kongruenzgenerator höchstens m verschiedene Zahlen $z_1, \dots, z_n$ erzeugt werden können.
 - Sobald sich eine Zahl z_k zum ersten Mal wiederholt, d.h., es gibt ein $m_0 > 0$ mit $z_k = z_{k-m_0}$,
 - dann beginnt erneut die gleiche Periode der Länge m_0 , die bereits einmal vollständig erzeugt worden ist, d.h., es gilt

$$z_{k+j} = z_{k-m_0+j} \quad \text{für jedes } j \geq 1.$$

- Bei einer ungünstigen Wahl der Parameter m, a, c bzw. z_0 kann die Länge der Periode m_0 sehr klein sein.
 - Beispielsweise gilt

$$m_0 = 2 \quad \text{für } a = c = z_0 = 5 \text{ und } m = 10,$$

wobei dann lediglich die Zahlen $5, 0, 5, 0, \dots$ generiert werden.

- Ein wünschenswertes Qualitätskriterium von linearen Kongruenzgeneratoren ist jedoch, dass die Länge der Periode m_0 möglichst nahe bei der Maximallänge m liegt.

Wir erwähnen nun (hinreichende und notwendige) Bedingungen an die Parameter m, a, c bzw. z_0 , die insbesondere sichern, dass die maximal mögliche Periode m erreicht wird.

Theorem 3.2

1. Falls $c > 0$, so erzeugt der in (1) definierte lineare Kongruenzgenerator genau dann für jeden Startwert $z_0 \in \{0, 1, \dots, m-1\}$ eine Zahlenfolge $z_1, \dots, z_n$ mit der maximal möglichen Periode m , wenn die folgenden Bedingungen erfüllt sind:
 - (a₁) Die Parameter c und m sind teilerfremd.
 - (a₂) Für jede Primzahl r , die m teilt, ist $a - 1$ ein Vielfaches von r .
 - (a₃) Falls m ein Vielfaches von 4 ist, dann ist auch $a - 1$ ein Vielfaches von 4.
 2. Falls $c = 0$, so gilt $m_0 = m - 1$ für jedes $z_0 \in \{1, \dots, m-1\}$ genau dann, wenn
 - (b₁) m eine Primzahl ist,
 - (b₂) für jede Primzahl r , die $m - 1$ teilt, die Zahl $a^{(m-1)/r} - 1$ nicht durch m teilbar ist.
 3. Falls $c = 0$ und falls es ein $k \in \mathbb{N}$ mit $m = 2^k \geq 16$ gibt, so gilt $m_0 = m/4$ genau dann, wenn z_0 eine ungerade Zahl ist und wenn $a \bmod (8) = 5$ oder $= 3$ gilt.
- Einen Beweis von Theorem 3.2, in dem Ergebnisse der Zahlentheorie (u.a. der kleine Satz von Fermat) verwendet werden, kann man beispielsweise
 - in Abschnitt 2.7 des Buches von B.D. Ripley (1987) *Stochastic Simulation*, J. Wiley & Sons, New York oder
 - in Abschnitt 3.2 von D.E. Knuth (1997) *The Art of Computer Programming*, Vol. II, Addison-Wesley, Reading MA finden.

- Darüber hinaus verweisen wir auch auf diese beiden Bücher hinsichtlich der Diskussion
 - von anderen Generatoren zur Erzeugung von Standard-Pseudozufallszahlen wie beispielsweise nicht-lineare Kongruenzgeneratoren, Schieberegister-Generatoren, Lagged-Fibonacci-Generatoren sowie die Kombination solcher Generatoren,
 - von alternativen Bedingungen an die Parameter m , a , c bzw. z_0 des linearen Kongruenzgenerators, der in (1) definiert wurde,
 - um Zahlenfolgen $z_1, \dots, z_n$ mit einer möglichst großen Periode m_0 und mit weiteren Güteeigenschaften zu erzeugen.
- Eine solche Güteeigenschaft besteht darin,
 - dass die Punkte $(u_1, u_2), \dots, (u_{n-1}, u_n)$ von jeweils aufeinanderfolgenden Pseudozufallszahlen u_{i-1}, u_i das Einheitsquadrat $[0, 1]^2$ möglichst gleichmäßig ausfüllen,
 - wobei die folgenden Zahlenbeispiele verdeutlichen, dass relativ geringfügige Änderungen der Parameter a und c zu völlig unterschiedlichen *Punktmustern* $(u_1, u_2), \dots, (u_{n-1}, u_n)$ führen können.
- Weitere Details hierzu sind beispielsweise in dem bereits erwähnten Buch von Ripley (1987) bzw. in dem Vorlesungsskript von H. Künsch unter <ftp://stat.ethz.ch/U/Kuensh/skript-sim.ps> zu finden, wo auch die folgenden Abbildungen enthalten sind.

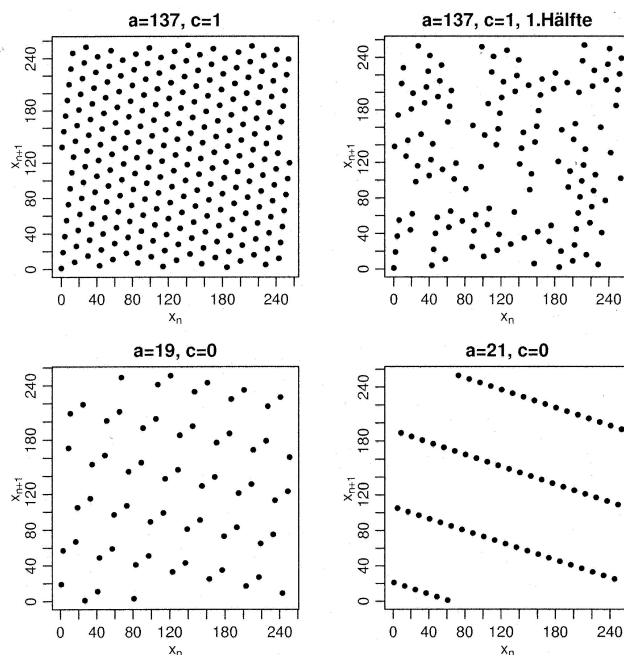


Abbildung 1: Punktmuster für Paare (u_{i-1}, u_i) aufeinanderfolgender Pseudozufallszahlen mit $m = 256$

3.1.3 Tests von Güteeigenschaften

- In der Literatur werden zahlreiche *statistische Signifikanztests* diskutiert, um Güteeigenschaften von Zufallszahlengeneratoren zu untersuchen, vgl. beispielsweise das Buch von G.S. Fishman (1996) *Monte Carlo: Concepts, Algorithms and Applications*, Springer-Verlag, New York.

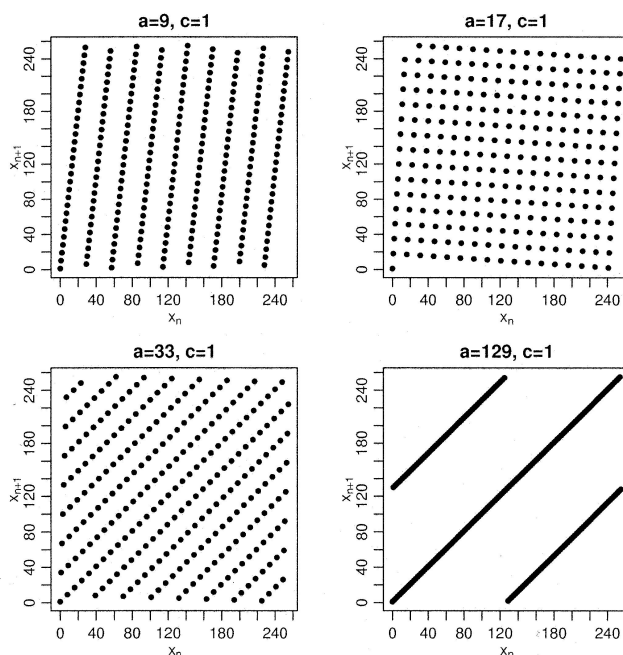


Abbildung 2: Punktmuster für Paare (u_{i-1}, u_i) aufeinanderfolgender Pseudozufallszahlen mit $m = 256$

- Wir erinnern hier lediglich an zwei solche Tests, die
 - im Zusammenhang mit der Untersuchung der Güteeigenschaften von linearen Kongruenzgeneratoren (oder von anderen Zufallszahlen-Generatoren) eine Rolle spielen und
 - bereits in der Vorlesung „Statistik I“ erwähnt wurden.
- Dabei wird der χ^2 -Anpassungstest von Pearson (vgl. Abschnitt 5.1 der Vorlesung „Statistik II“) verwendet, um zu prüfen,
 - ob die erzeugten Pseudozufallszahlen als Realisierungen von gleichverteilten Zufallsvariablen aufgefasst werden können und
 - ob davon ausgegangen werden kann, dass diese Zufallsvariablen unabhängig sind.
- Ein anderes Verfahren zur Erzeugung von Zahlenfolgen $u_1, u_2, \dots$ mit günstigen Güteeigenschaften
 - beruht auf der Minimierung des *Kolmogorov-Abstandes* (vgl. Abschnitt 1.5 der Vorlesung „Statistik I“)

$$D_n(u_1, \dots, u_n) = \sup_{x \in (0,1]} \left| \frac{1}{n} \# \{i : 1 \leq i \leq n, 0 < u_i \leq x\} - x \right|$$

zwischen der empirischen Verteilungsfunktion der „Stichprobe“ $u_1, \dots, u_n$ und der Verteilungsfunktion der $(0, 1]$ -Gleichverteilung, und zwar für *jede* natürliche Zahl n .

- Diese Vorgehensweise wird in der Literatur *Quasi-Monte-Carlo-Methode* genannt, vgl. zum Beispiel das Buch von H. Niederreiter (1992) *Random Number Generation and Quasi-Monte-Carlo Methods*, SIAM, Philadelphia.

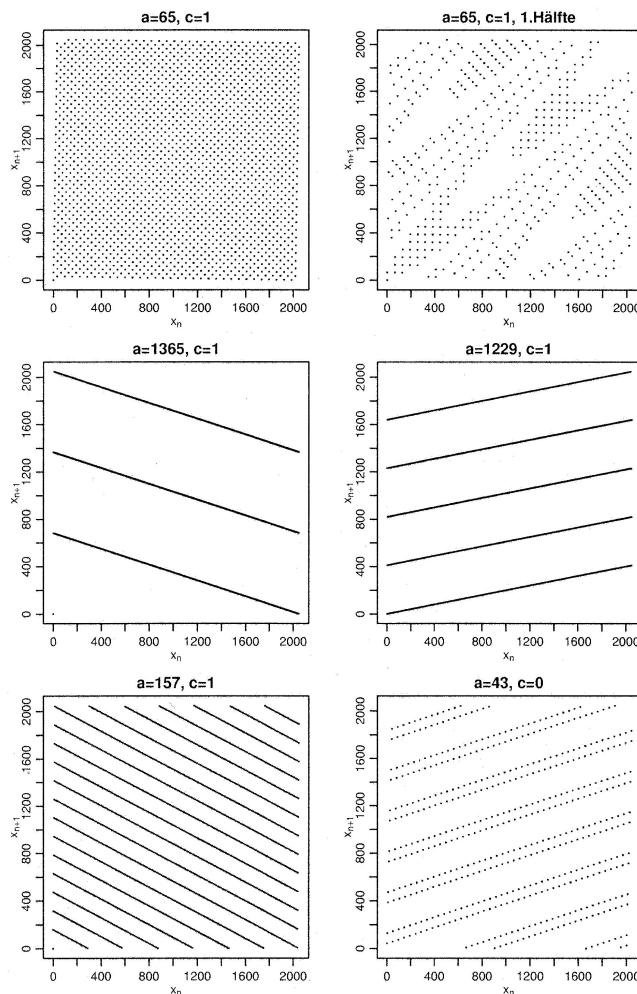


Abbildung 3: Punktmuster für Paare (u_{i-1}, u_i) aufeinanderfolgender Pseudozufallszahlen mit $m = 2048$

1. χ^2 -Anpassungstests auf Gleichverteilung

Wir betrachten zunächst den folgenden Test, um zu prüfen, ob die Pseudozufallszahlen $u_1, \dots, u_n$

- als Realisierungen von unabhängigen Stichprobenvariablen $U_1, \dots, U_n$ angesehen werden können, die gleichverteilt im Intervall $(0, 1]$ sind.
- Hierfür zerlegen wir das Intervall $(0, 1]$ in r gleichlange Teilintervalle $(0, 1/r], \dots, ((r-1)/r, 1]$ und
 - betrachten den $(r-1)$ -dimensionalen (hypothetischen) Parametervektor $\mathbf{p}_0 = (1/r, \dots, 1/r)$ bzw.
 - die Testgröße $T_n : \mathbb{R}^n \rightarrow [0, \infty)$ mit

$$T_n(u_1, \dots, u_n) = \sum_{j=1}^r \frac{(Z_j(u_1, \dots, u_n) - n/r)^2}{n/r},$$

wobei $Z_j(u_1, \dots, u_n) = \#\{i : 1 \leq i \leq n, j-1 < ru_i \leq j\}$ die Anzahl derjenigen Pseudozufallszahlen $u_1, \dots, u_n$ bezeichnet, die im Intervall $((j-1)/r, j/r]$ liegen.

- Weil die Testgröße T_n asymptotisch χ_{r-1}^2 -verteilt ist,
 - falls die Stichprobenvariablen $U_1, \dots, U_n$ unabhängig und im Intervall $(0, 1]$ gleichverteilt sind (vgl. Theorem 5.1 des Vorlesungsskriptes „Statistik II“),

- wird bei hinreichend großem n die Hypothese $H_0 : \mathbf{p} = \mathbf{p}_0$ abgelehnt, falls

$$T_n(u_1, \dots, u_n) > \chi_{r-1, 1-\alpha}^2,$$

wobei $\chi_{r-1, 1-\alpha}^2$ das $(1-\alpha)$ -Quantil der χ^2 -Verteilung mit $r-1$ Freiheitsgraden bezeichnet.

- Wir illustrieren dies mit Hilfe des folgenden Zahlenbeispiels. Für $\alpha = 0.05$, $n = 100\,000$ und $r = 10$ soll geprüft werden, ob
 - die Hypothese der Gleichverteilung mit einer (konkreten) Stichprobe $(u_1, \dots, u_{100\,000})$ von Pseudozufallszahlen vereinbar ist, für die sich der folgende Vektor der Klassenhäufigkeiten $(z_1, \dots, z_{10})$ ergibt:

z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
9 995	10 045	10 127	9 816	10 130	10 040	9 890	9 858	10 083	10 016

- In diesem Fall ist $T_{100\,000}(u_1, \dots, u_{100\,000}) = 10.99$, und es gilt somit

$$T_{100\,000}(u_1, \dots, u_{100\,000}) = 10.99 < \chi_{9, 0.95}^2 = 16.92.$$

- Die Hypothese der $(0, 1]$ -Gleichverteilung wird also nicht verworfen.

Beachte

- In Verallgemeinerung des bisher betrachteten χ^2 -Tests auf Gleichverteilung kann man auch prüfen,
 - ob für eine vorgegebene natürliche Zahl $d \geq 1$ (beispielsweise $d = 2$ oder $d = 3$) die Pseudozufallsvektoren $(u_1, \dots, u_d), \dots, (u_{(n-1)d+1}, \dots, u_{nd})$
 - als Realisierungen von unabhängigen Zufallsvektoren $(U_1, \dots, U_d), \dots, (U_{(n-1)d+1}, \dots, U_{nd})$ angesehen werden können, die gleichverteilt in $(0, 1]^d$ sind.
- Hierfür wird der Einheitswürfel $(0, 1]^d$
 - in r^d gleichgroße Teilwürfel B_j vom Typ $((i_1 - 1)/r, i_1/r] \times \dots \times ((i_d - 1)/r, i_d/r]$ zerlegt und
 - der $(r^d - 1)$ -dimensionale (hypothetische) Parametervektor $\mathbf{p}_0 = (1/r^d, \dots, 1/r^d)$ bzw.
 - die Testgröße $T_n : \mathbb{R}^{nd} \rightarrow [0, \infty)$ mit

$$T_n(\mathbf{u}_1, \dots, \mathbf{u}_n) = \sum_{j=1}^{r^d} \frac{(Z_j(\mathbf{u}_1, \dots, \mathbf{u}_n) - n/r^d)^2}{n/r^d}$$

betrachtet, wobei $\mathbf{u}_i = (u_{(i-1)d+1}, \dots, u_{id})$ und $Z_j(\mathbf{u}_1, \dots, \mathbf{u}_n) = \#\{i : 1 \leq i \leq n, \mathbf{u}_i \in B_j\}$ die Anzahl derjenigen Pseudozufallsvektoren $\mathbf{u}_1, \dots, \mathbf{u}_n$ bezeichnet, die in W_j liegen.

2. Run-Test

Zur Überprüfung der Güte von Zufallszahlengeneratoren gibt es noch weitere statistische Signifikanztests, durch die verifiziert werden kann,

- ob die erzeugten Pseudozufallszahlen $u_1, \dots, u_n$ als Realisierungen von unabhängigen Zufallsvariablen $U_1, \dots, U_n$ mit einer bestimmten Verteilung (im vorliegenden Fall die $(0, 1]$ -Gleichverteilung) angesehen werden können.
- Bei dem folgenden *Run-Test* wird insbesondere geprüft,
 - ob auch die Modellannahme der *Unabhängigkeit* der Stichprobenvariablen $U_1, \dots, U_n$ hinreichend gut durch die Pseudozufallszahlen $u_1, \dots, u_n$ widerspiegelt wird,
 - indem die Längen monoton wachsender Teilfolgen, sogenannte *Runs*, innerhalb der Folge der Pseudozufallszahlen $u_1, u_2, \dots$ untersucht werden.

- Hierfür definieren wir zunächst die Zufallsvariablen $V_1, V_2, \dots$ rekursiv durch

$$V_{j+1} = \min\{i : i > V_j + 1, U_i > U_{i+1}\}, \quad \forall j = 1, 2, \dots, \quad (5)$$

wobei $V_1 = \min\{i : i \geq 1, U_i > U_{i+1}\}$.

- Die Zufallsvariablen $W_1, W_2, \dots$ mit

$$W_1 = V_1 \quad \text{und} \quad W_{j+1} = V_{j+1} - (V_j + 1) \quad \text{für } j = 1, 2, \dots \quad (6)$$

werden dann die Runs der Folge $U_1, U_2, \dots$ genannt.

Der zu konstruierende Signifikanztest beruht auf der folgenden Eigenschaft der Runs $W_1, W_2, \dots$

Theorem 3.3 *Die in (6) eingeführten Zufallsvariablen $W_1, W_2, \dots$ sind unabhängig und identisch verteilt mit*

$$P(W_j = k) = \frac{k}{(k+1)!}, \quad \forall k = 1, 2, \dots, \quad (7)$$

falls die Zufallsvariablen $U_1, U_2, \dots$ unabhängig und gleichverteilt im Intervall $(0, 1]$ sind.

Beweis

- Seien $U_1, U_2, \dots$ unabhängige und in $(0, 1]$ gleichverteilte Zufallsvariablen.
 - Für jedes $n \geq 1$ und für beliebige natürliche Zahlen $k_1, \dots, k_n \geq 1$ gilt dann

$$\begin{aligned} P(W_1 = k_1, \dots, W_n = k_n) &= P(V_1 = k_1, V_2 - V_1 - 1 = k_2, \dots, V_n - V_{n-1} - 1 = k_n) \\ &= P(V_1 = k_1, V_2 = k_2 + k_1 + 1, \dots, V_n = k_n + \dots + k_1 + n - 1) \\ &= P(U_i \leq U_{i+1}, \forall i = 1, \dots, k_1 - 1, U_{k_1} > U_{k_1+1}, \\ &\quad U_i \leq U_{i+1}, \forall i = k_1 + 2, \dots, k_1 + 1 + k_2 - 1, U_{k_1+1+k_2} > U_{k_1+1+k_2+1}, \dots, \\ &\quad U_i \leq U_{i+1}, \forall i = k_1 + 1 + \dots + k_{n-1} + 2, \dots, k_1 + 1 + \dots + k_{n-1} + 1 + k_n - 1, \\ &\quad U_{k_1+1+\dots+k_{n-1}+1+k_n} > U_{k_1+1+\dots+k_{n-1}+1+k_n+1}) \\ &= P(U_i \leq U_{i+1}, \forall i = 1, \dots, k_1 - 1, U_{k_1} > U_{k_1+1}) \\ &\quad P(U_i \leq U_{i+1}, \forall i = k_1 + 2, \dots, k_1 + 1 + k_2 - 1, U_{k_1+1+k_2} > U_{k_1+1+k_2+1}) \\ &\quad \dots P(U_i \leq U_{i+1}, \forall i = k_1 + 1 + \dots + k_{n-1} + 2, \dots, k_1 + 1 + \dots + k_{n-1} + 1 + k_n - 1, \\ &\quad U_{k_1+1+\dots+k_{n-1}+1+k_n} > U_{k_1+1+\dots+k_{n-1}+1+k_n+1}) \\ &= P(U_i \leq U_{i+1}, \forall i = 1, \dots, k_1 - 1, U_{k_1} > U_{k_1+1}) \\ &\quad \dots P(U_i \leq U_{i+1}, \forall i = 1, \dots, k_n - 1, U_{k_n} > U_{k_n+1}). \end{aligned}$$

– Hieraus folgt, dass die Runs $W_1, W_2, \dots$ unabhängig und identisch verteilt sind.

- Außerdem ergibt sich mit vollständiger Induktion, dass für beliebige $k \in \mathbb{R}$ und $t \in (0, 1]$

$$P(U_1 \leq \dots \leq U_k \leq t) = \frac{t^k}{k!}. \quad (8)$$

– Für $k = 1$ ist (8) offenbar richtig, und aus der Formel der totalen Wahrscheinlichkeit ergibt

sich, dass

$$\begin{aligned}
 P(U_1 \leq \dots \leq U_{k+1} \leq t) &= \int_0^1 P(U_1 \leq \dots \leq U_k \leq U_{k+1} \leq t \mid U_{k+1} = x) P(U_{k+1} \in dx) \\
 &= \int_0^1 P(U_1 \leq \dots \leq U_k \leq x \leq t \mid U_{k+1} = x) P(U_{k+1} \in dx) \\
 &= \int_0^1 P(U_1 \leq \dots \leq U_k \leq x \leq t) dx,
 \end{aligned}$$

wobei sich letzte Gleichheit aus der Unabhängigkeit und $(0, 1]$ -Gleichverteiltheit der $U_1, U_2, \dots$ ergibt.

– Es gelte nun (8) für ein gewisses $k \geq 1$. Dann gilt auch

$$\begin{aligned}
 P(U_1 \leq \dots \leq U_{k+1} \leq t) &= \int_0^t P(U_1 \leq \dots \leq U_k \leq x) dx \\
 &= \int_0^t \frac{x^k}{k!} dx = \frac{t^{k+1}}{(k+1)!},
 \end{aligned}$$

wobei sich die zweite Gleichheit aus der Induktionsannahme ergibt.

– Damit ist die Gültigkeit von (8) für jedes $k \geq 1$ bewiesen.

- Aus (8) ergibt sich nun darüber hinaus, dass für jedes $k \in \mathbb{N}$

$$\begin{aligned}
 P(U_1 \leq \dots \leq U_k, U_k > U_{k+1}) &= \int_0^1 P(U_1 \leq \dots \leq U_k, U_k > x) dx \\
 &= \int_0^1 \left(P(U_1 \leq \dots \leq U_k \leq 1) - P(U_1 \leq \dots \leq U_k \leq x) \right) dx \\
 &= \int_0^1 \left(\frac{1}{k!} - \frac{x^k}{k!} \right) dx \\
 &= \frac{1}{k!} - \frac{1}{(k+1)!} = \frac{k}{(k+1)!}.
 \end{aligned}$$

□

Beachte

- Es mögen nun hinreichend viele Pseudozufallszahlen $u_1, u_2, \dots$ erzeugt werden, so dass sich aus ihnen gemäß (5) und (6) die n Runs $w_1, \dots, w_n$ ergeben.
- Wir zerlegen die positive Halbachse in r Intervalle $(a_1, b_1], \dots, (a_r, b_r]$, so dass
 - die Wahrscheinlichkeiten

$$p_{0,j} = \sum_{k \in \mathbb{N} \cap (a_j, b_j]} \frac{k}{(k+1)!}, \quad j = 1, \dots, r$$

annähernd gleich sind, und

- für diese Wahrscheinlichkeiten betrachten wir den $(r-1)$ -dimensionalen (hypothetischen) Vektor $\mathbf{p}_0 = (p_{0,1}, \dots, p_{0,r-1})$ bzw.

- die Testgröße $T_n : \mathbb{R}^n \rightarrow [0, \infty)$ mit

$$T_n(w_1, \dots, w_n) = \sum_{j=1}^r \frac{(Y_j(w_1, \dots, w_n) - np_{0,j})^2}{np_{0,j}},$$

wobei $Y_j(w_1, \dots, w_n) = \#\{i : 1 \leq i \leq n, a_j < w_i \leq b_j\}$ die Anzahl derjenigen Runlängen $w_1, \dots, w_n$ bezeichnet, die in der j -ten Klasse liegen.

- Gemäß Theorem 3.3 wird nun bei großem n , was die Erzeugung entsprechend vieler Pseudozufallszahlen $u_1, u_2, \dots$ erfordert, die Hypothese $H_0 : \mathbf{p} = \mathbf{p}_0$ abgelehnt, falls $T(w_1, \dots, w_n) > \chi_{r-1, 1-\alpha}^2$.

3.2 Transformation gleichverteilter Pseudozufallszahlen

- Ausgehend von Standard-Pseudozufallszahlen $u_1, u_2, \dots$, für deren Erzeugung in Abschnitt 3.1.2 eine grundlegende Methode diskutiert wurde, lassen sich
 - Pseudozufallszahlen $x_1, x_2, \dots$ erzeugen, die als Realisierungen von Zufallsvariablen $X_1, X_2, \dots$ mit anderen Verteilungen aufgefasst werden können,
 - zum Beispiel als Realisierungen $x_1, x_2, \dots$ von exponential-, Poisson-, binomial- oder normalverteilten Zufallsvariablen $X_1, X_2, \dots$
- Dabei werden insbesondere sogenannte *Inversionsmethoden* bzw. *Akzeptanz- und Verwerfungsmethoden* verwendet, deren Grundideen wir hier mit Hilfe einiger Beispiele erläutern wollen.
- Eine wesentlich umfassendere Diskussion solcher Algorithmen findet man beispielsweise in
 - L. Devroye (1986) *Nonuniform Random Variate Generation*. Springer-Verlag, New York,
 - G.S. Fishman (1996) *Monte Carlo: Concepts, Algorithms and Applications*. Springer-Verlag, New York,
 - C.P. Robert und G. Casella (1999) *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

3.2.1 Inversionsmethode

- Als eine Grundlage zur Erzeugung von Pseudozufallszahlen $x_1, x_2, \dots$, die als Realisierungen von Zufallsvariablen $X_1, X_2, \dots$ aufgefasst werden können,
 - deren Verteilungsfunktion $F : \mathbb{R} \rightarrow [0, 1]$ eine beliebige monoton nichtfallende und rechtsseitig stetige Funktion ist mit $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$,
 - kann die folgende Eigenschaft der verallgemeinerten Inversen F^{-1} von F verwendet werden.
- *Zur Erinnerung*
 - Sei $F : \mathbb{R} \rightarrow [0, 1]$ eine beliebige Verteilungsfunktion. Dann heißt die Funktion $F^{-1} : (0, 1] \rightarrow \mathbb{R} \cup \{\infty\}$ mit

$$F^{-1}(y) = \inf\{x : F(x) \geq y\} \quad (9)$$

die *verallgemeinerte inverse Funktion* der Verteilungsfunktion F (bzw. der zugehörigen Verteilung).

- Für beliebige $x \in \mathbb{R}$ und $y \in (0, 1)$ gilt

$$y \leq F(x) \quad \text{genau dann, wenn} \quad F^{-1}(y) \leq x, \quad (10)$$

vgl. Lemma WR-4.1.

Theorem 3.4

- Sei $U_1, U_2, \dots$ eine Folge von unabhängigen und $(0, 1]$ -gleichverteilten Zufallsvariablen, und sei $F : \mathbb{R} \rightarrow [0, 1]$ eine beliebige Verteilungsfunktion.
- Dann sind die Zufallsvariablen $X_1, X_2, \dots$ mit $X_i = F^{-1}(U_i)$ für jedes $i = 1, 2, \dots$ unabhängig und besitzen die Verteilungsfunktion F .

Beweis

- Die Unabhängigkeit von $X_1, X_2, \dots$ ergibt sich unmittelbar aus dem Transformationssatz für unabhängige Zufallsvariablen, vgl. Theorem WR-3.18.
- Außerdem ergibt sich aus (10), dass für beliebige $x \in \mathbb{R}$ und $i \in \mathbb{N}$

$$P(X_i \leq x) = P(F^{-1}(U_i) \leq x) \stackrel{(10)}{=} P(U_i \leq F(x)) = F(x).$$

□

Beispiele

- Wir diskutieren nun einige Beispiele, die verdeutlichen,
 - wie Theorem 3.4 genutzt werden kann, um Pseudozufallszahlen $x_1, x_2, \dots$ zu erzeugen,
 - die als Realisierungen von unabhängigen Zufallsvariablen $X_1, X_2, \dots$ mit einer vorgegebenen Verteilungsfunktion $F : \mathbb{R} \rightarrow [0, 1]$ aufgefasst werden können.
- Wir sprechen dann kurz von F -verteilten Pseudozufallszahlen $x_1, x_2, \dots$,
 - obwohl die empirische Verteilungsfunktion $\hat{F}_n$ der (konkreten) Stichprobe $x_1, \dots, x_n$
 - selbst für große n natürlich nur näherungsweise mit F übereinstimmt.
- Dabei kann Theorem 3.4 nur dann *direkt* angewendet werden, wenn
 - sich die verallgemeinerte inverse Funktion F^{-1} von F explizit (d.h. durch eine geschlossene Formel) bestimmen lässt,
 - was jedoch eher eine Ausnahmesituation ist.

1. Exponentialverteilung

- Sei $\lambda > 0$ und $F : \mathbb{R} \rightarrow [0, 1]$ die Verteilungsfunktion der $\text{Exp}(\lambda)$ -Verteilung mit

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

- Dann gilt $F^{-1}(u) = -\lambda^{-1} \log(1 - u)$ für jedes $u \in (0, 1]$.
- Gemäß Theorem 3.4 gilt also
 - für die Zufallsvariable $X = -\lambda^{-1} \log U \sim \text{Exp}(\lambda)$, falls U und somit auch $1 - U$ in $(0, 1]$ gleichverteilte Zufallsvariablen sind,
 - und die Pseudozufallszahlen $x_1, \dots, x_n$ mit

$$x_i = -\frac{\log u_i}{\lambda} \quad \text{für } i = 1, \dots, n$$

können als Realisierungen von unabhängigen und $\text{Exp}(\lambda)$ -verteilten Zufallsvariablen aufgefasst werden,

- falls $u_1, \dots, u_n$ die Realisierungen von unabhängigen, in $(0, 1]$ gleichverteilten Zufallsvariablen $U_1, \dots, U_n$ sind.

2. Erlang-Verteilung

- Sei $\lambda > 0$ und $r \in \mathbb{N}$, und sei $F : \mathbb{R} \rightarrow [0, 1]$ die Verteilungsfunktion der Erlang-Verteilung, d.h. der $\Gamma(\lambda, r)$ -Verteilung mit

$$F(x) = \begin{cases} \int_0^x \frac{\lambda e^{-\lambda v} (\lambda v)^{r-1}}{(r-1)!} dv, & \text{falls } x \geq 0, \\ 0, & \text{falls } x < 0. \end{cases} \quad (11)$$

- Dann lässt sich die verallgemeinerte inverse Funktion F^{-1} von F nicht explizit bestimmen, und somit kann Theorem 3.4 *nicht direkt* angewendet werden.
- In Abschnitt 1.3.1 der Vorlesung „Statistik I“ hatten wir jedoch gezeigt, dass $X_1 + \dots + X_r \sim \Gamma(\lambda, r)$, falls die Zufallsvariablen $X_1, \dots, X_r$ unabhängig und $\text{Exp}(\lambda)$ -verteilt sind.
- Gemäß Theorem 3.4 können also
 - die Pseudozufallszahlen $y_1, \dots, y_n$ mit

$$y_i = x_{r(i-1)+1} + \dots + x_{ri} = -\frac{\log(u_{r(i-1)+1} \cdot \dots \cdot u_{ri})}{\lambda} \quad \text{für } i = 1, \dots, n$$

als Realisierungen von unabhängigen und $\Gamma(\lambda, r)$ -verteilten Zufallsvariablen aufgefasst werden,

- falls $u_1, \dots, u_{rn}$ die Realisierungen von unabhängigen, in $(0, 1]$ gleichverteilten Zufallsvariablen $U_1, \dots, U_{rn}$ sind.
- Für $\lambda = 1/2$ können die Pseudozufallszahlen $y_1, \dots, y_n$ somit insbesondere als Realisierungen von unabhängigen und χ^2_{2r} -verteilten Zufallsvariablen aufgefasst werden.

3. Normalverteilung

- Zur Erzeugung von normalverteilten Pseudozufallszahlen gibt es beispielsweise den folgenden *Box-Muller-Algorithmus*, der ebenfalls exponentialverteilte Pseudozufallszahlen mit in die Überlegungen einbezieht.
- Die Zufallsvariablen U_1, U_2 seien unabhängig und im Intervall $(0, 1]$ gleichverteilt.
 - Gemäß Theorem 3.4 ist dann $X = -2 \log U_1$ eine $\text{Exp}(1/2)$ -verteilte Zufallsvariable, und
 - der Zufallsvektor (Y_1, Y_2) mit

$$Y_1 = \sqrt{X} \cos(2\pi U_2), \quad Y_2 = \sqrt{X} \sin(2\pi U_2)$$

ist $N(\mathbf{0}, \mathbf{I})$ -verteilt, d.h. Y_1, Y_2 sind unabhängige und $N(0, 1)$ -verteilte Zufallsvariablen,

- denn für beliebige $y_1, y_2 \in \mathbb{R}$ gilt

$$\begin{aligned} P(Y_1 \leq y_1, Y_2 \leq y_2) &= P\left(\sqrt{-2 \log U_1} \cos(2\pi U_2) \leq y_1, \sqrt{-2 \log U_1} \sin(2\pi U_2) \leq y_2\right) \\ &= \frac{1}{2} \int_0^1 \int_0^\infty \mathbb{I}(\sqrt{x} \cos(2\pi u) \leq y_1, \sqrt{x} \sin(2\pi u) \leq y_2) e^{-x/2} dx du \\ &= \frac{1}{2\pi} \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} e^{-(v^2+w^2)/2} dv dw \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_1} e^{-v^2/2} dv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_2} e^{-w^2/2} dw, \end{aligned}$$

wobei sich die vorletzte Gleichheit mit der Substitution

$$v = \sqrt{x} \cos(2\pi u), \quad w = \sqrt{x} \sin(2\pi u)$$

ergibt, deren Funktionaldeterminante gleich π ist.

- Die Pseudozufallszahlen $y_1, \dots, y_{2n}$ mit

$$y_{2k-1} = \sqrt{-2 \log u_{2k-1}} \cos(2\pi u_{2k}), \quad y_{2k} = \sqrt{-2 \log u_{2k-1}} \sin(2\pi u_{2k}) \quad (12)$$

- können somit als Realisierungen von unabhängigen und $N(0, 1)$ -verteilten Zufallsvariablen aufgefasst werden,
- falls $u_1, \dots, u_{2n}$ die Realisierungen von unabhängigen und im Intervall $(0, 1]$ gleichverteilten Zufallsvariablen $U_1, \dots, U_{2n}$ sind.
- Für beliebige $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ können dann die Pseudozufallszahlen $y'_1, \dots, y'_{2n}$ mit $y'_i = \sigma(y_i + \mu)$ als Realisierungen von unabhängigen und $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen aufgefasst werden.
- Beachte
 - Ein schnellerer Algorithmus zur Erzeugung von normalverteilten Pseudozufallszahlen ergibt sich,
 - wenn zusätzlich die im folgenden Abschnitt 3.2.3 diskutierte Akzeptanz- und Verwerfungsmethode angewendet wird,
 - weil dann die (relativ zeitaufwendige) Berechnung der trigonometrischen Funktionen in (12) vermieden werden kann.

3.2.2 Transformationsalgorithmen für diskrete Verteilungen

- Wenn Pseudozufallszahlen $x_1, x_2, \dots$ erzeugt werden sollen,
 - die als Realisierungen von diskreten Zufallsvariablen $X_1, X_2, \dots$ aufgefasst werden können,
 - welche jeweils die Werte $a_0, a_1, \dots \in \mathbb{R}$ mit den Wahrscheinlichkeiten $p_j = P(X_i = a_j) \geq 0$ für $j = 0, 1, \dots$ annehmen.
- Dann ist es manchmal zweckmäßig, wie folgt vorzugehen:
 - Sei U eine $(0, 1]$ -verteilte Zufallsvariable, und die Zufallsvariable X sei gegeben durch

$$X = \begin{cases} a_0, & \text{falls } U < p_0, \\ a_1, & \text{falls } p_0 \leq U < p_0 + p_1, \\ \vdots & \\ a_j, & \text{falls } p_0 + \dots + p_{j-1} \leq U < p_0 + \dots + p_j, \\ \vdots & \end{cases} \quad (13)$$

- Dann gilt $P(X = a_j) = p_j$ für jedes $j = 0, 1, \dots$
- Die Pseudozufallszahlen $x_1, \dots, x_n$ mit

$$x_i = \begin{cases} a_0, & \text{falls } u_i < p_0, \\ a_0, & \text{falls } p_0 \leq u_i < p_0 + p_1, \\ \vdots & \\ a_j, & \text{falls } p_0 + \dots + p_{j-1} \leq u_i < p_0 + \dots + p_j, \\ \vdots & \end{cases}$$

- können somit als Realisierungen von unabhängigen und $\mathbf{p}$ -verteilten Zufallsvariablen aufgefasst werden, wobei $\mathbf{p} = (p_0, p_1, \dots)^\top$,
- falls $u_1, \dots, u_n$ die Realisierungen von unabhängigen und im Intervall $(0, 1]$ gleichverteilten Zufallsvariablen $U_1, \dots, U_n$ sind.

Beispiel (*Geometrische Verteilung*)

- Wir betrachten die folgenden Werte a_j und die zugehörigen (Einzel-) Wahrscheinlichkeiten p_j .
 - Sei $a_j = j$ für $j = 0, 1, \dots$, und für $0 < p < 1$ mit $q = 1 - p$ sei

$$p_j = \begin{cases} 0, & \text{falls } j = 0, \\ p q^{j-1}, & \text{falls } j \geq 1. \end{cases}$$

- Dann gilt für jedes $j \geq 1$

$$1 - (p_1 + \dots + p_j) = p_{j+1} + p_{j+2} + \dots = p \sum_{i=j}^{\infty} q^i = q^j \quad (14)$$

bzw. $p_j = q^{j-1} - q^j$.

- Außerdem betrachten wir die Zufallsvariable

$$X = \left\lfloor \frac{\log U}{\log q} \right\rfloor + 1, \quad (15)$$

wobei U eine $(0, 1]$ -gleichverteilte Zufallsvariable ist und $\lfloor z \rfloor$ den ganzzahligen Teil von z bezeichnet.

- Dann gilt $P(X = j) = p q^{j-1}$ für jedes $j = 1, 2, \dots$, d.h. $X \sim \text{Geo}(p)$,
 - denn aus (14) und (15) ergibt sich, dass

$$\begin{aligned} X &\stackrel{(15)}{=} \min \left\{ j \geq 1 : j > \frac{\log U}{\log q} \right\} \\ &= \min \left\{ j \geq 1 : j \log q < \log U \right\} \\ &= \min \left\{ j \geq 1 : q^j < U \right\} \\ &= \sum_{j=1}^{\infty} j \mathbb{I}(q^j < U \leq q^{j-1}) \\ &\stackrel{(14)}{=} \sum_{j=1}^{\infty} j \mathbb{I}(p_1 + \dots + p_{j-1} \leq 1 - U < p_1 + \dots + p_j) \Bigg), \end{aligned}$$

- wobei die Zufallsvariable $1 - U$ ebenfalls $(0, 1]$ -gleichverteilt ist.

- Die Pseudozufallszahlen $x_1, \dots, x_n$ mit

$$x_i = \left\lfloor \frac{\log u_i}{\log q} \right\rfloor + 1$$

- können somit als Realisierungen von unabhängigen und geometrisch verteilten Zufallsvariablen $X_1, \dots, X_n \sim \text{Geo}(p)$ aufgefasst werden,

- falls $u_1, \dots, u_n$ die Realisierungen von unabhängigen und im Intervall $(0, 1]$ gleichverteilten Zufallsvariablen $U_1, \dots, U_n$ sind.

Für einige diskrete Verteilungen gibt es *spezielle* Transformationsalgorithmen, um Pseudozufallszahlen zu erzeugen, die als Realisierungen von Zufallsvariablen mit einer solchen Verteilung aufgefasst werden können.

Beispiele

1. Poisson-Verteilung (mit kleinem Erwartungswert λ)

- Falls $\lambda > 0$ eine kleine Zahl ist, dann ist das folgende Verfahren geeignet, um Poisson-verteilte Pseudozufallszahlen durch die Transformation
 - von exponentialverteilten Pseudozufallszahlen (auf ähnliche Weise wie in Abschnitt 3.2.1)
 - oder direkt aus $(0, 1]$ -gleichverteilten Pseudozufallszahlen zu erzeugen.
- Die Zufallsvariablen $X_1, X_2, \dots$ seien unabhängig und $\text{Exp}(\lambda)$ -verteilt.
 - Für die Zufallsvariable $Y = \max\{k \geq 0 : X_1 + \dots + X_k \leq 1\}$ ergibt sich dann aus der Darstellungsformel (11) für die Verteilungsfunktion der Erlang-Verteilung, dass für jedes $j \geq 0$

$$\begin{aligned}
 P(Y = j) &= P(Y \geq j) - P(Y \geq j + 1) \\
 &= P(X_1 + \dots + X_j \leq 1) - P(X_1 + \dots + X_{j+1} \leq 1) \\
 &= \int_0^1 \frac{\lambda e^{-\lambda v} (\lambda v)^{j-1}}{(j-1)!} dv - \int_0^1 \frac{\lambda e^{-\lambda v} (\lambda v)^j}{j!} dv \\
 &= \int_0^1 \frac{d}{dv} \left(\frac{e^{-\lambda v} (\lambda v)^j}{j!} \right) dv \\
 &= \frac{e^{-\lambda} \lambda^j}{j!}.
 \end{aligned}$$

- Mit anderen Worten: Es gilt $Y \sim \text{Poi}(\lambda)$.

- Die Pseudozufallszahlen $y_1, \dots, y_n$ mit

$$y_i = \max\{k \geq 0 : x_1 + \dots + x_k \leq i\} - y_{i-1}, \quad \forall i = 1, \dots, n, \quad (16)$$

bzw.

$$y_i = \max\{k \geq 0 : u_1 \cdot \dots \cdot u_k \geq e^{-i\lambda}\} - y_{i-1}, \quad \forall i = 1, \dots, n, \quad (17)$$

wobei $y_0 = 0$ und $x_j = -\lambda^{-1} \log u_j$ für $j = 1, 2, \dots$,

- können somit als Realisierungen von unabhängigen und $\text{Poi}(\lambda)$ -verteilten Zufallsvariablen aufgefasst werden,
- falls $x_1, x_2, \dots$ die Realisierungen von unabhängigen und $\text{Exp}(\lambda)$ -verteilten Zufallsvariablen $X_1, X_2, \dots$ sind bzw.
- falls $u_1, u_2, \dots$ die Realisierungen von unabhängigen und im Intervall $(0, 1]$ gleichverteilten Zufallsvariablen $U_1, U_2, \dots$ sind.
- Beachte
 - Weil der Erwartungswert der $\text{Poi}(\lambda)$ -Verteilung gleich λ ist, sind bei dem durch (16) bzw. (17) gegebenen Algorithmus im Mittel λ gleichverteilte Pseudozufallszahlen erforderlich, um eine neue $\text{Poi}(\lambda)$ -verteilte Pseudozufallszahl zu erzeugen.
 - Dieser Aufwand lässt sich reduzieren, wenn für große λ wie folgt vorgegangen wird.

2. Poisson-Verteilung (mit großem Erwartungswert λ)

- Falls $\lambda > 0$ eine große Zahl ist, wobei $a_j = j$ und $p_j = e^{-\lambda} \lambda^j / j!$ für $j = 0, 1, \dots$,
 - dann ist das unmittelbar auf dem Transformationsansatz (13) beruhende Verfahren besser geeignet, um $\text{Poi}(\lambda)$ -verteilte Pseudozufallszahlen zu erzeugen,
 - wenn dabei die Gültigkeit der Ungleichungen

$$U < p_0, \quad p_0 \leq U < p_0 + p_1, \dots, p_0 + \dots + p_{j-1} \leq U < p_0 + \dots + p_j, \dots \quad (18)$$

in der folgenden Reihenfolge überprüft wird,

- wobei die Rekursionsformel

$$p_{j+1} = \frac{\lambda}{j+1} p_j, \quad \forall j \geq 0,$$

zur Berechnung der Summen $P_j = \sum_{k=0}^j p_k$ für $j \geq 0$ verwendet wird.

- Sei $[\lambda] > 0$ der ganzzahlige Teil von λ . Dann wird zunächst geprüft, ob $U < P_{[\lambda]}$ gilt.
 - Falls diese Ungleichung gilt, dann wird geprüft, ob $U < P_{[\lambda]-1}, U < P_{[\lambda]-2}, \dots$ gilt, wobei in diesem Fall $X = \min\{k : U < P_k\}$ gesetzt wird.
 - Falls $U < P_{[\lambda]}$ nicht gilt, dann wird geprüft, ob $U < P_{[\lambda]+1}, U < P_{[\lambda]+2}, \dots$ gilt, wobei erneut $X = \min\{k : U < P_k\}$ gesetzt wird.
- Für den Erwartungswert $\mathbb{E}V$ der dabei erforderlichen Anzahl V von Überprüfungen gilt dann näherungsweise, dass

$$\begin{aligned} \mathbb{E}V &\approx 1 + \mathbb{E}|X - \lambda| \\ &= 1 + \sqrt{\lambda} \mathbb{E}\left(\frac{|X - \lambda|}{\sqrt{\lambda}}\right) \\ &\approx 1 + 0.798 \sqrt{\lambda}, \end{aligned}$$

- wobei in der letzten Approximationsformel die Tatsache genutzt wurde, dass die Zufallsvariable $(X - \lambda)/\sqrt{\lambda}$ für große λ näherungsweise $N(0, 1)$ -verteilt ist,
- was sich aus der Faltungstabilität der Poisson-Verteilung und aus dem zentralen Grenzwertsatz für Summen von unabhängigen und identisch verteilten Zufallsvariablen (Theorem WR-5.16) ergibt.
- Bei dieser Vorgehensweise wächst also
 - die im Mittel erforderliche Anzahl von Überprüfungen nur wie die Quadratwurzel $\sqrt{\lambda}$ von λ ,
 - wogegen bei dem vorher diskutierten Verfahren zur Erzeugung $\text{Poi}(\lambda)$ -verteilter Pseudozufallszahlen die jeweils im Mittel erforderliche Anzahl von Standard-Pseudozufallszahlen wie λ wächst.

3. Binomial-Verteilung

- Bei der Erzeugung von binomialverteilten Pseudozufallszahlen kann man ähnlich wie im Fall der Poisson-Verteilung vorgehen.
 - Für beliebige, jedoch fest vorgegebene Zahlen $n \in \mathbb{N}$ und $p \in (0, 1)$ mit $q = 1 - p$ sei

$$a_j = j \quad \text{und} \quad p_j = \frac{n!}{j!(n-j)!} p^j q^{n-j}, \quad \forall j = 0, 1, \dots, n.$$

- Für $j = 0, 1, \dots, n$ werden dann die Summen $P_j = \sum_{k=0}^j p_k$ mit der Rekursionsformel

$$p_{j+1} = \frac{n-j}{j+1} \frac{p}{q} p_j, \quad \forall j = 0, 1, \dots, n-1$$

berechnet.

- Falls $np > 0$ klein ist, dann wird
 - die Gültigkeit der Ungleichungen (18) in der natürlichen Reihenfolge überprüft,
 - wobei mit $U < p_0$ begonnen und $X = \min\{k : U < P_k\}$ gesetzt wird.
- Falls np groß ist,
 - dann ist es effizienter, die Gültigkeit der Ungleichungen (18) in der folgenden Reihenfolge zu verifizieren. Dabei wird zunächst geprüft, ob $U < P_{[np]}$ gilt.
 - Falls diese Ungleichung gilt, dann wird geprüft, ob $U < P_{[np]-1}, U < P_{[np]-2}, \dots$ gilt, wobei in diesem Fall $X = \min\{k : U < P_k\}$ gesetzt wird.
 - Falls $U < P_{[np]}$ nicht gilt, dann wird geprüft, ob $U < P_{[np]+1}, U < P_{[np]+2}, \dots$ gilt, wobei erneut $X = \min\{k : U < P_k\}$ gesetzt wird.

3.2.3 Akzeptanz- und Verwerfungsmethode

- Wir diskutieren nun eine weitere Methode, mit der man Pseudozufallszahlen $y_1, y_2, \dots$ erzeugen kann,
 - die als Realisierungen von unabhängigen und identisch verteilten Zufallsvariablen $Y_1, Y_2, \dots$ mit einer vorgegebenen Verteilungsfunktion G aufgefasst werden können,
 - wenn dabei von (nicht notwendig $(0, 1]$ -gleichverteilten) Pseudozufallszahlen $x_1, x_2, \dots$ ausgegangen wird,
 - die Realisierungen von unabhängigen und identisch verteilten Zufallsvariablen $X_1, X_2, \dots$ mit einer gewissen Verteilungsfunktion F sind,
 - so dass G *absolutstetig* bezüglich F ist, wobei die Dichte $g(x) = dG(x)/dF(x)$ beschränkt ist,
 - d.h., es gibt eine Konstante $c > 0$ mit

$$g(x) \leq c \quad \text{und} \quad G(y) = \int_{-\infty}^y g(x) dF(x), \quad \forall x, y \in \mathbb{R}. \quad (19)$$

- Wir betrachten zunächst den diskreten Fall.
 - Sei $a_j = j$ für jedes $j = 0, 1, \dots$, und seien $\mathbf{p} = (p_0, p_1, \dots)^\top$ bzw. $\mathbf{q} = (q_0, q_1, \dots)^\top$ zwei beliebige Wahrscheinlichkeitsfunktionen, so dass für jedes $j = 0, 1, \dots$ aus $p_j = 0$ die Gültigkeit von $q_j = 0$ folgt.
 - Sei $X : \Omega \rightarrow \{0, 1, \dots\}$ eine Zufallsvariable mit $P(X = j) = p_j$ für jedes $j = 0, 1, \dots$,
 - und sei $c > 0$ eine positive Zahl, so dass

$$\left(g(j) = \right) \quad \frac{q_j}{p_j} \leq c \quad \text{für jedes } j \geq 0 \text{ mit } p_j > 0. \quad (20)$$

Theorem 3.5

- Sei $(U_1, X_1), (U_2, X_2), \dots$ eine Folge von unabhängigen und identisch verteilten Zufallsvektoren, deren Komponenten unabhängig sind, wobei U_i eine $(0, 1]$ -gleichverteilte Zufallsvariable und X_i gemäß $\mathbf{p}$ verteilt ist.
- Dann ist

- die Zufallsvariable

$$I = \min \left\{ k \geq 1 : U_k < \frac{q_{X_k}}{c p_{X_k}} \right\} \quad (21)$$

geometrisch verteilt mit Erwartungswert c , d.h. $I \sim \text{Geo}(c^{-1})$,

- und die Zufallsvariable $Y = X_I$ ist gemäß $\mathbf{q}$ verteilt.

Beweis

- Aus der Definitionsgleichung (21) von I ergibt sich, dass für jedes $j \geq 1$

$$\begin{aligned}
 P(I = j) &= P\left(U_1 \geq \frac{q_{X_1}}{c p_{X_1}}, \dots, U_{j-1} \geq \frac{q_{X_{j-1}}}{c p_{X_{j-1}}}, U_j < \frac{q_{X_j}}{c p_{X_j}}\right) \\
 &= P\left(U_1 \geq \frac{q_{X_1}}{c p_{X_1}}\right) \dots P\left(U_{j-1} \geq \frac{q_{X_{j-1}}}{c p_{X_{j-1}}}\right) P\left(U_j < \frac{q_{X_j}}{c p_{X_j}}\right) \\
 &= p q^{j-1},
 \end{aligned}$$

- wobei $q = 1 - p$ und

$$\begin{aligned}
 p &= P\left(U_1 < \frac{q_{X_1}}{c p_{X_1}}\right) \\
 &= \sum_{k: p_k > 0} P\left(U_1 < \frac{q_{X_1}}{c p_{X_1}} \mid X_1 = k\right) P(X_1 = k) \\
 &= \sum_{k: p_k > 0} P\left(U_1 < \frac{q_k}{c p_k}\right) p_k \\
 &= \sum_{k: p_k > 0} \frac{q_k}{c p_k} p_k = \frac{1}{c}.
 \end{aligned}$$

- Hieraus folgt, dass $I \sim \text{Geo}(c^{-1})$.

- Außerdem gilt für jedes $j \geq 1$ mit $p_j > 0$

$$\begin{aligned}
 P(Y = j) &= P(X_I = j) = \sum_{k=1}^{\infty} P(X_I = j, I = k) \\
 &= \sum_{k=1}^{\infty} P(X_k = j, I = k) = \sum_{k=1}^{\infty} P(X_k = j) q^{k-1} P\left(U_k \leq \frac{q_j}{c p_j}\right) \\
 &= \sum_{k=1}^{\infty} p_j q^{k-1} \frac{q_j}{c p_j} = \frac{q_j}{c} \sum_{k=1}^{\infty} q^{k-1} \\
 &= \frac{q_j}{c} \frac{1}{1 - q} = q_j,
 \end{aligned}$$

und für jedes $j \geq 1$ mit $p_j = 0$ gilt

$$P(Y = j) = \sum_{k=1}^{\infty} P(X_k = j, I = k) \leq \sum_{k=1}^{\infty} P(X_k = j) = 0.$$

□

Beachte

- Aus Theorem 3.5 ergibt sich, dass im Mittel c Pseudozufallszahlen erforderlich sind, die F -verteilt sind, um eine G -verteilte Pseudozufallszahl zu erzeugen.
- Wenn mehrere Alternativen für die Wahl der Verteilungsfunktion F zur Verfügung stehen,
 - die ähnlich günstige Eigenschaften hinsichtlich der Erzeugung von F -verteilten Pseudozufallszahlen aufweisen,
 - dann sollte somit diejenige Verteilungsfunktion F gewählt werden, für die c am kleinsten ist.
- Außerdem ergibt sich aus Theorem 3.5,
 - dass die Werte $g(x)$ bzw. $g(j)$ der Dichte in (19) bzw. (20) *nicht vollständig*,
 - sondern nur bis auf einen konstanten Proportionalitätsfaktor bekannt sein müssen.

Im allgemeinen (d.h. nicht notwendig diskreten) Fall kann man ähnlich vorgehen, wobei das folgende Resultat als Grundlage zur Konstruktion von Akzeptanz- und Verwerfungsalgorithmen dient.

Theorem 3.6

- Seien $F, G : \mathbb{R} \rightarrow [0, 1]$ zwei beliebige Verteilungsfunktionen, so dass (19) gilt.
- Sei $(U_1, X_1), (U_2, X_2), \dots$ eine Folge von unabhängigen und identisch verteilten Zufallsvektoren, deren Komponenten unabhängig sind, wobei U_i eine $(0, 1]$ -gleichverteilte Zufallsvariable und X_i gemäß F verteilt ist.
- Dann ist die Zufallsvariable

$$I = \min \left\{ k \geq 1 : U_k < \frac{g(X_k)}{c} \right\} \quad (22)$$

geometrisch verteilt mit Erwartungswert c , d.h. $I \sim \text{Geo}(c^{-1})$, und die Zufallsvariable $Y = X_I$ ist gemäß G verteilt.

Beweis

- Ähnlich wie im Beweis von Theorem 3.5 ergibt sich, dass $P(I = j) = p q^{j-1}$ für jedes $j \geq 1$ gilt, wobei

$$\begin{aligned} p &= P\left(U_1 < \frac{g(X_1)}{c}\right) = \int_{\mathbb{R}} P\left(U_1 < \frac{g(x)}{c} \mid X_1 = x\right) dF(x) \\ &= \int_{\mathbb{R}} P\left(U_1 < \frac{g(x)}{c}\right) dF(x) = \int_{\mathbb{R}} \frac{g(x)}{c} dF(x) = \frac{1}{c}. \end{aligned}$$

- Außerdem gilt für jedes $y \in \mathbb{R}$

$$\begin{aligned} P(Y \leq y) &= P(X_I \leq y) = \sum_{k=1}^{\infty} P(X_I \leq y, I = k) \\ &= \sum_{k=1}^{\infty} P(X_k \leq y, I = k) \\ &= \sum_{k=1}^{\infty} \int_{-\infty}^y P(I = k \mid X_k = v) dF(v) \\ &= \sum_{k=1}^{\infty} q^{k-1} \int_{-\infty}^y P\left(U_k < \frac{g(v)}{c}\right) dF(v) \\ &= \frac{1}{1-q} \int_{-\infty}^y \frac{g(v)}{c} dF(v) \\ &= \int_{-\infty}^y g(v) dF(v) = G(y), \end{aligned}$$

wobei $1 - q = p = c^{-1}$. □

Auf die gleiche Weise ergibt sich die folgende vektorielle Version von Theorem 3.6.

Theorem 3.7

- Sei $m \geq 1$ eine beliebige, jedoch fest vorgegebene natürliche Zahl, und seien $F, G : \mathbb{R}^m \rightarrow [0, 1]$ zwei beliebige Verteilungsfunktionen (von m -dimensionalen Zufallsvektoren), so dass es gibt eine Konstante $c > 0$ mit

$$g(\mathbf{x}) \leq c \quad \text{und} \quad G(\mathbf{y}) = \int_{(-\infty, \mathbf{y}]} g(\mathbf{x}) dF(\mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^m. \quad (23)$$

- Sei $(U_1, \mathbf{X}_1), (U_2, \mathbf{X}_2), \dots$ eine Folge von unabhängigen und identisch verteilten Zufallsvektoren, deren Komponenten unabhängig sind, wobei U_i eine $(0, 1]$ -gleichverteilte Zufallsvariable und $\mathbf{X}_i$ gemäß F verteilt ist.
- Dann ist die Zufallsvariable

$$I = \min \left\{ k \geq 1 : U_k < \frac{g(\mathbf{X}_k)}{c} \right\} \quad (24)$$

geometrisch verteilt mit Erwartungswert c , d.h. $I \sim \text{Geo}(c^{-1})$, und der Zufallsvektor $\mathbf{Y} = \mathbf{X}_I$ ist gemäß G verteilt.

Beispiele

1. Gleichverteilung in beschränkten Borel-Mengen

- Der Zufallsvektor $\mathbf{X} : \Omega \rightarrow \mathbb{R}^m$ (mit der Verteilungsfunktion F) sei gleichverteilt in dem Quader $(-1, 1]^m$, und $B \in \mathcal{B}((-1, 1]^m)$ sei eine beliebige Borel-Menge in $(-1, 1]^m$ mit positivem Lebesgue-Maß $|B|$.
- Dann ist die Verteilungsfunktion $G : \mathbb{R}^m \rightarrow [0, 1]$ mit

$$G(\mathbf{y}) = \int_{(-\infty, \mathbf{y}]} \frac{\mathbb{I}(\mathbf{x} \in B)}{|B|} dF(\mathbf{x}), \quad \forall \mathbf{y} \in \mathbb{R}^m$$

absolutstetig bezüglich F , und für die (Radon-Nikodym-) Dichte $g : \mathbb{R}^m \rightarrow [0, \infty)$ gilt

$$g(\mathbf{x}) = \frac{\mathbb{I}(\mathbf{x} \in B)}{|B|} \leq c = |B|^{-1} \quad \text{bzw.} \quad \frac{g(\mathbf{x})}{c} = \mathbb{I}(\mathbf{x} \in B), \quad \forall \mathbf{x} \in \mathbb{R}^m.$$

- Gemäß Theorem 3.7 können nun in B gleichverteilte Pseudozufallsvektoren $\mathbf{y}_1, \mathbf{y}_2, \dots$ auf die folgende Weise erzeugt werden.
 1. Generiere m Pseudozufallszahlen $u_1, \dots, u_m$, die im Intervall $(0, 1]$ gleichverteilt sind.
 2. Falls $(2u_1 - 1, \dots, 2u_m - 1)^\top \notin B$ gilt, dann kehre zu Schritt 1 zurück.
 3. Setze $\mathbf{y} = (2u_1 - 1, \dots, 2u_m - 1)^\top$.

2. Normalverteilung

- Als eine Alternative zu dem in Abschnitt 3.2.1 diskutierten Box-Muller-Algorithmus betrachten wir noch ein anderes Verfahren zur Erzeugung von normalverteilten Pseudozufallszahlen,
 - das in der Literatur die *Polar-Methode* genannt wird,
 - wobei nun die Berechnung der trigonometrischen Funktionen in (12) vermieden wird.
- Der Zufallsvektor (V_1, V_2) sei im Einheitskreis $B = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$ gleichverteilt.
- Dann ist der Zufallsvektor (Y_1, Y_2) mit

$$Y_1 = \sqrt{-2 \log(V_1^2 + V_2^2)} \frac{V_1}{\sqrt{V_1^2 + V_2^2}}, \quad Y_2 = \sqrt{-2 \log(V_1^2 + V_2^2)} \frac{V_2}{\sqrt{V_1^2 + V_2^2}}$$

$N(\mathbf{0}, \mathbf{I})$ -verteilt, d.h. Y_1, Y_2 sind unabhängige und $N(0, 1)$ -verteilte Zufallsvariablen,

- denn mit der Substitution

$$v_1 = r \cos \theta, \quad v_2 = r \sin \theta,$$

- d.h. durch Übergang zu Polarkoordinaten, ergibt sich für beliebige $y_1, y_2 \in \mathbb{R}$

$$\begin{aligned} P(Y_1 \leq y_1, Y_2 \leq y_2) &= \frac{1}{\pi} \int_B \mathbb{I}\left(\frac{v_1 \sqrt{-2 \log(v_1^2 + v_2^2)}}{\sqrt{v_1^2 + v_2^2}} \leq y_1, \frac{v_2 \sqrt{-2 \log(v_1^2 + v_2^2)}}{\sqrt{v_1^2 + v_2^2}} \leq y_2\right) d(v_1, v_2) \\ &= \frac{1}{\pi} \int_0^{2\pi} \int_0^1 r \mathbb{I}\left(\sqrt{-2 \log(r^2)} \cos \theta \leq y_1, \sqrt{-2 \log(r^2)} \sin \theta \leq y_2\right) dr d\theta \\ &= \frac{1}{2} \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty \mathbb{I}\left(\sqrt{x} \cos \theta \leq y_1, \sqrt{x} \sin \theta \leq y_2\right) e^{-x/2} dx d\theta, \end{aligned}$$

- wobei die letzte Gleichheit das Ergebnis der folgenden Substitution ist:

$$x = -2 \log(r^2) \quad \text{bzw.} \quad -\frac{1}{2} e^{-x/2} dx = 2r dr.$$

- Hieraus ergibt sich nun die Behauptung (genauso wie bei der Begründung von Formel (12) in Abschnitt 3.2.1).
- Die Pseudozufallszahlen $y_1, \dots, y_{2n}$ mit

$$y_{2k-1} = \sqrt{-2 \log(v_{2k-1}^2 + v_{2k}^2)} \frac{v_{2k-1}}{\sqrt{v_{2k-1}^2 + v_{2k}^2}}, \quad y_{2k} = \sqrt{-2 \log(v_{2k-1}^2 + v_{2k}^2)} \frac{v_{2k}}{\sqrt{v_{2k-1}^2 + v_{2k}^2}}$$

- können somit als Realisierungen von unabhängigen und $N(0, 1)$ -verteilten Zufallsvariablen aufgefasst werden,
- falls $(v_1, v_2), \dots, (v_{2n-1}, v_{2n})$ die Realisierungen von unabhängigen und im Einheitskreis $B = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$ gleichverteilten Zufallsvektoren $(V_1, V_2), \dots, (V_{2n-1}, V_{2n})$ sind,
- die (wie im vorhergehenden Beispiel beschrieben) mit der Akzeptanz- und Verwerfungsmethode erzeugt werden können.

3.2.4 Quotienten von gleichverteilten Zufallsvariablen

In vielen Fällen lassen sich Zufallsvariablen mit absolutstetigen Verteilungen als Quotienten von gleichverteilten Zufallsvariablen darstellen.

- In Kombination mit der in Abschnitt 3.2.3 diskutierten Akzeptanz- und Verwerfungsmethode führt dies zu einem weiteren Typ von Simulationsalgorithmen.
- Um dies zu präzisieren, benötigen wir den folgenden Transformationssatz für die Dichte von absolutstetigen Zufallsvektoren, der bereits in Abschnitt 1.3.2 der Vorlesung „Statistik I“ erwähnt wurde.

Theorem 3.8

- Sei $\mathbf{X} = (X_1, \dots, X_n)^\top : \Omega \rightarrow \mathbb{R}^n$ ein absolutstetiger Zufallsvektor mit der (gemeinsamen) Dichte $f_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, \infty)$, und sei $\varphi = (\varphi_1, \dots, \varphi_n) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine Borel-messbare Abbildung mit stetigen partiellen Ableitungen $\partial \varphi_i / \partial x_j(x_1, \dots, x_n)$.

- Außerdem sei die Borel-Menge $C \in \mathcal{B}(\mathbb{R}^n)$ so gewählt, dass

$$\{\mathbf{x} \in \mathbb{R}^n : f_{\mathbf{X}}(\mathbf{x}) \neq 0\} \subset C$$

und

$$\det\left(\frac{\partial \varphi_i}{\partial x_j}(x_1, \dots, x_n)\right) \neq 0, \quad \forall \mathbf{x} = (x_1, \dots, x_n) \in C,$$

so daß die Einschränkung $\varphi : C \rightarrow D$ von φ auf die Menge C eine eindeutige Abbildung ist, wobei $D = \{\varphi(\mathbf{x}) : \mathbf{x} \in C\}$.

- Sei $\varphi^{-1} = (\varphi_1^{-1}, \dots, \varphi_n^{-1}) : D \rightarrow C$ die Umkehrung der Abbildung $\varphi : C \rightarrow D$.
- Dann ist auch der Zufallsvektor $\mathbf{Y} = \varphi(\mathbf{X})$ absolutstetig, und für die Dichte $f_{\mathbf{Y}}(\mathbf{y})$ von $\mathbf{Y}$ gilt

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} f_{\mathbf{X}}(\varphi_1^{-1}(\mathbf{y}), \dots, \varphi_n^{-1}(\mathbf{y})) \left| \det\left(\frac{\partial \varphi_i^{-1}}{\partial y_j}(y_1, \dots, y_n)\right) \right|, & \text{falls } \mathbf{y} = (y_1, \dots, y_n) \in D, \\ 0, & \text{falls } \mathbf{y} \notin D. \end{cases} \quad (25)$$

bzw.

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} f_{\mathbf{X}}(\varphi_1^{-1}(\mathbf{y}), \dots, \varphi_n^{-1}(\mathbf{y})) \left| \det\left(\frac{\partial \varphi_i}{\partial x_j}(\varphi^{-1}(y_1, \dots, y_n))\right) \right|^{-1}, & \text{falls } \mathbf{y} = (y_1, \dots, y_n) \in D, \\ 0, & \text{falls } \mathbf{y} \notin D. \end{cases} \quad (26)$$

Aus Theorem 3.8 ergibt sich nun das folgende Resultat über die Darstellung von Zufallsvariablen mit absolutstetigen Verteilungen als Quotienten von gleichverteilten Zufallsvariablen.

Theorem 3.9

- Sei $f' : \mathbb{R} \rightarrow [0, \infty)$ eine (Borel-messbare) beschränkte Funktion mit

$$0 < \int_{\mathbb{R}} f'(x) dx < \infty \quad \text{und} \quad \sup_{x \in \mathbb{R}} |x| \sqrt{f'(x)} < \infty. \quad (27)$$

- Der Zufallsvektor (V_1, V_2) sei gleichverteilt in der (beschränkten) Borel-Menge

$$B = \{(x_1, x_2) \in \mathbb{R}^2 : 0 < x_1 < \sqrt{f'(x_2/x_1)}\}. \quad (28)$$

- Dann ist der Quotient V_2/V_1 eine absolutstetige Zufallsvariable mit der Dichte $f : \mathbb{R} \rightarrow [0, \infty)$, wobei

$$f(x) = \frac{f'(x)}{\int_{\mathbb{R}} f'(y) dy}, \quad \forall x \in \mathbb{R}.$$

Beweis

- Aus (27) folgt, dass die in (28) definierte Borel-Menge B beschränkt ist mit $0 < |B| < \infty$.
- Denn
 - für $x_2 > 0$ ist die Ungleichung $x_1 < \sqrt{f'(x_2/x_1)}$ äquivalent mit $x_2 < x_2/x_1 \sqrt{f'(x_2/x_1)}$,
 - und für $x_2 < 0$ mit $x_2 > x_2/x_1 \sqrt{f'(x_2/x_1)}$.

- Somit gilt

$$B \subset [0, \sup_{x \in \mathbb{R}} \sqrt{f'(x)}] \times [\inf_{x < 0} x \sqrt{f'(x)}, \sup_{x > 0} x \sqrt{f'(x)}] \quad (29)$$

bzw.

$$B \subset [0, \sup_{x \in \mathbb{R}} \sqrt{f'(x)}] \times [-\sup_{x \in \mathbb{R}} |x| \sqrt{f'(x)}, \sup_{x \in \mathbb{R}} |x| \sqrt{f'(x)}].$$

- Die Dichte $f_{(V_1, V_2)}(v_1, v_2)$ des Zufallsvektors (V_1, V_2) ist somit gegeben durch

$$f_{(V_1, V_2)}(v_1, v_2) = |B|^{-1} \mathbf{1}(0 < v_1 < \sqrt{f'(v_2/v_1)}).$$

- Die Funktion $\varphi : C \rightarrow C$ mit $C = (0, \infty) \times \mathbb{R}$ und $\varphi(x_1, x_2) = (x_1, x_2/x_1)$
 - bildet die Menge C auf eindeutige Weise auf sich selbst ab,
 - und ihre Funktionaldeterminante ist gegeben durch

$$\det \left(\frac{\partial \varphi_i}{\partial x_j}(x_1, x_2) \right) = \det \begin{pmatrix} 1 & 0 \\ -\frac{x_2}{x_1^2} & \frac{1}{x_1} \end{pmatrix} = \frac{1}{x_1}, \quad \forall (x_1, x_2) \in C.$$

- Hieraus und aus Theorem 3.8 ergibt sich,
 - dass die Dichte $f_{(V_1, V_2/V_1)}(y_1, y_2)$ des Zufallsvektors $(V_1, V_2/V_1)$ gegeben ist durch

$$f_{(V_1, V_2/V_1)}(y_1, y_2) = |B|^{-1} y_1 \mathbf{1}(0 < y_1 < \sqrt{f'(y_2)})$$

- und dass somit die Randdichte $f_{V_2/V_1}(y_2)$ der zweiten Komponente V_2/V_1 von $(V_1, V_2/V_1)$ gegeben ist durch

$$f_{V_2/V_1}(y_2) = |B|^{-1} \int_0^{\sqrt{f'(y_2)}} y_1 dy_1 = \frac{f'(y_2)}{2|B|}. \quad \square$$

Beispiel (Normalverteilung)

- Aus Theorem 3.9 ergibt sich nun noch ein drittes Verfahren zur Erzeugung von $N(0, 1)$ -verteilten Pseudozufallszahlen (neben dem in Abschnitt 3.2.1 diskutierten Box-Muller-Algorithmus bzw. der in Abschnitt 3.2.3 erwähnten Polar-Methode).
- Dabei betrachten wir die Funktion $f' : \mathbb{R} \rightarrow [0, \infty)$ mit $f'(x) = \exp(-x^2/2)$ für jedes $x \in \mathbb{R}$, so dass für die Schranken in (29) gilt:

$$\sup_{x \in \mathbb{R}} \sqrt{f'(x)} = 1, \quad \inf_{x < 0} x \sqrt{f'(x)} = -\sqrt{2/e}, \quad \sup_{x > 0} x \sqrt{f'(x)} = \sqrt{2/e}.$$

- Gemäß Theorem 3.9 können nun $N(0, 1)$ -verteilte Pseudozufallszahlen $x_1, x_2, \dots$ wie folgt erzeugt werden.
 1. Generiere eine $(0, 1]$ -gleichverteilte Pseudozufallszahl u und eine $(-\sqrt{2/e}, \sqrt{2/e}]$ -gleichverteilte Pseudozufallszahl v .
 2. Falls $u \geq \sqrt{\exp(-v^2/(2u^2))}$ gilt, d.h., falls $\log u \geq -v^2/(4u^2)$ bzw. äquivalent hierzu $v^2 \geq -4u^2 \log u$, dann kehre zu Schritt 1 zurück.
 3. Setze $x = v/u$.

3.3 Simulationsmethoden mit Markov-Ketten

- Sei E eine beliebige endliche Menge, beispielsweise eine gewisse Menge von (potentiell möglichen) digitalisierten Binär- bzw. Graustufenbildern $\mathbf{x} = (x(v), v \in V)$,
 - wobei V eine endliche Menge von Pixeln ist,
 - jedem Pixel $v \in V$ aus dem „Beobachtungsfenster“ V ein Graustufenwert $x(v) \geq 0$ zugeordnet wird,
 - so dass eine „Matrix“ $(x(v), v \in V)$ mit bestimmten Eigenschaften entsteht.
- Sei $\pi : E \rightarrow (0, 1)$ eine beliebige (Wahrscheinlichkeits-) Funktion mit

$$\sum_{\mathbf{x} \in E} \pi_{\mathbf{x}} = 1 \quad \text{und} \quad \pi_{\mathbf{x}} > 0, \quad \forall \mathbf{x} \in E.$$

- Wenn die Anzahl $|E|$ der Elemente der Menge E groß ist,
 - dann führt die in Abschnitt 3.2 diskutierte Inversionsmethode bzw. die Akzeptanz- und Verwerfungsmethode typischerweise *nicht* zu effizienten Algorithmen,
 - um gemäß π verteilte Pseudozufallselemente $\mathbf{x}_1, \mathbf{x}_2, \dots$ aus E zu erzeugen.

Beachte

- Eine alternative Simulationsmethode beruht
 - auf der Konstruktion einer Markov-Kette $\mathbf{X}_0, \mathbf{X}_1, \dots$ mit dem Zustandsraum E
 - und mit einer (geeignet gewählten) irreduziblen und aperiodischen Übergangsmatrix $\mathbf{P}$,
 - so dass π die ergodische (Grenz-) Verteilung der Markov-Kette ist.
- Für hinreichend großes n
 - ist dann $\mathbf{X}_n$ näherungsweise π -verteilt und
 - kann somit in vielen Fällen zur effizienten Erzeugung von (näherungsweise) π -verteilten Pseudozufallselementen aus E dienen.
- Man spricht deshalb in diesem Zusammenhang von *Markov-Chain-Monte-Carlo-Simulation* (MCMC).

3.3.1 Beispiel: Hard-Core-Modell

(vgl. O. Häggström (2002) *Finite Markov Chains and Algorithmic Applications*. CU Press, Cambridge)

- Wir betrachten einen verbundenen Graph $G = (V, K)$
 - mit der Menge $V = \{v_1, \dots, v_{|V|}\}$ von (endlich vielen) Eckpunkten
 - und mit einer gewissen Menge $K \subset V^2$ von Kanten, die jeweils zwei Eckpunkte miteinander verbinden.
- Jedem Eckpunkt aus V wird entweder der Wert 0 oder 1 zugeordnet,
 - wobei die folgende Menge $E \subset \{0, 1\}^{|V|}$ von *zulässigen Konfigurationen* betrachtet wird,
 - für die keinem Paar von verbundenen Eckpunkten gleichzeitig der Wert 1 zugeordnet wird, vgl. die Abbildung 4.
- Um unter den zulässigen Konfigurationen $\mathbf{x} \in E$ „auf gut Glück“ auswählen zu können, betrachten wir die (diskrete) Gleichverteilung π auf der Menge E mit

$$\pi_{\mathbf{x}} = \frac{1}{\ell}, \quad \forall \mathbf{x} \in E, \quad (30)$$

wobei $\ell = |E|$ die Anzahl aller zulässigen Konfigurationen bezeichnet.

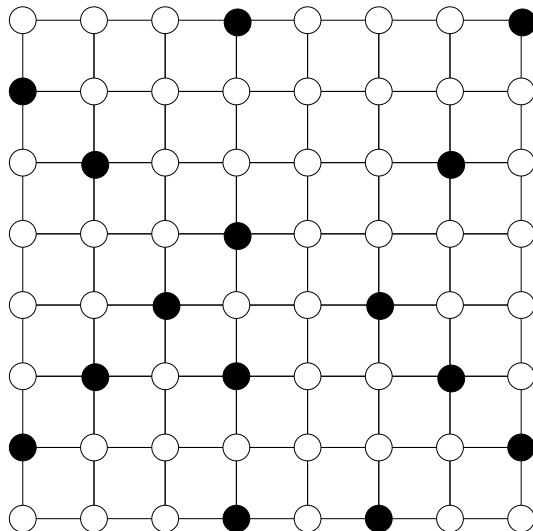


Abbildung 4: Quadratisches Gitter G der Größe 8×8 , wobei die schwarz gesetzten Pixel dem Wert 1 entsprechen

- Wenn die Anzahlen $|V|$ und $|K|$ der Eckpunkte und Kanten des verbundenen Graphen $G = (V, K)$ große Zahlen sind,
 - dann ist die explizite Beschreibung der Menge E der zulässigen Konfigurationen im allgemeinen schwierig,
 - die Anzahl ℓ aller zulässigen Konfigurationen ist deshalb typischerweise unbekannt,
 - und die Formel (30) kann somit nicht direkt zur Simulation von „zufällig ausgewählten“ zulässigen Konfigurationen genutzt werden.

MCMC-Simulationsalgorithmus

- Eine alternative Simulationsmethode besteht darin, eine Markov-Kette $\mathbf{X}_0, \mathbf{X}_1, \dots$ zu konstruieren,
 - mit dem Zustandsraum E und mit einer (geeignet gewählten) irreduziblen und aperiodischen Übergangsmatrix $\mathbf{P}$,
 - so dass die ergodische (Grenz-) Verteilung π der Markov-Kette $\mathbf{X}_0, \mathbf{X}_1, \dots$ durch (30) gegeben ist.
- Dabei erzeugen wir einen „Pfad“ $\mathbf{x}_0, \mathbf{x}_1, \dots$ der Markov-Kette, und zwar mit Hilfe der rekursiven Konstruktion von Markov-Ketten, die bereits in Abschnitt 2.1.3 diskutiert wurde:
 1. Wähle eine zulässige Anfangskonfiguration $\mathbf{x}_0 \in E$.
 2. Wähle „auf gut Glück“ einen beliebigen Eckpunkt $v \in V$ und wirf eine faire Münze.
 3. Falls das Ereignis „Wappen“ eintritt und falls $x_n(w) = 0$ für sämtliche Eckpunkte $w \in V$ gilt, die mit $v \in V$ verbunden sind, dann setze $x_{n+1}(v) = 1$; ansonsten setze $x_{n+1}(v) = 0$.
 4. Die Werte sämtlicher Eckpunkte $w \neq v$ bleiben unverändert, d.h., es gilt $x_{n+1}(w) = x_n(w)$ für jedes $w \neq v$.

Beachte

- Um die Schritte 2 – 4 dieses Algorithmus zu implementieren, muss die in (2.19) betrachtete „Update-Funktion“ $\varphi : E \times [0, 1] \rightarrow E$ entsprechend spezifiziert werden.

- Das Einheitsintervall $(0, 1]$ wird dabei in $2|V|$ gleichlange Teilintervalle (der Länge $1/2|V|$) zerlegt,
 - die den Ereignissen $(v_1, \text{Wappen}), (v_1, \text{Zahl}), \dots, (v_{|V|}, \text{Wappen}), (v_{|V|}, \text{Zahl})$ entsprechen,
 - und es gilt $\mathbf{x}' = \varphi(\mathbf{x}, z)$ mit

$$x'(v_i) = \begin{cases} 1, & \text{falls } z \in \left(\frac{2i-2}{2|V|}, \frac{2i-1}{2|V|}\right] \text{ und } x(w) = 0 \text{ für alle Eckpunkte } w \in V \text{ gilt,} \\ & \text{die mit } v \in V \text{ verbunden sind,} \\ 0, & \text{falls } z \in \left(\frac{2i-1}{2|V|}, \frac{2i}{2|V|}\right] \text{ oder falls } z \in \left(\frac{2i-2}{2|V|}, \frac{2i-1}{2|V|}\right] \text{ und } x(w) = 0 \\ & \text{nicht für alle Eckpunkte } w \in V \text{ gilt, die mit } v \in V \text{ verbunden sind,} \\ x(v_i), & \text{falls } z \notin \left(\frac{2i-2}{2|V|}, \frac{2i}{2|V|}\right]. \end{cases} \quad (31)$$

- Aus dem folgenden Theorem ergibt sich, dass für hinreichend großes n die n -te Rückgabe $\mathbf{x}_n = (x_n(v), v \in V)$ des Algorithmus näherungsweise als „zufällig ausgewählte“ (d.h. gemäß π -verteilte) zulässige Konfiguration angesehen werden kann.

Theorem 3.10

- Sei $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ die Übergangsmatrix des in (31) betrachteten MCMC-Algorithmus für das Hard-Core-Modell, und sei π die in (30) gegebene Wahrscheinlichkeitsfunktion.
- Dann ist $\mathbf{P}$ irreduzibel und aperiodisch, und das Paar $(\mathbf{P}, \pi)$ ist reversibel.

Beweis

- Um die Aperiodizität von $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ zu zeigen, genügt es zu beachten, dass sämtliche Diagonalelemente von $p_{\mathbf{x}\mathbf{x}}$ von $\mathbf{P}$ positiv sind.
- Die Irreduzibilität von $\mathbf{P}$ ergibt sich aus den folgenden Überlegungen.
 - Seien $\mathbf{x}, \mathbf{x}' \in E$ zwei zulässige Konfigurationen, und seien $m(\mathbf{x})$ bzw. $m(\mathbf{x}')$ die Anzahlen der auf 1 gesetzten Eckpunkte der Konfigurationen $\mathbf{x}$ bzw. $\mathbf{x}'$.
 - Dann ist zunächst der Übergang $\mathbf{x} \rightarrow \mathbf{x}_0$ in den „Nullzustand“ $\mathbf{x}_0 \in E$ in $m(\mathbf{x})$ Schritten mit positiver Wahrscheinlichkeit möglich, wobei $\mathbf{x}_0(v) = 0$ für jedes $v \in V$ gilt.
 - Dabei werden (jeweils mit positiver Wahrscheinlichkeit) die ursprünglich auf 1 gesetzten Eckpunkte von $\mathbf{x}$ der Reihe nach auf 0 gesetzt.
 - Anschließend kann man auf ähnliche Weise in $m(\mathbf{x}')$ Schritten mit jeweils positiver Wahrscheinlichkeit vom „Nullzustand“ $\mathbf{x}_0$ in den Zustand $\mathbf{x}'$ übergehen.
 - Hieraus ergibt sich insgesamt, dass der Übergang $\mathbf{x} \rightarrow \mathbf{x}'$ in endlich vielen Schritten mit positiver Wahrscheinlichkeit möglich ist.
- Es ist nun noch die Gültigkeit der Detailed-Balance-Gleichung (2.85) zu zeigen, d.h., dass

$$\pi_{\mathbf{x}} p_{\mathbf{x}\mathbf{x}'} = \pi_{\mathbf{x}'} p_{\mathbf{x}'\mathbf{x}}, \quad \forall \mathbf{x}, \mathbf{x}' \in E. \quad (32)$$

- Wenn die Konfigurationen $\mathbf{x}, \mathbf{x}' \in E$ übereinstimmen, dann ist die Gültigkeit von (32) offensichtlich.
- Wenn $x(v) \neq x'(v)$ für mehr als einen Eckpunkt $v \in V$, dann gilt $p_{\mathbf{x}\mathbf{x}'} = p_{\mathbf{x}'\mathbf{x}} = 0$, und (32) gilt somit auch in diesem Fall.
- Es gelte nun $x(v) \neq x'(v)$ für ein $v \in V$ und $x(w) = x'(w)$ für alle $w \neq v$.

- Dann gilt $x(w) = x'(w) = 0$ für sämtliche Eckpunkte $w \neq v$, die mit v verbunden sind, und folglich ist

$$\pi_{\mathbf{x}} p_{\mathbf{x}\mathbf{x}'} = \frac{1}{\ell} \frac{1}{2|V|} = \pi_{\mathbf{x}'} p_{\mathbf{x}'\mathbf{x}}. \quad \square$$

Beachte

- Für jedes $\mathbf{x} \in E$ sei $m(\mathbf{x})$ die Anzahl der auf 1 gesetzten Eckpunkte der zulässigen Konfiguration $\mathbf{x}$.
- Wenn die zulässige Konfiguration „auf gut Glück“ ausgewählt wird, dann gilt für den Erwartungswert $\mathbb{E}Y$ der zufälligen Anzahl Y der auf 1 gesetzten Eckpunkte, dass

$$\mathbb{E}Y = \frac{1}{\ell} \sum_{\mathbf{x} \in E} m(\mathbf{x}). \quad (33)$$

- Wenn ℓ eine große Zahl ist, dann ist jedoch die direkte Bestimmung des Erwartungswertes $\mathbb{E}Y$ mit Hilfe von Formel (33) im allgemeinen nicht möglich, weil die analytische Bestimmung der Anzahlen $m(\mathbf{x})$ schwierig ist.
- Eine alternative Methode zur (näherungsweisen) Bestimmung des Erwartungswertes $\mathbb{E}Y$ beruht auf der Erzeugung von k „zufällig ausgewählten“ zulässigen Konfigurationen $\mathbf{x}_n^{(1)}, \mathbf{x}_n^{(2)}, \dots, \mathbf{x}_n^{(k)} \in E$ mit k Runs des oben beschriebenen Algorithmus der MCMC-Simulation.
- Das arithmetische Mittel $(m(\mathbf{x}_n^{(1)}) + m(\mathbf{x}_n^{(2)}) + \dots + m(\mathbf{x}_n^{(k)}))/k$ liegt wegen des starken Gesetzes der großen Zahlen (mit hoher Wahrscheinlichkeit) nahe bei $\mathbb{E}Y$, falls die Runlänge n und der Stichprobenumfang k jeweils hinreichend groß sind.

3.3.2 Gibbs-Sampler

Der in Abschnitt 3.3.1 betrachtete MCMC-Algorithmus, um „zufällig ausgewählte“ zulässige Konfigurationen des Hard-Core-Modells zu erzeugen, ist ein Spezialfall des sogenannten *Gibbs-Samplers* zur MCMC-Simulation von diskreten (hochdimensionalen) Zufallsvektoren.

- Sei V eine endliche (nichtleere) Index-Menge, und sei $\mathbf{X} = (X(v), v \in V)$ ein diskreter Zufallsvektor,
 - der mit Wahrscheinlichkeit 1 seine Werte in der endlichen (Zustands-) Menge $E \subset \mathbb{R}^{|V|}$ annimmt, wobei wir voraussetzen,
 - dass es für jedes Paar $\mathbf{x}, \mathbf{x}' \in E$ eine (endliche) Folge von Zuständen $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_n \in E$ gibt, so dass

$$\mathbf{y}_0 = \mathbf{x}, \quad \mathbf{y}_n = \mathbf{x}' \quad \text{und} \quad \#\{v \in V : y_i(v) \neq y_{i+1}(v)\} = 1, \quad \forall i = 0, \dots, n-1. \quad (34)$$

- Sei $\pi = (\pi_{\mathbf{x}}, \mathbf{x} \in E)$ die Wahrscheinlichkeitsfunktion des Zufallsvektors $\mathbf{X}$ mit $\pi_{\mathbf{x}} > 0$ für jedes $\mathbf{x} \in E$, und für jedes $v \in V$ sei

$$\pi_{x(v)|\mathbf{x}(-v)} = P(X(v) = x(v) \mid \mathbf{X}(-v) = \mathbf{x}(-v)) \quad (35)$$

- die bedingte Wahrscheinlichkeit, dass die Komponente $X(v)$ von $\mathbf{X}$ den Wert $x(v)$ annimmt,
- unter der Bedingung, dass der Vektor $\mathbf{X}(-v) = (X(w), w \in V \setminus \{v\})$ der übrigen Komponenten gleich $\mathbf{x}(-v)$ ist, wobei vorausgesetzt wird, dass $(x(v), \mathbf{x}(-v)) \in E$.

MCMC-Simulationsalgorithmus

- Ähnlich wie in Abschnitt 3.3.1 konstruieren wir eine Markov-Kette $\mathbf{X}_0, \mathbf{X}_1, \dots$
 - mit dem Zustandsraum E und mit einer (geeignet gewählten) irreduziblen und aperiodischen Übergangsmatrix $\mathbf{P}$,
 - so dass π die ergodische (Grenz-) Verteilung der zu konstruierenden Markov-Kette $\mathbf{X}_0, \mathbf{X}_1, \dots$ ist.

- Dabei erzeugen wir erneut mit Hilfe der rekursiven Konstruktion, die bereits in Abschnitt 2.1.3 diskutiert wurde, einen „Pfad“ $\mathbf{x}_0, \mathbf{x}_1, \dots$ der Markov-Kette:

1. Wähle einen Anfangszustand $\mathbf{x}_0 \in E$.
2. Wähle eine Komponente $v \in V$ gemäß einer (vorgegebenen) Wahrscheinlichkeitsfunktion $\mathbf{q} = (q_v, v \in V)$, so dass $q_v > 0$ für jedes $v \in V$.
3. Generiere den (aktualisierten) Wert $x_{n+1}(v)$ der v -ten Komponente gemäß der (bedingten) Wahrscheinlichkeitsfunktion

$$\pi_{\cdot | \mathbf{x}_n(-v)} = (\pi_{x(v) | \mathbf{x}_n(-v)}, \forall x(v) \text{ mit } (x(v), \mathbf{x}(-v)) \in E).$$

4. Die Werte sämtlicher Komponenten $w \neq v$ bleiben unverändert, d.h., es gilt $x_{n+1}(w) = x_n(w)$ für jedes $w \neq v$.

Theorem 3.11 Die Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ sei gegeben durch

$$p_{\mathbf{x}\mathbf{x}'} = \sum_{v \in V} q_v \pi_{x'(v) | \mathbf{x}(-v)} \mathbb{I}(\mathbf{x}(-v) = \mathbf{x}'(-v)), \quad \forall \mathbf{x}, \mathbf{x}' \in E, \quad (36)$$

wobei die bedingten Wahrscheinlichkeiten $\pi_{x'(v) | \mathbf{x}(-v)}$ in (35) gegeben sind. Dann ist $\mathbf{P}$ irreduzibel und aperiodisch, und das Paar $(\mathbf{P}, \pi)$ ist reversibel.

Beweis Die Behauptung ergibt sich auf ähnliche Weise wie im Beweis von Theorem 3.10.

- Um die Aperiodizität von $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ zu zeigen, genügt es zu beachten,
 - dass für jedes $\mathbf{x} \in E$

$$p_{\mathbf{x}, \mathbf{x}} = \sum_{v \in V} q_v \pi_{x(v) | \mathbf{x}(-v)} = \sum_{v \in V} q_v \frac{\pi_{\mathbf{x}}}{\sum_{\mathbf{z} \in E: \mathbf{z}(-v) = \mathbf{x}(-v)} \pi_{\mathbf{z}}} > 0$$

- und dass somit sämtliche Diagonalelemente $p_{\mathbf{x}, \mathbf{x}}$ von $\mathbf{P}$ positiv sind.
- Die Irreduzibilität von $\mathbf{P}$ ergibt sich aus den folgenden Überlegungen.
 - Für beliebige, jedoch fest vorgegebene $\mathbf{x}, \mathbf{x}' \in E$ sei $k \leq |V|$ die Anzahl derjenigen Komponenten $v \in V$ mit $x(v) \neq x'(v)$.
 - Für $k = 0$, d.h. $\mathbf{x} = \mathbf{x}'$, hatten wir bereits beim Nachweis der Aperiodizität gezeigt, dass $p_{\mathbf{x}\mathbf{x}} > 0$.
 - Sei nun $k > 0$, wobei wir o.B.d.A. annehmen, dass die Komponenten linear durchnummeriert sind und dass die ersten k Komponenten von $\mathbf{x}$ und $\mathbf{x}'$ voneinander verschieden sind.
 - Die Voraussetzung (34) über die Struktur des Zustandsraumes E impliziert, dass es eine Folge $\mathbf{y}_0, \dots, \mathbf{y}_k \in E$ gibt, so dass $\mathbf{y}_0 = \mathbf{x}$ und

$$\mathbf{y}_1 = (x'(v_1), x(v_2), \dots, x(v_{|V|})) , \dots , \mathbf{y}_k = (x'(v_1), \dots, x'(v_k), x(v_{k+1}), \dots, x(v_{|V|})) = \mathbf{x}'.$$

- Für jedes $i = 0, \dots, k-1$ gilt dann

$$p_{\mathbf{y}_i \mathbf{y}_{i+1}} = q_{v_i} \pi_{y_{i+1}(v_i) | \mathbf{y}_i(-v_i)} = q_{v_i} \frac{\pi_{\mathbf{y}_{i+1}}}{\sum_{\mathbf{z} \in E: \mathbf{z}(-v_i) = \mathbf{y}_i(-v_i)} \pi_{\mathbf{z}}} > 0 \quad (37)$$

und somit auch $p_{\mathbf{x}\mathbf{x}'}^{(k)} \geq \prod_{i=0}^{k-1} p_{\mathbf{y}_i \mathbf{y}_{i+1}} > 0$.

- Es ist nun noch die Gültigkeit der Detailed-Balance-Gleichung (2.85) zu zeigen, d.h., dass

$$\pi_{\mathbf{x}} p_{\mathbf{x}\mathbf{x}'} = \pi_{\mathbf{x}'} p_{\mathbf{x}'\mathbf{x}}, \quad \forall \mathbf{x}, \mathbf{x}' \in E. \quad (38)$$

- Wenn $\mathbf{x} = \mathbf{x}'$ gilt, dann ist die Gültigkeit von (38) offensichtlich.

- Wenn $x(v) \neq x'(v)$ für mehr als eine Komponente $v \in V$, dann gilt $p_{\mathbf{x}\mathbf{x}'} = p_{\mathbf{x}'\mathbf{x}} = 0$, und (38) gilt somit auch in diesem Fall.
- Es gelte nun $x(v) \neq x'(v)$ für ein $v \in V$ und $x(w) = x'(w)$ für alle $w \neq v$. Dann ergibt sich aus (37), dass

$$\pi_{\mathbf{x}} p_{\mathbf{x}\mathbf{x}'} = \pi_{\mathbf{x}} \frac{q_v \pi_{\mathbf{x}'}}{\sum_{\mathbf{z} \in E: \mathbf{z}(-v)=\mathbf{x}(-v)} \pi_{\mathbf{z}}} = \pi_{\mathbf{x}'} \frac{q_v \pi_{\mathbf{x}}}{\sum_{\mathbf{z} \in E: \mathbf{z}(-v)=\mathbf{x}'(-v)} \pi_{\mathbf{z}}} = \pi_{\mathbf{x}'} p_{\mathbf{x}'\mathbf{x}}. \quad \square$$

Sei $\mathbf{X}_0, \mathbf{X}_1, \dots$ eine Markov-Kette mit dem Zustandsraum E und mit der in (36) gegebenen Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$. Aus Theorem 3.11 ergibt sich dann, dass

$$\lim_{n \rightarrow \infty} d_{\text{TV}}(\boldsymbol{\alpha}_n, \boldsymbol{\pi}) = 0 \quad (39)$$

für jede Anfangsverteilung $\boldsymbol{\alpha}_0$, wobei $\boldsymbol{\alpha}_n$ die Verteilung von $\mathbf{X}_n$ bezeichnet. Außerdem gilt für den Gibbs-Sampler die folgende *Monotonie-Eigenschaft*.

Theorem 3.12 Für jedes $n = 0, 1, \dots$ gilt

$$d_{\text{TV}}(\boldsymbol{\alpha}_n, \boldsymbol{\pi}) \geq d_{\text{TV}}(\boldsymbol{\alpha}_{n+1}, \boldsymbol{\pi}). \quad (40)$$

Beweis

- Für beliebige $v \in V$ und $\mathbf{x}' \in E$ ergibt sich aus (35), dass

$$\begin{aligned} \sum_{\mathbf{x} \in E: \mathbf{x}(-v)=\mathbf{x}'(-v)} \pi_{\mathbf{x}'(v)|\mathbf{x}(-v)} \pi_{\mathbf{x}} &\stackrel{(35)}{=} \sum_{\mathbf{x} \in E: \mathbf{x}(-v)=\mathbf{x}'(-v)} \pi_{\mathbf{x}'} \pi_{\mathbf{x}(v)|\mathbf{x}'(-v)} \\ &= \pi_{\mathbf{x}'} \underbrace{\sum_{\mathbf{x} \in E: \mathbf{x}(-v)=\mathbf{x}'(-v)} \pi_{\mathbf{x}(v)|\mathbf{x}'(-v)}}_{=1} = \pi_{\mathbf{x}'} . \end{aligned} \quad (41)$$

- Hieraus und aus der Definitionsgleichung (36) der Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ ergibt sich, dass

$$\begin{aligned} 2 d_{\text{TV}}(\boldsymbol{\alpha}_{n+1}, \boldsymbol{\pi}) &= \sum_{\mathbf{x}' \in E} |\alpha_{n+1, \mathbf{x}'} - \pi_{\mathbf{x}'}| \\ &= \sum_{\mathbf{x}' \in E} \left| \sum_{\mathbf{x} \in E} \alpha_{n, \mathbf{x}} p_{\mathbf{x}\mathbf{x}'} - \pi_{\mathbf{x}'} \right| \\ &\stackrel{(36)}{=} \sum_{\mathbf{x}' \in E} \left| \sum_{\mathbf{x} \in E} \alpha_{n, \mathbf{x}} \sum_{v \in V} q_v \pi_{\mathbf{x}'(v)|\mathbf{x}(-v)} \mathbb{I}(\mathbf{x}(-v) = \mathbf{x}'(-v)) - \pi_{\mathbf{x}'} \right| \\ &= \sum_{\mathbf{x}' \in E} \left| \sum_{v \in V} q_v \sum_{\mathbf{x} \in E: \mathbf{x}(-v)=\mathbf{x}'(-v)} \pi_{\mathbf{x}'(v)|\mathbf{x}(-v)} \alpha_{n, \mathbf{x}} - \pi_{\mathbf{x}'} \right| \\ &\stackrel{(41)}{=} \sum_{\mathbf{x}' \in E} \left| \sum_{v \in V} q_v \sum_{\mathbf{x} \in E: \mathbf{x}(-v)=\mathbf{x}'(-v)} \pi_{\mathbf{x}'(v)|\mathbf{x}(-v)} (\alpha_{n, \mathbf{x}} - \pi_{\mathbf{x}}) \right| \\ &\leq \sum_{\mathbf{x}' \in E} \sum_{v \in V} q_v \sum_{\mathbf{x} \in E: \mathbf{x}(-v)=\mathbf{x}'(-v)} \pi_{\mathbf{x}'(v)|\mathbf{x}(-v)} |\alpha_{n, \mathbf{x}} - \pi_{\mathbf{x}}| \\ &= \underbrace{\sum_{\mathbf{x} \in E} \sum_{v \in V} q_v \sum_{\mathbf{x}' \in E: \mathbf{x}'(-v)=\mathbf{x}(-v)} \pi_{\mathbf{x}'(v)|\mathbf{x}(-v)} |\alpha_{n, \mathbf{x}} - \pi_{\mathbf{x}}|}_{\stackrel{(36)}{=} \sum_{\mathbf{x}' \in E} p_{\mathbf{x}\mathbf{x}'} = 1} \\ &= \sum_{\mathbf{x} \in E} |\alpha_{n, \mathbf{x}} - \pi_{\mathbf{x}}| = 2 d_{\text{TV}}(\boldsymbol{\alpha}_n, \boldsymbol{\pi}). \quad \square \end{aligned}$$

Beachte

- Eine modifizierte Version des bisher in diesem Abschnitt betrachteten Gibbs-Samplers ist der *zyklische Gibbs-Sampler*, bei dem die jeweils zu aktualisierende Komponente $v \in V$
 - *nicht* gemäß einer (vorgegebenen) Wahrscheinlichkeitsfunktion $\mathbf{q} = (q_v, v \in V)$, so dass $q_v > 0$ für jedes $v \in V$, ausgewählt wird,
 - sondern die Komponenten $v \in V$ werden durchnummeriert und der Reihe nach (auf deterministische Weise) ausgewählt.
- Falls $k = n|V| + i$ für gewisse Zahlen $n = 0, 1, \dots$ und $i = 1, \dots, |V|$, dann ist die Matrix $\mathbf{P}(k) = (p_{\mathbf{x}\mathbf{x}'}(k))$ der Übergangswahrscheinlichkeiten $p_{\mathbf{x}\mathbf{x}'}(k)$ im k -ten Schritt gegeben durch

$$p_{\mathbf{x}\mathbf{x}'}(k) = \pi_{x'(v_i)|\mathbf{x}(-v_i)} \mathbb{I}(\mathbf{x}(-v_i) = \mathbf{x}'(-v_i)), \quad \forall \mathbf{x}, \mathbf{x}' \in E. \quad (42)$$

- Für einen gesamten (Scan-) Zyklus, bei dem jede Komponente genau einmal aktualisiert wird, ergibt sich die Übergangsmatrix

$$\mathbf{P} = \mathbf{P}(1) \cdot \dots \cdot \mathbf{P}(|V|). \quad (43)$$

- Man kann sich leicht überlegen, dass die in (42) und (43) gegebene Matrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$
 - irreduzibel und aperiodisch ist
 - und dass π die stationäre (Grenz-) Verteilung von $\mathbf{P}$ ist,
 - denn für jedes $i = 1, \dots, |V|$ und für jedes $\mathbf{x}' \in E$ ergibt sich aus (41) und (42), dass

$$\sum_{\mathbf{x} \in E} \pi_{\mathbf{x}} p_{\mathbf{x}\mathbf{x}'}(i) \stackrel{(42)}{=} \sum_{\mathbf{x} \in E} \pi_{\mathbf{x}} \pi_{x'(v_i)|\mathbf{x}(-v_i)} \mathbb{I}(\mathbf{x}(-v_i) = \mathbf{x}'(-v_i)) \stackrel{(41)}{=} \pi_{\mathbf{x}'}$$

- und dass somit auch

$$\sum_{\mathbf{x} \in E} \pi_{\mathbf{x}} p_{\mathbf{x}\mathbf{x}'} = \pi_{\mathbf{x}'}.$$

- Das Paar $(\mathbf{P}, \pi)$ ist im allgemeinen nicht reversibel. In Abschnitt 2.3.4 hatten wir jedoch gezeigt, dass das Paar $(\mathbf{M}, \pi)$ reversibel ist, wobei

$$\mathbf{M} = \mathbf{P}\tilde{\mathbf{P}} \quad \text{mit} \quad \tilde{\mathbf{P}} = \text{diag}(\pi_{\mathbf{x}}^{-1})\mathbf{P}^{\top} \text{diag}(\pi_{\mathbf{x}}) \quad (44)$$

die multiplikativ reversible Version von $\mathbf{P}$ bezeichnet.

Theorem 3.13 *Es gilt*

$$\mathbf{M} = \mathbf{P}(1) \cdot \dots \cdot \mathbf{P}(|V|) \cdot \mathbf{P}(|V|) \cdot \dots \cdot \mathbf{P}(1), \quad (45)$$

d.h., die multiplikativ reversible Version $\mathbf{M}$ der „Vorwärts-Scan-Matrix“ $\mathbf{P}$ stimmt mit der „Vorwärts-Rückwärts-Scan-Matrix“ überein.

Beweis

- Es genügt zu zeigen, dass $\tilde{\mathbf{P}} = \mathbf{P}(|V|) \cdot \dots \cdot \mathbf{P}(1)$ für die in (44) definierte Matrix $\tilde{\mathbf{P}} = (\tilde{p}_{\mathbf{x}\mathbf{x}'})$ gilt.
- Aus (42)–(44) ergibt sich, dass für beliebige $\mathbf{x}, \mathbf{x}' \in E$

$$\begin{aligned} \tilde{p}_{\mathbf{x}\mathbf{x}'} &= \left(\text{diag}(\pi_{\mathbf{x}}^{-1})\mathbf{P}^{\top} \text{diag}(\pi_{\mathbf{x}}) \right)_{\mathbf{x}\mathbf{x}'} \\ &= \left(\text{diag}(\pi_{\mathbf{x}}^{-1})\mathbf{P}^{\top}(|V|) \cdot \dots \cdot \mathbf{P}^{\top}(1) \text{diag}(\pi_{\mathbf{x}}) \right)_{\mathbf{x}\mathbf{x}'} \\ &= \sum_{\mathbf{y}_1, \dots, \mathbf{y}_{|V|-1} \in E} \frac{1}{\pi_{\mathbf{x}}} \pi_{x(v_{|V|})|\mathbf{y}_1(-v_{|V|})} \mathbb{I}(\mathbf{x}(-v_{|V|}) = \mathbf{y}_1(-v_{|V|})) \pi_{y_1(v_{|V|-1})|\mathbf{y}_2(-v_{|V|-1})} \\ &\quad \times \mathbb{I}(\mathbf{y}_1(-v_{|V|-1}) = \mathbf{y}_2(-v_{|V|-1})) \dots \pi_{y_{|V|-1}(v_1)|\mathbf{x}'(-v_1)} \mathbb{I}(\mathbf{y}_{|V|-1}(-v_1) = \mathbf{x}'(-v_1)) \pi_{\mathbf{x}'} . \end{aligned}$$

- Hieraus und aus (35) ergibt sich nun (ähnlich wie beim Beweis von (38)), dass

$$\begin{aligned}\tilde{\mathbf{p}}_{\mathbf{x}\mathbf{x}'} &= \sum_{\mathbf{y}_1, \dots, \mathbf{y}_{|V|-1} \in E} \pi_{y_1(v_{|V|})|\mathbf{x}(-v_{|V|})} \mathbb{I}(\mathbf{x}(-v_{|V|}) = \mathbf{y}_1(-v_{|V|})) \pi_{y_2(v_{|V|-1})|\mathbf{y}_1(-v_{|V|-1})} \\ &\quad \times \mathbb{I}(\mathbf{y}_1(-v_{|V|-1}) = \mathbf{y}_2(-v_{|V|-1})) \cdots \pi_{x'(v_1)|\mathbf{y}_{|V|-1}(-v_1)} \mathbb{I}(\mathbf{y}_{|V|-1}(-v_1) = \mathbf{x}'(-v_1)) \\ &= \left(\mathbf{P}(|V|) \cdots \mathbf{P}(1) \right)_{\mathbf{x}\mathbf{x}'} .\end{aligned}$$

□

Beachte

- Bei der praktischen Anwendung von Gibbs-Samplern wird stets vorausgesetzt,
 - dass die in (36) bzw. (42) betrachteten bedingten Wahrscheinlichkeiten

$$\pi_{x(v)|\mathbf{x}(-v)} = P(X(v) = x(v) \mid \mathbf{X}(-v) = \mathbf{x}(-v))$$

für jedes $v \in V$ lediglich von dem Vektor $(x(w), w \in \mathcal{N}(v))$ derjenigen Werte abhängen,

- die der Zufallsvektor $\mathbf{X} = (X(w), w \in V)$ in einer gewissen (möglichst kleinen) Nachbarschaft $\mathcal{N}(v) \subset V$ von $v \in V$ annimmt.
- Dabei wird eine Familie $\mathcal{N} = \{\mathcal{N}(v), v \in V\}$ von Teilmengen von V ein *System von Nachbarschaften* genannt, falls für beliebige $v, w \in V$
 - (a) $v \notin \mathcal{N}(v)$,
 - (b) aus $w \in \mathcal{N}(v)$ stets $v \in \mathcal{N}(w)$ folgt.
- Für das in Abschnitt 3.3.1 diskutierte Beispiel des Hard-Core-Modells ist $\mathcal{N}(v)$ für jeden Eckpunkt $v \in V$ die Menge derjenigen Eckpunkte $w \neq v$, die (direkt) mit v durch *eine* Kante verbunden sind.

3.3.3 Metropolis-Hastings-Algorithmus

- Wir zeigen nun, dass der in Abschnitt 3.3.2 betrachtete Gibbs-Sampler ein Spezialfall einer größeren Klasse von MCMC-Algorithmen ist, und zwar der sogenannten *Algorithmen vom Metropolis-Hastings-Typ*, wobei die Verallgemeinerung in zweierlei Hinsicht erfolgt:

1. Der in (36) betrachtete Ansatz

$$p_{\mathbf{x}\mathbf{x}'} = \sum_{v \in V} q_v \pi_{x'(v)|\mathbf{x}(-v)} \mathbb{I}(\mathbf{x}(-v) = \mathbf{x}'(-v)), \quad \forall \mathbf{x}, \mathbf{x}' \in E \quad (46)$$

für die Matrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ der Übergangswahrscheinlichkeiten $p_{\mathbf{x}\mathbf{x}'}$ wird verallgemeinert.

2. Zusätzlich wird ein *Verfahren zur Akzeptanz bzw. Verwerfung* der Updates $\mathbf{x} \rightarrow \mathbf{x}'$ in den Algorithmus integriert, dass auf einer ähnlichen Idee wie die in Abschnitt 3.2.3 diskutierte Akzeptanz- und Verwerfungsmethode beruht, vgl. insbesondere Theorem 3.5.

- Sei V eine endliche (nichtleere) Index-Menge, und sei $\mathbf{X} = (X(v), v \in V)$ ein diskreter Zufallsvektor,
 - der mit Wahrscheinlichkeit 1 seine Werte in der endlichen (Zustands-) Menge $E \subset \mathbb{R}^{|V|}$ annimmt,
 - wobei wir sowie bisher für die Wahrscheinlichkeitsfunktion $\pi = (\pi_{\mathbf{x}}, \mathbf{x} \in E)$ des Zufallsvektors $\mathbf{X}$ voraussetzen, dass $\pi_{\mathbf{x}} > 0$ für jedes $\mathbf{x} \in E$.
- Die Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ der zu konstruierenden Markov-Kette $\mathbf{X}_0, \mathbf{X}_1, \dots$ mit der ergodischen (Grenz-) Verteilung π sei gegeben durch den Ansatz

$$p_{\mathbf{x}\mathbf{x}'} = q_{\mathbf{x}\mathbf{x}'} a_{\mathbf{x}\mathbf{x}'}, \quad \forall \mathbf{x}, \mathbf{x}' \in E \text{ mit } \mathbf{x} \neq \mathbf{x}', \quad (47)$$

- wobei $\mathbf{Q} = (q_{\mathbf{x}\mathbf{x}'})$ eine beliebige stochastische Matrix ist, die irreduzibel und aperiodisch sei, so dass $q_{\mathbf{x}\mathbf{x}'} = 0$ genau dann, wenn $q_{\mathbf{x}'\mathbf{x}} = 0$.
- die Matrix $\mathbf{A} = (a_{\mathbf{x}\mathbf{x}'})$ gegeben ist durch

$$a_{\mathbf{x}\mathbf{x}'} = \frac{s_{\mathbf{x}\mathbf{x}'}}{1 + t_{\mathbf{x}\mathbf{x}'}} , \quad (48)$$

so dass

$$t_{\mathbf{x}\mathbf{x}'} = \begin{cases} \frac{\pi_{\mathbf{x}} q_{\mathbf{x}\mathbf{x}'}}{\pi_{\mathbf{x}'} q_{\mathbf{x}'\mathbf{x}}} , & \text{falls } q_{\mathbf{x}\mathbf{x}'} > 0, \\ 0 , & \text{falls } q_{\mathbf{x}\mathbf{x}'} = 0, \end{cases} \quad (49)$$

- und $\mathbf{S} = (s_{\mathbf{x}\mathbf{x}'})$ eine beliebige symmetrische Matrix ist mit

$$0 < s_{\mathbf{x}\mathbf{x}'} \leq 1 + \min \{t_{\mathbf{x}\mathbf{x}'}, t_{\mathbf{x}'\mathbf{x}}\} . \quad (50)$$

Beachte

- Der in (47) betrachtete Strukturansatz für die Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ lässt sich wie folgt interpretieren.
 - Zunächst wird gemäß $\mathbf{Q} = (q_{\mathbf{x}\mathbf{x}'})$ ein Kandidat $\mathbf{x}' \in E$ für den Update $\mathbf{x} \rightarrow \mathbf{x}'$ erzeugt.
 - Falls $\mathbf{x}' \neq \mathbf{x}$, dann wird $\mathbf{x}'$ mit der Wahrscheinlichkeit $a_{\mathbf{x}\mathbf{x}'}$ akzeptiert,
 - d.h., mit der Wahrscheinlichkeit $1 - a_{\mathbf{x}\mathbf{x}'}$ wird der Update $\mathbf{x} \rightarrow \mathbf{x}'$ verworfen (und der jeweils aktuelle Zustand $\mathbf{x}$ wird somit nicht verändert).
- Um den in (47)–(50) definierten Metropolis-Hastings-Algorithmus anwenden zu können, müssen für eine gegebene „potentielle“ Übergangsmatrix $\mathbf{Q} = (q_{\mathbf{x}\mathbf{x}'})$ lediglich die Quotienten $\pi_{\mathbf{x}}/\pi_{\mathbf{x}'}$ für diejenigen Paare $\mathbf{x}, \mathbf{x}' \in E$ von Zuständen bekannt sein, für die $q_{\mathbf{x}\mathbf{x}'} > 0$ gilt.
- Der in Abschnitt 3.3.2 betrachtete Spezialfall des Gibbs-Samplers ergibt sich,
 - wenn die „potentiellen“ Übergangswahrscheinlichkeiten $q_{\mathbf{x}\mathbf{x}'}$ so wie in (46) gewählt werden.
 - Für beliebige $\mathbf{x}, \mathbf{x}' \in E$ mit $\#\{v \in V : x(v) \neq x'(v)\} \leq 1$ gilt dann

$$\pi_{\mathbf{x}} q_{\mathbf{x}\mathbf{x}'} = \pi_{\mathbf{x}'} q_{\mathbf{x}'\mathbf{x}} \quad \text{und somit} \quad t_{\mathbf{x}\mathbf{x}'} = 1 .$$

- Mit der Wahl $s_{\mathbf{x}\mathbf{x}'} = 1 + \min \{t_{\mathbf{x}\mathbf{x}'}, t_{\mathbf{x}'\mathbf{x}}\}$ ergibt sich nun, dass $a_{\mathbf{x}\mathbf{x}'} = 1$ für beliebige $\mathbf{x}, \mathbf{x}' \in E$ mit $\#\{v \in V : x(v) \neq x'(v)\} \leq 1$.

Theorem 3.14 *Die in (47)–(50) gegebene Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ ist irreduzibel und aperiodisch, und das Paar $(\mathbf{P}, \pi)$ ist reversibel.*

Beweis

- Weil die in (48)–(50) gegebenen Akzeptanzwahrscheinlichkeiten $a_{\mathbf{x}\mathbf{x}'}$ für beliebige $\mathbf{x}, \mathbf{x}' \in E$ positiv sind, ergibt sich die Irreduzibilität und Aperiodizität $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ aus den entsprechenden Eigenschaften von $\mathbf{Q} = (q_{\mathbf{x}\mathbf{x}'})$.
- Um die Gültigkeit der Detailed-Balance-Gleichung (2.85) zu zeigen, d.h., dass

$$\pi_{\mathbf{x}} p_{\mathbf{x}\mathbf{x}'} = \pi_{\mathbf{x}'} p_{\mathbf{x}'\mathbf{x}} , \quad \forall \mathbf{x}, \mathbf{x}' \in E , \quad (51)$$

unterscheiden wir zwei Fälle.

- Wenn $q_{\mathbf{x}\mathbf{x}'} = q_{\mathbf{x}'\mathbf{x}} = 0$, dann gilt $p_{\mathbf{x}\mathbf{x}'} = p_{\mathbf{x}'\mathbf{x}} = 0$ und somit auch (51).
- Wenn $q_{\mathbf{x}\mathbf{x}'} > 0$, dann gilt auch $q_{\mathbf{x}'\mathbf{x}} > 0$, und aus (47)–(50) ergibt sich, dass

$$\begin{aligned}\pi_{\mathbf{x}} p_{\mathbf{x}\mathbf{x}'} &= \pi_{\mathbf{x}} q_{\mathbf{x}\mathbf{x}'} a_{\mathbf{x}\mathbf{x}'} \\ &= \pi_{\mathbf{x}} q_{\mathbf{x}\mathbf{x}'} \frac{s_{\mathbf{x}\mathbf{x}'} \pi_{\mathbf{x}'} q_{\mathbf{x}'\mathbf{x}}}{\pi_{\mathbf{x}'} q_{\mathbf{x}'\mathbf{x}} + \pi_{\mathbf{x}} q_{\mathbf{x}\mathbf{x}'}} \\ &= \pi_{\mathbf{x}'} p_{\mathbf{x}'\mathbf{x}},\end{aligned}$$

wobei sich die letzte Gleichheit aus der Symmetrie der Matrix $\mathbf{S} = (s_{\mathbf{x}\mathbf{x}'})$ ergibt. $\square$

Beispiele

1. Metropolis-Algorithmus

- Der klassische Metropolis-Algorithmus ergibt sich, wenn in (50) die Gleichheit betrachtet wird, d.h., wenn

$$s_{\mathbf{x}\mathbf{x}'} = 1 + \min \{t_{\mathbf{x}\mathbf{x}'}, t_{\mathbf{x}'\mathbf{x}}\}, \quad \forall \mathbf{x}, \mathbf{x}' \in E.$$

- Für die Akzeptanzwahrscheinlichkeiten $a_{\mathbf{x}\mathbf{x}'}$ gilt dann für beliebige $\mathbf{x}, \mathbf{x}' \in E$ mit $q_{\mathbf{x}\mathbf{x}'} > 0$, dass

$$\begin{aligned}a_{\mathbf{x}\mathbf{x}'} &= \frac{1 + \min \{t_{\mathbf{x}\mathbf{x}'}, t_{\mathbf{x}'\mathbf{x}}\}}{1 + t_{\mathbf{x}'\mathbf{x}}} \\ &= \frac{\min \{1 + t_{\mathbf{x}\mathbf{x}'}, 1 + t_{\mathbf{x}'\mathbf{x}}\}}{1 + t_{\mathbf{x}\mathbf{x}'}} = \min \left\{ 1, \frac{1 + t_{\mathbf{x}'\mathbf{x}}}{1 + t_{\mathbf{x}\mathbf{x}'}} \right\} \\ &= \min \left\{ 1, \frac{\pi_{\mathbf{x}'} q_{\mathbf{x}'\mathbf{x}}}{\pi_{\mathbf{x}} q_{\mathbf{x}\mathbf{x}'}} \right\},\end{aligned}$$

d.h.,

$$a_{\mathbf{x}\mathbf{x}'} = \min \left\{ 1, \frac{\pi_{\mathbf{x}'} q_{\mathbf{x}'\mathbf{x}}}{\pi_{\mathbf{x}} q_{\mathbf{x}\mathbf{x}'}} \right\}, \quad \forall \mathbf{x}, \mathbf{x}' \in E \text{ mit } q_{\mathbf{x}\mathbf{x}'} > 0. \quad (52)$$

- Wenn die Matrix $\mathbf{Q} = (q_{\mathbf{x}\mathbf{x}'})$ der „potentiellen“ Übergangswahrscheinlichkeiten symmetrisch ist, dann ergibt sich aus (52), dass

$$a_{\mathbf{x}\mathbf{x}'} = \min \left\{ 1, \frac{\pi_{\mathbf{x}'}}{\pi_{\mathbf{x}}} \right\}, \quad \forall \mathbf{x}, \mathbf{x}' \in E \text{ mit } q_{\mathbf{x}\mathbf{x}'} > 0. \quad (53)$$

- Die Akzeptanzwahrscheinlichkeiten $a_{\mathbf{x}\mathbf{x}'}$ sind also insbesondere dann durch (53) gegeben, wenn die „potentiellen“ Updates $\mathbf{x} \rightarrow \mathbf{x}'$ „auf gut Glück“ gewählt werden, d.h., wenn

$$q_{\mathbf{x}\mathbf{x}'} = \frac{1}{|E|}, \quad \forall \mathbf{x}, \mathbf{x}' \in E.$$

2. Barker-Algorithmus

- Der sogenannte Barker-Algorithmus ergibt sich, wenn die Matrix $\mathbf{S} = (s_{\mathbf{x}\mathbf{x}'})$ mit $s_{\mathbf{x}\mathbf{x}'} = 1$ für beliebige $\mathbf{x}, \mathbf{x}' \in E$ betrachtet wird.
- Für die Akzeptanzwahrscheinlichkeiten $a_{\mathbf{x}\mathbf{x}'}$ gilt dann

$$a_{\mathbf{x}\mathbf{x}'} = \frac{\pi_{\mathbf{x}'} q_{\mathbf{x}'\mathbf{x}}}{\pi_{\mathbf{x}'} q_{\mathbf{x}'\mathbf{x}} + \pi_{\mathbf{x}} q_{\mathbf{x}\mathbf{x}'}} , \quad \forall \mathbf{x}, \mathbf{x}' \in E \text{ mit } q_{\mathbf{x}\mathbf{x}'} > 0. \quad (54)$$

- Wenn die Matrix $\mathbf{Q} = (q_{\mathbf{x}\mathbf{x}'})$ der „potentiellen“ Übergangswahrscheinlichkeiten symmetrisch ist, dann gilt insbesondere

$$a_{\mathbf{x}\mathbf{x}'} = \frac{\pi_{\mathbf{x}'}}{\pi_{\mathbf{x}'} + \pi_{\mathbf{x}}}, \quad \forall \mathbf{x}, \mathbf{x}' \in E \text{ mit } q_{\mathbf{x}\mathbf{x}'} > 0. \quad (55)$$

MCMC-Simulationsalgorithmus

- Genauso wie bei dem in Abschnitt 3.3.2 diskutierten Gibbs-Sampler konstruieren wir eine Markov-Kette $\mathbf{X}_0, \mathbf{X}_1, \dots$
 - mit dem Zustandsraum E und mit der in (47)–(50) gegebenen (irreduziblen und aperiodischen) Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$,
 - so dass π die ergodische (Grenz-) Verteilung der Markov-Kette $\mathbf{X}_0, \mathbf{X}_1, \dots$ ist.
- Für hinreichend großes n stimmt dann die Verteilung α_n von $\mathbf{X}_n$ näherungsweise mit π überein.
- Um Fehlerbetrachtungen für MCMC-Simulationsalgorithmen durchführen zu können, ist es nützlich,
 - den *Variationsabstand* $d_{TV}(\alpha_n, \pi)$ zwischen den Verteilungen α_n und π
 - bzw. *Abschätzungen* hierfür zu kennen, vgl. Abschnitt 3.4.1.

3.4 Fehleranalyse bei MCMC-Simulation

3.4.1 Abschätzung der Konvergenzgeschwindigkeit

- Wir zeigen nun, wie die in Abschnitt 2.3 hergeleiteten allgemeinen Abschätzungen für den Variationsabstand $d_{TV}(\alpha_n, \pi)$ bzw. für den zweitgrößten Betrag $|\theta_2| = \max\{|\lambda_2|, |\lambda_\ell|\}$ der Eigenwerte $\lambda_1, \dots, \lambda_\ell$ der Übergangsmatrix $\mathbf{P}$ genutzt werden können,
 - um obere Schranken für die Größe von $d_{TV}(\alpha_n, \pi)$ zu bestimmen, die sich im n -ten Iterationsschritt bei der MCMC-Simulation mit dem Metropolis-Algorithmus ergibt,
 - wenn die zu simulierende Verteilung π den folgenden Bedingungen genügt.
- Und zwar nehmen wir an,
 - dass $\pi_{\mathbf{x}} \neq \pi_{\mathbf{x}'}$ für beliebige $\mathbf{x}, \mathbf{x}' \in E$ mit $\mathbf{x} \neq \mathbf{x}'$,
 - und dass die Zustände $\mathbf{x}_1, \dots, \mathbf{x}_\ell \in E$ so numeriert sind, dass $\pi_{\mathbf{x}_1} > \dots > \pi_{\mathbf{x}_\ell}$.
- Wir können deshalb (o.B.d.A.) zu der bereits in Abschnitt 2.3 verwendeten Schreibweise zurückkehren und identifizieren die Zustände $\mathbf{x}_1, \dots, \mathbf{x}_\ell \in E$ mit den ersten ℓ natürlichen Zahlen, d.h. $E = \{1, \dots, \ell\}$.
- Die Wahrscheinlichkeiten π_i ($= \pi_{\mathbf{x}_i}$) lassen sich dann in der folgenden Form schreiben:

$$\pi_i = \frac{b^{h(i)}}{z(b)}, \quad \forall i = 1, \dots, \ell, \quad (56)$$

- wobei $h : \{1, \dots, \ell\} \rightarrow (1, \infty)$ eine monoton wachsende Funktion ist,
- $b \in (0, 1)$ so gewählt wird, dass für eine gewisse Konstante $c \geq 1$

$$h(i+1) - h(i) \geq c, \quad \forall i = 1, \dots, \ell - 1 \quad (57)$$

- und $z(b) = \sum_{i=1}^{\ell} b^{h(i)}$ ein (im allgemeinen unbekannter) Normierungsfaktor ist.
- Um einen Metropolis-Algorithmus zur MCMC-Simulation von $\pi = (\pi_1, \dots, \pi_\ell)^\top$ definieren zu können, setzen wir außerdem voraus,
 - dass die Basis b und die Differenzen $h(i+1) - h(i)$ für jedes $i = 1, \dots, \ell - 1$ bekannt sind,
 - d.h. insbesondere, dass die Quotienten π_{i+1}/π_i für jedes $i = 1, \dots, \ell - 1$ bekannt sind.

- Für die Matrix $\mathbf{Q} = (q_{ij})$ der „potentiellen“ Übergänge $i \rightarrow j$ gelte

$$q_{ij} = \begin{cases} \frac{1}{2}, & \text{falls } i = 1, j = 1, 2 \text{ oder } i = \ell, j = \ell, \ell - 1, \\ \frac{1}{2}, & \text{falls } i = 2, \dots, \ell - 1 \text{ und } j = i - 1, i + 1, \\ 0, & \text{sonst.} \end{cases} \quad (58)$$

- Die Akzeptanzwahrscheinlichkeiten a_{ij} seien durch (53) gegeben, d.h.

$$a_{ij} = \min \left\{ 1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \right\} = \min \{ 1, b^{h(j)-h(i)} \}, \quad \forall i, j \in \{1, \dots, \ell\} \text{ mit } q_{ij} = q_{ji} > 0.$$

- Für die Eintragungen $p_{ij} = q_{ij} a_{ij}$ der Übergangsmatrix $\mathbf{P} = (p_{ij})$ der MCMC-Simulation ergibt sich dann aus (56) und (58), dass

$$p_{11} = 1 - \frac{b^{h(2)-h(1)}}{2}, \quad p_{12} = \frac{b^{h(2)-h(1)}}{2}, \quad p_{\ell, \ell-1} = p_{\ell\ell} = \frac{1}{2} \quad (59)$$

und für $i = 2, \dots, \ell - 1$

$$p_{i, i-1} = \frac{1}{2}, \quad p_{i, i+1} = \frac{b^{h(i+1)-h(i)}}{2}, \quad p_{ii} = 1 - p_{i, i-1} - p_{i, i+1}. \quad (60)$$

Theorem 3.15 Für den zweitgrößten Eigenwert λ_2 der in (59)–(60) definierten Übergangsmatrix $\mathbf{P} = (p_{ij})$ gilt

$$\lambda_2 \leq 1 - \frac{(1 - b^{c/2})^2}{2}. \quad (61)$$

Beweis

- In Theorem 3.14 hatten wir gezeigt, dass das Paar $(\mathbf{P}, \pi)$ reversibel ist.
- Aus dem *Theorem von Rayleigh* (vgl. Theorem 2.17) ergibt sich somit die Darstellungsformel

$$\lambda_2 = 1 - \inf_{\mathbf{x} \in \mathbb{R}_{\neq}^{\ell}} \frac{D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x})}{\text{Var}_{\pi}(\mathbf{x})}, \quad (62)$$

- wobei $\mathbb{R}_{\neq}^{\ell} = \{\mathbf{x} = (x_1, \dots, x_{\ell})^{\top} \in \mathbb{R}^{\ell} : x_i \neq x_j \text{ für ein Paar } i, j \in E\}$ die Menge derjenigen Vektoren aus $\mathbb{R}^{\ell}$ bezeichnet, für die nicht alle Komponenten gleich sind,
- $\text{Var}_{\pi}(\mathbf{x}) = \|\mathbf{x}\|_{\pi}^2 - (\mathbf{x})_{\pi}^2$ die Varianz der Komponenten von $\mathbf{x}$ bezüglich π
- und $D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x}) = ((\mathbf{I} - \mathbf{P})\mathbf{x}, \mathbf{x})_{\pi}$ die Dirichlet-Form des reversiblen Paares $(\mathbf{P}, \pi)$ ist.
- Wegen (62) genügt es zu zeigen, dass

$$\text{Var}_{\pi}(\mathbf{x}) \leq a D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^{\ell} \quad (63)$$

für eine Konstante a mit

$$0 < a \leq \frac{2}{(1 - b^{c/2})^2}. \quad (64)$$

- Ähnlich wie im Beweis von Theorem 2.18 ergibt sich mit den dort eingeführten Bezeichnungen, dass für jedes $\theta \in (0, 1)$

$$\begin{aligned} 2 \text{Var}_{\pi}(\mathbf{x}) &= \sum_{i, j \in E} (x_i - x_j)^2 \pi_i \pi_j \\ &= \sum_{i, j \in E} \left(\sum_{e \in \gamma_{ij}} \frac{1}{Q(e)^{\theta}} Q(e)^{\theta} (x_{e-} - x_{e+}) \right)^2 \pi_i \pi_j \\ &\leq \sum_{i, j \in E} \left(\sum_{e \in \gamma_{ij}} Q(e)^{2\theta} (x_{e-} - x_{e+})^2 \right) \left(\sum_{e \in \gamma_{ij}} \frac{1}{Q(e)^{2\theta}} \right) \pi_i \pi_j, \end{aligned}$$

wobei der „gerichteten Kante“ $e = (e^-, e^+)$ die „Kantenwahrscheinlichkeit“ $Q(e) = \pi_{e^-} p_{e^- e^+}$ zugeordnet wird und γ_{ij} den „Pfad“ von i nach j bezeichnet.

- Mit der Schreibweise $|\gamma_{ij}|_\theta = \sum_{e \in \gamma_{ij}} Q(e)^{-2\theta}$ erhalten wir somit, dass

$$\begin{aligned} 2 \text{Var}_{\pi}(\mathbf{x}) &\leq \sum_{i,j \in E} |\gamma_{ij}|_\theta \sum_{e \in \gamma_{ij}} Q(e)^{2\theta} (x_{e^-} - x_{e^+})^2 \pi_i \pi_j \\ &= \sum_{e \in \mathcal{E}} (x_{e^-} - x_{e^+})^2 Q(e) Q(e)^{2\theta-1} \sum_{\gamma_{ij} \ni e} \pi_i \pi_j |\gamma_{ij}|_\theta. \end{aligned}$$

- Hieraus ergibt sich die Gültigkeit von (63) für

$$a = \max_{e \in \mathcal{E}} \left\{ Q(e)^{2\theta-1} \sum_{\gamma_{ij} \ni e} \pi_i \pi_j |\gamma_{ij}|_\theta \right\}, \quad (65)$$

weil wir in Lemma 2.8 gezeigt hatten, dass

$$2 D_{(\mathbf{P}, \pi)}(\mathbf{x}, \mathbf{x}) = \sum_{e \in \mathcal{E}} (x_{e^-} - x_{e^+})^2 Q(e).$$

- Es ist nun noch zu zeigen, dass die in (65) betrachtete Konstante a der Ungleichung (64) genügt.
 - Hierfür wählen wir den Pfad $\gamma_{ij} = (i, i+1, \dots, j-1, j)$ für jedes Paar $i, j \in E$ mit $i < j$.
 - Dann ergibt sich aus (56) und (59)–(60), dass

$$Q(i, i+1) = \pi_i p_{i, i+1} = \frac{b^{h(i)}}{z(b)} \frac{b^{h(i+1)-h(i)}}{2} = \frac{\pi_{i+1}}{2}.$$

- Aus der in Theorem 3.14 gezeigten Reversibilität des Paares $(\mathbf{P}, \pi)$ folgt somit, dass

$$Q(i+1, i) = Q(i, i+1) = \frac{\pi_{i+1}}{2}.$$

- Wegen (56) und (57) ergibt sich hieraus, dass für beliebige $i, j \in E$ mit $i < j$

$$\begin{aligned} |\gamma_{ij}|_\theta &= \left(\left(\frac{\pi_{i+1}}{\pi_j} \right)^{-2\theta} + \dots + \left(\frac{\pi_j}{\pi_j} \right)^{-2\theta} \right) \left(\frac{\pi_j}{2} \right)^{-2\theta} \leq \left(b^{2(j-i-1)c\theta} + \dots + b^{2c\theta} + 1 \right) \left(\frac{\pi_j}{2} \right)^{-2\theta} \\ &\leq \frac{2^{2\theta} \pi_j^{-2\theta}}{1 - b^{2c\theta}}. \end{aligned}$$

- Außerdem haben sämtliche Kanten $e \in \mathcal{E}$ die Form $e = (i, i+1)$ oder $e = (i, i-1)$, denn für die in (58)–(60) gegebenen Eintragungen der Übergangsmatrix $\mathbf{P} = (p_{ij})$ gilt $p_{ij} = 0$, falls $|i - j| > 1$.
- Für die in (65) betrachtete Konstante a ergibt sich somit insgesamt, dass für $\theta < 1/2$

$$\begin{aligned} a &= \max_{e \in \mathcal{E}} \left\{ Q(e)^{2\theta-1} \sum_{\gamma_{ij} \ni e} \pi_i \pi_j |\gamma_{ij}|_\theta \right\} \\ &\leq \max_{k=1, \dots, \ell-1} \left\{ Q(k, k+1)^{2\theta-1} \sum_{1 \leq i \leq k, k+1 \leq j \leq \ell} 2^{2\theta} \frac{\pi_i \pi_j^{1-2\theta}}{1 - b^{2c\theta}} \right\} \\ &\leq \frac{2}{(1 - b^{2c\theta})(1 - b^{c(1-2\theta)})}, \end{aligned}$$

weil $Q(k, k+1)^{2\theta-1} = (\pi_{k+1}/2)^{2\theta-1}$ bzw. $\sum_{1 \leq i \leq k} \pi_i \leq 1$ und

$$\sum_{k+1 \leq j \leq \ell} \pi_j^{1-2\theta} = \left(\left(\frac{\pi_{k+1}}{\pi_{k+1}} \right)^{1-2\theta} + \dots + \left(\frac{\pi_\ell}{\pi_{k+1}} \right)^{1-2\theta} \right) \pi_{k+1}^{1-2\theta} \leq \frac{\pi_{k+1}^{1-2\theta}}{1 - b^{c(1-2\theta)}}.$$

- Für $\theta = 1/4$ ergibt sich hieraus die Abschätzung (64). □

Bei der Herleitung einer unteren Schranke für den kleinsten Eigenwert λ_ℓ der in (59)–(60) definierten Übergangsmatrix $\mathbf{P} = (p_{ij})$ ist der folgende Hilfssatz nützlich.

Lemma 3.1

- Sei $\mathbf{A} = (a_{ij})$ eine beliebige $\ell \times \ell$ -Matrix, und für jedes $i = 1, \dots, \ell$ sei $r_i = \sum_{j: 1 \leq j \leq \ell, j \neq i} |a_{ij}|$.
- Sei λ ein beliebiger Eigenwert von $\mathbf{A}$, sei $\phi = (\phi_1, \dots, \phi_\ell)^\top \neq \mathbf{0}$ ein rechter Eigenvektor, der zu λ gehört, und sei k die Nummer derjenigen Komponente ϕ_k mit

$$|\phi_k| = \max_{i=1, \dots, \ell} |\phi_i| > 0.$$

- Dann gilt

$$|\lambda - a_{kk}| \leq r_k. \quad (66)$$

Beweis

- Aus der Definitionsgleichung $\mathbf{A}\phi = \lambda\phi$ von λ bzw. ϕ ergibt sich insbesondere, dass

$$\sum_{j=1}^{\ell} a_{kj} \phi_j = \lambda \phi_k \quad \text{bzw.} \quad (\lambda - a_{kk}) \phi_k = \sum_{j: 1 \leq j \leq \ell, j \neq k} a_{kj} \phi_j.$$

- Hieraus folgt, dass

$$|\lambda - a_{kk}| |\phi_k| \leq \sum_{j: 1 \leq j \leq \ell, j \neq k} |a_{kj}| |\phi_j| \leq r_k |\phi_k| \quad \text{bzw.} \quad |\lambda - a_{kk}| \leq r_k. \quad \square$$

Theorem 3.16 Für den kleinsten Eigenwert λ_ℓ der in (59)–(60) definierten Übergangsmatrix $\mathbf{P} = (p_{ij})$ gilt

$$\lambda_\ell \geq -b^c. \quad (67)$$

Beweis

- Aus Lemma 3.1 ergibt sich für $\mathbf{A} = \mathbf{P}$ (und für den in Lemma 3.1 betrachteten, zu λ_ℓ gehörenden Index k), dass

$$|\lambda_\ell - p_{kk}| \leq \sum_{j: 1 \leq j \leq \ell, j \neq k} p_{kj} = 1 - p_{kk} \quad \text{bzw.} \quad \lambda_\ell \geq -1 + 2p_{kk}.$$

- Hieraus und aus (59)–(60) folgt, dass

$$\lambda_\ell \geq -1 + 2 \min_{i=1, \dots, \ell} p_{ii} \geq -1 + 2 \left(1 - \frac{1}{2} - \frac{b^c}{2} \right) = -b^c. \quad \square$$

Beachte Insgesamt ergibt sich aus den Theoremen 3.15 und 3.16 die Abschätzung

$$|\theta_2| = \max\{\lambda_2, |\lambda_\ell|\} \leq \max\left\{ 1 - \frac{(1 - b^{c/2})^2}{2}, b^c \right\} = 1 - \frac{(1 - b^{c/2})^2}{2}. \quad (68)$$

3.4.2 MCMC-Schätzer; Bias und Fundamentalmatrix

In diesem Abschnitt betrachten wir Güteeigenschaften von Monte-Carlo-Schätzern für Erwartungswerte.

- Beispiele für ähnliche Fragestellungen wurden bereits in Abschnitt 3.1.1 diskutiert,
 - und zwar im Zusammenhang mit der statistischen Schätzung der Zahl π
 - bzw. des Wertes von Integralen mittels Monte-Carlo-Simulation.
- Dabei hatten wir in Abschnitt 3.1.1 jedoch vorausgesetzt,
 - dass die verwendeten Pseudozufallszahlen als Realisierungen von unabhängigen und identisch verteilten Stichprobenvariablen aufgefasst werden können,
 - wogegen wir jetzt annehmen, dass die Stichprobenvariablen eine (geeignet gewählte) Markov-Kette bilden.
- Deshalb sprechen wir nun auch von *Markov-Chain-Monte-Carlo-Schätzern* bzw. kurz von MCMC-Schätzern.

Statistisches Modell

- Sei V eine endliche (nichtleere) Index-Menge, und sei $\mathbf{X} = (X(v), v \in V)$ ein diskreter Zufallsvektor,
 - der mit Wahrscheinlichkeit 1 seine Werte in der endlichen (Zustands-) Menge $E \subset \mathbb{R}^{|V|}$ annimmt,
 - wobei wir den Wertebereich E mit der Menge $E = \{1, \dots, \ell\}$ der ersten $\ell = |E|$ natürlichen Zahlen identifizieren
 - und über die Wahrscheinlichkeitsfunktion $\boldsymbol{\pi} = (\pi_i, i \in E)$ des Zufallsvektors $\mathbf{X}$ voraussetzen, dass $\pi_i > 0$ für jedes $i \in E$.
- Für eine beliebige, jedoch fest vorgegebene Funktion $\varphi : E \rightarrow \mathbb{R}$
 - soll der Erwartungswert $\theta = \mathbb{E} \varphi(\mathbf{X})$ mittels MCMC-Simulation geschätzt werden, wobei

$$\theta = \boldsymbol{\pi}^\top \boldsymbol{\varphi} \quad (69)$$

- und $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_\ell)^\top$ den Vektor der Funktionswerte von $\varphi : E \rightarrow \mathbb{R}$ bezeichnet.
- Als Schätzer für θ betrachten wir die Zufallsvariable

$$\hat{\theta}_n = \frac{1}{n} \sum_{k=0}^{n-1} \varphi(\mathbf{X}_k), \quad \forall n \geq 1, \quad (70)$$

- wobei $\mathbf{X}_0, \mathbf{X}_1, \dots$ eine Markov-Kette ist mit dem Zustandsraum E , mit einer (beliebigen, jedoch fest vorgegebenen) Anfangsverteilung $\boldsymbol{\alpha}$ sowie
- mit einer irreduziblen und aperiodischen Übergangsmatrix $\mathbf{P} = (p_{ij})$, so dass $\boldsymbol{\pi}$ die ergodische (Grenz-) Verteilung bezüglich $\mathbf{P}$ ist.

Beachte

- Weil die Anfangsverteilung $\boldsymbol{\alpha}$ typischerweise *nicht* mit der zu simulierenden Verteilung $\boldsymbol{\pi}$ übereinstimmt,
 - ist der in (70) definierte MCMC-Schätzer $\hat{\theta}_n$ bei fest vorgegebenem (endlichen) Stichprobenumfang n *nicht* erwartungstreu,
 - d.h., im allgemeinen gilt $\mathbb{E} \hat{\theta}_n \neq \theta$ für jedes $n \geq 1$.
- Bei der Bestimmung des *Bias* $\mathbb{E} \hat{\theta}_n - \theta$ ist die folgende Darstellungsformel nützlich.

Theorem 3.17 Für jedes $n \geq 1$ gilt

$$\mathbb{E} \hat{\theta}_n = \frac{1}{n} \boldsymbol{\alpha}^\top \sum_{k=0}^{n-1} \mathbf{P}^k \boldsymbol{\varphi}. \quad (71)$$

Beweis

- In Theorem 2.3 hatten wir gezeigt, dass für jedes $k \geq 1$ die Verteilung α_k von $\mathbf{X}_k$ gegeben ist durch $\alpha_k^\top = \alpha^\top \mathbf{P}^k$.
- Somit ergibt sich aus der Definitionsgleichung (70) des MCMC-Schätzers $\hat{\theta}_n$, dass

$$\mathbb{E} \hat{\theta}_n = \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E} \varphi(\mathbf{X}_k) = \frac{1}{n} \sum_{k=0}^{n-1} \alpha_k^\top \varphi = \frac{1}{n} \sum_{k=0}^{n-1} \alpha^\top \mathbf{P}^k \varphi = \frac{1}{n} \alpha^\top \sum_{k=0}^{n-1} \mathbf{P}^k \varphi. \quad \square$$

Beachte

- Aus Theorem 3.17 und aus der Ergodizität der Übergangsmatrix $\mathbf{P}$ ergibt sich unter Berücksichtigung von (69) sofort, dass

$$\lim_{n \rightarrow \infty} \mathbb{E} \hat{\theta}_n = \theta,$$

- d.h., der in (70) definierte MCMC-Schätzer $\hat{\theta}_n$ für θ ist asymptotisch erwartungstreu.

Außerdem kann die Asymptotik des Ausdruckes $n(\mathbb{E} \hat{\theta}_n - \theta)$ für $n \rightarrow \infty$ bestimmt werden. Hierfür leiten wir zunächst zwei elementare Hilfssätze her.

Lemma 3.2 *Sei Π die $\ell \times \ell$ Matrix, die aus den ℓ identischen (Zeilen-) Vektoren $\pi^\top$ besteht. Dann gilt*

$$(\mathbf{P} - \Pi)^n = \mathbf{P}^n - \Pi \quad (72)$$

für jedes $n \geq 1$ und insbesondere

$$\lim_{n \rightarrow \infty} (\mathbf{P} - \Pi)^n = \mathbf{0}. \quad (73)$$

Beweis

- Offenbar gilt (72) für $n = 1$.
 - Wenn nun angenommen wird, dass (72) für ein $n - 1 \geq 1$ gilt, dann gilt auch

$$\begin{aligned} (\mathbf{P} - \Pi)^n &= (\mathbf{P} - \Pi)^{n-1} (\mathbf{P} - \Pi) = (\mathbf{P}^{n-1} - \Pi)(\mathbf{P} - \Pi) \\ &= \mathbf{P}^n - \Pi \mathbf{P} - \mathbf{P}^{n-1} \Pi + \Pi^2 = \mathbf{P}^n - \Pi, \end{aligned}$$

wobei sich die letzte Gleichheit aus der Tatsache ergibt, dass

$$\pi^\top \mathbf{P} = \pi^\top \quad \text{und somit} \quad \Pi \mathbf{P} = \mathbf{P} \Pi = \Pi = \Pi^2.$$

- Damit ist die Gültigkeit von (72) für jedes $n \geq 1$ bewiesen.
- Weil vorausgesetzt wird, dass $\mathbf{P}$ irreduzibel und aperiodisch ist,
 - ergibt sich aus den Theoremen 2.4 und 2.9, dass $\mathbf{P}^n - \Pi \rightarrow \mathbf{0}$, wenn $n \rightarrow \infty$.
 - Hieraus und aus (72) folgt, dass auch $(\mathbf{P} - \Pi)^n \rightarrow \mathbf{0}$ gilt, wenn $n \rightarrow \infty$. $\square$

Beachte

- Aus der in Lemma 3.2 gezeigten Null-Konvergenz $(\mathbf{P} - \Pi)^n \rightarrow \mathbf{0}$ für $n \rightarrow \infty$, ergibt sich nun mit Hilfe von Lemma 2.4, dass die Matrix $\mathbf{I} - (\mathbf{P} - \Pi)$ invertierbar ist.
- Um dies zu zeigen, genügt es, in Lemma 2.4 die Matrix $\mathbf{A} = \mathbf{P} - \Pi$ zu betrachten.

- Die inverse Matrix

$$\mathbf{Z} = (\mathbf{I} - (\mathbf{P} - \mathbf{\Pi}))^{-1} \quad (74)$$

ist also wohldefiniert und wird die *Fundamentalmatrix* von $\mathbf{P}$ genannt.

Lemma 3.3 Für die Fundamentalmatrix $\mathbf{Z} = (\mathbf{I} - (\mathbf{P} - \mathbf{\Pi}))^{-1}$ der irreduziblen und aperiodischen Übergangsmatrix $\mathbf{P}$ gelten die Darstellungsformeln

$$\mathbf{Z} = \mathbf{I} + \sum_{k=1}^{\infty} (\mathbf{P}^k - \mathbf{\Pi}) \quad (75)$$

und

$$\mathbf{Z} = \mathbf{I} + \lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} \frac{n-k}{n} (\mathbf{P}^k - \mathbf{\Pi}). \quad (76)$$

Beweis

- Die Gültigkeit von (75) ergibt sich aus den Lemma 2.4 und Lemma 3.2, denn für $\mathbf{A} = \mathbf{P} - \mathbf{\Pi}$ gilt, dass

$$\begin{aligned} \mathbf{Z} &= (\mathbf{I} - \mathbf{A})^{-1} \\ &= (\mathbf{I} - \mathbf{A})^{-1} \lim_{n \rightarrow \infty} (\mathbf{I} - \mathbf{A}^n) \\ &= \lim_{n \rightarrow \infty} \left((\mathbf{I} - \mathbf{A})^{-1} (\mathbf{I} - \mathbf{A}^n) \right) \\ &\stackrel{(2.79)}{=} \lim_{n \rightarrow \infty} \left(\mathbf{I} + \mathbf{A} + \dots + \mathbf{A}^{n-1} \right) \\ &= \mathbf{I} + \sum_{k=1}^{\infty} \mathbf{A}^k \quad \left(= \mathbf{I} + \sum_{k=1}^{\infty} (\mathbf{P} - \mathbf{\Pi})^k \right) \\ &\stackrel{(72)}{=} \mathbf{I} + \sum_{k=1}^{\infty} (\mathbf{P}^k - \mathbf{\Pi}). \end{aligned}$$

- Um die Gültigkeit von (76) zu zeigen, genügt es zu beachten, dass

$$\sum_{k=1}^n (\mathbf{P}^k - \mathbf{\Pi}) - \sum_{k=1}^{n-1} \frac{n-k}{n} (\mathbf{P}^k - \mathbf{\Pi}) = \sum_{k=1}^n \frac{k}{n} (\mathbf{P}^k - \mathbf{\Pi}) = \frac{1}{n} \sum_{k=1}^n k (\mathbf{P} - \mathbf{\Pi})^k$$

und dass der letzte Ausdruck gegen $\mathbf{0}$ strebt, wenn $n \rightarrow \infty$.

- Dabei ergibt sich diese Null-Konvergenz aus der Tatsache, dass jede $\ell \times \ell$ Matrix $\mathbf{A}$ der Identität

$$(\mathbf{I} - \mathbf{A}) \sum_{k=1}^n k \mathbf{A}^k = \sum_{k=1}^n \mathbf{A}^k - n \mathbf{A}^{n+1}$$

genügt und dass somit für $\mathbf{A} = \mathbf{P} - \mathbf{\Pi}$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n k (\mathbf{P} - \mathbf{\Pi})^k &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{Z} \sum_{k=1}^n (\mathbf{P} - \mathbf{\Pi})^k - \mathbf{Z} (\mathbf{P} - \mathbf{\Pi})^{n+1} \right) \\ &\stackrel{(72)}{=} \mathbf{Z} \left(\underbrace{\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (\mathbf{P}^k - \mathbf{\Pi})}_{\stackrel{(75)}{\rightarrow} \mathbf{Z} - \mathbf{I}} - \underbrace{\lim_{n \rightarrow \infty} (\mathbf{P} - \mathbf{\Pi})^{n+1}}_{\stackrel{(73)}{\rightarrow} \mathbf{0}} \right) \\ &= \mathbf{0}. \end{aligned}$$

□

Mit Hilfe von Theorem 3.17 und Lemma 3.3 können wir nun das asymptotische Verhalten des Bias $\mathbb{E} \hat{\theta}_n - \theta$ genauer bestimmen.

Theorem 3.18

- Sei $a = (\alpha^\top \mathbf{Z} - \pi^\top) \varphi$, wobei $\mathbf{Z}$ die in (74) eingeführte Fundamentalmatrix von $\mathbf{P}$ ist.
- Für jedes $n \geq 1$ gilt dann

$$n(\mathbb{E} \hat{\theta}_n - \theta) = a + e_n, \quad (77)$$

wobei e_n ein Restglied ist mit $e_n \rightarrow 0$ für $n \rightarrow \infty$.

Beweis

- Aus der Darstellungsformel (75) in Lemma 3.3 ergibt sich, dass

$$\begin{aligned} \alpha^\top \mathbf{Z} \varphi &= \alpha^\top \varphi + \alpha^\top \lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} (\mathbf{P}^k - \Pi) \varphi \\ &= \alpha^\top \varphi + \lim_{n \rightarrow \infty} \left(\alpha^\top \left(\sum_{k=1}^{n-1} \mathbf{P}^k \right) \varphi - (n-1) \underbrace{\alpha^\top \Pi \varphi}_{=\pi^\top \varphi} \right) \\ &= \lim_{n \rightarrow \infty} \left(\alpha^\top \left(\sum_{k=0}^{n-1} \mathbf{P}^k \right) \varphi - (n-1) \pi^\top \varphi \right). \end{aligned}$$

- Hieraus und aus Theorem 3.17 ergibt sich für eine gewisse Folge $\{e_n\}$ mit $e_n \rightarrow 0$, dass

$$\begin{aligned} a &= (\alpha^\top \mathbf{Z} - \pi^\top) \varphi \\ &= \alpha^\top \left(\sum_{k=0}^{n-1} \mathbf{P}^k \right) \varphi - n \pi^\top \varphi - e_n \\ &\stackrel{(71)}{=} n \mathbb{E} \hat{\theta}_n - n \pi^\top \varphi - e_n \\ &\stackrel{(69)}{=} n(\mathbb{E} \hat{\theta}_n - \theta) - e_n. \end{aligned} \quad \square$$

3.4.3 Asymptotische Schätzvarianz; mittlerer quadratischer Fehler

Für das in Abschnitt 3.4.2 eingeführte statistische Modell untersuchen wir nun das asymptotische Verhalten der Varianz $\text{Var} \hat{\theta}_n$, wenn $n \rightarrow \infty$.

Theorem 3.19 *Es gilt*

$$\lim_{n \rightarrow \infty} n \text{Var} \hat{\theta}_n = \sigma^2 + 2\pi^\top \text{diag}(\varphi)(\mathbf{Z} - \mathbf{I})\varphi, \quad (78)$$

wobei $\sigma^2 = \sum_{i=1}^{\ell} \pi_i(\varphi_i - \theta)^2$ und $\mathbf{Z} = (\mathbf{I} - (\mathbf{P} - \Pi))^{-1}$ die in (74) definierte Fundamentalmatrix von $\mathbf{P}$ ist.

Beweis

- Offenbar gilt

$$n^2 \text{Var} \hat{\theta}_n = \mathbb{E} \left(\sum_{k=0}^{n-1} \varphi(\mathbf{X}_k) \right)^2 - \left(\sum_{k=0}^{n-1} \mathbb{E} \varphi(\mathbf{X}_k) \right)^2 \quad (79)$$

und somit

$$n^2 \text{Var} \hat{\theta}_n = \sum_{k=0}^{n-1} \mathbb{E} \varphi^2(\mathbf{X}_k) + 2 \sum_{0 \leq k < k' \leq n-1} \mathbb{E} \left(\varphi(\mathbf{X}_k) \varphi(\mathbf{X}_{k'}) \right) - \left(\sum_{k=0}^{n-1} \mathbb{E} \varphi(\mathbf{X}_k) \right)^2.$$

- Wir nutzen nun diese Darstellungsformel, um die Gültigkeit von (78) zunächst für den Fall $\alpha_0 = \pi$ zu zeigen.

– In diesem Fall gilt nämlich, dass

$$\left(\sum_{k=0}^{n-1} \mathbb{E} \varphi(\mathbf{X}_k) \right)^2 = (n\theta)^2 \quad \text{und} \quad \sum_{k=0}^{n-1} \mathbb{E} \varphi^2(\mathbf{X}_k) = n \sum_{i=1}^{\ell} \pi_i \varphi_i^2.$$

– Außerdem ergibt sich dann aus der Stationarität der Markov-Kette $\{\mathbf{X}_n\}$, dass

$$\sum_{0 \leq k < k' \leq n-1} \mathbb{E} \left(\varphi(\mathbf{X}_k) \varphi(\mathbf{X}_{k'}) \right) = \sum_{k=1}^{n-1} (n-k) \mathbb{E} \left(\varphi(\mathbf{X}_0) \varphi(\mathbf{X}_k) \right),$$

wobei

$$\mathbb{E} \left(\varphi(\mathbf{X}_0) \varphi(\mathbf{X}_k) \right) = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \pi_i \varphi_i p_{ij}^{(k)} \varphi_j = \boldsymbol{\pi}^\top \text{diag}(\varphi) \mathbf{P}^k \boldsymbol{\varphi}$$

und $\mathbf{P}^k = \mathbf{P}^{(k)} = (p_{ij}^{(k)})$ die Matrix der k -stufigen Übergangswahrscheinlichkeiten bezeichnet.

– Hieraus folgt nun insgesamt, dass

$$\begin{aligned} \frac{1}{n} \text{Var} \left(\sum_{k=0}^{n-1} \varphi(\mathbf{X}_k) \right) &= \sum_{i=1}^{\ell} \pi_i \varphi_i^2 + 2 \boldsymbol{\pi}^\top \text{diag}(\varphi) \sum_{k=1}^{n-1} \frac{n-k}{n} \mathbf{P}^k \boldsymbol{\varphi} - n\theta^2 \\ &= \sigma^2 + 2 \boldsymbol{\pi}^\top \text{diag}(\varphi) \left(\sum_{k=1}^{n-1} \frac{n-k}{n} \mathbf{P}^k \boldsymbol{\varphi} - \frac{n-1}{2} \boldsymbol{\Pi} \boldsymbol{\varphi} \right) \\ &= \sigma^2 + 2 \boldsymbol{\pi}^\top \text{diag}(\varphi) \left(\sum_{k=1}^{n-1} \frac{n-k}{n} (\mathbf{P}^k - \boldsymbol{\Pi}) \right) \boldsymbol{\varphi}, \end{aligned}$$

– wobei in der zweiten Gleichheit die folgende Identität genutzt wurde:

$$\theta^2 = \boldsymbol{\pi}^\top \text{diag}(\varphi) \boldsymbol{\Pi} \boldsymbol{\varphi}.$$

– Hieraus ergibt sich (78), wenn dabei die Darstellungsformel (76) für $\mathbf{Z} - \mathbf{I}$ berücksichtigt wird.

- Wir zeigen nun noch, dass (78) auch für jede beliebige Anfangsverteilung α gilt.

– Dabei präzisieren wir die Bezeichnungsweise wie folgt: Wir schreiben $\mathbf{X}_0^{(\alpha)}, \mathbf{X}_1^{(\alpha)}, \dots$ anstelle von $\mathbf{X}_0, \mathbf{X}_1, \dots$ bzw. $\hat{\theta}_n^{(\alpha)}$ anstelle von $\hat{\theta}_n$.

– Es genügt zu zeigen, dass

$$\lim_{n \rightarrow \infty} n \left(\text{Var} \hat{\theta}_n^{(\pi)} - \text{Var} \hat{\theta}_n^{(\alpha)} \right) = 0. \quad (80)$$

– Hierfür führen wir noch die folgenden Bezeichnungen ein: Für $0 < r < n-1$ sei

$$Y_r^{(\cdot)} = \sum_{k=0}^{r-1} \varphi(\mathbf{X}_k^{(\cdot)}) \quad \text{und} \quad Z_{rn}^{(\cdot)} = \sum_{k=r}^{n-1} \varphi(\mathbf{X}_k^{(\cdot)}).$$

– Dann ergibt sich aus (79), dass

$$\begin{aligned} &n^2 \left(\text{Var} \hat{\theta}_n^{(\pi)} - \text{Var} \hat{\theta}_n^{(\alpha)} \right) \\ &= \left(\mathbb{E} (Y_r^{(\pi)} + Z_{rn}^{(\pi)})^2 - \mathbb{E} (Y_r^{(\alpha)} + Z_{rn}^{(\alpha)})^2 \right) - \left((\mathbb{E} Y_r^{(\pi)} + \mathbb{E} Z_{rn}^{(\pi)})^2 - (\mathbb{E} Y_r^{(\alpha)} + \mathbb{E} Z_{rn}^{(\alpha)})^2 \right) \\ &= \left(\mathbb{E} (Y_r^{(\pi)})^2 - (\mathbb{E} Y_r^{(\pi)})^2 - \mathbb{E} (Y_r^{(\alpha)})^2 + (\mathbb{E} Y_r^{(\alpha)})^2 \right) \\ &\quad + 2 \mathbb{E} \left((Y_r^{(\pi)} - \mathbb{E} (Y_r^{(\pi)})) (Z_{rn}^{(\pi)} - \mathbb{E} (Z_{rn}^{(\pi)})) \right) - 2 \mathbb{E} \left((Y_r^{(\alpha)} - \mathbb{E} (Y_r^{(\alpha)})) (Z_{rn}^{(\alpha)} - \mathbb{E} (Z_{rn}^{(\alpha)})) \right) \\ &\quad + \left(\mathbb{E} (Z_{rn}^{(\pi)})^2 - (\mathbb{E} Z_{rn}^{(\pi)})^2 \right) - \left(\mathbb{E} (Z_{rn}^{(\alpha)})^2 - (\mathbb{E} Z_{rn}^{(\alpha)})^2 \right), \end{aligned}$$

wobei wir die drei Summanden des letzten Ausdruckes mit I_r , II_{rn} bzw. III_{rn} bezeichnen.

- Der Summand I_r hängt nicht von n ab, weshalb $\lim_{n \rightarrow \infty} n^{-1} I_r = 0$.
- Weil der Zustandsraum E endlich ist, gilt für die Konstante $c = \max_{i \in E} |\varphi(i)| < \infty$, dass

$$\frac{1}{n} II_{rn} \leq 4rc \mathbb{E} \left(\frac{1}{n} |Z_{rn}^{(\pi)} - \mathbb{E}(Z_{rn}^{(\pi)})| \right) + 4rc \mathbb{E} \left(\frac{1}{n} |Z_{rn}^{(\alpha)} - \mathbb{E}(Z_{rn}^{(\alpha)})| \right).$$

- Hieraus folgt, dass $\lim_{n \rightarrow \infty} n^{-1} II_{rn} = 0$ für jedes $r > 0$, weil mit Wahrscheinlichkeit 1

$$\frac{1}{n} |Z_{rn}^{(\cdot)} - \mathbb{E}(Z_{rn}^{(\cdot)})| \leq 2c$$

für jedes $n > r$ und

$$\lim_{n \rightarrow \infty} \frac{1}{n} |Z_{rn}^{(\pi)} - \mathbb{E}(Z_{rn}^{(\pi)})| = \lim_{n \rightarrow \infty} \frac{1}{n} |Z_{rn}^{(\alpha)} - \mathbb{E}(Z_{rn}^{(\alpha)})| = 0.$$

- Außerdem gilt für $n > r > 0$ die Abschätzung

$$\begin{aligned} \frac{1}{n} III_{rn} &\leq \frac{1}{n} \sum_{i=1}^{\ell} \left(\mathbb{E}(Z_{0,n-r}^{(\delta_i)})^2 - (\mathbb{E} Z_{0,n-r}^{(\delta_i)})^2 \right) |\pi_i - \alpha_{ri}| \\ &\leq \underbrace{\sup_{n>0} \max_{j \in \{1, \dots, \ell\}} \frac{1}{n+r} \mathbb{E} \left(Z_{0n}^{(\delta_j)} - \mathbb{E} Z_{0n}^{(\delta_j)} \right)^2}_{< \infty} \sum_{i=1}^{\ell} |\pi_i - \alpha_{ri}|, \end{aligned}$$

wobei man sich leicht überlegen kann, dass das Supremum endlich ist.

- Weil die letzte Summe wegen der Ergodizität der Markov-Kette $\mathbf{X}_0^{(\alpha)}, \mathbf{X}_1^{(\alpha)}, \dots$ beliebig klein wird, falls r hinreichend groß ist, ist damit die Gültigkeit von (80) bewiesen. $\square$

Beachte

- Zur Erinnerung
 - Aus Theorem 2.1 des Vorlesungsskriptes „Statistik I“ ergibt sich für den *mittleren quadratischen Fehler* $\mathbb{E}((\hat{\theta}_n - \theta)^2)$ des in (70) definierten MCMC-Schätzers $\hat{\theta}_n = \hat{\theta}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ für θ , dass

$$\mathbb{E}((\hat{\theta}_n - \theta)^2) = (\mathbb{E} \hat{\theta}_n - \theta)^2 + \text{Var} \hat{\theta}_n, \quad (81)$$
 - d.h., der MQ-Fehler des MCMC-Schätzers $\hat{\theta}_n$ ist die Summe aus quadriertem Bias $(\mathbb{E} \hat{\theta}_n - \theta)^2$ und Varianz $\text{Var} \hat{\theta}_n$ des Schätzers $\hat{\theta}_n$.
- Die beiden Summanden auf der rechten Seite von (81) konvergieren unterschiedlich schnell gegen 0, wenn $n \rightarrow \infty$.
 - In Theorem 3.19 haben wir gezeigt, dass $\text{Var} \hat{\theta}_n = O(n^{-1})$ gilt.
 - Andererseits ergibt sich aus Theorem 3.18, dass $(\mathbb{E} \hat{\theta}_n - \theta)^2 = O(n^{-2})$.
- Hieraus folgt, dass
 - *nicht* der Bias $\mathbb{E} \hat{\theta}_n - \theta$, sondern die asymptotische Varianz $\text{Var} \hat{\theta}_n$ des Schätzers $\hat{\theta}_n$
 - die entscheidende Rolle bei der asymptotischen Analyse des MQ-Fehlers $\mathbb{E}((\hat{\theta}_n - \theta)^2)$ des in (70) definierten MCMC-Schätzers $\hat{\theta}$ spielt, wenn $n \rightarrow \infty$.
- Mit anderen Worten: Es kann sinnvoll sein, die Simulationsmatrix $\mathbf{P}$ so zu wählen,
 - dass die asymptotische Schätzvarianz $\lim_{n \rightarrow \infty} n \text{Var} \hat{\theta}_n$ möglichst klein ist,
 - gegebenenfalls bei einer gewissen Vergrößerung des asymptotischen Bias $\lim_{n \rightarrow \infty} n(\mathbb{E} \hat{\theta}_n - \theta)$.

Um diese Problematik näher zu untersuchen, führen wir die folgende Bezeichnung ein: Sei

$$V(\varphi, \mathbf{P}, \boldsymbol{\pi}) = \lim_{n \rightarrow \infty} n \text{Var } \hat{\theta}_n,$$

wobei $\varphi : E \rightarrow \mathbb{R}$ eine beliebige Funktion und $(\mathbf{P}, \boldsymbol{\pi})$ ein beliebiges reversibles Paar ist.

Theorem 3.20

- Seien $\mathbf{P}_1 = (p_{1,ij})$ und $\mathbf{P}_2 = (p_{2,ij})$ zwei Übergangsmatrizen auf E , so dass $(\mathbf{P}_1, \boldsymbol{\pi})$ und $(\mathbf{P}_2, \boldsymbol{\pi})$ jeweils reversible Paare sind.
 - Für beliebige $i, j \in E$ mit $i \neq j$ gelte $p_{1,ij} \geq p_{2,ij}$,
 - d.h. außerhalb der Hauptdiagonalen seien sämtliche Eintragungen der Übergangsmatrix $\mathbf{P}_1$ gleich oder größer als die entsprechenden Eintragungen der Übergangsmatrix $\mathbf{P}_2$.
- Dann gilt für jede Funktion $\varphi : E \rightarrow \mathbb{R}$

$$V(\varphi, \mathbf{P}_1, \boldsymbol{\pi}) \leq V(\varphi, \mathbf{P}_2, \boldsymbol{\pi}). \quad (82)$$

Beweis

- Sei $\mathbf{P} = (p_{ij})$ eine Übergangsmatrix, so dass das Paar $(\mathbf{P}, \boldsymbol{\pi})$ reversibel ist. Es genügt zu zeigen, dass

$$\frac{\partial}{\partial p_{ij}} V(\varphi, \mathbf{P}, \boldsymbol{\pi}) \leq 0, \quad \forall i, j \in E \text{ mit } i \neq j. \quad (83)$$

- Aus Theorem 3.19 ergibt sich, dass

$$\frac{\partial}{\partial p_{ij}} V(\varphi, \mathbf{P}, \boldsymbol{\pi}) = 2\boldsymbol{\pi}^\top \text{diag}(\varphi) \frac{\partial \mathbf{Z}}{\partial p_{ij}} \boldsymbol{\varphi}, \quad (84)$$

wobei $\mathbf{Z}$ die in (74) eingeführte Fundamentalmatrix von $\mathbf{P}$ ist.

- Andererseits folgt aus $\mathbf{Z}\mathbf{Z}^{-1} = \mathbf{I}$, dass

$$\left(\frac{\partial \mathbf{Z}}{\partial p_{ij}} \right) \mathbf{Z}^{-1} + \mathbf{Z} \left(\frac{\partial \mathbf{Z}^{-1}}{\partial p_{ij}} \right) = \mathbf{0}$$

und somit

$$\frac{\partial \mathbf{Z}}{\partial p_{ij}} = -\mathbf{Z} \frac{\partial \mathbf{Z}^{-1}}{\partial p_{ij}} \mathbf{Z}.$$

- Hieraus und aus (84) ergibt sich, dass

$$\frac{\partial}{\partial p_{ij}} V(\varphi, \mathbf{P}, \boldsymbol{\pi}) = -2\boldsymbol{\pi}^\top \text{diag}(\varphi) \mathbf{Z} \frac{\partial \mathbf{Z}^{-1}}{\partial p_{ij}} \mathbf{Z} \boldsymbol{\varphi}. \quad (85)$$

- Weil das Paar $(\mathbf{P}, \boldsymbol{\pi})$ reversibel ist, ergibt sich aus der in Lemma 3.3 hergeleiteten Darstellungsformel (75) für die Fundamentalmatrix $\mathbf{Z} = (z_{ij})$, dass für beliebige $i, j \in E$

$$\pi_i z_{ij} = \pi_i \delta_{ij} + \sum_{k=1}^{\infty} (\pi_i p_{ij}^{(k)} - \pi_i \pi_j) = \pi_j \delta_{ji} + \sum_{k=1}^{\infty} (\pi_j p_{ji}^{(k)} - \pi_j \pi_i) = \pi_j z_{ji}.$$

- Hieraus folgt, dass

$$\begin{aligned} \boldsymbol{\pi}^\top \text{diag}(\varphi) \mathbf{Z} &= \left(\sum_{i=1}^{\ell} \pi_i \varphi_i z_{i1}, \dots, \sum_{i=1}^{\ell} \pi_i \varphi_i z_{i\ell} \right) \\ &= \left(\pi_1 \sum_{i=1}^{\ell} z_{1i} \varphi_i, \dots, \pi_\ell \sum_{i=1}^{\ell} z_{\ell i} \varphi_i \right) \\ &= (\mathbf{Z} \boldsymbol{\varphi})^\top \text{diag}(\boldsymbol{\pi}). \end{aligned}$$

– Somit ergibt sich aus (85), dass

$$\frac{\partial}{\partial p_{ij}} V(\varphi, \mathbf{P}, \boldsymbol{\pi}) = -2(\mathbf{Z}\boldsymbol{\varphi})^\top \text{diag}(\boldsymbol{\pi}) \frac{\partial \mathbf{Z}^{-1}}{\partial p_{ij}} \mathbf{Z}\boldsymbol{\varphi} = 2(\mathbf{Z}\boldsymbol{\varphi})^\top \text{diag}(\boldsymbol{\pi}) \frac{\partial \mathbf{P}}{\partial p_{ij}} \mathbf{Z}\boldsymbol{\varphi}, \quad (86)$$

wobei in der letzten Gleichheit die Identität

$$\frac{\partial \mathbf{Z}^{-1}}{\partial p_{ij}} = -\frac{\partial \mathbf{P}}{\partial p_{ij}}$$

genutzt wurde, die sich unmittelbar aus der Definitionsgleichung (74) von $\mathbf{Z}$ ergibt.

- Weil $\mathbf{P} = (p_{ij})$ eine stochastische Matrix ist und weil das Paar $(\mathbf{P}, \boldsymbol{\pi})$ reversibel ist,
 - sind nur die Eintragungen p_{ij} mit $i < j$ (oder nur die Eintragungen p_{ij} mit $i > j$) frei wählbar,
 - denn für jedes Index-Paar $i, j \in E$ mit $i \neq j$ lassen sich die Eintragungen p_{ji} , p_{ii} und p_{jj} wie folgt durch p_{ij} ausdrücken: Es gilt

$$p_{ji} = \frac{\pi_i}{\pi_j} p_{ij}, \quad p_{ii} = c - p_{ij}, \quad p_{jj} = c' - \frac{\pi_i}{\pi_j} p_{ij},$$

wobei c und c' Konstanten sind, die nicht von p_{ij} abhängen.

- Für beliebige $i', j' \in E$ ist somit die Eintragung $(\text{diag}(\boldsymbol{\pi})(\partial \mathbf{P} / \partial p_{ij}))_{i', j'}$ des Matrixproduktes $\text{diag}(\boldsymbol{\pi})(\partial \mathbf{P} / \partial p_{ij})$ gegeben durch

$$\left(\text{diag}(\boldsymbol{\pi}) \frac{\partial \mathbf{P}}{\partial p_{ij}} \right)_{i', j'} = \begin{cases} -\pi_i, & \text{falls } (i', j') = (i, i) \text{ oder } (i', j') = (j, j), \\ \pi_i, & \text{falls } (i', j') = (i, j) \text{ oder } (i', j') = (j, i), \\ 0, & \text{sonst.} \end{cases}$$

- Hieraus folgt, dass die Matrix $\text{diag}(\boldsymbol{\pi})(\partial \mathbf{P} / \partial p_{ij})$ nichtpositiv definit ist, d.h., für jedes $\mathbf{x} \in \mathbb{R}^\ell$ gilt

$$\mathbf{x}^\top \text{diag}(\boldsymbol{\pi}) \frac{\partial \mathbf{P}}{\partial p_{ij}} \mathbf{x} \leq 0.$$

- Aus (86) ergibt sich nun dass für beliebige $i, j \in E$ mit $i \neq j$

$$\frac{\partial}{\partial p_{ij}} V(\varphi, \mathbf{P}, \boldsymbol{\pi}) = 2(\mathbf{Z}\boldsymbol{\varphi})^\top \text{diag}(\boldsymbol{\pi}) \frac{\partial \mathbf{P}}{\partial p_{ij}} \mathbf{Z}\boldsymbol{\varphi} \leq 0.$$

- Damit ist die Gültigkeit von (83) bewiesen. □

Beachte Aus Theorem 3.20 folgt insbesondere,

- dass durch die Simulationsmatrix $\mathbf{P}$ des Metropolis-Algorithmus (d.h. wenn in (50) die Gleichheit betrachtet wird) die asymptotische Varianz $V(\varphi, \mathbf{P}, \boldsymbol{\pi})$ minimiert wird,
- und zwar in der Klasse aller Metropolis-Hastings-Algorithmen mit einer beliebigen, jedoch fest vorgegebenen „potentiellen Übergangsmatrix“ $\mathbf{Q} = (q_{ij})$.

3.5 Kopplungsalgorithmen; perfekte MCMC-Simulation

- Wir diskutieren nun Algorithmen,
 - die (ähnlich wie in den Abschnitten 3.3 und 3.4) auf Markov-Ketten beruhen und
 - die gleichzeitig eine vorgegebene (diskrete) Verteilung $\boldsymbol{\pi}$ nicht nur näherungsweise, sondern in einem gewissen Sinne exakt simulieren.
- Man spricht deshalb in diesem Zusammenhang von „perfekter“ MCMC-Simulation.

3.5.1 Vorwärtskopplung; Gegenbeispiel

- Zunächst betrachten wir eine Methode zur „Kopplung“ der Pfade von Markov-Ketten, bei der die „Zeit“
 - auf die allgemein als natürlich empfundene Weise „nach vorn“ voranschreitet,
 - weshalb man in diesem Zusammenhang auch von *Vorwärtskopplung* spricht.
- Für jedes $i \in \{1, \dots, \ell\}$ sei $\mathbf{X}^{(i)} = (\mathbf{X}_0^{(i)}, \mathbf{X}_1^{(i)}, \dots)$ eine homogene Markov-Kette mit dem endlichen Zustandsraum $E = \{\mathbf{x}_1, \dots, \mathbf{x}_\ell\}$
 - mit dem (deterministischen) Anfangszustand $\mathbf{X}_0^{(i)} = \mathbf{x}_i$ und mit der irreduziblen und aperiodischen Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$,
 - so dass $\boldsymbol{\pi} = (\pi_{\mathbf{x}}, \mathbf{x} \in E)$ die ergodische (Grenz-) Verteilung der Markov-Kette $\mathbf{X}^{(i)}$ ist.

Definitionen

- Für jedes $k \in \{1, \dots, \ell\}$ betrachten wir
 - eine Folge $\mathbf{U}^{(k)} = (U_1^{(k)}, U_2^{(k)}, \dots)$ von unabhängigen und $(0, 1]$ -gleichverteilten Zufallsvariablen $U_n^{(k)}$,
 - die wir *Innovation* im n -ten Schritt bei Vorliegen des Zustandes $\mathbf{x}_k \in E$ nennen.
- Dabei betrachten wir zwei verschiedene Fälle:
 - Wir nehmen entweder an, dass die Folgen $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(\ell)}$ unabhängig sind,
 - oder wir betrachten lediglich *eine* Folge $\mathbf{U} = (U_1, U_2, \dots)$ und setzen $\mathbf{U}^{(1)} = \dots = \mathbf{U}^{(\ell)} = \mathbf{U}$.
- Die Markov-Kette $\mathbf{X}^{(i)}$ sei rekursiv gegeben durch

$$\mathbf{X}_n^{(i)} = \varphi(\mathbf{x}_k, U_n^{(k)}), \quad \text{falls } \mathbf{X}_{n-1}^{(i)} = \mathbf{x}_k, \quad (87)$$

wobei $\varphi : E \times (0, 1] \rightarrow E$ eine sogenannte *gültige Update-Funktion* ist, d.h.,

- die Funktion $\varphi(\mathbf{x}, \cdot) : (0, 1] \rightarrow E$ sei für jedes $\mathbf{x} \in E$ stückweise konstant,
- und für beliebige $\mathbf{x}, \mathbf{x}' \in E$ mit $p_{\mathbf{x}\mathbf{x}'} > 0$ sei die Gesamtlänge des Menge $\{u \in (0, 1] : \varphi(\mathbf{x}, u) = \mathbf{x}'\}$ gleich $p_{\mathbf{x}\mathbf{x}'}$.
- Die Zufallsvariable $\tau = \min\{n \geq 1 : \mathbf{X}_n^{(1)} = \dots = \mathbf{X}_n^{(\ell)}\}$ heißt *Kopplungszeit*, wobei $\tau = \infty$ gesetzt wird, falls es keine natürliche Zahl n gibt mit $\mathbf{X}_n^{(1)} = \dots = \mathbf{X}_n^{(\ell)}$.

Theorem 3.21 *Falls die Innovationen-Folgen $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(\ell)}$ unabhängig sind, dann gilt $\tau < \infty$ mit Wahrscheinlichkeit 1, und für jedes $n > \tau$ gilt $\mathbf{X}_n^{(1)} = \dots = \mathbf{X}_n^{(\ell)}$.*

Beweis

- Unmittelbar aus der rekursiven Definitionsgleichung (87) der Markov-Ketten $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(\ell)}$ ergibt sich, dass $\mathbf{X}_n^{(1)} = \dots = \mathbf{X}_n^{(\ell)}$ für jedes $n > \tau$.
- Es ist also noch zu zeigen, dass $P(\tau < \infty) = 1$. Hierfür genügt es zu zeigen, dass für beliebige $i \neq i'$

$$\lim_{r \rightarrow \infty} P(\max\{n : \mathbf{X}_n^{(i)} \neq \mathbf{X}_n^{(i')}\} \leq r) = 1.$$

– Wegen

$$\begin{aligned} P(\max\{n : \mathbf{X}_n^{(i)} \neq \mathbf{X}_n^{(i')}\} \leq r) &= 1 - P(\max\{n : \mathbf{X}_n^{(i)} \neq \mathbf{X}_n^{(i')}\} > r) \\ &= 1 - P(\mathbf{X}_r^{(i)} \neq \mathbf{X}_r^{(i')}) \end{aligned}$$

ist dies äquivalent mit

$$\lim_{r \rightarrow \infty} P(\mathbf{X}_r^{(i)} \neq \mathbf{X}_r^{(i')}) = 0.$$

- Sei nun $n_0 \geq 1$ eine natürliche Zahl mit

$$\min_{\mathbf{x}, \mathbf{x}' \in E} p_{\mathbf{x}\mathbf{x}'}^{(n_0)} = c > 0,$$

und sei $r = m(r)n_0 + k$ für gewisse Zahlen $m(r) = 0, 1, \dots$ und $k = 0, 1, \dots, n_0 - 1$.

- Dann ergibt sich aus der Unabhängigkeit der Innovationen-Folgen $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(\ell)}$, dass für $r \rightarrow \infty$

$$\begin{aligned} P(\mathbf{X}_r^{(i)} \neq \mathbf{X}_r^{(i')}) &= P(\mathbf{X}_{n_0}^{(i)} \neq \mathbf{X}_{n_0}^{(i')}, \mathbf{X}_r^{(i)} \neq \mathbf{X}_r^{(i')}) \\ &= \sum_{j=1}^{\ell} \sum_{j' \neq j} P(\mathbf{X}_{n_0}^{(i)} = \mathbf{x}_j, \mathbf{X}_{n_0}^{(i')} = \mathbf{x}_{j'}) P(\mathbf{X}_r^{(i)} \neq \mathbf{X}_r^{(i')} \mid \mathbf{X}_{n_0}^{(i)} = \mathbf{x}_j, \mathbf{X}_{n_0}^{(i')} = \mathbf{x}_{j'}) \\ &= \sum_{j=1}^{\ell} \sum_{j' \neq j} P(\mathbf{X}_{n_0}^{(i)} = \mathbf{x}_j) P(\mathbf{X}_{n_0}^{(i')} = \mathbf{x}_{j'}) P(\mathbf{X}_{r-n_0}^{(j)} \neq \mathbf{X}_{r-n_0}^{(j')}) \\ &= \underbrace{\sum_{j=1}^{\ell} p_{\mathbf{x}_i \mathbf{x}_j}^{(n_0)} \sum_{j' \neq j} p_{\mathbf{x}_{i'} \mathbf{x}_{j'}}^{(n_0)}}_{\leq 1-c} P(\mathbf{X}_{r-n_0}^{(j)} \neq \mathbf{X}_{r-n_0}^{(j')}) \\ &\leq (1-c) \max_{j \neq j'} P(\mathbf{X}_{r-n_0}^{(j)} \neq \mathbf{X}_{r-n_0}^{(j')}) \\ &\vdots \\ &\leq (1-c)^{m(r)} \rightarrow 0. \end{aligned}$$

□

Beachte

- Unter zusätzlichen Annahmen über die (irreduzible und aperiodische) Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ kann die Endlichkeit der Kopplungszeit τ auch dann gezeigt werden,
 - wenn nur *eine* Folge $\mathbf{U} = (U_1, U_2, \dots)$ von Innovationen, d.h. $\mathbf{U} = \mathbf{U}^{(1)} = \dots = \mathbf{U}^{(\ell)}$,
 - und wenn die folgende (gültige) Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ betrachtet wird, wobei für jedes $\mathbf{x} \in E$

$$\varphi(\mathbf{x}, u) = \mathbf{x}_j, \quad \text{falls } \sum_{r=1}^{j-1} p_{\mathbf{x}\mathbf{x}_r} < u \leq \sum_{r=1}^j p_{\mathbf{x}\mathbf{x}_r}. \quad (88)$$

- Eine solche Zusatzbedingung an $\mathbf{P}$ diskutieren wir in dem folgenden Theorem, vgl. auch die Monotonie-Bedingungen in Abschnitt 3.5.3.

Theorem 3.22

- Sei $\mathbf{U}^{(1)} = \dots = \mathbf{U}^{(\ell)} = \mathbf{U}$, die Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ sei durch (88) gegeben, und für ein $\mathbf{x}_{i_0} \in E$ gelte

$$\max_{\mathbf{x} \in E} \sum_{r=1}^{i_0-1} p_{\mathbf{x}\mathbf{x}_r} < \min_{\mathbf{x} \in E} \sum_{r=1}^{i_0} p_{\mathbf{x}\mathbf{x}_r}. \quad (89)$$

- Dann gilt $\tau < \infty$ mit Wahrscheinlichkeit 1, und für jedes $n > \tau$ gilt $\mathbf{X}_n^{(1)} = \dots = \mathbf{X}_n^{(\ell)}$.

Beweis

- Um zu beweisen, dass $P(\tau < \infty) = 1$, genügt es ähnlich wie im Beweis von Theorem 3.21 zu zeigen, dass für beliebige $i \neq i'$

$$\lim_{r \rightarrow \infty} P(\mathbf{X}_r^{(i)} \neq \mathbf{X}_r^{(i')}) = 0.$$

- Es gilt

$$\begin{aligned}
P(\mathbf{X}_r^{(i)} \neq \mathbf{X}_r^{(i')}) &= P(\mathbf{X}_1^{(i)} \neq \mathbf{X}_1^{(i')}, \mathbf{X}_r^{(i)} \neq \mathbf{X}_r^{(i')}) \\
&= \sum_{j=1}^{\ell} \sum_{j' \neq j} P(\mathbf{X}_1^{(i)} = \mathbf{x}_j, \mathbf{X}_1^{(i')} = \mathbf{x}_{j'}) P(\mathbf{X}_r^{(i)} \neq \mathbf{X}_r^{(i')} \mid \mathbf{X}_1^{(i)} = \mathbf{x}_j, \mathbf{X}_1^{(i')} = \mathbf{x}_{j'}) \\
&= \sum_{j=1}^{\ell} \sum_{j' \neq j} P(\mathbf{X}_1^{(i)} = \mathbf{x}_j, \mathbf{X}_1^{(i')} = \mathbf{x}_{j'}) P(\mathbf{X}_{r-1}^{(j)} \neq \mathbf{X}_{r-1}^{(j')}) \\
&\leq \underbrace{\left(1 - P(\mathbf{X}_1^{(i)} = \mathbf{X}_1^{(i')})\right)}_{\geq d > 0} \max_{j' \neq j} P(\mathbf{X}_{r-1}^{(j)} \neq \mathbf{X}_{r-1}^{(j')}) \\
&\vdots \\
&\leq (1-d)^r \longrightarrow 0,
\end{aligned}$$

wobei sich aus (87) – (89) ergibt, dass

$$0 < d = \max_{\mathbf{x} \in E} \sum_{r=1}^{i_0-1} p_{\mathbf{x}\mathbf{x}_r} - \min_{\mathbf{x} \in E} \sum_{r=1}^{i_0} p_{\mathbf{x}\mathbf{x}_r} \leq P(\mathbf{X}_1^{(i)} = \mathbf{X}_1^{(i')} = \mathbf{x}_{i_0}) \leq P(\mathbf{X}_1^{(i)} = \mathbf{X}_1^{(i')}). \quad \square$$

Beachte

- Aus $P(\tau < \infty) = 1$ folgt im allgemeinen *nicht*, dass $\mathbf{X}_\tau^{(i)} \sim \pi$,
 - d.h., zum Kopplungszeitpunkt τ stimmt die Zustandsverteilung der Markov-Kette $\mathbf{X}^{(i)}$ im allgemeinen nicht mit der stationären Grenzverteilung π überein, was man jedoch zunächst vermuten könnte.
- Das folgende *Gegenbeispiel* verdeutlicht dieses Paradoxon.
 - Wir betrachten den Zustandsraum $E = \{1, 2\}$ sowie die irreduzible und aperiodische Übergangsmatrix

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 \\ 1 & 0 \end{pmatrix}$$

mit der stationären (Grenz-) Verteilung $\pi = (2/3, 1/3)^\top$.

- Aus $\mathbf{X}_{\tau-1}^{(1)} \neq \mathbf{X}_{\tau-1}^{(2)}$ folgt dann aber zwangsläufig, dass $\mathbf{X}_{\tau-1}^{(1)} = 2$ oder $\mathbf{X}_{\tau-1}^{(2)} = 2$ und somit $\mathbf{X}_\tau^{(1)} = \mathbf{X}_\tau^{(2)} = 1$ gelten muss.

3.5.2 Propp-Wilson-Algorithmus; Coupling-from-the-Past

- Zur Erinnerung
 - Bei der in Abschnitt 3.5.1 diskutierten Vorwärtskopplung ist der Anfangspunkt 0 der Simulation *deterministisch*, wogegen der Endpunkt, d.h. der Kopplungszeitpunkt τ *zufällig* ist.
 - Dabei stimmt die Zustandsverteilung der Markov-Kette $\mathbf{X}^{(i)}$ zum Kopplungszeitpunkt τ im allgemeinen *nicht* mit der stationären Grenzverteilung π überein.
- Deshalb betrachten wir nun ein anderes Kopplungsverfahren,
 - das in der Literatur *Coupling-from-the-Past* (CFTP) bzw. *Rückwärtskopplung* genannt wird und
 - das in der Mitte der 90-er Jahre von Propp und Wilson am *Massachusetts Institute of Technology* (MIT) entwickelt worden ist.

- Dabei gehen wir zwar auf ähnliche Weise wie in Abschnitt 3.5.1 vor, wählen jedoch jetzt den Anfangspunkt der Simulation *zufällig*, wogegen der Endpunkt *deterministisch* ist.
 - Mit anderen Worten: Wir „starten“ die Markov-Ketten $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(\ell)}$ *nicht* zum „Zeitpunkt“ 0,
 - sondern hinreichend weit in der „Vergangenheit“, so dass spätestens bis zum „Zeitpunkt“ 0 sämtliche Pfade miteinander verschmolzen sind.

Um diese Vorgehensweise mathematisch präzise formulieren zu können, führen wir die folgenden Bezeichnungen ein.

- Für jeden potentiellen „Startzeitpunkt“ $m \in \{-1, -2, \dots\}$ und für jedes $i \in \{1, \dots, \ell\}$ sei

$$\mathbf{X}^{(m,i)} = (\mathbf{X}_m^{(m,i)}, \mathbf{X}_{m+1}^{(m,i)}, \dots)$$

- eine homogene Markov-Kette mit dem endlichen Zustandsraum $E = \{\mathbf{x}_1, \dots, \mathbf{x}_\ell\}$
- mit dem (deterministischen) Anfangszustand $\mathbf{X}_m^{(m,i)} = \mathbf{x}_i$ und mit der irreduziblen und aperiodischen Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$,
- so dass $\boldsymbol{\pi} = (\pi_{\mathbf{x}}, \mathbf{x} \in E)$ die ergodische (Grenz-) Verteilung der Markov-Kette $\mathbf{X}^{(m,i)}$ ist.
- Für jedes $k \in \{1, \dots, \ell\}$ betrachten wir
 - eine Folge $\mathbf{U}^{(k)} = (U_0^{(k)}, U_{-1}^{(k)}, \dots)$ von unabhängigen und $(0, 1]$ -gleichverteilten Zufallsvariablen,
 - wobei wir $U_{-n}^{(k)}$ (ähnlich wie in Abschnitt 3.5.1) *Innovation* im $(-n)$ -ten Schritt bei Vorliegen des Zustandes $\mathbf{x}_k \in E$ nennen.
- Dabei betrachten wir erneut die beiden Fälle,
 - dass die Innovationen-Folgen $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(\ell)}$ entweder unabhängig sind oder
 - dass $\mathbf{U}^{(1)} = \dots = \mathbf{U}^{(\ell)} = \mathbf{U}$ gilt.
- Die Markov-Kette $\mathbf{X}^{(m,i)}$ sei rekursiv durch eine gültige Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ gegeben, d.h., es gelte

$$\mathbf{X}_n^{(m,i)} = \varphi(\mathbf{x}_k, U_n^{(k)}), \quad \text{falls } \mathbf{X}_{n-1}^{(m,i)} = \mathbf{x}_k. \quad (90)$$

Definition Die Zufallsvariable $\zeta = \min\{-m \geq 1 : \mathbf{X}_0^{(m,1)} = \dots = \mathbf{X}_0^{(m,\ell)}\}$ heißt *CFTP-Kopplungszeit*, wobei $\zeta = \infty$ gesetzt wird, falls es keine natürliche Zahl $-m$ gibt mit $\mathbf{X}_0^{(m,1)} = \dots = \mathbf{X}_0^{(m,\ell)}$.

Theorem 3.23 Sei $P(\zeta < \infty) = 1$. Für jedes $m \leq -\zeta$ gilt dann

$$\mathbf{X}_0^{(m,1)} = \dots = \mathbf{X}_0^{(m,\ell)}.$$

Außerdem gilt für beliebige $m \leq -\zeta$ und $i, j \in \{1, \dots, \ell\}$, dass

$$\mathbf{X}_0^{(m,i)} = \mathbf{X}_0^{(-\zeta,j)} \sim \boldsymbol{\pi}.$$

Beweis

- Unmittelbar aus der rekursiven Definitionsgleichung (90) der Markov-Ketten $\mathbf{X}^{(m,1)}, \dots, \mathbf{X}^{(m,\ell)}$ ergibt sich, dass $\mathbf{X}_0^{(m,1)} = \dots = \mathbf{X}_0^{(m,\ell)}$ und $\mathbf{X}_0^{(m,i)} = \mathbf{X}_0^{(-\zeta,j)}$ für beliebige $m \leq -\zeta$ und $i, j \in \{1, \dots, \ell\}$.

- Weil $P(\zeta < \infty) = 1$ vorausgesetzt wird, gilt außerdem für jedes $k \in \{1, \dots, \ell\}$, dass

$$\begin{aligned}
P\left(\mathbf{X}_0^{(-\zeta, i)} = \mathbf{x}_k\right) &= \lim_{m \rightarrow -\infty} P\left(\mathbf{X}_0^{(-\zeta, i)} = \mathbf{x}_k, \zeta \leq -m\right) \\
&= \lim_{m \rightarrow -\infty} P\left(\mathbf{X}_0^{(m, i)} = \mathbf{x}_k, \zeta \leq -m\right) \\
&= \lim_{m \rightarrow -\infty} P\left(\mathbf{X}_0^{(m, i)} = \mathbf{x}_k\right) - \underbrace{\lim_{m \rightarrow -\infty} P\left(\mathbf{X}_0^{(m, i)} = \mathbf{x}_k, \zeta > -m\right)}_{=0} \\
&= \lim_{m \rightarrow -\infty} P\left(\mathbf{X}_0^{(m, i)} = \mathbf{x}_k\right) \\
&= \lim_{m \rightarrow -\infty} P\left(\mathbf{X}_{-m}^{(0, i)} = \mathbf{x}_k\right) = \pi_{\mathbf{x}_k},
\end{aligned}$$

wobei sich die vorletzte Gleichheit aus der Homogenität der Markov-Kette $\mathbf{X}^{(m, i)}$ ergibt. $\square$

Beachte

- Wenn die Anzahl ℓ der Elemente des Zustandsraumes $E = \{\mathbf{x}_1, \dots, \mathbf{x}_\ell\}$ groß ist,
 - dann kann die MCMC-Simulation von π auf der Grundlage des CFTP-Algorithmus von Propp und Wilson mit großem Rechenaufwand verbunden sein,
 - der zur Erzeugung sämtlicher „Pfade“ mit den Anfangszuständen $\mathbf{x}_1, \dots, \mathbf{x}_\ell$ erforderlich ist.
- Dieser Aufwand lässt sich jedoch in manchen Fällen verringern. Beispiele hierfür werden in den folgenden Abschnitten 3.5.3 und 3.5.4 diskutiert,
 - wobei der Zustandsraum $E = \{\mathbf{x}_1, \dots, \mathbf{x}_\ell\}$ und die Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ gewisse Monotonie-Eigenschaften besitzen
 - und dann lediglich *eine* Folge $\mathbf{U} = (U_0, U_{-1}, \dots)$ von unabhängigen und $(0, 1]$ -gleichverteilten Innovationen betrachtet wird
 - sowie nur *zwei* verschiedene Pfade generiert werden müssen.

3.5.3 Monotone Kopplungsalgorithmen

- Wir setzen nun zusätzlich voraus, dass auf dem Zustandsraum $E = \{\mathbf{x}_1, \dots, \mathbf{x}_\ell\}$ eine *Halbordnung* $\preceq$ definiert ist mit einem *Minimalelement* $\mathbf{0} \in E$ und einem *Maximalelement* $\mathbf{1} \in E$, d.h., es gelte

- (a) $\mathbf{x} \preceq \mathbf{x}, \quad \forall \mathbf{x} \in E,$
- (b) aus $\mathbf{x} \preceq \mathbf{y}$ und $\mathbf{y} \preceq \mathbf{z}$ folgt $\mathbf{x} \preceq \mathbf{z}, \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in E,$
- (c) aus $\mathbf{x} \preceq \mathbf{y}$ und $\mathbf{y} \preceq \mathbf{x}$ folgt $\mathbf{x} = \mathbf{y}, \quad \forall \mathbf{x}, \mathbf{y} \in E,$
- (d) $\mathbf{0} \preceq \mathbf{x} \preceq \mathbf{1}, \quad \forall \mathbf{x} \in E.$

- Außerdem setzen wir voraus,

- dass die Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ *isoton* bezüglich der Halbordnung $\preceq$ ist, d.h., für beliebige $\mathbf{x}, \mathbf{y} \in E$ mit $\mathbf{x} \preceq \mathbf{y}$ gelte

$$\varphi(\mathbf{x}, u) \preceq \varphi(\mathbf{y}, u), \quad \forall u \in (0, 1]. \quad (91)$$

- Die Innovationen $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(\ell)}$ seien mit Wahrscheinlichkeit 1 identisch,

- d.h., wir betrachten lediglich *eine* Folge $\mathbf{U} = (U_0, U_{-1}, \dots)$ von unabhängigen und $(0, 1]$ -gleichverteilten Zufallsvariablen und setzen $\mathbf{U}^{(1)} = \dots = \mathbf{U}^{(\ell)} = \mathbf{U},$

- wobei die Markov-Kette $\mathbf{X}^{(m,i)}$ für beliebige $m \in \{-1, -2, \dots\}$ und $i \in \{1, \dots, \ell\}$ rekursiv gegeben sei durch

$$\mathbf{X}_n^{(m,i)} = \varphi(\mathbf{X}_{n-1}^{(m,i)}, U_n), \quad \forall n = m+1, m+2, \dots \quad (92)$$

Beachte

- Wenn $\mathbf{x}_i \preceq \mathbf{x}_j$, dann ergibt sich aus (91) und (92), dass für jedes $n \geq m$

$$\mathbf{X}_n^{(m,i)} \preceq \mathbf{X}_n^{(m,j)}. \quad (93)$$

- Insbesondere gilt für beliebige $n \geq m$ und $i \in \{1, \dots, \ell\}$, dass

$$\mathbf{X}_n^{(m,\min)} \preceq \mathbf{X}_n^{(m,i)} \preceq \mathbf{X}_n^{(m,\max)}. \quad (94)$$

- Dabei seien $\mathbf{X}^{(m,\min)}$ und $\mathbf{X}^{(m,\max)}$ die durch (92) rekursiv definierten Markov-Ketten

$$\mathbf{X}^{(m,\min)} = (\mathbf{X}_m^{(m,\min)}, \mathbf{X}_{m+1}^{(m,\min)}, \dots) \quad \text{und} \quad \mathbf{X}^{(m,\max)} = (\mathbf{X}_m^{(m,\max)}, \mathbf{X}_{m+1}^{(m,\max)}, \dots)$$

- mit $\mathbf{X}_m^{(m,\min)} = \mathbf{0}$ bzw. $\mathbf{X}_m^{(m,\max)} = \mathbf{1}$.
- Wegen (94) genügt es, den Startpunkt der Simulation so weit in die Vergangenheit zu legen,
 - dass die Pfade der *Minorante* $\mathbf{X}^{(m,\min)}$ und der *Majorante* $\mathbf{X}^{(m,\max)}$ bis zum Zeitpunkt 0 miteinander verschmelzen,
 - d.h., wir betrachten die *CFTP-Kopplungszeit*

$$\zeta = \min\{-m \geq 1 : \mathbf{X}_0^{(m,\min)} = \mathbf{X}_0^{(m,\max)}\}. \quad (95)$$

Theorem 3.24 Die Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ genüge der Isotonie-Bedingung (91).

- Für die in (95) definierte CFTP-Kopplungszeit gilt dann $\zeta < \infty$ mit Wahrscheinlichkeit 1.
- Außerdem gilt für beliebige $m \leq -\zeta$ und $i, j \in \{1, \dots, \ell\}$, dass $\mathbf{X}_0^{(m,i)} = \mathbf{X}_0^{(-\zeta,j)} \sim \pi$.

Beweis

- Der Nachweis, dass $\mathbf{X}_0^{(m,i)} = \mathbf{X}_0^{(-\zeta,j)} \sim \pi$ für beliebige $m \leq -\zeta$ und $i, j \in \{1, \dots, \ell\}$ gilt, falls $P(\zeta < \infty) = 1$, verläuft ähnlich wie im Beweis von Theorem 3.23 und wird deshalb hier weggelassen.
- Wir zeigen lediglich, dass $P(\zeta < \infty) = 1$ gilt.
 - Dabei nutzen wir die Tatsache, dass für jedes $r \geq 1$

$$\{\zeta > r\} \subset \left\{ \mathbf{X}_{-r+1}^{(-r,\min)} \neq \mathbf{1}, \dots, \mathbf{X}_0^{(-r,\min)} \neq \mathbf{1} \right\}, \quad (96)$$

denn aus (94) ergibt sich, dass

$$\begin{aligned} \{\zeta > r\} &= \left\{ \mathbf{X}_0^{(-r,\min)} \neq \mathbf{X}_0^{(-r,\max)} \right\} \\ &= \left\{ \mathbf{X}_{-r+1}^{(-r,\min)} \neq \mathbf{X}_{-r+1}^{(-r,\max)}, \dots, \mathbf{X}_0^{(-r,\min)} \neq \mathbf{X}_0^{(-r,\max)} \right\} \\ &\stackrel{(94)}{\subset} \left\{ \mathbf{X}_{-r+1}^{(-r,\min)} \neq \mathbf{1}, \dots, \mathbf{X}_0^{(-r,\min)} \neq \mathbf{1} \right\}. \end{aligned}$$

- So wie im Beweis von Theorem 3.21 sei nun $n_0 \geq 1$ eine natürliche Zahl mit

$$\min_{\mathbf{x}, \mathbf{x}' \in E} p_{\mathbf{x}\mathbf{x}'}^{(n_0)} = c > 0, \quad (97)$$

und sei $r = m(r)n_0 + k$ für gewisse Zahlen $m(r) = 0, 1, \dots$ und $k = 0, 1, \dots, n_0 - 1$.

– Dann ergibt sich aus (96) und (97), dass

$$\begin{aligned}
 P(\zeta = \infty) &= \lim_{r \rightarrow \infty} P(\zeta > r) \\
 &\stackrel{(96)}{\leq} \lim_{r \rightarrow \infty} P(\mathbf{X}_{-r+1}^{(-r, \min)} \neq \mathbf{1}, \dots, \mathbf{X}_0^{(-r, \min)} \neq \mathbf{1}) \\
 &\leq \lim_{r \rightarrow \infty} \sum_{\mathbf{x}_1, \dots, \mathbf{x}_{m(r)} \neq \mathbf{1}} p_{\mathbf{0}\mathbf{x}_1}^{(n_0)} p_{\mathbf{x}_1\mathbf{x}_2}^{(n_0)} \cdots p_{\mathbf{x}_{m(r)-1}\mathbf{x}_{m(r)}}^{(n_0)} \\
 &\stackrel{(97)}{\leq} \lim_{r \rightarrow \infty} (1 - c)^{m(r)} = 0.
 \end{aligned}$$

□

Beachte

- Manchmal ist die Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ nicht isoton, sondern *antiton* bezüglich der Halbordnung $\preceq$, d.h., für beliebige $\mathbf{x}, \mathbf{y} \in E$ mit $\mathbf{x} \preceq \mathbf{y}$ gilt

$$\varphi(\mathbf{x}, u) \succeq \varphi(\mathbf{y}, u), \quad \forall u \in (0, 1]. \quad (98)$$

- In diesem Fall ist die folgende *Cross-Over-Technik* nützlich.
 - Dabei konstruieren wir mit Hilfe der Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ eine neue (isotone) Update-Funktion $\varphi' : E \times (0, 1]^2 \rightarrow E$, die gegeben ist durch

$$\varphi'(\mathbf{x}; u_1, u_2) = \varphi(\varphi(\mathbf{x}, u_1), u_2), \quad \forall \mathbf{x} \in E; u_1, u_2 \in (0, 1]. \quad (99)$$

- Aus (98) und (99) ergibt sich nämlich, dass für beliebige $\mathbf{x}, \mathbf{y} \in E$ mit $\mathbf{x} \preceq \mathbf{y}$

$$\varphi'(\mathbf{x}; u_1, u_2) = \varphi(\varphi(\mathbf{x}, u_1), u_2) \preceq \varphi(\varphi(\mathbf{y}, u_1), u_2) = \varphi'(\mathbf{y}; u_1, u_2), \quad \forall u_1, u_2 \in (0, 1],$$

d.h., $\varphi' : E \times (0, 1]^2 \rightarrow E$ ist isoton, falls $\varphi : E \times (0, 1] \rightarrow E$ antiton ist.

- Sei nun $\varphi : E \times (0, 1] \rightarrow E$ eine gültige Update-Funktion bezüglich der irreduziblen und aperiodischen Übergangsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ mit der ergodischen (Grenz-) Verteilung $\boldsymbol{\pi} = (\pi_{\mathbf{x}}, \mathbf{x} \in E)$.
 - Dann ist die in (99) definierte Abbildung $\varphi' : E \times (0, 1]^2 \rightarrow E$ eine gültige Update-Funktion bezüglich der (irreduziblen und aperiodischen) Zwei-Schritt-Übergangsmatrix $\mathbf{P}^{(2)} = (p_{\mathbf{x}\mathbf{x}'}^{(2)})$ mit der gleichen ergodischen (Grenz-) Verteilung $\boldsymbol{\pi} = (\pi_{\mathbf{x}}, \mathbf{x} \in E)$,
 - und für die Kopplungszeit $\zeta' = \min\{-m \geq 1 : \mathbf{X}_0^{(2m, \min)} = \mathbf{X}_0^{(2m, \max)}\}$ ergibt sich auf die gleiche Weise wie im Beweis von Theorem 3.24, dass $\zeta' < \infty$ mit Wahrscheinlichkeit 1 und $\mathbf{X}_0^{(-2\zeta', i)} \sim \boldsymbol{\pi}$ für jedes $i \in \{1, \dots, \ell\}$, falls $\varphi : E \times (0, 1] \rightarrow E$ antiton ist.

3.5.4 Beispiele: Geburts- und Todesprozesse; Ising-Modell

1. Geburts- und Todesprozesse

- Die in (88) definierte Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ genügt insbesondere dann der Isotonie-Bedingung (91),
 - wenn der Zustandsraum mit der Menge $E = \{1, \dots, \ell\}$ identifiziert wird und dabei die natürliche Ordnung $\leq$ der Zahlen $1, \dots, \ell$ betrachtet wird
 - und wenn die Simulationsmatrix $\mathbf{P} = (p_{ij})$ *isoton* bezüglich der Ordnung $\leq$ ist, d.h., für beliebige $i, j \in E$ mit $i \leq j$ gelte

$$\sum_{r=k}^{\ell} p_{ir} \leq \sum_{r=k}^{\ell} p_{jr}, \quad \forall k = 1, \dots, \ell. \quad (100)$$

- Eine Klasse von Übergangsmatrizen $\mathbf{P} = (p_{ij})$, für die die Isotonie-Bedingung (100) erfüllt ist, ist durch die tridiagonalen Matrizen von *Geburts- und Todesprozessen* gegeben, so dass

$$\mathbf{P} = \begin{pmatrix} 1 - p_{12} & p_{12} & 0 & \dots & 0 \\ p_{21} & 1 - p_{21} - p_{23} & p_{23} & \dots & 0 \\ 0 & p_{32} & 1 - p_{32} - p_{34} & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & p_{\ell-1,\ell} \\ 0 & 0 & 0 & \dots & 1 - p_{\ell,\ell-1} \end{pmatrix},$$

wobei $0 < p_{i,i+1} \leq 1/2$ für jedes $i = 1, \dots, \ell - 1$ und $0 < p_{i,i-1} \leq 1/2$ für jedes $i = 2, \dots, \ell$.

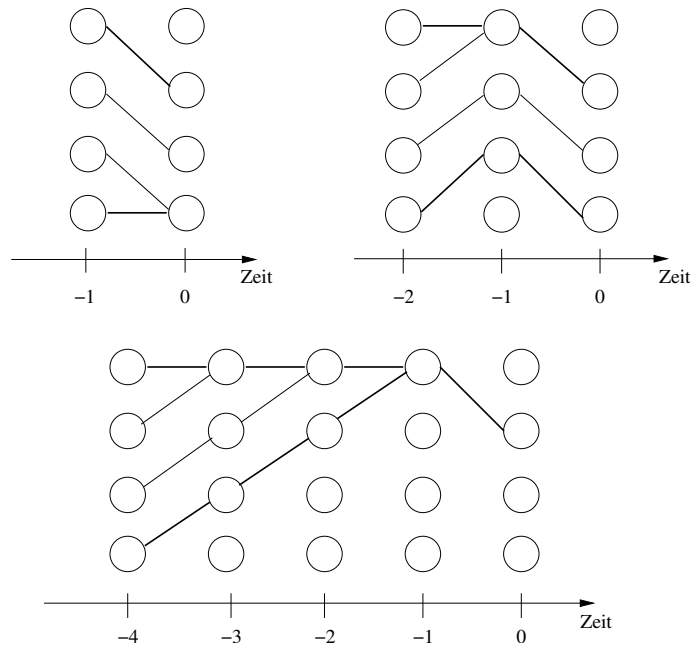


Abbildung 5: Monotone Rückwärtskopplung bei isotonen Geburts- und Todesprozessen

- Andererseits genügt die in (88) definierte Update-Funktion $\varphi : E \times (0, 1] \rightarrow E$ der Antitonie-Bedingung (98),
 - wenn $\mathbf{P} = (p_{ij})$ *antiton* bezüglich der Ordnung $\leq$ ist,
 - d.h., wenn für beliebige $i, j \in E$ mit $i \leq j$

$$\sum_{r=k}^{\ell} p_{ir} \geq \sum_{r=k}^{\ell} p_{jr}, \quad \forall k = 1, \dots, \ell. \quad (101)$$

- Man kann leicht zeigen, dass es keine tridiagonale Übergangsmatrix $\mathbf{P} = (p_{ij})$ gibt, so dass die Bedingung (101) erfüllt ist, d.h. Geburts- und Todesprozessen sind niemals antiton.

- Die Antitonia-Bedingung (101) ist jedoch beispielsweise für die folgende Matrix erfüllt:

$$\mathbf{P} = \begin{pmatrix} 0 & \dots & 0 & 0 & 1 \\ 0 & \dots & 0 & 1/2 & 1/2 \\ 0 & \dots & 1/3 & 1/3 & 1/3 \\ \vdots & & \vdots & \vdots & \vdots \\ 0 & \dots & 1/(\ell-1) & 1/(\ell-1) & 1/(\ell-1) \\ 1/\ell & \dots & 1/\ell & 1/\ell & 1/\ell \end{pmatrix}$$

2. Ising-Modell

- Ähnlich wie bei dem in Abschnitt 3.3.1 diskutierten Hard-Core-Modell
 - betrachten wir einen verbundenen Graph $G = (V, K)$ mit der Menge $V = \{v_1, \dots, v_{|V|}\}$ von (endlich vielen) Eckpunkten
 - und mit einer gewissen Menge $K \subset V^2$ von Kanten $e = (v_i, v_j)$, die jeweils zwei Eckpunkte v_i, v_j miteinander verbinden.
- Jedem Eckpunkt aus V wird entweder der Wert -1 oder 1 zugeordnet,
 - wobei der Zustandsraum $E = \{-1, 1\}^{|V|}$ aller Konfigurationen $\mathbf{x} = (x(v), v \in V)$ betrachtet wird, d.h., es gilt entweder $x(v) = -1$ oder $x(v) = 1$ für jedes $v \in V$
 - mit der „Bild-Interpretation“, dass $x(v) = -1$ ein weißes Pixel und $x(v) = 1$ ein schwarzes Pixel bedeutet.
- Für jedes $\mathbf{x} \in E$ sei die Wahrscheinlichkeit $\pi_{\mathbf{x}}$ der Konfiguration $\mathbf{x}$ gegeben durch den Ansatz (vgl. auch Übungsaufgabe 11.1)

$$\pi_{\mathbf{x}} = \frac{1}{z_{G,J}} \exp \left(J \sum_{e=(v_i, v_j) \in K} x(v_i) x(v_j) \right) \quad (102)$$

für einen gewissen Parameter $J \geq 0$, der in der Physik als „inverse Temperatur“ interpretiert wird:

- Für $J = 0$ (unendliche Temperatur) ist die in (102) gegebene Verteilung $\pi = (\pi_{\mathbf{x}}, \mathbf{x} \in E)$ die (diskrete) Gleichverteilung auf E .
- Für $J \gg 0$ (niedrige Temperatur) besitzen diejenigen Konfigurationen große Wahrscheinlichkeiten, für die nur wenige benachbarte (d.h. durch Kanten verbundene) Paare von Eckpunkten unterschiedlich gefärbt sind.
- Für $J \rightarrow \infty$ (Null-Temperatur) konvergiert die in (102) gegebene Verteilung $\pi = (\pi_{\mathbf{x}}, \mathbf{x} \in E)$ gegen die „Zwei-Punkt-Gleichverteilung“ $(\delta_{\mathbf{0}} + \delta_{\mathbf{1}})/2$,
- wobei $\mathbf{0}$ und $\mathbf{1}$ die beiden (extremen) Konfigurationen sind, die entweder nur aus weißen oder nur aus schwarzen Pixeln bestehen, d.h. $0(v) = -1$ bzw. $1(v) = 1$ für jedes $v \in V$.
- Dabei ist $z_{G,J} > 0$ eine (im allgemeinen unbekannte) Normierungskonstante mit

$$z_{G,J} = \sum_{\mathbf{x} \in E} \exp \left(-J \sum_{e=(v_i, v_j) \in K} x(v_i) x(v_j) \right).$$

- Die folgenden Abbildungen wurde dem Buch von O. Häggström (2002) *Finite Markov Chains and Algorithmic Applications*, CU Press, Cambridge entnommen.

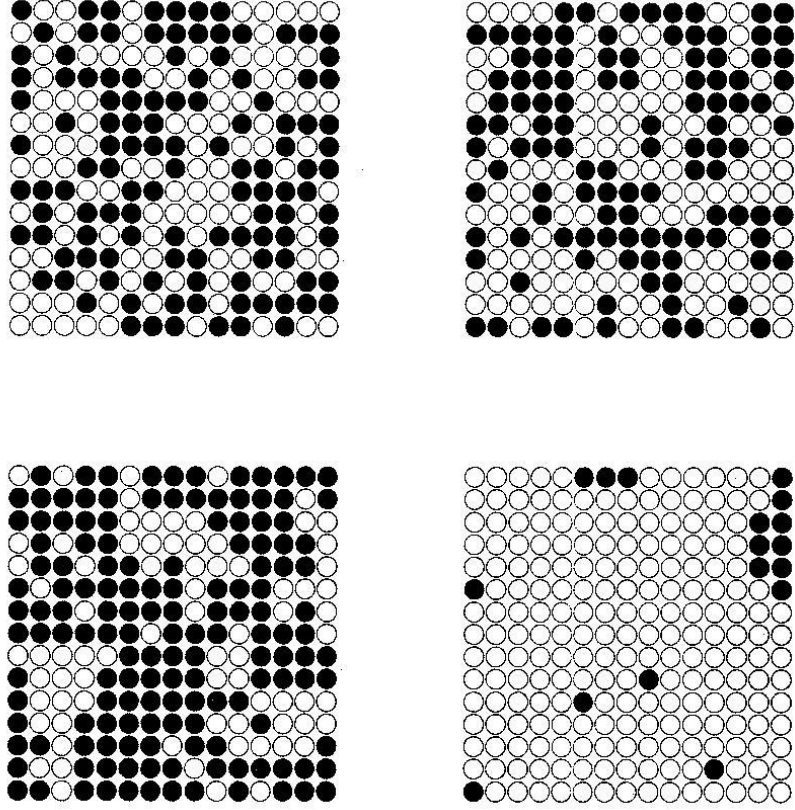


Abbildung 6: Typische Konfigurationen des Ising-Modells für $J = 0$ (oben links), $J = 0.15$ (oben rechts), $J = 0.3$ (unten links) bzw. $J = 0.5$ (unten rechts)

- Sie verdeutlichen die oben erwähnte Rolle des Parameters J ,
 - d.h., mit wachsendem J werden die zusammenhängenden „Klumpen“ von jeweils identisch gefärbten Pixeln größer.
- Die Simulationsmatrix $\mathbf{P} = (p_{\mathbf{x}\mathbf{x}'})$ sei durch den Gibbs-Sampler gegeben. Es gelte also (36), d.h.

$$p_{\mathbf{x}\mathbf{x}'} = \sum_{v \in V} q_v \pi_{x'(v) | \mathbf{x}(-v)} \mathbb{I}(\mathbf{x}(-v) = \mathbf{x}'(-v)), \quad \forall \mathbf{x}, \mathbf{x}' \in E.$$

- Dabei gilt für beliebige $\mathbf{x}, \mathbf{x}' \in E$ mit $\mathbf{x}(-v) = \mathbf{x}'(-v)$

$$\pi_{x'(v) | \mathbf{x}(-v)} = \begin{cases} \frac{\pi_{\mathbf{x}_+}}{\pi_{\mathbf{x}_+} + \pi_{\mathbf{x}_-}}, & \text{falls } x'(v) = 1, \\ \frac{\pi_{\mathbf{x}_-}}{\pi_{\mathbf{x}_+} + \pi_{\mathbf{x}_-}}, & \text{falls } x'(v) = -1, \end{cases}$$

wobei $x_-(v) = -1$ und $\mathbf{x}_-(-v) = \mathbf{x}(-v)$ bzw. $x_+(v) = 1$ und $\mathbf{x}_+(-v) = \mathbf{x}(-v)$.

- Aus (102) ergibt sich somit für $x'(v) = 1$

$$\pi_{x'(v) | \mathbf{x}(-v)} = \frac{\exp\left(J(k_+(\mathbf{x}(-v)) - k_-(\mathbf{x}(-v)))\right)}{\exp\left(J(k_-(\mathbf{x}(-v)) - k_+(\mathbf{x}(-v)))\right) + \exp\left(J(k_+(\mathbf{x}(-v)) - k_-(\mathbf{x}(-v)))\right)}$$

und analog für $x'(v) = -1$

$$\pi_{x'(v)|\mathbf{x}(-v)} = \frac{\exp\left(J(k_-(\mathbf{x}(-v)) - k_+(\mathbf{x}(-v)))\right)}{\exp\left(J(k_+(\mathbf{x}(-v)) - k_-(\mathbf{x}(-v)))\right) + \exp\left(J(k_-(\mathbf{x}(-v)) - k_+(\mathbf{x}(-v)))\right)}$$

bzw.

$$\pi_{x'(v)|\mathbf{x}(-v)} = \begin{cases} \frac{1}{1 + \exp\left(-2J(k_+(\mathbf{x}(-v)) - k_-(\mathbf{x}(-v)))\right)}, & \text{falls } x'(v) = 1, \\ \frac{1}{1 + \exp\left(-2J(k_-(\mathbf{x}(-v)) - k_+(\mathbf{x}(-v)))\right)}, & \text{falls } x'(v) = -1, \end{cases} \quad (103)$$

wobei $k_+(\mathbf{x}(-v))$ bzw. $k_-(\mathbf{x}(-v))$ die Anzahl derjenigen mit v verbundenen Eckpunkte ist, die auf 1 bzw. -1 gesetzt sind.

- Für den Zustandsraum $E = \{-1, 1\}^{|V|}$ definieren wir nun die Halbordnungsrelation $\preceq$
 - mit $\mathbf{x} \preceq \mathbf{y}$, falls $x(v) \leq y(v)$ für jedes $v \in V$, so dass $\mathbf{0} \preceq \mathbf{x} \preceq \mathbf{1}$ für jedes $\mathbf{x} \in E$,
 - wobei die Elemente des Zustandsraumes $E = \{\mathbf{x}_1, \dots, \mathbf{x}_\ell\}$ so numeriert sein mögen, dass $i \leq j$ gilt, falls $\mathbf{x}_i \preceq \mathbf{x}_j$ (was beispielsweise bei der *lexikographischen Nummerung* der Elemente von E zutreffend ist).
- Aus (103) ergibt sich dann, dass für beliebige $\mathbf{x}, \mathbf{y} \in E$ mit $\mathbf{x} \preceq \mathbf{y}$

$$\pi_{1|\mathbf{x}(-v)} \leq \pi_{1|\mathbf{y}(-v)} \quad \text{bzw.} \quad \pi_{-1|\mathbf{x}(-v)} \geq \pi_{-1|\mathbf{y}(-v)}, \quad (104)$$

weil $1/(1 + e^{-a}) \leq 1/(1 + e^{-b})$ für beliebige Zahlen $a, b \in \mathbb{R}$ mit $a \leq b$.

- Die Update-Funktion $\varphi : E \times (0, 1]^2 \rightarrow E$ sei gegeben durch $\varphi(\mathbf{x}; u_1, u_2) = \mathbf{x}'$, wobei $\mathbf{x}' = (\mathbf{x}'(v), v \in V)$ und für jedes $i = 1, \dots, |V|$

$$\mathbf{x}'(v_i) = \begin{cases} 1, & \text{falls } \sum_{j=1}^{i-1} q(v_j) < u_1 \leq \sum_{j=1}^i q(v_j) \text{ und } u_2 < \pi_{1|\mathbf{x}(-v_i)}, \\ -1, & \text{falls } \sum_{j=1}^{i-1} q(v_j) < u_1 \leq \sum_{j=1}^i q(v_j) \text{ und } u_2 \geq \pi_{1|\mathbf{x}(-v_i)}, \\ \mathbf{x}(v_i), & \text{sonst.} \end{cases}$$

- Für diese Update-Funktion $\varphi : E \times (0, 1]^2 \rightarrow E$ ergibt sich aus (104), dass für beliebige $\mathbf{x}, \mathbf{y} \in E$ mit $\mathbf{x} \preceq \mathbf{y}$

$$\varphi(\mathbf{x}; u_1, u_2) \preceq \varphi(\mathbf{y}; u_1, u_2), \quad \forall u_1, u_2 \in (0, 1].$$

- D.h., die (91) entsprechende Isotonie-Bedingung ist bezüglich der Halbordnung $\preceq$ erfüllt.

3.5.5 Read-Once-Modifikation des CFTP-Algorithmus

- Eine Problematik des „monotonen“ CFTP-Algorithmus, der in den Abschnitten 3.5.3 und 3.5.4 diskutiert worden ist, besteht darin, dass
 - *sämtliche* Innovationen $U_0, U_{-1}, \dots, U_{-\zeta}$ gespeichert werden müssen, wobei ζ die in (95) definierte „monotone“ CFTP-Kopplungszeit ist mit

$$\zeta = \min\{-m \geq 1 : \mathbf{X}_0^{(m, \min)} = \mathbf{X}_0^{(m, \max)}\}.$$

- Deshalb schlug David Wilson im Jahre 2000 die folgende Modifikation des CFTP-Algorithmus mit dem Ziel vor, den erforderlichen Speicherbedarf weiter zu verringern.

- Die Hauptidee der Modifikation besteht darin, die in den Abschnitten 3.5.2 – 3.5.4 betrachtete Rückwärtskopplung
 - mit Hilfe einer Folge von unabhängigen und identisch verteilten „Vorwärts-Simulationsblöcken“ darzustellen, wobei
 - die (potentiellen) „Startzeitpunkte“ $m \in \{-1, -2, \dots\}$ der Markov-Kette $\mathbf{X}^{(m,i)} = (\mathbf{X}_m^{(m,i)}, \mathbf{X}_{m+1}^{(m,i)}, \dots)$ auch *zufällig* gewählt werden können.
- Dabei seien die Innovationen-Folgen $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(\ell)}$ mit Wahrscheinlichkeit 1 identisch,
 - d.h., es wird lediglich *eine* Folge $\mathbf{U} = (\dots, U_{-1}, U_0, U_1, \dots)$ von unabhängigen und gleichverteilten Zufallsvariablen betrachtet und $\mathbf{U}^{(1)} = \dots = \mathbf{U}^{(\ell)} = \mathbf{U}$ gesetzt.
 - Außerdem wird vorausgesetzt, dass die in (87) und (90) definierten Markov-Ketten $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(\ell)}$ bzw. $\mathbf{X}^{(m,1)}, \dots, \mathbf{X}^{(m,\ell)}$ so beschaffen sind,
 - dass die (Vorwärts- bzw. Rückwärts-) Kopplungszeiten

$$\tau = \min\{n \geq 1 : \mathbf{X}_n^{(1)} = \dots = \mathbf{X}_n^{(\ell)}\} \quad \text{bzw.} \quad \zeta = \min\{-m \geq 1 : \mathbf{X}_0^{(m,1)} = \dots = \mathbf{X}_0^{(m,\ell)}\}$$

jeweils mit Wahrscheinlichkeit 1 endlich sind.

- Wir betrachten nun Vorwärts-Simulationsblöcke mit der (zunächst deterministischen) Länge T für ein $T \geq 1$:
 - Für beliebige $k \geq 0$ bzw. $i = 1, \dots, \ell$ sei $\mathbf{X}_{kT}^{(kT,i)} = \mathbf{x}_i$ und

$$\mathbf{X}_n^{(kT,i)} = \varphi(\mathbf{X}_{n-1}^{(kT,i)}, U_n), \quad \forall n = kT + 1, kT + 2, \dots$$

- Außerdem betrachten wir für jedes $k \geq 0$ das Ereignis

$$C_{kT} = \left\{ \mathbf{X}_{(k+1)T}^{(kT,i)} = \mathbf{X}_{(k+1)T}^{(kT,j)}, \quad \forall i \neq j \in \{1, \dots, \ell\} \right\},$$

wobei die Blocklänge T so gewählt sei, dass

$$0 < P(C_T) \quad \left(= P(C_{kT}), \quad \forall k \geq 0 \right). \quad (105)$$

- Bei Start mit $k = 0$ ist die Read-Once-Modifikation des CFTP-Algorithmus dann gegeben durch:
 1. Simuliere $\mathbf{X}_n^{(kT,i)}$ über φ und $\mathbf{U}$ für $n = kT + 1, \dots, (k+1)T$.
 2. Setze $m = k$ und $k = k + 1$. Falls das Ereignis C_{mT} eingetreten ist, fahre fort in Schritt 3, anderenfalls kehre zu Schritt 1 zurück.
 3. Wiederhole die Schritte 1 und 2 bis das Ereignis $C_{m'T}$ für ein $m' > m$ eintritt und gib den Wert von $\mathbf{X}_{m'T}^{(mT,i)}$ für ein beliebiges $i \in \{1, \dots, \ell\}$ als Realisierung gemäß $\boldsymbol{\pi}$ aus.

Beispiel

- Für $\ell = 3$ Zustände betrachten wir die (irreduzible und aperiodische) Übergangsmatrix

$$\mathbf{P} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 \end{pmatrix}$$

- Für die Blocklänge $T = 2$ und für die $(0, 1]$ -gleichverteilten Pseudozufallszahlen

$$\mathbf{u} = (0.01, 0.60, 0.82, 0.47, 0.36, 0.59, 0.34, 0.89, \dots)$$

ergibt sich dabei der folgende Simulationslauf:

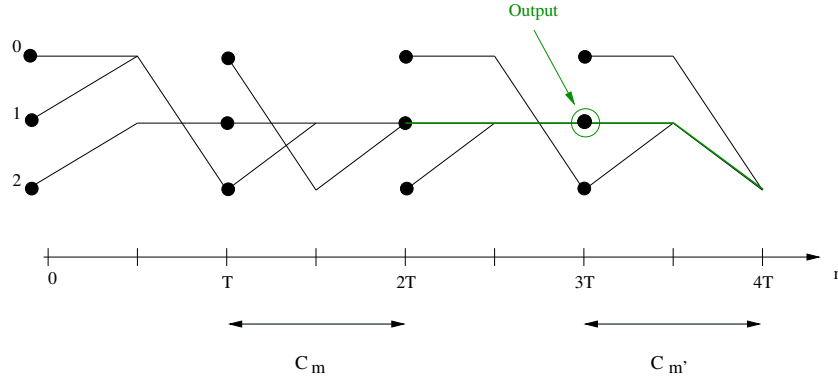


Abbildung 7: Read-Once-Algorithmus

Beachte

- Weil die Simulationsblöcke und damit auch die Ereignisse $C_T, C_{2T}, \dots$ unabhängig sind und weil $P(C_T) = P(C_{2T}) = \dots$ gilt,
 - liefern die ersten m' Vorwärts-Simulationsblöcke des oben beschriebenen Algorithmus, wenn sie in umgekehrter Reihenfolge betrachtet werden, insbesondere das in Abschnitt 3.5.2 diskutierte Coupling-from-the-Past.
 - Hieraus ergibt sich, dass $\mathbf{X}_{m'T}^{(mT,i)} \sim \pi$ für jedes $i \in \{1, \dots, l\}$.
 - Der letzte, d.h. der $(m' + 1)$ -te Vorwärts-Simulationsblock dient lediglich zur Definition einer Stopprege.

- Die Read-Once-Modifikation des CFTP-Algorithmus terminiert mit Wahrscheinlichkeit 1,
 - wenn die Bedingung (105) erfüllt ist, d.h., wenn $P(C_T) > 0$.
 - Für isotone Update-Funktionen gilt dies, falls $T \geq n_0$, wobei $n_0 \geq 1$ eine natürliche Zahl mit

$$\min_{\mathbf{x}, \mathbf{x}' \in E} p_{\mathbf{x}\mathbf{x}'}^{(n_0)} = c > 0,$$

vgl. den Beweis von Theorem 3.24.

- Wenn T eine Zufallsvariable ist,
 - die so wie die Vorwärtskopplungszeit τ verteilt ist und unabhängig von der Innovationen-Folge $\mathbf{U} = (U_0, U_{-1}, \dots)$ ist,
 - dann ergibt sich aus den folgenden elementaren, jedoch nützlichen Eigenschaften der Kopplungszeiten τ und ζ , dass $P(C_T) \geq 1/2$.

Theorem 3.25 *Es gilt $\tau \stackrel{d}{=} \zeta$. Wenn die Kopplungszeiten τ und ζ darüber hinaus unabhängig sind, dann gilt*

$$P(\zeta \leq \tau) \geq \frac{1}{2}. \quad (106)$$

Beweis

- Aus der Homogenität der Markov-Ketten $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(\ell)}$ bzw. $\mathbf{X}^{(m,1)}, \dots, \mathbf{X}^{(m,\ell)}$ ergibt sich für jede natürliche Zahl $k \geq 1$, dass

$$\begin{aligned} P(\tau = k) &= P\left(\min\{n \geq 1 : \mathbf{X}_n^{(1)} = \dots = \mathbf{X}_n^{(\ell)}\} = k\right) \\ &= P\left(\min\{-m \geq 1 : \mathbf{X}_0^{(m,1)} = \dots = \mathbf{X}_0^{(m,\ell)}\} = k\right) \\ &= P(\zeta = k). \end{aligned}$$

- Die Kopplungszeiten τ und ζ seien nun unabhängig und mit Wahrscheinlichkeit 1 endlich.
 - Hieraus ergibt sich, dass

$$\begin{aligned}
 P(\zeta \leq \tau) &= \sum_{k=1}^{\infty} P(\zeta \leq \tau \mid \tau = k) P(\tau = k) = \sum_{k=1}^{\infty} P(\zeta \leq k \mid \tau = k) P(\tau = k) \\
 &= \sum_{k=1}^{\infty} P(\zeta \leq k) P(\tau = k) = \sum_{k=1}^{\infty} P(\tau \leq k) P(\zeta = k) \\
 &\quad \vdots \\
 &= P(\zeta \geq \tau),
 \end{aligned}$$

- wobei sich die vorletzte Gleichheit aus der identischen Verteiltheit $\tau \stackrel{d}{=} \zeta$ von τ und ζ ergibt, die im ersten Teil des Beweises gezeigt wurde.
- Somit gilt

$$\begin{aligned}
 2 P(\zeta \leq \tau) &= P(\zeta \leq \tau) + P(\zeta \geq \tau) \\
 &= 1 - P(\zeta > \tau) + 1 - P(\zeta < \tau) \\
 &= 2 - P(\zeta \neq \tau) \\
 &\geq 2 - 1 = 1.
 \end{aligned}$$

□