

Studiengang Diplom-Mathematik

# **DIPLOMARBEIT**

## **Die Markov Ketten Monte Carlo Methode zum Testen stochastischer Unabhängigkeit**

eingereicht von  
Christina Gunkel  
am 3. Juni 2008

Erste Gutachterin: Prof. Dr. Christine Müller  
Zweiter Gutachter: Prof. Dr. Rolf Biehler

# Inhaltsverzeichnis

<b>1</b>	<b>Markov Ketten</b>	<b>3</b>
1.1	Definition und Eigenschaften . . . . .	3
1.2	Computer Simulation von Markov Ketten . . . . .	15
1.2.1	Beispiele . . . . .	18
1.3	Starke Markov Eigenschaft . . . . .	22
1.4	Invariante Wahrscheinlichkeitsverteilung und das Grenzwerttheorem . . . . .	23
1.5	Rekurrenz und Transienz . . . . .	35
1.6	Der Ergodensatz . . . . .	46
<b>2</b>	<b>Markov Ketten Monte Carlo Methode</b>	<b>50</b>
2.1	Hastings Algorithmus . . . . .	50
2.2	Metropolis Algorithmus . . . . .	53
2.2.1	Komponentenweiser Metropolis-Hastings Algorithmus . . . . .	55
2.2.2	Gibbs Sampler . . . . .	56
<b>3</b>	<b>Unabhängigkeitstest für 2-dimensionale Kontingenztafeln</b>	<b>57</b>
3.1	Bezeichnungen . . . . .	57
3.2	Stochastische Unabhängigkeit und der $\chi^2$ -Test . . . . .	58
3.2.1	$\chi^2$ -Unabhängigkeitstest . . . . .	59
3.3	Die Exponentialfamilie . . . . .	59
3.4	Die Hypergeometrische Verteilung . . . . .	61
3.5	Bedingter p-Wert . . . . .	64
3.6	Kontingenztafeln mit gegebenen Zeilen- und Spaltensummen . . . . .	66
3.7	Modifizierter Metropolis Algorithmus . . . . .	70
<b>4</b>	<b>Anwendungsbeispiele</b>	<b>74</b>
4.1	Automarkt . . . . .	74
4.2	Scheidungsrecht . . . . .	76
4.3	Tod durch Pferdetritte . . . . .	78
4.3.1	Ablehnung der Nullhypothese . . . . .	80
4.3.2	p-Wert im Bereich des Signifikanzniveaus . . . . .	82
4.4	Ähnlichkeit zweier Gedichte . . . . .	83
4.4.1	Annahme der Nullhypothese . . . . .	85
4.4.2	p-Wert im Bereich des Signifikanzniveaus . . . . .	87

4.5 Zusammenfassung der Ergebnisse . . . . .	88
<b>A Programme</b>	<b>96</b>
A.1 Computer Simulation von Markov Ketten . . . . .	96
A.2 Modifizierter Metropolis Algorithmus . . . . .	98

# Einleitung

Der  $\chi^2$ -Unabhängigkeitstest ist ein häufig angewandter Test. Da dieser jedoch eine befriedigende Approximation der  $\chi^2$ -Verteilung lediglich für Kontingenztafeln mit großen Einträgen, d.h. mit den erwarteten Zelhäufigkeiten  $> 5$ , aufzeigt, ist es notwendig eine Alternative für Kontingenztafeln mit kleinen Einträgen zu finden.

Ziel dieser Diplomarbeit ist es aufbauend auf den Artikel von P. Diaconis und B. Sturmfels [7] einen Unabhängigkeitstest für  $K \times L$ -Kontingenztafeln mit Hilfe der Markov Ketten Monte Carlo Methode durchzuführen<sup>1</sup> und auf diese Weise die nicht zufriedenstellende Approximation der  $\chi^2$ -Verteilung zu umgehen.

Im ersten Kapitel findet man dafür zunächst eine Einführung von Markov Ketten, die u.a. durch die Bücher von O. Häggström, P. Brémaud und J.R. Norris ([12], [3] und [20]) zusätzlich vertieft werden kann. Nach grundlegenden Definitionen und Eigenschaften wird in Abschnitt 1.4 die Existenz und Eindeutigkeit einer invarianten Wahrscheinlichkeitsverteilung für jede aperiodische und irreduzible Markov Kette gezeigt und ergänzend dazu das Grenzwerttheorem für Markov Ketten bewiesen, in dem das Langzeitverhalten der Verteilung der Zustände charakterisiert wird. Am Ende des ersten Kapitels in Abschnitt 1.6 findet man schließlich den Ergodensatz, der eine Art Gesetz der großen Zahlen darstellt und in Simulationen dazu dienen kann Größen vom Typ  $E_\pi(f(X_0))$  zu schätzen, wenn die invariante Verteilung  $\pi$  unbekannt ist.

Die Markov Ketten Monte Carlo Methode (MCMC) wird anschließend in Kapitel 2 vorgestellt. Die Grundlage hierfür bilden vor allem die Artikel von W.K. Hastings [13] und S. Chib und E. Greenberg [6]. Wichtig für das weitere Vorgehen ist dabei vor allem der in Abschnitt 2.2 beschriebene Metropolis(-Hastings) Algorithmus, mit dem sich basierend auf dem Grenzwerttheorem eine Markov Kette mit einer gewünschten invarianten Wahrscheinlichkeitsverteilung simulieren lässt.

Der Bezug zu Kontingenztafeln und der Unabhängigkeit wird in Kapitel 3 hergestellt. Orientiert an P. Diaconis und B. Sturmfels [7], sowie F. Rapallo [22] wird nach einer kurzen Einführung in Abschnitt 3.4 gezeigt, dass bei Zufallsvariablen aus der Exponentialfamilie die gemeinsame Verteilung unabhängig vom unbekanntem Parameter(vektor)  $\theta$  ist, wenn man sie mit der gemeinsamen suffizienten Statistik bedingt. Auf der Menge  $\mathcal{F}_t$  der Häufigkeitsverteilungen, die man durch die Stichproben mit festem Wert dieser suffizienten Statistik erhält, ergibt sich damit eine von  $\theta$  unabhängige Verteilung  $H_t$ . Um einen Unabhängigkeitstest mit Hilfe dieser bedingten Verteilung durchführen zu können, wird in Abschnitt 3.5 der u.a. von S. Sullivant [24] definierte bedingte p-Wert eingeführt. Er ist ein Wert der Form  $E_{H_t}(g(f(Y)))$ , der allerdings nicht exakt berechnet werden kann, wenn  $\mathcal{F}_t$  sehr groß ist. In diesem Fall kann er aber aufgrund des Ergodensatzes mit Hilfe der Markov Ketten Monte Carlo Methode approximiert werden, indem man eine Markov Kette von Häufigkeitsverteilungen mit festem Wert der suffizienten Statistik simuliert und den entsprechenden Schätzwert berechnet. In Abschnitt 3.6 werden zu diesem Zweck die Ergebnisse aus Abschnitt 3.4 auf Häufigkeitstabellen mit festen Zeilen- und Spaltensummen übertragen. Es wird gezeigt, dass die Zeilen- und Spaltensummen eine suffiziente Statistik für  $\theta$  darstellen, dass unter der Nullhypothese der Unabhängigkeit

---

<sup>1</sup>Eine entsprechende Übertragung auf den mehrdimensionalen Fall ist z.B. in [7] zu finden.

die erforderlichen Voraussetzungen erfüllt und somit die bereits gewonnenen Erkenntnisse auf diesen Fall übertragbar sind. Am Ende von Kapitel 3 wird schließlich in Abschnitt 3.7 der Metropolis(-Hastings) Algorithmus so modifiziert, dass mit diesem Markov Ketten simuliert werden können, die als Zustandsraum die Menge der Häufigkeitstabellen mit festen Zeilen- und Spaltensummen besitzen. D.h. mit diesem „modifizierten Metropolis Algorithmus“ lassen sich Markov Ketten mit einer von  $\theta$  unabhängigen invarianten Verteilung simulieren, mit denen man schließlich den bedingten p-Wert approximativ berechnen kann.

Im vierten und letzten Kapitel findet dieser theoretisch erarbeitete Algorithmus anhand von aus der Literatur (vgl. u.a. Bishop *et al.* [5]) ausgewählten Beispielen eine praktische Anwendung. U.a. wird hier für relativ große Matrizen mit kleinen Einträgen (erwartete Zellhäufigkeiten  $< 5$ ) der bedingte p-Wert approximativ berechnet und mit dem p-Wert des  $\chi^2$ -Unabhängigkeitstest verglichen. Um sich ein Bild davon zu machen wie sich der Unterschied zwischen diesen beiden Werten in den verschiedenen Wertebereichen verhält, werden die jeweiligen Matrizen anschließend so abgewandelt, dass pro Beispiel je eine Annahme und eine Ablehnung der Unabhängigkeitshypothese auf dem Niveau 0,05 vorliegt. Zusätzlich wird untersucht, ob der Unterschied zwischen bedingten p-Wert und p-Wert des  $\chi^2$ -Unabhängigkeitstests im Bereich des Signifikanzniveaus so groß ist, dass er in bestimmten Fällen einen Unterschied in Annahme und Ablehnung der Nullhypothese auf dem Niveau 0,05 bewirken könnte.

# Kapitel 1

## Markov Ketten

### 1.1 Definition und Eigenschaften

#### Definition 1.1.

Sei  $(\Omega, \mathcal{A}, \mathcal{P})$  ein Wahrscheinlichkeitsraum,  $I$  eine beliebige nichtleere Indexmenge und  $(S, \mathcal{S})$  ein messbarer Raum.

Eine Familie  $\{X_t; t \in I\}$  von Zufallsvariablen mit Werten in  $S$  heißt **stochastischer Prozess** mit Parameterbereich  $I$  und Zustandsraum  $S$ .

#### Definition 1.2.

Eine **Markov Kette** ist ein stochastischer Prozess  $\{X_t; t \in \mathbb{Z}^+\}$  mit abzählbarem Zustandsraum  $S$ , der die folgende Markov Eigenschaft besitzt:

Für alle  $n \in \mathbb{Z}^+$ , alle  $s_{i_0}, s_{i_1}, \dots, s_{i_{n-1}} \in S$  und für alle  $s_i, s_j \in S$  mit

$$P(X_0 = s_{i_0}, X_1 = s_{i_1}, \dots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) > 0$$

ist

$$P(X_{n+1} = s_j \mid X_0 = s_{i_0}, \dots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) = P(X_{n+1} = s_j \mid X_n = s_i). \quad (1.1)$$

#### Korollar 1.3.

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S$ .

Dann gilt für alle  $n$  und alle  $s_{i_0}, \dots, s_{i_n} \in S$

$$\begin{aligned} P(X_0 = s_{i_0}, \dots, X_n = s_{i_n}) \\ = P(X_0 = s_{i_0})P(X_1 = s_{i_1} \mid X_0 = s_{i_0}) \cdots P(X_n = s_{i_n} \mid X_{n-1} = s_{i_{n-1}}). \end{aligned}$$

*Beweis.* Es gilt:

$$\begin{aligned}
& P(X_0 = s_{i_0}, \dots, X_n = s_{i_n}) \\
&= P(X_0 = s_{i_0})P(X_1 = s_{i_1} \mid X_0 = s_{i_0})P(X_2 = s_{i_2} \mid X_0 = s_{i_0}, X_1 = s_{i_1}) \cdots \\
&\quad \cdots P(X_n = s_{i_n} \mid X_0 = s_{i_0}, \dots, X_{n-1} = s_{i_{n-1}}) \\
&\stackrel{(1.1)}{=} P(X_0 = s_{i_0})P(X_1 = s_{i_1} \mid X_0 = s_{i_0})P(X_2 = s_{i_2} \mid X_1 = s_{i_1}) \cdots \\
&\quad \cdots P(X_n = s_{i_n} \mid X_{n-1} = s_{i_{n-1}})
\end{aligned}$$

□

**Lemma 1.4.**

Es seien  $C_1, C_2, \dots$  disjunkte Ereignisse mit  $\bigcup_{i=1}^{\infty} C_i = C$  und die bedingten Wahrscheinlichkeiten  $P(A \mid B \cap C_r)$  seien unabhängig von  $r$ .

Dann gilt:

$$P(A \mid B \cap C_r) = P(A \mid B \cap C).$$

*Beweis.* Es gilt:

$$\begin{aligned}
P(A \mid B \cap C_r)P(B \cap C) &= P(A \mid B \cap C_r) \sum_{i=1}^{\infty} P(B \cap C_i) \\
&= \sum_{i=1}^{\infty} P(A \mid B \cap C_i)P(B \cap C_i) \\
&= \sum_{i=1}^{\infty} P(A \cap B \cap C_i) \\
&= P(A \cap B \cap C) \\
&= P(A \mid B \cap C)P(B \cap C).
\end{aligned}$$

Dividiert man dies durch  $P(B \cap C)$  ergibt sich schließlich

$$P(A \mid B \cap C_r) = P(A \mid B \cap C).$$

[15]

□

**Satz 1.5.**

Es seien  $0 < n < N$  und  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette.

Dann gilt für alle  $s_{i_n} \in S$  und alle Teilmengen  $E \subset S^n, F \subset S^{N-n}$

$$\begin{aligned}
& P((X_{n+1}, \dots, X_N) \in F \mid X_n = s_{i_n}, (X_0, \dots, X_{n-1}) \in E) \\
&= P((X_{n+1}, \dots, X_N) \in F \mid X_n = s_{i_n}).
\end{aligned} \tag{1.2}$$

*Beweis.* Wegen der  $\sigma$ -Additivität von  $P$  reicht es aus anzunehmen, dass  $F$  aus nur einem Element  $(s_{i_{n+1}}, \dots, s_{i_N}) \in S^{N-n}$  besteht.

Nach Korollar 1.3 gilt für beliebige  $s_{i_0}, \dots, s_{i_{n-1}} \in S$ :

$$\begin{aligned}
& P((X_{n+1}, \dots, X_N) \in F \mid X_n = s_{i_n}, (X_0, \dots, X_{n-1}) = (s_{i_0}, \dots, s_{i_{n-1}})) \\
&= \frac{P(X_0 = s_{i_0}, \dots, X_N = s_{i_N})}{P(X_0 = s_{i_0}, \dots, X_n = s_{i_n})} \\
&= \frac{P(X_0 = s_{i_0})P(X_0 = s_{i_0} \mid X_1 = s_{i_1}) \cdots P(X_N = s_{i_N} \mid X_{N-1} = s_{i_{N-1}})}{P(X_0 = s_{i_0})P(X_0 = s_{i_0} \mid X_1 = s_{i_1}) \cdots P(X_n = s_{i_n} \mid X_{n-1} = s_{i_{n-1}})} \\
&= P(X_{n+1} = s_{i_{n+1}} \mid X_n = s_{i_n}) \cdots P(X_N = s_{i_N} \mid X_{N-1} = s_{i_{N-1}}) \\
&=: p
\end{aligned}$$

Da  $p$  unabhängig von  $(s_{i_0}, \dots, s_{i_{n-1}})$  ist, kann man Lemma 1.4 anwenden. Sei hierfür  $C$  eine beliebige disjunkte Vereinigung von Mengen der Form  $\{(X_0, \dots, X_{n-1}) = (s_{i_0}, \dots, s_{i_{n-1}})\}$ . Dann gilt:

$$P((X_{n+1}, \dots, X_N) \in F \mid \{X_n = s_{i_n}\} \cap C) = p.$$

Setzt man zunächst  $C = \{(X_0, \dots, X_{n-1}) \in E\}$ , ergibt sich

$$\begin{aligned}
p &= P((X_{n+1}, \dots, X_N) \in F \mid \{X_n = s_{i_n}\} \cap C) \\
&= P((X_{n+1}, \dots, X_N) \in F \mid X_n = s_{i_n}, (X_0, \dots, X_{n-1}) \in E)
\end{aligned}$$

Für  $C = \Omega$  andererseits erhält man:

$$\begin{aligned}
p &= P((X_{n+1}, \dots, X_N) \in F \mid \{X_n = s_{i_n}\} \cap C) \\
&= P((X_{n+1}, \dots, X_N) \in F \mid X_n = s_{i_n}),
\end{aligned}$$

womit also die rechte sowie die linke Seite von (1.2) gleich  $p$  sind. [15] □

**Satz 1.6 (Chapman-Kolmogorow-Gleichung).**

Es seien  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette und  $k < l < m$ .

Dann gilt für alle  $s_h, s_j \in S$

$$P(X_m = s_j \mid X_k = s_h) = \sum_{s_i \in S} P(X_l = s_i \mid X_k = s_h)P(X_m = s_j \mid X_l = s_i) \quad (1.3)$$

*Beweis.* Es gilt:

$$\begin{aligned}
P(X_k = s_h, X_m = s_j) &= \sum_{s_i \in S} P(X_k = s_h, X_l = s_i, X_m = s_j) \\
&= \sum_{s_i \in S} P(X_k = s_h, X_l = s_i)P(X_m = s_j \mid X_k = s_h, X_l = s_i) \\
&\stackrel{(1.2)}{=} \sum_{s_i \in S} P(X_k = s_h, X_l = s_i)P(X_m = s_j \mid X_l = s_i).
\end{aligned}$$



Die Division beider Seiten durch  $P(X_k = s_h)$  ergibt

$$\frac{P(X_k = s_h, X_m = s_j)}{P(X_k = s_h)} = P(X_m = s_j \mid X_k = s_h)$$

und

$$\begin{aligned} \sum_{s_i \in S} \frac{P(X_k = s_h, X_l = s_i)}{P(X_k = s_h)} P(X_m = s_j \mid X_l = s_i) \\ = \sum_{s_i \in S} P(X_l = s_i \mid X_k = s_h) P(X_m = s_j \mid X_l = s_i). \end{aligned}$$

[15]

□

**Definition 1.7.**

Eine Markov Kette mit abzählbarem Zustandsraum  $S$  heißt **homogen** oder Kette mit **stationären Übergangswahrscheinlichkeiten**, wenn für alle  $s_i, s_j \in S$   $P(X_{n+1} = s_j \mid X_n = s_i)$  unabhängig von  $n$  ist.

Für homogene Markov Ketten gilt für alle  $s_i, s_j \in S$  und alle  $n \in \mathbb{Z}^+$

$$P(X_{n+1} = s_j \mid X_n = s_i) = P(X_1 = s_j \mid X_0 = s_i) =: p_{ij}.$$

Man bezeichnet  $p_{ij}$  als **Übergangswahrscheinlichkeit** und die (stochastische) Matrix  $P := (p_{ij})$  als **Übergangsmatrix**.

Entsprechend bezeichnet

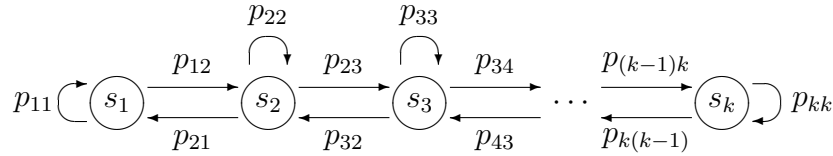
$$p_{ij}^{(m)} := P(X_{n+m} = s_j \mid X_n = s_i)$$

die **m-Schritt-Übergangswahrscheinlichkeit** von  $s_i$  nach  $s_j$ . Dass diese nicht von  $n$  abhängt, ist für  $m = 1$  Definition der Homogenität und ergibt sich für  $m \geq 2$  induktiv mittels der Chapman-Kolmogorow-Gleichung (1.3), die sich ausdrücken lässt in der Form:

$$p_{ij}^{(n+m)} = \sum_{s_l \in S} p_{il}^{(n)} p_{lj}^{(m)}.$$

Setzt man  $p_{ij}^{(0)} = \delta_{ij}$ , so ist diese Gleichung auch für  $n = 0$  und  $m = 0$  erfüllt. (vgl. [15, S. 200])

Markov Ketten werden oft durch Diagramme dargestellt, die das Verhalten der Ketten anschaulicher machen. Wie in [12] z.B. sei zunächst ein allgemeines Beispiel gegeben:



**Im Folgenden beziehe ich mich ausschließlich auf homogene Markov Ketten, was der Einfachheit halber nicht mehr explizit angegeben werden wird.**

**Lemma 1.8.**

Sei eine Markov Kette mit Zustandsraum  $S$  und Übergangsmatrix  $P$  gegeben. Dann gilt:

$$(p_{ij}^{(m)}) = P^m.$$

*Beweis.* Sei zunächst  $m = 1$ . Dann gilt:  $(p_{ij}^{(1)}) = (p_{ij}) = P$ .

Sei nun  $(p_{ij}^{(n)}) = P^n$  erfüllt für  $n \in \mathbb{N}$  beliebig aber fest.

Dann ergibt sich:

$$p_{ij}^{(n+1)} = \sum_{l \in \{1, \dots, k\}} p_{il}^{(n)} p_{lj}^{(1)} = (P^n P)_{ij} = (P^{n+1})_{ij} \quad ,$$

wobei  $(P^{n+1})_{ij}$  den Eintrag  $(i, j)$  der Matrix  $P^{n+1}$  bezeichnet. □

**Definition 1.9.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$ .

Der Zeilenvektor  $\mu^{(0)}$  mit

$$\mu^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_k^{(0)}) = (P(X_0 = s_1), P(X_0 = s_2), \dots, P(X_0 = s_k))$$

heißt **Startverteilung** der Markov Kette.

Entsprechend dieser Definition bezeichnen  $\mu^{(1)}, \mu^{(2)}, \dots$  die Verteilungen der Markov Kette zum Zeitpunkt  $n = 1, 2, \dots$ , so dass

$$\mu^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \dots, \mu_k^{(n)}) = (P(X_n = s_1), P(X_n = s_2), \dots, P(X_n = s_k)).$$

**Satz 1.10.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$ , Startverteilung  $\mu^{(0)}$  und Übergangsmatrix  $P$ .

Dann gilt für jedes  $n \in \mathbb{Z}^+$

$$\mu^{(n)} = \mu^{(0)} P^n. \quad (1.4)$$

*Beweis.* Sei zunächst  $n = 1$ .

Für  $j = 1, \dots, k$  gilt mit dem Satz von der totalen Wahrscheinlichkeit:

$$\begin{aligned} \mu_j^{(1)} &= P(X_1 = s_j) \\ &= \sum_{i=1}^k P(X_0 = s_i) P(X_1 = s_j \mid X_0 = s_i) \\ &= \sum_{i=1}^k \mu_i^{(0)} p_{ij} \\ &= (\mu^{(0)} P)_j, \end{aligned}$$

wobei  $(\mu^{(0)} P)_j$  das  $j$ -te Element des Zeilenvektors  $(\mu^{(0)} P)$  bezeichnet.

Somit gilt:  $\mu^{(1)} = \mu^{(0)} P$

Per Induktion wird nun der allgemeine Fall von (1.4) bewiesen:

Seien dafür  $m \in \mathbb{N}$  beliebig aber fest und (1.4) erfüllt für  $n = m$ .

Für  $n = m + 1$  erhält man so:

$$\begin{aligned} \mu_j^{(m+1)} &= P(X_{m+1} = s_j) \\ &= \sum_{i=1}^k P(X_m = s_i) P(X_{m+1} = s_j \mid X_m = s_i) \\ &= \sum_{i=1}^k \mu_i^{(m)} p_{ij} \\ &= (\mu^{(m)} P)_j, \end{aligned}$$

d.h.  $\mu^{(m+1)} = \mu^{(m)} P$

Nach Induktionsvoraussetzung gilt aber  $\mu^{(m)} = \mu^{(0)} P^m$ , womit sich ergibt:

$$\mu^{(m+1)} = \mu^{(m)} P = \mu^{(0)} P^m P = \mu^{(0)} P^{m+1}$$

[12]

□

**Definition 1.11.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ .

Man sagt, dass Zustand  $s_i$  in  $n$  Schritten zu Zustand  $s_j$  führt und schreibt  $s_i \rightarrow s_j[n]$ , wenn

$$P(X_{m+n} = s_j \mid X_m = s_i) > 0$$

ist. Existiert ein  $n \geq 1$  mit  $s_i \rightarrow s_j[n]$ , so sagt man  $s_i$  führe zu  $s_j$  und schreibt  $s_i \rightarrow s_j$ .

Gilt  $s_i \rightarrow s_j$  und  $s_j \rightarrow s_i$ , so sagt man  $s_i$  **kommuniziere** mit  $s_j$  ( $s_i \leftrightarrow s_j$ ).

**Definition 1.12.**

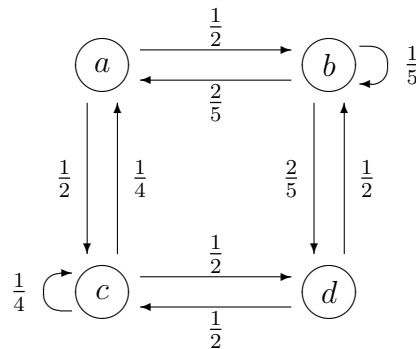
Eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  heißt **irreduzibel**,

wenn für alle  $s_i, s_j$  gilt  $s_i \leftrightarrow s_j$ , d.h. die Kette ist irreduzibel, falls für alle  $s_i, s_j \in S$  ein  $n$  existiert, so dass  $p_{ij}^{(n)} > 0$ .

Ist dies nicht der Fall, heißt die Kette **reduzibel**.

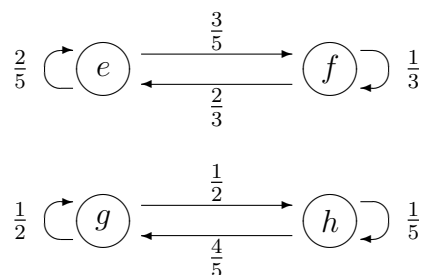
**Beispiel 1.13.**

a) Irreduzibel:



Dieses Diagramm beschreibt eine irreduzible Markov Kette. Hier kommunizieren alle Zustände miteinander, so dass man von jedem beliebigen Zustand zu jedem anderen gelangen kann.

b) Reduzibel:



Startet man bei der durch dieses Diagramm beschriebenen Markov Kette dagegen in Zustand  $e$  oder  $f$ , ist die Kette auf die Zustände  $e$  und  $f$  beschränkt. Startet man in Zustand  $g$  oder  $h$ , gilt dies entsprechend für die Zustände  $g$  und  $h$ . Eine Eigenschaft von reduziblen Markov Ketten ist es, dass sie sich auf lange Zeit gesehen wie eine Kette mit kleinerem Zustandsraum verhalten. (vgl. [12, S. 25])

Für dieses Beispiel ergibt sich die Übergangsmatrix

$$P = \begin{pmatrix} \frac{2}{5} & \frac{3}{5} & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{4}{5} & \frac{1}{5} \end{pmatrix}.$$

Startet man in Zustand  $e$  oder  $f$ , verhält sich die Kette so, als läge eine Markov Kette mit Zustandsraum  $\{e, f\}$  und Übergangsmatrix

$$P = \begin{pmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

vor. Bei Start in Zustand  $g$  oder  $h$  verhält sie sich entsprechend wie eine Kette mit Zustandsraum  $\{g, h\}$  und Übergangsmatrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{4}{5} & \frac{1}{5} \end{pmatrix}.$$

**Definition 1.14.**

Die **Periode**  $d(s_i)$  eines Zustands  $s_i \in S$  ist definiert als

$$\begin{aligned} d(s_i) &:= \text{ggT}\{n \geq 1 : p_{ii}^{(n)} > 0\} \\ &= \text{ggT}\{n \geq 1 : s_i \rightarrow s_i[n]\} \end{aligned}$$

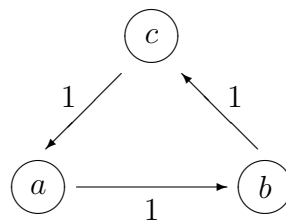
Gilt nicht  $s_i \rightarrow s_i$ , so sei  $d(s_i) = \infty$ .

**Definition 1.15.**

Gilt  $d(s_i) = 1$ , dann heißt der Zustand  $s_i$  **aperiodisch**. Eine Markov Kette heißt aperiodisch, wenn alle ihre Zustände aperiodisch sind. Sie heißt **periodisch** mit Periode  $d$ , wenn alle  $d(s_i) = d \geq 2$  sind.

**Beispiel 1.16.**

a) Periodische Markov Kette der Periode 3:



Diese Diagramm beschreibt eine Markov Kette mit Zustandsraum  $S = \{a, b, c\}$  und Übergangsmatrix

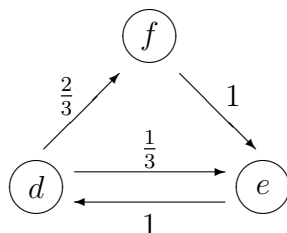
$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Beim Start in einem beliebigen Zustand aus  $S$  werden bei dieser Kette drei Schritte benötigt, um zu eben diesen Zustand zurückzukehren und man erreicht ihn wieder in jedem dritten Schritt. Es gilt somit

$$d(s_i) = \text{ggT}\{n \geq 1 : p_{ii}^{(n)} > 0\} = \text{ggT}\{n \geq 1 : s_i \rightarrow s_i[n]\} = 3 \quad \text{für } i = 1, 2, 3,$$

d.h. diese Markov Kette besitzt die Periode  $d = 3$ .

b) Aperiodische Markov Kette:



Die hier beschriebene Kette ist eine aperiodische Markov Kette mit Zustandsraum  $S = \{d, e, f\}$  und Übergangsmatrix

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Startet man bei dieser Kette in einem Zustand aus  $S$ , gibt es mehrere Möglichkeiten wieder zu diesem Zustand zurückzukehren.

Betrachte man z.B. den Zustand  $d$ . Von diesem Zustand aus ist es zum Einen möglich über  $e$  zu  $d$  zurückzukehren. Eine andere Möglichkeit wäre es jedoch von  $d$  erst nach  $f$ , anschließend nach  $e$  und von dort dann schließlich wieder zu  $d$  zu gelangen. Es ergibt sich

$$p_{11}^{(2)} = \frac{1}{3} \cdot 1 \quad \text{bzw.} \quad p_{11}^{(3)} = \frac{2}{3} \cdot 1 \cdot 1$$

und hiermit

$$d(d) = \text{ggT}\{n \geq 1 : p_{11}^{(n)} > 0\} = 1.$$

Der Zustand  $d$  ist also ein aperiodischer Zustand.

Analoge Überlegungen ergeben für  $e$

$$p_{22}^{(2)} = 1 \cdot \frac{1}{3} \quad \text{bzw.} \quad p_{22}^{(3)} = 1 \cdot \frac{2}{3} \cdot 1$$

und für  $f$  z.B.

$$p_{33}^{(3)} = \frac{2}{3} \cdot 1 \cdot 1 \quad \text{oder} \quad p_{33}^{(5)} = 1 \cdot 1 \cdot \frac{1}{3} \cdot 1 \cdot \frac{2}{3}.$$

Zusammengefasst gilt

$$d(d) = d(e) = d(f) = 1,$$

d.h. alle drei Zustände (und damit die gesamte Markov Kette) sind aperiodisch.

**Satz 1.17.**

Gegeben sei eine aperiodische Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$  mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ .

Dann existiert ein  $N < \infty$  so, dass

$$p_{ii}^{(n)} > 0 \quad \text{für alle } i \in \{1, \dots, k\} \text{ und alle } n \geq N.$$

Zum Beweis dieses Satzes soll ein Lemma aus der Zahlentheorie helfen:

**Lemma 1.18.**

Sei  $A = \{a_1, a_2, \dots\}$  eine Menge natürlicher Zahlen, für die gilt:

- (i)  $ggT\{a_1, a_2, \dots\} = 1$ ,
- (ii)  $A$  ist abgeschlossen bezüglich der Addition,  
d.h. ist  $a \in A$  und  $a' \in A$ , so gilt  $a + a' \in A$ .

Dann existiert eine ganze Zahl  $N < \infty$  so, dass  $n \in A$  für alle  $n \geq N$ .

*Beweis.* siehe [12, S.25] □

*Beweis von Satz 1.17.* Für  $s_i \in S$  sei  $A_i = \{n \geq 1 : p_{ii}^{(n)} > 0\}$  die Menge der möglichen Rückkehrzeiten zum Zustand  $s_i$  bei Start in  $s_i$ .

Da die Markov Kette nach Voraussetzung aperiodisch ist, ist auch der Zustand  $s_i$  aperiodisch. Somit gilt:  $ggT\{n \geq 1 : p_{ii}^{(n)} > 0\} = 1$

Weiter ist  $A_i$  abgeschlossen bezüglich der Addition, denn:

Sind  $a, a' \in A_i$ , so gilt  $P(X_a = s_i \mid X_0 = s_i) > 0$  und  $P(X_{a+a'} = s_i \mid X_a = s_i) > 0$ .

Damit ergibt sich:

$$\begin{aligned} P(X_{a+a'} = s_i \mid X_0 = s_i) &\geq P(X_a = s_i, X_{a+a'} = s_i \mid X_0 = s_i) \\ &= \frac{P(X_a = s_i, X_{a+a'} = s_i, X_0 = s_i)}{P(X_0 = s_i)} \\ &= \frac{P(X_0 = s_i)}{P(X_0 = s_i)} P(X_a = s_i \mid X_0 = s_i) P(X_{a+a'} = s_i \mid X_a = s_i) \\ &> 0 \end{aligned}$$

und somit gilt  $a + a' \in A_i$ .

Mit Lemma 1.18 existiert ein  $N_i < \infty$  so, dass  $p_{ii}^{(n)} > 0$  für alle  $n \geq N_i$ .

Setzt man schließlich  $N = \max\{N_1, \dots, N_k\}$ , so erhält man die gewünschte Aussage. □

[12]



**Korollar 1.19.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine aperiodische und irreduzible Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ .

Dann existiert ein  $M < \infty$ , so dass

$$p_{ij}^{(n)} > 0 \quad \text{für alle } i, j \in \{1, \dots, k\} \text{ und alle } n \geq M.$$

*Beweis.* Aus der vorausgesetzten Aperiodizität und mit Satz 1.17 folgt, dass es eine ganze Zahl  $N < \infty$  gibt, so dass

$$p_{ii}^{(n)} > 0 \text{ für alle } i \in \{1, \dots, k\} \text{ und alle } n \geq N.$$

Seien nun  $s_i, s_j \in S$  fest. Da die Markov Kette nach Voraussetzung irreduzibel ist, findet man ein  $n_{ij}$ , so dass  $p_{ij}^{(n_{ij})} > 0$ .

Sei  $M_{ij} = N + n_{ij}$ . Dann gilt für jedes  $m \geq M_{ij}$ :

$$\begin{aligned} & P(X_m = s_j \mid X_0 = s_i) \\ & \geq P(X_{m-n_{ij}} = s_i, X_m = s_j \mid X_0 = s_i) \\ & = \frac{P(X_{m-n_{ij}} = s_i, X_m = s_j, X_0 = s_i)}{P(X_0 = s_i)} \\ & = \frac{P(X_0 = s_i)P(X_{m-n_{ij}} = s_i \mid X_0 = s_i)P(X_m = s_j \mid X_{m-n_{ij}})}{P(X_0 = s_i)} \\ & = \underbrace{P(X_{m-n_{ij}} = s_i \mid X_0 = s_i)}_{>0, \text{ da } m-n_{ij} \geq N} \underbrace{P(X_m = s_j \mid X_{m-n_{ij}} = s_i)}_{>0, \text{ da } p_{ij}^{(n_{ij})} > 0} > 0 \end{aligned}$$

Somit ist  $p_{ij}^{(m)} > 0$  für alle  $m \geq M_{ij}$ .

Setzt man schließlich

$$M = \max \{M_{11}, M_{12}, \dots, M_{1k}, M_{21}, \dots, M_{kk}\},$$

folgt die Behauptung. [12] □

## 1.2 Computer Simulation von Markov Ketten

In diesem Abschnitt soll eine Methode beschrieben werden, mit der man eine Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$  mit bekanntem Zustandsraum  $S = \{s_1, \dots, s_k\}$ , gegebener Startverteilung  $\mu^{(0)}$  und Übergangsmatrix  $P$  simulieren kann. (vgl. [12, Kapitel 3])

Für die gewünschte Simulation wird im Wesentlichen folgendes benötigt:

- eine Folge  $\{U_t; t \in \mathbb{Z}^+\}$  von unabhängigen und identisch verteilten Zufallsvariablen, die auf dem Intervall  $[0, 1]$  gleichverteilt sind,
- eine Einführungsfunktion  $\psi$  und
- eine Update-Funktion  $\phi$ .

### Definition 1.20.

Eine **Einführungsfunktion** ist eine Funktion  $\psi : [0, 1] \rightarrow S$ , die wir benutzen um den Anfangswert  $X_0$  zu erzeugen. Sie soll folgende Eigenschaften erfüllen:

- (i)  $\psi$  ist stückweise konstant, d.h. man kann das Intervall  $[0, 1]$  in endlich viele Teilintervalle aufteilen so, dass  $\psi$  auf jedem dieser Intervalle konstant ist.
- (ii) für jedes  $s \in S$  ist  $\mu^{(0)}(s)$  die Länge des Intervalls, auf dem  $\psi(x) = s$  gilt. Das entspricht:

$$\int_0^1 1_{\{s\}}(\psi(x)) dx = \mu^{(0)}(s) \quad \text{für jedes } s \in S. \quad (1.5)$$

Liegt eine solche Einführungsfunktion vor, kann man mit der ersten Zufallsvariablen  $U_0$  den Wert  $X_0$  erzeugen, indem man  $X_0 = \psi(U_0)$  setzt. Auf diese Weise ist die korrekte Verteilung von  $X_0$  gegeben, denn für alle  $s \in S$  gilt:

$$P(X_0 = s) = P(\psi(U_0) = s) = \int_0^1 1_{\{s\}}(\psi(x)) dx \stackrel{(1.5)}{=} \mu^{(0)}(s).$$

### Definition 1.21.

Eine Einführungsfunktion  $\psi$  heißt **gültig** für die Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$ , wenn sie (1.5) erfüllt für alle  $s \in S$ .

**Konstruktion einer gültigen Einführungsfunktion:**

Seien  $S = \{s_1, \dots, s_k\}$  der Zustandsraum und  $\mu^{(0)}$  die Startverteilung der zu simulierenden Markov Kette. Dann kann man definieren:

$$\psi(x) = \begin{cases} s_1 & \text{für } x \in [0, \mu^{(0)}(s_1)) \\ s_2 & \text{für } x \in [\mu^{(0)}(s_1), \mu^{(0)}(s_1) + \mu^{(0)}(s_2)) \\ \vdots & \vdots \\ s_i & \text{für } x \in \left[ \sum_{j=1}^{i-1} \mu^{(0)}(s_j), \sum_{j=1}^i \mu^{(0)}(s_j) \right) \\ \vdots & \vdots \\ s_k & \text{für } x \in \left[ \sum_{j=1}^{k-1} \mu^{(0)}(s_j), 1 \right] \end{cases} \quad (1.6)$$

**Lemma 1.22.**

Die Funktion  $\psi$ , definiert in (1.6), ist eine gültige Einführungsfunktion für die Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$ .

*Beweis.* Es ist zu zeigen, dass  $\psi$  die Eigenschaften (i) und (ii) aus Definition 1.20 erfüllt. Eigenschaft (i) ist mit (1.6) offensichtlich.

Zusätzlich gilt:

$$\int_0^1 1_{\{s_i\}}(\psi(x)) dx = \sum_{j=1}^i \mu^{(0)}(s_j) - \sum_{j=1}^{i-1} \mu^{(0)}(s_j) = \mu^{(0)}(s_i) \quad \text{für } i = 1, \dots, k,$$

was bedeutet, dass auch (ii) erfüllt ist und  $\psi$  somit eine gültige Einführungsfunktion darstellt.  $\square$

Mit Hilfe der Einführungsfunktion und der ersten Zufallsvariablen  $U_0$  erhält man also den Anfangswert  $X_0$ . Findet man nun einen Weg den Wert  $X_{n+1}$  mit Hilfe des Wertes  $X_n$  für jedes  $n$  zu erzeugen, kann man diese Vorgehensweise iterativ anwenden um die gesamte Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$  zu erhalten.

**Definition 1.23.**

Eine **Update-Funktion** ist eine Funktion  $\phi : S \times [0, 1] \rightarrow S$ , mit der wir für jedes  $n$  aus dem Wert  $X_n$  und der Zufallsvariablen  $U_{n+1}$  den Wert  $X_{n+1}$  erzeugen. Folgende Eigenschaften muss eine solche Funktion erfüllen:

- (i) für festes  $s_i$  ist die Funktion  $\phi(s_i, x)$  stückweise konstant (, fasst man sie als Funktion von  $x$  auf,) und

(ii) für alle feste  $s_i, s_j \in S$  ist  $p_{ij}$  die Länge des Intervalls, auf dem  $\phi(s_i, x) = s_j$  gilt. Dies entspricht folgender Gleichung:

$$\int_0^1 1_{\{s_j\}}(\phi(s_i, x))dx = p_{ij} \quad \text{für alle } s_i, s_j \in S. \quad (1.7)$$

Erfüllt  $\phi$  die Gleichung (1.7), so gilt:

$$\begin{aligned} P(X_{n+1} = s_j \mid X_n = s_i) &= P(\phi(s_i, U_{n+1}) = s_j \mid X_n = s_i) \\ &= P(\phi(s_i, U_{n+1}) = s_j) \\ &= \int_0^1 1_{\{s_j\}}(\phi(s_i, x))dx \\ &\stackrel{(1.7)}{=} p_{ij}. \end{aligned} \quad (1.8)$$

In (1.8) gilt  $P(\phi(s_i, U_{n+1}) = s_j \mid X_n = s_i) = P(\phi(s_i, U_{n+1}) = s_j)$ , da  $U_{n+1}$  unabhängig von  $(U_1, \dots, U_n)$  und somit auch unabhängig von  $X_n$  ist. Mit dem gleichen Argument bleibt die Wahrscheinlichkeit auch gleich, wenn man sie mit den Werten  $(X_0, \dots, X_{n-1})$  bedingt. Aus diesem Grund stellt die beschriebene Vorgehensweise eine korrekte Simulation der Markov Kette dar.

**Definition 1.24.**

Eine Update-Funktion  $\phi$  heißt **gültig** für die Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$ , wenn sie für alle  $s_i, s_j \in S$  die Gleichung (1.7) erfüllt.

**Konstruktion einer gültigen Update-Funktion:**

Die Konstruktion einer gültigen Update-Funktion läuft analog zu der einer gültigen Einführungsfunktion. Für jedes  $s_i \in S$  setzt man:

$$\phi(s_i, x) = \begin{cases} s_1 & \text{für } x \in [0, p_{i1}) \\ s_2 & \text{für } x \in [p_{i1}, p_{i1} + p_{i2}) \\ \vdots & \vdots \\ s_j & \text{für } x \in \left[ \sum_{l=1}^{j-1} p_{il}, \sum_{l=1}^j p_{il} \right) \\ \vdots & \vdots \\ s_k & \text{für } x \in \left[ \sum_{l=1}^{k-1} p_{il}, 1 \right] \end{cases} \quad (1.9)$$

**Lemma 1.25.**

Die Funktion  $\phi$ , definiert in (1.9), ist eine gültige Update-Funktion für die Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$ .

*Beweis.* Eigenschaft (i) aus Definition 1.23 ist mit (1.9) offensichtlich.  
Es gilt zusätzlich:

$$\int_0^1 1_{\{s_j\}}(\phi(s_i, x)) dx = \sum_{l=1}^j p_{il} - \sum_{l=1}^{j-1} p_{il} = p_{ij} \quad \text{für alle } s_i, s_j \in S,$$

weswegen auch Eigenschaft (ii) aus Definition 1.23 erfüllt ist.  $\phi$  ist somit eine gültige Update-Funktion für die Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$ .  $\square$

**Zusammenfassung:**

Zur Simulation einer homogenen Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$  mit Zustandsraum  $S$ , Startverteilung  $\mu^{(0)}$  und Übergangsmatrix  $P$  konstruiert man zunächst eine gültige Einführungsfunktion  $\psi$  (siehe z.B. (1.6)) und eine gültige Update-Funktion  $\phi$  (siehe z.B. (1.9)). Mit Hilfe von unabhängigen und identisch verteilten Zufallsvariablen  $\{U_t; t \in \mathbb{Z}^+\}$ , die auf dem Intervall  $[0, 1]$  gleichverteilt sind, setzt man schließlich:

$$\begin{aligned} X_0 &= \psi(U_0) \\ X_1 &= \phi(X_0, U_1) \\ X_2 &= \phi(X_1, U_2) \\ X_3 &= \phi(X_2, U_3) \\ &\vdots \end{aligned} \tag{1.10}$$

**1.2.1 Beispiele****Beispiel: Wetter**

Man nehme wie in [12, S. 12ff] an, eine sicherere Methode das Wetter vorauszusagen als dem Wetterbericht im Fernsehen zu glauben, wäre es zu raten, dass das morgige Wetter dem von heute entsprechen wird. Dann könnte man das Wetter durch eine Markov Kette simulieren.

Um dies zu vereinfachen seien lediglich zwei Arten von Wetter gegeben, nämlich „Sonne“ und „Regen“. Liegt man beim Raten zu 75 % richtig, egal, ob es gerade regnet oder die Sonne scheint, ergibt sich eine Markov Kette mit dem Zustandsraum  $S = \{\text{Sonne}, \text{Regen}\}$  und der Übergangsmatrix

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

Geht man von einem Regentag aus, gilt für die Anfangsverteilung  $\mu^{(0)} = (0, 1)$ . Mittels der in Anhang A.1 gegebenen Funktion `mk(n,s,mu,p,d)` mit den Parametern  $s = S$ ,  $p = P$  und  $\mu = \mu^{(0)}$  ( $d = s \in S$  beliebig) lässt sich somit Folgendes in  $R$  simulieren:

```
> s<-c("Sonne", "Regen")
> p<-matrix(c(0.75,0.25,0.25,0.75),nrow=2)
> mu<-c(0,1)

> set.seed(12345)
> mk(50,s,mu,p,"Sonne")
 [1] "Regen" "Regen" "Regen" "Regen" "Regen" "Regen" "Sonne" "Sonne" "Sonne" "Sonne"
[10] "Regen" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne"
[19] "Sonne" "Regen" "Regen" "Regen" "Regen" "Regen" "Regen" "Regen" "Regen" "Regen"
[28] "Regen" "Sonne" "Sonne" "Regen" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne"
[37] "Regen" "Regen" "Regen" "Sonne" "Regen" "Regen" "Regen" "Regen" "Regen" "Regen"
[46] "Regen" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne" "Regen"
```

Für den Winter ist zusätzlich der mögliche Schneefall zu beachten. Für diese Jahreszeit könnte man also den Zustandsraum  $S_W = \{Sonne, Regen, Schnee\}$  mit der Übergangsmatrix

$$P_W = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.15 & 0.7 & 0.15 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

annehmen<sup>1</sup>. Geht man auch hier von einem Regentag aus, ergibt sich entsprechend die Anfangsverteilung  $\mu_W^{(0)} = (0, 1, 0)$ . Mit den Parametern  $sw = S_W$ ,  $pw = P_W$  und  $\mu_w = \mu_W^{(0)}$  ( $d = s \in S_W$  beliebig) lässt sich für dieses Modell die folgende Markov Kette simulieren:

```
> sw<-c("Regen", "Sonne", "Schnee")
> pw<-matrix(c(0.5,0.15,0.2,0.3,0.7,0.3,0.2,0.15,0.5),nrow=3)
> muw<-c(0,1,0)

> set.seed(12345)
> mk(50,sw,muw,pw,"Sonne")
 [1] "Sonne" "Schnee" "Schnee" "Schnee" "Sonne" "Sonne" "Sonne" "Sonne"
 [9] "Sonne" "Schnee" "Regen" "Regen" "Sonne" "Regen" "Regen" "Regen"
[17] "Regen" "Regen" "Regen" "Schnee" "Sonne" "Sonne" "Schnee" "Schnee"
[25] "Schnee" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne" "Sonne" "Regen"
[33] "Regen" "Sonne" "Sonne" "Sonne" "Schnee" "Schnee" "Schnee" "Regen"
[41] "Sonne" "Sonne" "Schnee" "Schnee" "Sonne" "Sonne" "Regen" "Regen"
[49] "Regen" "Sonne" "Schnee"
```

---

<sup>1</sup>Für den Sommer ergäbe sich entsprechend die Übergangsmatrix  $P_S = \begin{pmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.75 & 0 \\ 0.5 & 0.5 & 0 \end{pmatrix}$ .

Dieses Modell würde sich exakt wie das obige Beispiel mit  $S = \{Sonne, Regen\}$  verhalten.

## Beispiel: Anteil eines Zustandes an der Gesamtlänge

Angenommen, zu dem Zustandsraum  $S = \{1, \dots, 10\}$  sei folgende Übergangsmatrix  $P$

```
> P
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,] 0.000 0.000 0.300 0.300 0.100 0.100 0.000 0.100 0.050 0.050
[2,] 0.310 0.030 0.046 0.090 0.170 0.150 0.050 0.050 0.020 0.084
[3,] 0.080 0.030 0.120 0.000 0.210 0.190 0.150 0.130 0.054 0.036
[4,] 0.090 0.010 0.000 0.050 0.190 0.210 0.270 0.140 0.036 0.004
[5,] 0.064 0.026 0.090 0.050 0.190 0.210 0.170 0.140 0.036 0.024
[6,] 0.080 0.010 0.150 0.004 0.080 0.260 0.170 0.160 0.066 0.020
[7,] 0.030 0.130 0.080 0.000 0.004 0.130 0.230 0.320 0.050 0.026
[8,] 0.060 0.031 0.009 0.050 0.105 0.045 0.036 0.342 0.268 0.054
[9,] 0.090 0.080 0.040 0.017 0.003 0.000 0.000 0.330 0.280 0.160
[10,] 0.230 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.570 0.200
```

sowie die Anfangsverteilung  $\mu^{(0)} = (0.1, \dots, 0.1)$  gegeben. Dann lässt sich mit Hilfe der Funktion `mk(n,s,mu,P,d)` (Anhang A.1) mit den Parametern  $\mathbf{s} = S$ ,  $P = P$  und  $\mu = \mu^{(0)}$  ( $d = s \in S$  beliebig) folgende Markov Kette simulieren:

```
>P<-matrix(c(0,0,0.3,0.3,0.1,0.1,0,0.1,0.05,0.05,0.31,0.03,0.046,0.09,0.17,0.15,
+0.05,0.05,0.02,0.084,0.08,0.03,0.12,0,0.21,0.19,0.15,0.13,0.054,0.036,0.09,0.01,
+0,0.05,0.19,0.21,0.27,0.14,0.036,0.004,0.064,0.026,0.09,0.05,0.19,0.21,0.17,
+0.14,0.036,0.024,0.08,0.01,0.15,0.004,0.08,0.26,0.17,0.16,0.066,0.02,0.03,0.13,
+0.08,0,0.004,0.13,0.23,0.32,0.05,0.026,0.06,0.031,0.009,0.05,0.105,0.045,0.036,
+0.342,0.268,0.054,0.09,0.08,0.04,0.017,0.003,0,0,0.33,0.28,0.16,0.23,0,0,0,0,0,
+0,0,0.57,0.2),nrow=10,byrow=TRUE)
> s<-c(1:10)
> mu<-rep(0.1,10)

> set.seed(12345)
> mk(50,s,mu,P,1)
 [1] 8 9 9 10 9 2 2 5 7 10 1 3 7 1 4 6 6 6 3 9 8 7 9 9 9
[26] 8 9 8 5 6 8 1 3 7 6 6 8 9 9 2 6 6 9 9 8 7 2 1 3 6
[51] 9
```

Interessant wäre es nun, zu wissen, welchen Anteil die einzelnen Zustände der Kette an der Gesamtlänge  $n$  haben.

Für  $n \geq 0$  sei

$$Y(n) = \frac{1}{n+1} \sum_{i=0}^n 1_{\{s\}}(X_i) \quad \text{für ein } s \in S.$$

Mit Hilfe der Funktion `mk(n,s,mu,P,d)` wird neben der simulierten Markov Kette auch der Verlauf von  $Y(n)$  für den gewählten Zustand  $d = s \in S$  gezeigt. Für die Zustände 1, 2, 4 und 9 z.B. ergeben sich für  $n = 10000$  die auf Seite 21 gezeigten Bilder. Hier scheinen sich die Werte von  $Y(n)$  jeweils mit steigendem  $n$  mehr und mehr an einen festen Wert

anzunähern. Dass dies unter bestimmten Voraussetzungen bei Markov Ketten wirklich der Fall ist und um welchen Wert es sich handelt, wird am Ende dieses Kapitels mit Hilfe des sogenannten Ergodensatzes gezeigt.

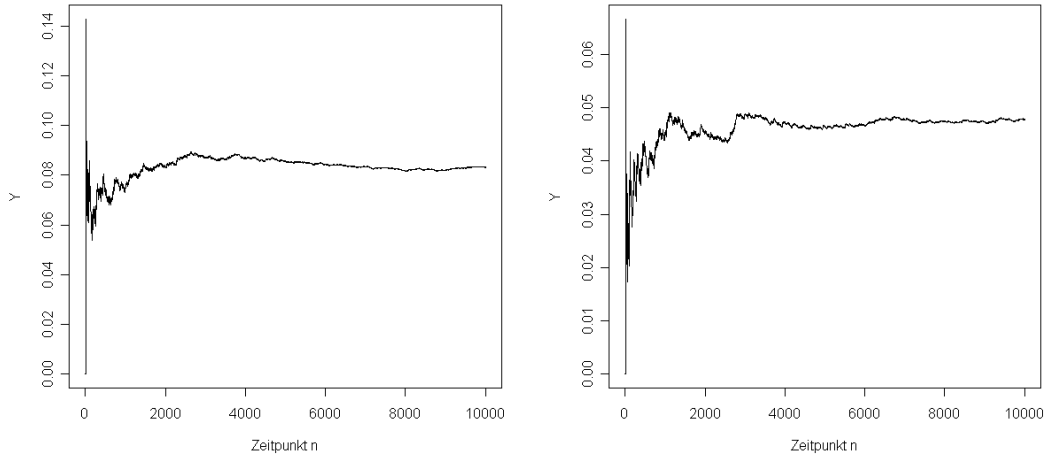


Abbildung 1.1: Verlauf von  $Y(n)$  für  $n = 10000$ , links für  $d=1$  ; rechts für  $d=4$

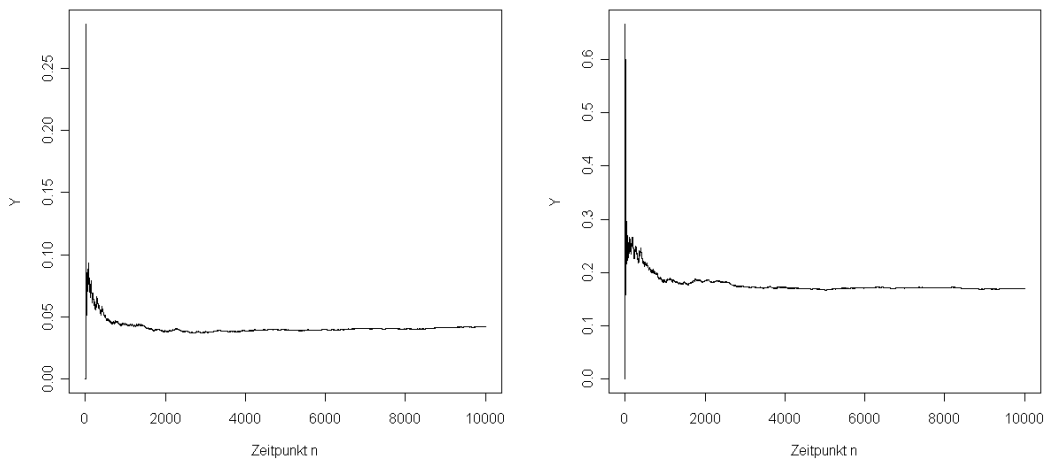


Abbildung 1.2: Verlauf von  $Y(n)$  für  $n = 10000$ , links für  $d=2$  ; rechts für  $d=9$



## 1.3 Starke Markov Eigenschaft

### Definition 1.26.

Eine Zufallsvariable  $T : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$  heißt **Stoppzeit**, wenn für alle  $n \geq 0$  das Ereignis  $\{\omega : T(\omega) = n\}$  nur von  $X_0, \dots, X_n$  abhängt.

### Satz 1.27 (Starke Markov Eigenschaft).

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S$  und Übergangsmatrix  $P$ .  $T$  sei eine Stoppzeit dieser Markov Kette.

Dann gilt für alle  $0 \leq k \leq T$ , alle  $m \in \mathbb{N}$  und jeden Zustand  $s_i \in S$ , gegeben, dass  $X_T = s_i$  ist, und  $P(X_T = s_i, X_{T-1} = s_{i_1}, \dots, X_{T-k} = s_{i_k}) > 0$  vorausgesetzt:

$$\begin{aligned} P(X_{T+1} = s_{j_1}, \dots, X_{T+m} = s_{j_m} \mid X_T = s_i, X_{T-1} = s_{i_1}, \dots, X_{T-k} = s_{i_k}) \\ = P(X_1 = s_{j_1}, \dots, X_m = s_{j_m} \mid X_0 = s_i). \end{aligned}$$

*Beweis.* Nach Multiplikation mit  $P(X_T = s_i, X_{T-1} = s_{i_1}, \dots, X_{T-k} = s_{i_k})$  bleibt folgende Gleichung zu zeigen:

$$\begin{aligned} P(X_{T+1} = s_{j_1}, \dots, X_{T+m} = s_{j_m}, X_T = s_i, X_{T-1} = s_{i_1}, \dots, X_{T-k} = s_{i_k}) \\ = P(X_1 = s_{j_1}, \dots, X_m = s_{j_m} \mid X_0 = s_i) P(X_T = s_i, X_{T-1} = s_{i_1}, \dots, X_{T-k} = s_{i_k}). \end{aligned}$$

Spaltet man die linke Seite dieser Gleichung nach allen (endlichen) Werten auf, die  $T$  annehmen kann, ergibt sich:

$$\begin{aligned} P(X_{T+1} = s_{j_1}, \dots, X_{T+m} = s_{j_m}, X_T = s_i, X_{T-1} = s_{i_1}, \dots, X_{T-k} = s_{i_k}) \\ = \sum_{n \in \mathbb{N}_0} P(T = n, X_{n+1} = s_{j_1}, \dots, X_{n+m} = s_{j_m}, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}). \end{aligned}$$

Da das Ereignis  $\{X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}\} \cap \{T = n\}$  ausschließlich von  $X_0, \dots, X_n$  abhängt, lässt sich Satz 1.5 anwenden:

$$\begin{aligned} P(T = n, X_{n+1} = s_{j_1}, \dots, X_{n+m} = s_{j_m}, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \\ = P(X_{n+1} = s_{j_1}, \dots, X_{n+m} = s_{j_m} \mid T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \\ \cdot P(T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \end{aligned}$$

$$\begin{aligned} \stackrel{\text{Satz}}{\stackrel{1.5}{=}} P(X_{n+1} = s_{j_1}, \dots, X_{n+m} = s_{j_m} \mid T = n, X_n = s_i) \\ \cdot P(T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \\ = \frac{P(T = n, X_n = s_i, X_{n+1} = s_{j_1}, \dots, X_{n+m} = s_{j_m})}{P(T = n, X_n = s_i)} \dots \\ \dots P(T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \end{aligned}$$

$$\begin{aligned}
&\stackrel{\text{Korollar 1.3}}{=} \frac{P(T = n, X_n = s_i)}{P(T = n, X_n = s_i)} P(X_{n+1} = s_{j_1} \mid X_n = s_i, T = n) P(X_{n+2} = s_{j_2} \mid X_{n+1} = s_{j_1}) \cdots \\
&\quad \cdots P(X_{n+m} = s_{j_m} \mid X_{n+m-1} = s_{j_{m-1}}) P(T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \\
&= p_{ij_1} p_{j_1 j_2} \cdots p_{j_{m-1} j_m} P(T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \\
&= \frac{P(X_0 = s_i)}{P(X_0 = s_i)} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{m-1} j_m} P(T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \\
&\stackrel{\text{Korollar 1.3}}{=} \frac{P(X_0 = s_i, X_1 = s_{j_1}, \dots, X_m = s_{j_m})}{P(X_0 = s_i)} P(T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \\
&= P(X_1 = s_{j_1}, \dots, X_m = s_{j_m} \mid X_0 = s_i) P(T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k})
\end{aligned}$$

Durch Summation über  $n \in \mathbb{N}_0$  erhält man schließlich:

$$\begin{aligned}
&P(X_{T+1} = s_{j_1}, \dots, X_{T+m} = s_{j_m}, X_T = s_i, X_{T-1} = s_{i_1}, \dots, X_{T-k} = s_{i_k}) \\
&= \sum_{n \in \mathbb{N}_0} P(T = n, X_n = s_i, X_{n+1} = s_{j_1}, \dots, X_{n+m} = s_{j_m}, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) \\
&= \sum_{n \in \mathbb{N}_0} P(T = n, X_n = s_i, X_{n-1} = s_{i_1}, \dots, X_{n-k} = s_{i_k}) P(X_1 = s_{j_1}, \dots, X_m = s_{j_m} \mid X_0 = s_i) \\
&= P(X_T = s_i, X_{T-1} = s_{i_1}, \dots, X_{T-k} = s_{i_k}) P(X_1 = s_{j_1}, \dots, X_m = s_{j_m} \mid X_0 = s_i)
\end{aligned}$$

[14]

□

## 1.4 Invariante Wahrscheinlichkeitsverteilung und das Grenzwerttheorem

### Definition 1.28.

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ .

Eine Wahrscheinlichkeitsverteilung  $\pi = (\pi_1, \dots, \pi_k)$  heißt **invariante Wahrscheinlichkeitsverteilung** der homogenen Markov Kette, falls gilt:

$$\pi = \pi P, \tag{1.11}$$

d.h.

$$\sum_{i=1}^k \pi_i p_{ij} = \pi_j \quad \text{für } j = 1, \dots, k. \quad (\text{„general balance“})$$

**Definition 1.29.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette wie in Definition 1.10.

Startet die Kette im Zustand  $s_i$  (d.h.  $X_0 = s_i$ ), so ist der **Zeitpunkt des ersten Eintreffens (hitting time)** definiert durch

$$T_{ij} = \min \{n \geq 1 : X_n = s_j\}.$$

Wird der Zustand  $s_j$  nie besucht, so sei  $T_{ij} = \infty$ .

Die erwartete Zeit, die man benötigt um zu Zustand  $s_j$  zu gelangen, nennt man **erwarteter Eintreffszeitpunkt (mean hitting time)** und definiert sie durch

$$\tau_{ij} = E(T_{ij}).$$

Gilt  $i = j$ , so nennt man  $\tau_{ii}$  die **erwartete Rückkehrzeit (mean return time)**.

**Beispiel 1.30.**

Der Zeitpunkt des ersten Eintreffens  $T_{ij}$  ist eine Stoppzeit, da

$$\{T_{ij} = n\} = \{X_0 = s_i, X_1 \neq s_j, \dots, X_{n-1} \neq s_j, X_n = s_j\}$$

gilt.

**Lemma 1.31.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine aperiodische und irreduzible Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ .

Für jedes Paar  $s_i, s_j \in S$  gilt, falls die Kette in  $s_i$  startet:

$$P(T_{ij} < \infty) = 1. \tag{1.12}$$

Ferner gilt:

$$E(T_{ij}) < \infty. \tag{1.13}$$

*Beweis.* Nach Korollar 1.19 existiert ein  $M < \infty$  so, dass  $p_{ij}^{(M)} > 0$  für alle  $i, j \in \{1, \dots, k\}$ . Sei ein solches  $M$  fest gegeben und sei

$$\alpha = \min \{p_{ij}^{(M)} : i, j \in \{1, \dots, k\}\}.$$

Dann gilt  $\alpha > 0$ .

Seien ebenfalls  $s_i, s_j \in S$  fest und es starte die Kette in  $s_i$ . Dann gilt:

$$\begin{aligned} P(T_{ij} > M) &= P(X_1 \neq s_j, X_2 \neq s_j, \dots, X_M \neq s_j) \\ &\leq P(X_M \neq s_j) \\ &= P(X_M \neq s_j \mid X_0 = s_i) \\ &= 1 - P(X_M = s_j \mid X_0 = s_i) \\ &\leq 1 - \alpha \end{aligned}$$

Ferner gilt:

$$\begin{aligned}
& P(T_{ij} > 2M) \\
&= P(X_{2M} \neq s_j, \dots, X_M \neq s_j, \dots, X_1 \neq s_j) \\
&= \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_{2M} \neq s_j, \dots, X_M = s_i, \dots, X_1 \neq s_j) \\
&= \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_{2M} \neq s_j, \dots, X_{M+1} \neq s_j \mid X_M = s_i, \dots, X_1 \neq s_j) P(X_M = s_i, \dots, X_1 \neq s_j) \\
&= P(X_{2M} \neq s_j, \dots, X_{M+1} \neq s_j \mid X_M \neq s_j, \dots, X_1 \neq s_j) \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_M = s_i, \dots, X_1 \neq s_j) \\
&= \frac{P(X_{2M} \neq s_j, \dots, X_{M+1} \neq s_j, X_M \neq s_j, \dots, X_1 \neq s_j)}{P(X_M \neq s_j, \dots, X_1 \neq s_j)} \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_M = s_i, \dots, X_1 \neq s_j) \\
&\leq P(X_{2M} \neq s_j, \dots, X_1 \neq s_j) \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_M = s_i, \dots, X_1 \neq s_j) \\
&\leq P(X_M \neq s_j, \dots, X_1 \neq s_j) \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_M = s_i, \dots, X_1 \neq s_j) \\
&= P(T_{ij} > M) \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_M = s_i, \dots, X_1 \neq s_j) \\
&\leq (1 - \alpha) \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_M = s_i, X_{M-1} \neq s_j, \dots, X_1 \neq s_j) \\
&= (1 - \alpha) P(X_M \neq s_i, \dots, X_1 \neq s_j) \\
&= (1 - \alpha) P(T_{ij} > M) \\
&\leq (1 - \alpha)^2
\end{aligned}$$

Nimmt man nun an, dass

$$P(T_{ij} > (l-1)M) \leq (1 - \alpha)^{l-1} \quad (1.14)$$

gilt, so ergibt sich für jedes  $l$ :

$$\begin{aligned}
& P(T_{ij} > lM) \\
&= P(X_{lM} \neq s_j, \dots, X_{(l-1)M} \neq s_j, \dots, X_1 \neq s_j) \\
&= \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_{lM} \neq s_j, \dots, X_{(l-1)M} = s_i, \dots, X_1 \neq s_j) \\
&= \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_{lM} \neq s_j, \dots, X_{(l-1)M+1} \neq s_j \mid X_{(l-1)M} = s_i, \dots, X_1 \neq s_j) \cdot \\
&\quad \cdot P(X_{(l-1)M} = s_i, \dots, X_1 \neq s_j)
\end{aligned}$$

$$\begin{aligned}
&= P(X_{lM} \neq s_j, \dots, X_{(l-1)M+1} \neq s_j \mid X_{(l-1)M} \neq s_j, \dots, X_1 \neq s_j) \cdot \\
&\quad \cdot \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_{(l-1)M} = s_i, \dots, X_1 \neq s_j) \\
&\leq P(X_{lM} \neq s_j, \dots, X_{(l-1)M+1} \neq s_j, X_{(l-1)M} \neq s_j, \dots, X_1 \neq s_j) \cdot \\
&\quad \cdot \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_{(l-1)M} = s_i, \dots, X_1 \neq s_j) \\
&\leq P(T_{ij} > M) \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_{(l-1)M} = s_i, \dots, X_1 \neq s_j) \\
&\leq (1 - \alpha) \sum_{\substack{s_i \in S \\ s_i \neq s_j}} P(X_{(l-1)M} = s_i, \dots, X_1 \neq s_j) \\
&= (1 - \alpha) P(X_{(l-1)M} \neq s_j, \dots, X_1 \neq s_j) \\
&= (1 - \alpha) P(T_{ij} > (l-1)M) \\
&\stackrel{(1.14)}{\leq} (1 - \alpha)(1 - \alpha)^{l-1} \\
&= (1 - \alpha)^l,
\end{aligned}$$

was mit  $l \rightarrow \infty$  gegen 0 strebt. Hieraus folgt  $P(T_{ij} = \infty) = 0$ , womit (1.12) gezeigt ist. Zum Beweis von (1.13) sei darauf hingewiesen, dass gilt:

$$E(X) = \sum_{k=1}^{\infty} kP(X = k) = \sum_{k=1}^{\infty} P(X \geq k).$$

Man erhalt:

$$\begin{aligned}
E(T_{ij}) &= \sum_{n=1}^{\infty} P(T_{ij} \geq n) = \sum_{n=0}^{\infty} P(T_{ij} > n) & (1.15) \\
&= \sum_{l=0}^{\infty} \sum_{n=lM}^{(l+1)M-1} P(T_{ij} > n) \\
&\leq \sum_{l=0}^{\infty} \sum_{n=lM}^{(l+1)M-1} P(T_{ij} > lM) \\
&= M \sum_{l=0}^{\infty} P(T_{ij} > lM) \\
&\leq M \sum_{l=0}^{\infty} (1 - \alpha)^l = M \frac{1}{1 - (1 - \alpha)} = \frac{M}{\alpha} < \infty.
\end{aligned}$$

[12]

□

**Satz 1.32 (Existenz invarianter Wahrscheinlichkeitsverteilungen).**

Für jede aperiodische, irreduzible Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$  existiert eine invariante Wahrscheinlichkeitsverteilung.

*Beweis.* Bezeichne  $S = \{s_1, \dots, s_k\}$  wie gewöhnlich den Zustandsraum und  $P$  die Übergangsmatrix der Markov Kette. Diese starte in  $s_1$  und es sei

$$\rho_i := \sum_{n=0}^{\infty} P(X_n = s_i, T_{11} > n) \quad \text{für } i=1, \dots, k,$$

d.h.  $\rho_i$  bezeichne die erwartete Anzahl an Besuchen im Zustand  $s_i$  bis zum Zeitpunkt  $T_{11} - 1$ . Da nach Lemma 1.31  $\tau_{11} = E(T_{11}) < \infty$  gilt und

$$\rho_i = \sum_{n=0}^{\infty} P(X_n = s_i, T_{11} > n) < \sum_{n=0}^{\infty} \sum_{j=1}^k P(X_n = s_j, T_{11} > n) = \sum_{n=0}^{\infty} P(T_{11} > n) = E(T_{11})$$

ist, ergibt sich, dass auch  $\rho_i < \infty$  ist.

Für eine invariante Wahrscheinlichkeitsverteilung kommt also in Frage:

$$\pi = (\pi_1, \dots, \pi_k) = \left( \frac{\rho_1}{\tau_{11}}, \dots, \frac{\rho_k}{\tau_{11}} \right).$$

Zu zeigen ist also, dass  $\pi$  zum einen eine Wahrscheinlichkeitsverteilung ist und zum anderen die Eigenschaft (1.11) erfüllt ( $\pi_j = \sum_{i=1}^k \pi_i p_{ij}$ ).

Offensichtlich gilt  $\pi_i \geq 0$  für  $i = 1, \dots, k$ .

Zudem gilt:

$$\begin{aligned} \tau_{11} = E(T_{11}) &\stackrel{(1.15)}{=} \sum_{n=0}^{\infty} P(T_{11} > n) \\ &= \sum_{n=0}^{\infty} \sum_{i=1}^k P(X_n = s_i, T_{11} > n) \\ &= \sum_{i=1}^k \sum_{n=0}^{\infty} P(X_n = s_i, T_{11} > n) \\ &= \sum_{i=1}^k \rho_i, \end{aligned}$$

womit sich ergibt:

$$\sum_{i=1}^k \pi_i = \frac{1}{\tau_{11}} \sum_{i=1}^k \rho_i = 1.$$

$\pi$  erfüllt somit die Eigenschaften einer Wahrscheinlichkeitsverteilung.  
Zum Beweis der Eigenschaft (1.11) sei zunächst  $j \neq 1$ . Dann gilt:

$$\begin{aligned}
\pi_j &= \frac{\rho_j}{\tau_{11}} \\
&= \frac{1}{\tau_{11}} \sum_{n=0}^{\infty} P(X_n = s_j, T_{11} > n) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} P(X_n = s_j, T_{11} \geq n) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} P(X_n = s_j, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} \sum_{i=2}^k P(X_{n-1} = s_i, X_n = s_j, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} \sum_{i=2}^k P(X_{n-1} = s_i, T_{11} > n-1) P(X_n = s_j \mid X_{n-1} = s_i, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} \sum_{i=2}^k P(X_{n-1} = s_i, T_{11} > n-1) P(X_n = s_j \mid X_{n-1} = s_i, X_{n-2} \neq s_1, \dots, X_1 \neq s_1) \\
&\stackrel{(1.2)}{=} \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} \sum_{i=2}^k P(X_{n-1} = s_i, T_{11} > n-1) P(X_n = s_j \mid X_{n-1} = s_i) \\
&= \frac{1}{\tau_{11}} \sum_{n=1}^{\infty} \sum_{i=2}^k p_{ij} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{i=2}^k p_{ij} \sum_{n=1}^{\infty} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \left( p_{1j} \underbrace{\sum_{n=1}^{\infty} P(X_{n-1} = s_1, T_{11} > n-1)}_{=0} + \sum_{i=2}^k p_{ij} \sum_{n=1}^{\infty} P(X_{n-1} = s_i, T_{11} > n-1) \right) \\
&= \frac{1}{\tau_{11}} \sum_{i=1}^k p_{ij} \sum_{n=1}^{\infty} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \frac{1}{\tau_{11}} \sum_{i=1}^k p_{ij} \sum_{m=0}^{\infty} P(X_m = s_i, T_{11} > m) \\
&= \sum_{i=1}^k \frac{1}{\tau_{11}} \rho_i p_{ij} = \sum_{i=1}^k \pi_i p_{ij} \tag{1.16}
\end{aligned}$$

Um nun auch den Fall  $j = 1$  zu beweisen, sei bemerkt, dass aus der Definition von  $\rho_i$

direkt folgt:

$$\begin{aligned}
\rho_1 &= \sum_{n=0}^{\infty} P(X_n = s_1, T_{11} > n) \\
&= \sum_{n=1}^{\infty} P(X_n = s_1, X_n \neq s_1, X_{n-1} \neq s_1, \dots, X_1 \neq s_1) + P(X_0 = s_1, T_{11} > 0) \\
&= \underbrace{P(X_0 = s_1, T_{11} > 0)}_{=1} + \underbrace{P(X_1 = s_1, X_1 \neq s_1)}_{=0} + \underbrace{P(X_2 = s_1, X_2 \neq s_1, X_1 \neq s_1)}_{=0} + \dots \\
&\quad + \underbrace{P(X_j = s_1, X_j \neq s_1, X_{j-1} \neq s_1, \dots, X_1 \neq s_1)}_{=0} + \dots \\
&= 1
\end{aligned}$$

Damit erhält man:

$$\begin{aligned}
\rho_1 = 1 &\stackrel{(1.12)}{=} P(T_{11} < \infty) = \sum_{n=1}^{\infty} P(T_{11} = n) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^k P(X_{n-1} = s_i, T_{11} = n) \\
&= \sum_{i=1}^k \underbrace{P(X_0 = s_i, T_{11} = 1)}_{=0 \text{ für } i=2, \dots, k} + \sum_{n=2}^{\infty} \sum_{i=1}^k P(X_{n-1} = s_i, T_{11} = n) \\
&= P(X_0 = s_1, T_{11} = 1) + \sum_{n=2}^{\infty} \left( \sum_{i=2}^k P(X_{n-1} = s_i, T_{11} = n) + \underbrace{P(X_{n-1} = s_1, T_{11} = n)}_{=0 \text{ für } n=2, \dots, \infty} \right) \\
&= P(X_0 = s_1, X_1 = s_1) + \sum_{n=2}^{\infty} \sum_{i=2}^k P(X_{n-1} = s_i, T_{11} = n) \\
&= p_{11} + \sum_{n=2}^{\infty} \sum_{i=2}^k P(X_{n-1} = s_i, X_n = s_1, X_{n-1} \neq s_1, \dots, X_1 \neq s_1) \\
&= p_{11} + \sum_{n=2}^{\infty} \sum_{i=2}^k P(X_{n-1} = s_i, T_{11} > n-1) P(X_n = s_1 \mid X_{n-1} = s_i, \dots, X_1 \neq s_1) \\
&\stackrel{(1.2)}{=} p_{11} + \sum_{n=2}^{\infty} \sum_{i=2}^k P(X_{n-1} = s_i, T_{11} > n-1) P(X_n = s_1 \mid X_{n-1} = s_i) \tag{1.17} \\
&= p_{11} + \sum_{n=2}^{\infty} \sum_{i=2}^k p_{i1} P(X_{n-1} = s_i, T_{11} > n-1)
\end{aligned}$$



$$\begin{aligned}
&= p_{11} \left[ \sum_{n=2}^{\infty} \underbrace{P(X_{n-1} = s_1, T_{11} > n-1)}_{=0 \text{ für } n=2, \dots, \infty} + \underbrace{P(X_0 = s_1, T_{11} > 0)}_{=1} \right] \\
&\quad + \sum_{i=2}^k p_{i1} \sum_{n=2}^{\infty} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= p_{11} \sum_{n=1}^{\infty} P(X_{n-1} = s_1, T_{11} > n-1) \\
&\quad + \sum_{i=2}^k p_{i1} \sum_{n=2}^{\infty} P(X_{n-1} = s_i, T_{11} > n-1) + \underbrace{P(X_0 = s_i, T_{11} > 0)}_{=0 \text{ für } i=2, \dots, k} \\
&= p_{11} \sum_{n=1}^{\infty} P(X_{n-1} = s_1, T_{11} > n-1) + \sum_{i=2}^k p_{i1} \sum_{n=1}^{\infty} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \sum_{i=1}^k p_{i1} \sum_{n=1}^{\infty} P(X_{n-1} = s_i, T_{11} > n-1) \\
&= \sum_{i=1}^k p_{i1} \sum_{m=0}^{\infty} P(X_m = s_i, T_{11} > m) \\
&= \sum_{i=1}^k \rho_i p_{i1} \tag{1.18}
\end{aligned}$$

wobei man in (1.17)  $X_{n-1} \neq s_1$  vernachlässigen kann, da ausschließlich über  $i = 2, \dots, k$  summiert wird.

Mit (1.18) folgt schließlich:

$$\pi_1 = \frac{\rho_1}{\tau_{11}} = \sum_{i=1}^k \frac{1}{\tau_{11}} \rho_i p_{i1} = \sum_{i=1}^k \pi_i p_{i1}$$

und zusammen mit (1.16) ist somit (1.11) für  $\pi$  erfüllt. [12] □

### Definition 1.33.

Seien  $\nu^{(1)} = (\nu_1^{(1)}, \dots, \nu_k^{(1)})$ ,  $\nu^{(2)} = (\nu_1^{(2)}, \dots, \nu_k^{(2)})$  und  $\nu = (\nu_1, \dots, \nu_k)$  Wahrscheinlichkeitsverteilungen auf  $S = \{s_1, \dots, s_k\}$ .

Der **totale Variationsabstand** zwischen  $\nu^{(1)}$  und  $\nu^{(2)}$  ist definiert als

$$d_{TV}(\nu^{(1)}, \nu^{(2)}) = \frac{1}{2} \sum_{i=1}^k |\nu_i^{(1)} - \nu_i^{(2)}|.$$

Durch  $d_{TV}$  wird eine Metrik auf Wahrscheinlichkeitsverteilungen beschrieben, weswegen man folgendes definieren kann:

$\nu^{(n)}$  konvergiert mit  $n \rightarrow \infty$  in **totaler Variation** gegen  $\nu$  ( $\nu^{(n)} \xrightarrow{TV} \nu$ ), falls

$$\lim_{n \rightarrow \infty} d_{TV}(\nu^{(n)}, \nu) = 0.$$

Wegen der Konstanten  $\frac{1}{2}$  nimmt der **totale Variationsabstand**  $d_{TV}$  ausschließlich Werte aus dem Intervall  $[0, 1]$  an.

Gilt  $d_{TV}(\nu^{(1)}, \nu^{(2)}) = 0$ , so ist  $\nu^{(1)} = \nu^{(2)}$ .

Gilt  $d_{TV}(\nu^{(1)}, \nu^{(2)}) = 1$ , so sind  $\nu^{(1)}$  und  $\nu^{(2)}$  disjunkt in dem Sinne, dass man  $S$  in zwei disjunkte Untermengen  $S'$  und  $S''$  teilen kann, so dass  $\nu^{(1)}$  sich vollständig in  $S'$  und  $\nu^{(2)}$  sein gesamtes Wahrscheinlichkeitsmaß in  $S''$  erstreckt. (vgl. [12, S.34])

**Satz 1.34 (Grenzwerttheorem für Markov Ketten).**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine aperiodische und irreduzible Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$ , Übergangsmatrix  $P$  und beliebiger Startverteilung  $\mu^{(0)}$ .

Dann gilt für jede invariante Wahrscheinlichkeitsverteilung  $\pi$  der Markov Kette:

$$\mu^{(n)} \xrightarrow{TV} \pi.$$

*Beweis.*  $\{X_t; t \in \mathbb{Z}^+\}$  habe man durch Simulation erhalten (siehe Abschnitt 1.2).

Also:

$$\begin{aligned} X_0 &= \psi_{\mu^{(0)}}(U_0) \\ X_1 &= \phi(X_0, U_1) \\ X_2 &= \phi(X_1, U_2) \\ &\vdots \end{aligned}$$

wobei  $\psi_{\mu^{(0)}}$  eine gültige Einführungsfunktion für  $\mu^{(0)}$  und  $\phi$  eine gültige Update-Funktion für  $P$  darstellt.  $\{U_t; t \in \mathbb{Z}^+\}$  ist dabei eine unabhängige und identisch verteilte Folge von Zufallsvariablen gleichverteilt auf  $[0, 1]$ .

Sei eine zweite Markov Kette  $\{X'_t; t \in \mathbb{Z}^+\}$  erzeugt worden mit der gültigen Einführungsfunktion  $\psi_\pi$  für die Verteilung  $\pi$  und einer weiteren unabhängigen und identisch verteilten Folge  $\{U'_t; t \in \mathbb{Z}^+\}$  von Zufallsvariablen gleichverteilt auf  $[0, 1]$  (unabhängig von  $\{U_t; t \in \mathbb{Z}^+\}$ ).

D.h.:

$$\begin{aligned} X'_0 &= \psi_\pi(U'_0) \\ X'_1 &= \phi(X'_0, U'_1) \\ X'_2 &= \phi(X'_1, U'_2) \\ &\vdots \end{aligned}$$

Es gilt:

$X'_n$  ist für jedes  $n$   $\pi$ -verteilt, da  $\pi$  nach Voraussetzung eine invariante Wahrscheinlichkeitsverteilung ist. Und weil  $\{U_t; t \in \mathbb{Z}^+\}$  und  $\{U'_t; t \in \mathbb{Z}^+\}$  als voneinander unabhängig angenommen wurden, sind auch  $\{X_t; t \in \mathbb{Z}^+\}$  und  $\{X'_t; t \in \mathbb{Z}^+\}$  voneinander unabhängig.

Zu zeigen ist nun, dass es ein  $n$  gibt so, dass  $X_n = X'_n$  (, d.h. dass die beiden Ketten mit der Wahrscheinlichkeit 1 aufeinander treffen).

Hierfür sei der Zeitpunkt des ersten Treffens definiert durch:

$$T = \min \{n : X_n = X'_n\}.$$

Es gelte  $T = \infty$ , falls die beiden Ketten nie aufeinander treffen.

Da  $\{X_t; t \in \mathbb{Z}^+\}$  aperiodisch und irreduzibel ist, findet man mit Hilfe von Korollar 1.19 ein  $M < \infty$  so, dass

$$p_{ij}^{(M)} > 0 \quad \text{für alle } i, j \in \{1, \dots, k\}.$$

Sei

$$\alpha = \min \{p_{ij}^{(M)} : i \in \{1, \dots, k\}\}.$$

Es gilt offensichtlich  $\alpha > 0$ .

Man erhält:

$$\begin{aligned} P(T \leq M) &\geq P(X_M = X'_M) \\ &\geq P(X_M = s_1, X'_M = s_1) \\ &= P(X_M = s_1)P(X'_M = s_1) \\ &= \left( \sum_{i=1}^k P(X_0 = s_i, X_M = s_1) \right) \left( \sum_{i=1}^k P(X'_0 = s_i, X'_M = s_1) \right) \\ &= \left( \sum_{i=1}^k P(X_0 = s_i)P(X_M = s_1 | X_0 = s_i) \right) \left( \sum_{i=1}^k P(X'_0 = s_i)P(X'_M = s_1 | X'_0 = s_i) \right) \\ &\geq \left( \alpha \sum_{i=1}^k P(X_0 = s_i) \right) \left( \alpha \sum_{i=1}^k P(X'_0 = s_i) \right) = \alpha^2 \end{aligned}$$

und somit

$$P(T > M) \leq 1 - \alpha^2.$$

Damit folgt:

$$\begin{aligned} P(T > 2M) &= P(X_{2M} \neq X'_{2M}, \dots, X_0 \neq X'_0) \\ &= \frac{P(X_{2M} \neq X'_{2M}, \dots, X_0 \neq X'_0)}{P(X_M \neq X'_M, \dots, X_0 \neq X'_0)} \cdot P(X_M \neq X'_M, \dots, X_0 \neq X'_0) \\ &\leq P(X_{2M} \neq X'_{2M}, \dots, X_0 \neq X'_0)P(X_M \neq X'_M, \dots, X_0 \neq X'_0) \\ &\leq P(X_M \neq X'_M, \dots, X_0 \neq X'_0)P(X_M \neq X'_M, \dots, X_0 \neq X'_0) \\ &= P(T > M)P(X_M \neq X'_M, \dots, X_0 \neq X'_0) \\ &\leq (1 - \alpha^2)P(X_M \neq X'_M, \dots, X_0 \neq X'_0) \\ &= (1 - \alpha^2)P(T > M) \\ &\leq (1 - \alpha^2)^2. \end{aligned}$$

Unter der Annahme, dass

$$P(T > (l-1)M) \leq (1 - \alpha^2)^{l-1} \quad (1.19)$$

gilt, ergibt sich analog für jedes  $l$ :

$$\begin{aligned} P(T > lM) &= P(X_{lM} \neq X'_{lM}, \dots, X_{(l-1)M} \neq X'_{(l-1)M}, \dots, X_0 \neq X'_0) \\ &= \frac{P(X_{lM} \neq X'_{lM}, \dots, X_0 \neq X'_0)}{P(X_{(l-1)M} \neq X'_{(l-1)M}, \dots, X_0 \neq X'_0)} \cdot P(X_{(l-1)M} \neq X'_{(l-1)M}, \dots, X_0 \neq X'_0) \\ &\leq P(X_{lM} \neq X'_{lM}, \dots, X_0 \neq X'_0) P(X_{(l-1)M} \neq X'_{(l-1)M}, \dots, X_0 \neq X'_0) \\ &\leq P(T > M) P(X_{(l-1)M} \neq X'_{(l-1)M}, \dots, X_0 \neq X'_0) \\ &\leq (1 - \alpha^2) P(X_{(l-1)M} \neq X'_{(l-1)M}, \dots, X_0 \neq X'_0) \\ &= (1 - \alpha^2) P(T_{ij} > (l-1)M) \\ &\leq (1 - \alpha^2)(1 - \alpha^2)^{l-1} \\ &= (1 - \alpha^2)^l, \end{aligned}$$

was mit  $l \rightarrow \infty$  gegen 0 strebt. Also:

$$\lim_{n \rightarrow \infty} P(T > n) = 0, \quad (1.20)$$

was bedeutet, dass die beiden Markov Ketten mit der Wahrscheinlichkeit 1 aufeinander treffen.

Als nächstes wird eine dritte Markov Kette  $\{X''_t; t \in \mathbb{Z}^+\}$  konstruiert. Hierfür soll gelten:

$$X''_0 = X_0$$

und für jedes  $n$ :

$$X''_{n+1} = \begin{cases} \phi(X''_n, U_{n+1}) & , \text{ falls } X''_n \neq X'_n \\ \phi(X''_n, U'_{n+1}) & , \text{ falls } X''_n = X'_n. \end{cases}$$

D.h. die Kette  $\{X''_t; t \in \mathbb{Z}^+\}$  verhält sich exakt wie die Kette  $\{X_t; t \in \mathbb{Z}^+\}$ , bis sie zum Zeitpunkt  $T$  auf die Kette  $\{X'_t; t \in \mathbb{Z}^+\}$  trifft. Ab diesem Zeitpunkt nimmt sie das Verhalten der Kette  $\{X'_t; t \in \mathbb{Z}^+\}$  an. Diese dritte Kette ist tatsächlich eine Markov Kette mit Übergangsmatrix  $P$ , weil es sich bei den verwendeten Zufallsvariablen um „neue“ auf  $[0, 1]$  gleichverteilte Variablen handelt. Sie sind somit unabhängig von allen vorherigen Zufallsvariablen. Ob die neue Kette sich auf  $U_{n+1}$  oder  $U'_{n+1}$  bezieht, hängt zwar von den vorherigen Zufallsvariablen ab, dies spielt jedoch keine Rolle, da  $U_{n+1}$  und  $U'_{n+1}$  zum einen dieselbe Verteilung vorweisen und zum anderen unabhängig von allen vorherigen Werten bis zum Zeitpunkt  $n$  sind. (vgl.[12, S. 37]) Wegen  $X''_0 = X_0$ , ist  $X''_0$   $\mu^{(0)}$ -verteilt und somit hat  $X''_n$  für jedes  $n$  die Verteilung  $\mu^{(n)}$ .

Hiermit erhält man nun für jedes  $i \in \{1, \dots, k\}$ :

$$\begin{aligned} \mu_i^{(n)} - \pi_i &= P(X''_n = s_i) - P(X'_n = s_i) \\ &\leq P(X''_n = s_i) - P(X''_n = s_i, X'_n = s_i) \\ &= P(X''_n = s_i, X'_n \neq s_i) \\ &\leq P(X''_n \neq X'_n) \\ &= P(T > n), \end{aligned}$$

was nach (1.20) mit  $n \rightarrow \infty$  gegen 0 strebt.  
 Auf dieselbe Weise ergibt sich:

$$\pi_i - \mu_i^{(n)} \leq P(T > n),$$

was ebenfalls mit  $n \rightarrow \infty$  gegen 0 strebt.  
 Hieraus folgt schließlich

$$\lim_{n \rightarrow \infty} |\mu_i^{(n)} - \pi_i| = 0,$$

was wiederum bedeutet, dass

$$\lim_{n \rightarrow \infty} d_{TV}(\mu^{(n)}, \pi) = \lim_{n \rightarrow \infty} \frac{1}{2} \sum_{i=1}^k |\mu_i^{(n)} - \pi_i| = 0$$

und somit  $\mu^{(n)} \xrightarrow{TV} \pi$  gilt. [12] □

**Satz 1.35 (Eindeutigkeit der invarianten Wahrscheinlichkeitsverteilung).**

Jede aperiodische, irreduzible Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$  besitzt genau eine invariante Wahrscheinlichkeitsverteilung.

*Beweis.* Sei  $P$  die Übergangsmatrix der Markov Kette. Nach Satz 1.32 existiert mindestens eine invariante Wahrscheinlichkeitsverteilung der gegebenen Markov Kette.

Seien  $\pi$  und  $\pi'$  zwei (möglicherweise verschiedene) invariante Wahrscheinlichkeitsverteilungen von  $\{X_t; t \in \mathbb{Z}^+\}$ . Angenommen, die Kette startet mit der Startverteilung  $\mu^{(0)} = \pi'$ . Da  $\pi'$  invariant ist, gilt:

$$\mu^{(1)} = \mu^{(0)}P = \pi'P = \pi'.$$

Durch Iteration erhält man  $\mu^{(n)} = \pi'$  für alle  $n \in \mathbb{Z}^+$ . Nach Satz 1.34 gilt auf der anderen Seite jedoch:

$$\mu^{(n)} \xrightarrow{TV} \pi,$$

d.h.

$$\lim_{n \rightarrow \infty} d_{TV}(\mu^{(n)}, \pi) = 0.$$

Da aber  $\mu^{(n)} = \pi'$  ist, ergibt sich:

$$\lim_{n \rightarrow \infty} d_{TV}(\pi', \pi) = 0.$$

Und weil  $d_{TV}(\pi', \pi)$  nicht von  $n$  abhängt, gilt  $d_{TV}(\pi', \pi) = 0$ , was bedeutet, dass  $\pi' = \pi$  ist. [12] □

**Definition 1.36.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ .

Eine Wahrscheinlichkeitsverteilung  $\pi$  auf  $S$  heißt **reversibel** für die Markov Kette (oder für die Übergangsmatrix  $P$ ), falls für alle  $i, j \in \{1, \dots, k\}$  gilt

$$\pi_i p_{ij} = \pi_j p_{ji} \quad (\text{„detailed balance“}) \quad (1.21)$$

Die Markov Kette heißt reversibel, falls eine reversible Verteilung für sie existiert.

**Satz 1.37.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ .

Ist  $\pi$  eine reversible Verteilung der Markov Kette, so ist  $\pi$  ebenfalls eine invariante Wahrscheinlichkeitsverteilung dieser Kette.

*Beweis.* Es gilt:

$$\pi_j = \pi_j \sum_{i=1}^k p_{ji} = \sum_{i=1}^k \pi_j p_{ji} \stackrel{(1.21)}{=} \sum_{i=1}^k \pi_i p_{ij}.$$

[12]

□

## 1.5 Rekurrenz und Transienz

**Definition 1.38.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$ .

Ein Zustand  $s_i \in S$  heißt **rekurrent**, wenn

$$P(X_n = s_i \text{ für } \infty\text{-viele } n \mid X_0 = s_i) = 1$$

gilt und er heißt **transient**, wenn

$$P(X_n = s_i \text{ für } \infty\text{-viele } n \mid X_0 = s_i) = 0$$

ist.

Ein rekurrenter Zustand heißt **positiv rekurrent**, wenn

$$E(T_{ii}) < \infty$$

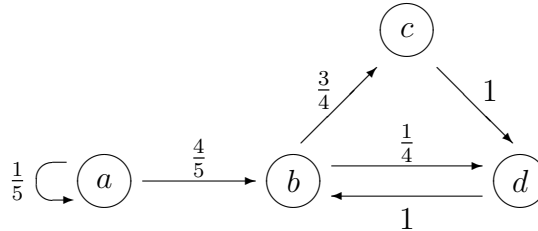
gilt. Anderenfalls nennt man  $s_i$  **null rekurrent**.

Eine Markov Kette heißt (positiv) rekurrent bzw. (null rekurrent) transient, wenn jeder ihrer Zustände (positiv) rekurrent bzw. (null rekurrent) transient ist.

**Beispiel 1.39.**

Ein Zustand ist also rekurrent, wenn man unendlich oft zu ihm zurückkehrt. Transient bedeutet, dass man diesen Zustand eventuell für immer verlässt.

Folgendes Diagramm veranschaulicht dies:



Bei Zustand  $a$  handelt es sich hier z.B. um einen transienten Zustand, der nicht mehr erreicht wird, wenn er ersteinmal verlassen wurde. Es gilt also

$$P(X_n = a \quad \text{für } \infty\text{-viele } n \mid X_0 = a) = 0$$

Die Zustände  $b, c$  und  $d$  sind dagegen rekurrent, da man immer wieder zu ihnen zurückkehrt, d.h. für  $s = a, b, d$  gilt

$$P(X_n = s \quad \text{für } \infty\text{-viele } n \mid X_0 = s) = 1.$$

**Definition 1.40.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$ . Die Kette starte in  $s_i$ .

Entsprechend des ersten Eintreffzeitpunkts (hitting time)

$$T_{ij} = \min\{n \geq 1 : X_n = s_j\}$$

ist der **Zeitpunkt des r-ten Eintreffens (rth passage time)** definiert durch:

$$T_{ij}^{(0)} = 0, \quad T_{ij}^{(1)} = T_{ij}$$

und, für  $r = 0, 1, 2, \dots$ ,

$$T_{ij}^{(r+1)} = \min\{n \geq T_{ij}^{(r)} + 1 : X_n = s_j\}.$$

Entsprechend der letzten Definition beträgt die Länge des r-ten „Ausflugs“:

$$S_{ij}^{(r)} = \begin{cases} T_{ij}^{(r)} - T_{ij}^{(r-1)} & \text{falls } T_{ij}^{(r-1)} < \infty \\ 0 & \text{sonst} \end{cases}$$

**Lemma 1.41.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$  und es sei  $X_0 = s_i$ .

Für  $r = 2, 3, \dots$  ist  $S_{ij}^{(r)}, T_{ij}^{(r-1)} < \infty$  gegeben, unabhängig von  $\{X_m : m \leq T_{ij}^{(r-1)}\}$  und es gilt:

$$P(S_{ij}^{(r)} = n \mid T_{ij}^{(r-1)} < \infty) = P(T_{jj} = n).$$

*Beweis.* Es sei  $T = T_{ij}^{(r-1)}$ . Dann ist  $T$  eine Stoppzeit von  $\{X_t; t \in \mathbb{Z}^+\}$  und mit  $T < \infty$  gilt automatisch  $X_T = s_j$ . Mit Hilfe der starken Markov Eigenschaft (Satz 1.27) ist  $\{X_{T+t}; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Übergangsmatrix  $P$  und unabhängig von  $X_0, X_1, \dots, X_T$ . Folgend aus der Definition gilt aber

$$S_{ij}^{(r)} = \inf \{n \geq 1 : X_{T+n} = s_j\}$$

und somit ist  $S_{ij}^{(r)}$  der Zeitpunkt des ersten Eintreffens der Kette  $\{X_{T+t}; t \in \mathbb{Z}^+\}$  im Zustand  $s_j$ . Damit ergibt sich:

$$P(S_{ij}^{(r)} = n \mid T_{ij}^{(r-1)} < \infty) = P(T_{jj} = n).$$

[20]

□

**Definition 1.42.**

Für die Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$  mit Zustandsraum  $S$  bezeichne

$$V_i := \sum_{l=0}^{\infty} 1_{\{s_i\}}(X_l)$$

die Anzahl der Besuche im Zustand  $s_i \in S$ .

**Lemma 1.43.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ . Die Kette starte in  $s_i$ .

Ist weiter  $f_i := P(T_{ii} < \infty)$ , dann gilt für  $r = 0, 1, 2, \dots$ :

$$P(V_{ii} > r) = P(T_{ii}^{(r)} < \infty) = f_i^r.$$

*Beweis.* Da  $X_0 = s_i$  gilt, ist auch  $\{V_{ii} > r\} = \{T_{ii}^{(r)} < \infty\}$ .

Für  $r = 0$  ist somit

$$1 = P(V_{ii} > 0) = P(T_{ii}^{(0)} < \infty) = f_i^0$$



Man zeige nun die allgemeine Aussage mit Hilfe der Induktion:

Es gelte  $P(V_{ii} > r) = P(T_{ii}^{(r)} < \infty) = f_i^r$  für jedes  $r$ .

Dann gilt:

$$\begin{aligned}
 P(V_{ii} > r + 1) &= P(T_{ii}^{(r+1)} < \infty) \\
 &= P(T_{ii}^{(r)} < \infty, S_{ii}^{(r+1)} < \infty) \\
 &= P(S_{ii}^{(r+1)} < \infty \mid T_{ii}^{(r)} < \infty) P(T_{ii}^{(r)} < \infty) \\
 &\stackrel{\text{Lemma 1.41}}{=} P(T_{ii} < \infty) P(T_{ii}^{(r)} < \infty) \\
 &= f_i f_i^r \\
 &= f_i^{r+1}
 \end{aligned}$$

[20]

□

**Satz 1.44.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$  und die Kette starte in  $s_i$ .

Dann gelten folgende Aussagen:

- (i) ist  $P(T_{ii} < \infty) = 1$ , dann ist  $s_i$  rekurrent und es gilt  $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$ ;
- (ii) ist  $P(T_{ii} < \infty) < 1$ , dann ist  $s_i$  transient und es gilt  $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$ .

Insbesondere ist jeder Zustand entweder rekurrent oder transient.

*Beweis.* Ist

$$P(T_{ii} < \infty) = 1 \tag{1.22}$$

so gilt mit Hilfe von Lemma 1.43

$$P(V_{ii} = \infty) = \lim_{r \rightarrow \infty} P(V_{ii} > r) = \lim_{r \rightarrow \infty} P(T_{ii}^{(r)} < \infty) = \lim_{r \rightarrow \infty} f_i^r = \lim_{r \rightarrow \infty} P(T_{ii} < \infty)^r \stackrel{(1.22)}{=} 1.$$

$s_i$  ist somit rekurrent und es gilt weiter:

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \sum_{n=0}^{\infty} P(X_n = s_i \mid X_0 = s_i) = \sum_{n=0}^{\infty} E(1_{\{s_i\}}(X_n)) = E\left(\sum_{n=0}^{\infty} 1_{\{s_i\}}(X_n)\right) = E(V_{ii}) = \infty.$$

Ist andererseits

$$P(T_{ii} < \infty) < 1,$$

dann gilt mit Lemma 1.43:

$$\begin{aligned}
\sum_{n=0}^{\infty} p_{ii}^{(n)} &= E(V_{ii}) \\
&= \sum_{l=1}^{\infty} lP(V_{ii} = l) \\
&= \sum_{l=1}^{\infty} \sum_{r=0}^{l-1} P(V_{ii} = l) \\
&= \sum_{r=0}^{\infty} \sum_{l=r+1}^{\infty} P(V_{ii} = l) \\
&= \sum_{r=0}^{\infty} P(V_{ii} > r) \\
&= \sum_{r=0}^{\infty} f_i^r = \frac{1}{1 - f_i} < \infty
\end{aligned}$$

und somit ist  $P(V_{ii} = \infty) = 0$  und der Zustand  $s_i$  ist transient. [20] □

**Satz 1.45.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ .

Wenn  $s_i \in S$  und  $s_j \in S$  miteinander kommunizieren, dann sind sie entweder beide rekurrent oder beide sind transient.

*Beweis.* Nach Definition 1.11 kommunizieren die beiden Zustände  $s_i, s_j \in S$  miteinander, wenn es ein  $m > 0$  und ein  $n > 0$  gibt so, dass  $p_{ij}^{(m)} > 0$  und  $p_{ji}^{(n)} > 0$  gilt.

Geht man von  $s_i$  nach  $s_j$  in  $m$  Schritten, danach von  $s_j$  nach  $s_j$  in  $N$  Schritten und anschließend von  $s_j$  nach  $s_i$  in  $n$  Schritten, so hat man einen Weg von  $s_i$  zurück nach  $s_i$  in  $m + N + n$  Schritten. Aus diesem Grund gilt mit Hilfe der Chapman-Kolmogorow-Gleichung (1.3):

$$p_{ii}^{(m+N+n)} = \sum_{s_j \in S} p_{ij}^{(m)} p_{ji}^{(N+n)} = \sum_{s_j \in S} \sum_{s_l \in S} p_{ij}^{(m)} p_{jl}^{(N)} p_{li}^{(n)} \geq p_{ij}^{(m)} p_{jj}^{(N)} p_{ji}^{(n)} \quad \text{für alle } N \geq 0.$$

Und daraus ergibt sich

$$\sum_{N=0}^{\infty} p_{jj}^{(N)} \leq \frac{1}{p_{ij}^{(m)} p_{ji}^{(n)}} \sum_{N=0}^{\infty} p_{ii}^{(m+N+n)}.$$

Sei nun  $s_i$  transient. Dann gilt mit Satz 1.44

$$\sum_{N=0}^{\infty} p_{jj}^{(N)} \leq \frac{1}{p_{ij}^{(m)} p_{ji}^{(n)}} \sum_{N=0}^{\infty} p_{ii}^{(m+N+n)} < \infty.$$

Nach Satz 1.44 ist somit  $s_j$  ebenfalls transient.

Ersetzt man „<“ durch „=“, ergibt sich mit Satz 1.44 entsprechendes für die Rekurrenz. [3], [20]  $\square$

**Definition 1.46.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette. Für einen festen Zustand  $s_r$  sei die erwartete Anzahl der Zeitpunkte, die zwischen den Besuchen in  $s_r$  in  $s_i$  verbracht werden, bezeichnet durch:

$$\gamma_i^r := E \left( \sum_{n=0}^{T_{rr}-1} 1_{\{s_i\}}(X_n) \right).$$

D.h.  $\gamma_i^r$  bezeichnet die erwartete Anzahl der Besuche in  $s_i$  vor dem ersten Eintreffen in  $s_r$ .

**Lemma 1.47.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine irreduzible und rekurrente Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ . Und es sei  $s_r \in S$  beliebig aber fest.

Dann gilt:

- (i)  $\gamma_r^r = 1$  ;
- (ii)  $\gamma^r = (\gamma_i^r : i \in \{1, \dots, k\})$  genügt der Gleichung  $\gamma^r P = \gamma^r$ ;
- (iii)  $0 < \gamma_i^r < \infty$  für alle  $i \in \{1, \dots, k\}$ .

*Beweis.* (i) Da  $\{T_{rr} = n\} = \{X_0 = s_r, X_1 \neq s_r, \dots, X_{n-1} \neq s_r, X_n = s_r\}$  gilt, ist

$$\gamma_r^r = E \left( \sum_{n=0}^{T_{rr}-1} 1_{\{s_r\}}(X_n) \right) = 1$$

offensichtlich.

(ii) a) Sei zunächst  $s_r \neq s_j$ .

Für  $n = 2, \dots$  gilt dann:

$$\begin{aligned} & P(X_{n-1} = s_i, X_n = s_j, T_{rr} \geq n \mid X_0 = s_r) \\ &= \frac{P(X_{n-1} = s_i, X_n = s_j, X_0 = s_r, T_{rr} \geq n)}{P(X_0 = s_r)} \\ &= \frac{P(X_n = s_j \mid X_{n-1} = s_i, X_0 = s_r, T_{rr} > n-1) P(X_{n-1} = s_i, X_0 = s_r, T_{rr} > n-1)}{P(X_0 = s_r)} \\ &\stackrel{(1.2)}{=} \frac{P(X_n = s_j \mid X_{n-1} = s_i) P(X_{n-1} = s_i, X_0 = s_r, T_{rr} > n-1)}{P(X_0 = s_r)} \\ &= p_{ij} P(X_{n-1} = s_i, T_{rr} > n-1 \mid X_0 = s_r) \end{aligned} \tag{1.23}$$

Für  $n = 1$  ergibt sich:

$$\begin{aligned}
& P(X_{n-1} = s_i, X_n = s_j, T_{rr} \geq n \mid X_0 = s_r) \\
&= P(X_0 = s_i, X_n = s_j, T_{rr} \geq n \mid X_0 = s_r) \\
&= 0 \\
&= p_{ij}P(X_0 = s_i, T_{rr} > 0 \mid X_0 = s_r) \\
&= p_{ij}P(X_{n-1} = s_i, T_{rr} > n - 1 \mid X_0 = s_r)
\end{aligned} \tag{1.24}$$

b) Es sei nun  $s_r = s_i$ .

Dann gilt für alle  $n = 2, 3, \dots$

$$P(X_{n-1} = s_r, X_n = s_j, T_{rr} \geq n \mid X_0 = s_r) = 0 = p_{rj}P(X_{n-1} = s_r, T_{rr} > n - 1 \mid X_0 = s_r).$$

Für  $n = 1$  erhält man:

$$\begin{aligned}
& P(X_{n-1} = s_i, X_n = s_j, T_{rr} \geq n \mid X_0 = s_r) \\
&= P(X_0 = s_r, X_1 = s_j, T_{rr} \geq 1 \mid X_0 = s_r) \\
&= \frac{P(X_0 = s_r, X_1 = s_j, X_0 = s_r, T_{rr} \geq 1)}{P(X_0 = s_r)} \\
&= \frac{P(X_1 = s_j \mid X_0 = s_r, T_{rr} > 0)P(X_0 = s_r, X_0 = s_r, T_{rr} > 0)}{P(X_0 = s_r)} \\
&\stackrel{(1.2)}{=} \frac{P(X_1 = s_j \mid X_0 = s_r)P(X_0 = s_r, X_0 = s_r, T_{rr} > 0)}{P(X_0 = s_r)} \\
&= p_{rj}P(X_0 = s_r, T_{rr} > 0 \mid X_0 = s_r) \\
&= p_{ij}P(X_{n-1} = s_i, T_{rr} > n - 1 \mid X_0 = s_r)
\end{aligned}$$

Da  $\{X_t; t \in \mathbb{Z}^+\}$  nach Voraussetzung rekurrent ist, ist  $T_{rr} < \infty$  und es gilt  $X_0 = X_{T_{rr}} = s_r$ . Deshalb erhält man:

$$\begin{aligned}
\gamma_j^r &= E \left( \sum_{n=0}^{T_{rr}-1} 1_{\{s_j\}}(X_n) \right) \\
&\stackrel{X_0=X_{T_{rr}}}{=} E \left( \sum_{n=1}^{T_{rr}} 1_{\{s_j\}}(X_n) \right) \\
&= E \left( \sum_{n=1}^{\infty} 1_{\{s_j, T_{rr} \geq n\}}(X_n) \right) \\
&= \sum_{n=1}^{\infty} P(X_n = s_j, T_{rr} \geq n \mid X_0 = s_r) \cdot P(X_0 = s_r) \\
&= \sum_{s_i \in S} \sum_{n=1}^{\infty} P(X_{n-1} = s_i, X_n = s_j, T_{rr} \geq n \mid X_0 = s_r) \cdot P(X_0 = s_r) \\
&\stackrel{a) \text{ und } b)}{=} \sum_{s_i \in S} p_{ij} \sum_{n=1}^{\infty} P(X_{n-1} = s_i, T_{rr} \geq n \mid X_0 = s_r) \cdot P(X_0 = s_r)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{s_i \in S} p_{ij} \sum_{m=0}^{\infty} E(1_{\{s_i, T_{rr}-1 \geq m\}}(X_m)) \\
&= \sum_{s_i \in S} p_{ij} E\left(\sum_{m=0}^{T_{rr}-1} 1_{\{s_i\}}(X_m)\right) \\
&= \sum_{s_i \in S} \gamma_i^r p_{ij}
\end{aligned}$$

und somit ist  $\gamma^r P = \gamma^r$ .

(iii) Wegen der vorausgesetzten Irreduzibilität existieren für alle  $s_i \in S$  ein  $n \geq 0$  und ein  $m \geq 0$  so, dass  $p_{ir}^{(n)} > 0$  und  $p_{ri}^{(m)} > 0$ .

Mit (ii) gilt:  $\gamma^r = \gamma^r P^m$  und somit  $\gamma_j^r = \sum_{s_i \in S} \gamma_i^r p_{ij}^{(m)}$ .

Daraus ergibt sich für alle  $s_i \in S$ :

$$\gamma_i^r \stackrel{(ii)}{=} \sum_{s_j \in S} \gamma_j^r p_{ji}^{(m)} \geq \gamma_r^r p_{ri}^{(m)} \stackrel{(i)}{=} p_{ri}^{(m)} > 0$$

und

$$\gamma_i^r p_{ir}^{(n)} \leq \sum_{s_j \in S} \gamma_j^r p_{jr}^{(n)} \stackrel{(ii)}{=} \gamma_r^r \stackrel{(i)}{=} 1,$$

woraus schließlich  $0 < \gamma_i^r < \infty$  folgt. [20] □

**Lemma 1.48.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine irreduzible Markov Kette mit Zustandsraum  $S$  und Übergangsmatrix  $P$ .  $s_r \in S$  sei beliebig aber fest und  $\lambda = (\lambda_1, \lambda_2, \dots)$  sei eine invariante Verteilung der Kette mit  $\lambda_r = 1$ .

Dann gilt:

$$\lambda \geq \gamma^r,$$

d.h.

$$\lambda_i \geq \gamma_i^r \quad \text{für alle } i = 1, 2, \dots$$

Ist  $\{X_t; t \in \mathbb{Z}^+\}$  zusätzlich rekurrent, dann gilt:

$$\lambda = \gamma^r.$$

*Beweis.* Für jedes  $s_j \in S$  gilt:

$$\begin{aligned}
\lambda_j &= \sum_{s_{i_1} \in S} \lambda_{i_1} p_{i_1 j} \\
&= \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \lambda_{i_1} p_{i_1 j} + \lambda_r p_{r j}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \underbrace{\sum_{s_{i_2} \in S} \lambda_{i_2} p_{i_2 i_1} p_{i_1 j}}_{=\lambda_{i_1}} + 1 \cdot p_{rj} \\
&= \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \sum_{\substack{s_{i_2} \in S \\ s_{i_2} \neq s_r}} \lambda_{i_2} p_{i_2 i_1} p_{i_1 j} + \lambda_r p_{r i_1} p_{i_1 j} + p_{rj} \\
&= \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \sum_{\substack{s_{i_2} \in S \\ s_{i_2} \neq s_r}} \lambda_{i_2} p_{i_2 i_1} p_{i_1 j} + \left( p_{rj} + \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} p_{r i_1} p_{i_1 j} \right) \\
&\vdots \\
&= \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \dots \sum_{\substack{s_{i_{n-1}} \in S \\ s_{i_{n-1}} \neq s_r}} \lambda_{i_{n-1}} p_{i_{n-1} i_{n-2}} p_{i_{n-2} i_{n-3}} \dots p_{i_2 i_1} p_{i_1 j} \\
&\quad + \left( p_{rj} + \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} p_{r i_1} p_{i_1 j} + \dots + \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \dots \sum_{\substack{s_{i_{n-1}} \in S \\ s_{i_{n-1}} \neq s_r}} p_{r i_{n-1}} \dots p_{i_2 i_1} p_{i_1 j} \right)
\end{aligned}$$

Weiter ist für alle  $n \geq 1$ :

$$\begin{aligned}
&P(X_n = s_j, T_{rr} \geq n \mid X_0 = s_r) \\
&= \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \dots \sum_{\substack{s_{i_{n-1}} \in S \\ s_{i_{n-1}} \neq s_r}} P(X_1 = s_{i_1}, \dots, X_n = s_j \mid X_0 = s_r) \\
&= \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \dots \sum_{\substack{s_{i_{n-1}} \in S \\ s_{i_{n-1}} \neq s_r}} \frac{P(X_0 = s_r, X_1 = s_{i_1}, \dots, X_n = s_j)}{P(X_0 = s_r)} \\
&\stackrel{\text{Korollar 1.3}}{=} \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \dots \sum_{\substack{s_{i_{n-1}} \in S \\ s_{i_{n-1}} \neq s_r}} \frac{P(X_0 = s_r)}{P(X_0 = s_r)} P(X_1 = s_{i_1} \mid X_0 = s_r) \dots P(X_n = s_j \mid X_{n-1} = s_{i_{n-1}}) \\
&= \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \dots \sum_{\substack{s_{i_{n-1}} \in S \\ s_{i_{n-1}} \neq s_r}} p_{r i_1} \dots p_{i_{n-1} j}
\end{aligned}$$

Und somit ergibt sich für  $s_j \neq s_r$ :

$$\begin{aligned}
\lambda_j &\geq p_{rj} + \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} p_{r i_1} p_{i_1 j} + \dots + \sum_{\substack{s_{i_1} \in S \\ s_{i_1} \neq s_r}} \dots \sum_{\substack{s_{i_{n-1}} \in S \\ s_{i_{n-1}} \neq s_r}} p_{r i_{n-1}} \dots p_{i_2 i_1} p_{i_1 j} \\
&= P(X_1 = s_j, T_{rr} \geq 1 \mid X_0 = s_r) + \dots + P(X_n = s_j, T_{rr} \geq n \mid X_0 = s_r),
\end{aligned}$$

was mit  $n \rightarrow \infty$  gegen  $\gamma_j^r$  konvergiert, weswegen

$$\lambda \geq \gamma^k$$

ist.

Ist die Markov Kette rekurrent, so ist  $\gamma^k$  nach Lemma 1.47 invariant. Setzt man  $\mu = \lambda - \gamma^k$ , ist  $\mu \geq 0$  und wegen

$$\begin{aligned} \mu_j &= \lambda_j - \gamma_j^k \\ &= \sum_{s_i \in S} \lambda_i p_{ij} - \sum_{s_i \in S} \gamma_i^k p_{ij} \\ &= \sum_{s_i \in S} (\lambda_i - \gamma_i^k) p_{ij} \\ &= \sum_{s_i \in S} \mu_i p_{ij}, \end{aligned}$$

ist  $\mu$  ebenfalls invariant.

Da  $\{X_t; t \in \mathbb{Z}^+\}$  nach Voraussetzung irreduzibel ist, gilt  $p_{ir}^{(n)} > 0$  für einige  $n$ ,  $s_i$  gegeben. Zusätzlich ist

$$0 = \underbrace{\lambda_r}_{=1} - \underbrace{\gamma_r^r}_{=1} = \mu_r = \sum_{s_j \in S} \mu_j p_{jr}^{(n)} \geq \mu_i p_{ir}^{(n)},$$

woraus  $\mu_i = 0$  und somit  $\lambda_i = \gamma_i^k$  folgt. [20] □

**Satz 1.49.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine irreduzible Markov Kette mit Zustandsraum  $S$  und Übergangsmatrix  $P$ .

Dann sind folgende Aussagen äquivalent:

- (i) jeder Zustand ist positiv rekurrent;
- (ii) einige Zustände  $s_i$  sind positiv rekurrent;
- (iii)  $\{X_t; t \in \mathbb{Z}^+\}$  besitzt eine invariante Wahrscheinlichkeitsverteilung  $\pi$ .

Ist ferner (iii) erfüllt, gilt  $\tau_{ii} = E(T_{ii}) = \frac{1}{\pi_i}$  für alle  $s_i \in S$ .

*Beweis.* (i)  $\Rightarrow$  (ii) Das ist offensichtlich.

(ii)  $\Rightarrow$  (iii) Ist der Zustand  $s_i$  positiv rekurrent, d.h.  $E(T_{ii}) < \infty$ , so gilt  $P(T_{ii} < \infty) = 1$  und laut Satz 1.44 ist  $s_i$  somit auch rekurrent. Weil  $\{X_t; t \in \mathbb{Z}^+\}$  irreduzibel ist, sind mit

Satz 1.45 alle Zustände und damit die gesamte Markov Kette rekurrent.  
 Nach Satz 1.47 ist  $\gamma^i$  dann invariant. Die Kette starte in  $s_i$ . Dann gilt:

$$\begin{aligned}
 \sum_{s_j \in S} \gamma_j^i &= \sum_{s_j \in S} E \left( \sum_{n=0}^{T_{ii}-1} 1_{\{s_j\}}(X_n) \right) \\
 &= \sum_{s_j \in S} E \left( \sum_{n \geq 0} 1_{\{s_j\}}(X_n) 1_{\{n \leq T_{ii}-1\}} \right) \\
 &= \sum_{s_j \in S} \sum_{n \geq 0} E(1_{\{s_j\}}(X_n) 1_{\{n \leq T_{ii}-1\}}) \\
 &= \sum_{n \geq 0} \sum_{s_j \in S} P(X_n = s_j, T_{ii} - 1 \geq n) \\
 &= \sum_{n \geq 0} P(T_{ii} \geq n + 1) \\
 &= \sum_{n=1}^{\infty} P(T_{ii} \geq n) \\
 &= E(T_{ii}) = \tau_{ii} < \infty
 \end{aligned} \tag{1.25}$$

und somit definiert  $\pi_j = \frac{\gamma_j^i}{\tau_{ii}}$  eine invariante Wahrscheinlichkeitsverteilung.

(iii)  $\Rightarrow$  (i) Sei  $s_r \in S$  beliebig aber fest. Weil die Kette  $\{X_t; t \in \mathbb{Z}^+\}$  irreduzibel ist und  $\sum_{s_i \in S} \pi_i = 1$  gilt, ergibt sich

$$\pi_r = \sum_{s_i \in S} \pi_i p_{ir}^{(n)} > 0 \quad \text{für einige } n.$$

Sei  $\lambda_i = \frac{\pi_i}{\pi_r}$ . Dann ist  $\lambda$  eine invariante Verteilung mit  $\lambda_r = 1$  und mit Lemma 1.48 gilt  $\lambda \geq \gamma^r$ . Damit ist

$$\tau_{rr} = \sum_{s_i \in S} \gamma_i^r \leq \sum_{s_i \in S} \frac{\pi_i}{\pi_r} = \frac{1}{\pi_r} < \infty \tag{1.26}$$

und  $s_r$  somit rekurrent.

Mit dem Wissen, dass die gesamte Markov Kette also rekurrent ist und zusammen mit Lemma 1.48 ist  $\lambda = \gamma^r$  und in (1.26) gilt schließlich  $\tau_{rr} = \frac{1}{\pi_r}$ . [20]  $\square$



## 1.6 Der Ergodensatz

**Satz 1.50 (Starkes Gesetz der großen Zahlen).**

Sei  $\{Y_t; t \in \mathbb{N}\}$  eine Folge von unabhängigen und identisch verteilten, nichtnegativen Zufallsvariablen mit  $E(Y_1) < \infty$ .

Dann gilt

$$\frac{Y_1 + \cdots + Y_n}{n} \longrightarrow E(Y_1) \quad \text{mit } n \longrightarrow \infty \quad \text{fast sicher.}$$

*Beweis.* siehe z.B. [3, S.47-49] □

**Lemma 1.51 (Regenerative Cycle Theorem).**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine Markov Kette mit Anfangswert  $s_1$ , der fast sicher rekurrent ist, d.h.

$$P(X_n = s_1 \quad \text{für } \infty\text{-viele } n) = 1 \quad \text{fast sicher.}$$

$T_1 = t_1, t_2, \dots$  bezeichne die erfolgreichen Rückkehrzeiten zum Zustand  $s_1$  und es sei  $t_1 = 0$ .

Dann sind

$$\{X_{t_k}, X_{t_k+1}, \dots, X_{t_{k+1}-1}; k \geq 0\}$$

unabhängig und identisch verteilt. Vgl. [3, S. 87]

**Lemma 1.52.**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine irreduzible und rekurrente Markov Kette mit Zustandsraum  $S = \{s_1, \dots, s_k\}$  und Übergangsmatrix  $P$ . Die Kette starte in  $s_1$ .

$\gamma^1 = \{\gamma_1^1, \dots, \gamma_k^1\}$  wird als das kanonische invariante Maß assoziiert mit  $s_1$  bezeichnet.

Es gilt

$$\gamma_i^1 = E \left( \sum_{n \geq 1} 1_{\{s_j\}}(X_n) 1_{\{n \leq T_{11}\}} \right) = E \left( \sum_{n=1}^{T_{11}} 1_{\{s_j\}}(X_n) \right) \quad (1.27)$$

Für  $n \geq 1$  sei

$$v(n) = \sum_{j=1}^n 1_{\{s_1\}}(X_j) \quad (1.28)$$

die Anzahl der Besuche im Zustand  $s_1$  bis zum Zeitpunkt  $n$ .

Weiter sei  $f : S \longrightarrow \mathbb{R}$  so, dass

$$\sum_{s_j \in S} |f(s_j)| \gamma_j^1 < \infty. \quad (1.29)$$

Dann gilt für jede Startverteilung  $\mu^{(0)}$ :

$$\lim_{N \uparrow \infty} \frac{1}{v(N)} \sum_{l=1}^N f(X_l) = \sum_{s_j \in S} f(s_j) \gamma_j^1 \quad \text{fast sicher.} \quad (1.30)$$

*Beweis.* Es seien  $T_1 = t_1, t_2, t_3, \dots$  die erfolgreichen Rückkehrzeiten zum Zustand  $s_1$  und es sei

$$U_p = \sum_{n=t_p+1}^{t_{p+1}} f(X_n).$$

Nach Lemma 1.51 ist  $\{U_p; p \in \mathbb{N}\}$  eine unabhängige und identisch verteilte Folge. Man nehme weiter an, dass  $f \geq 0$  gilt. Zusammen mit der starken Markov Eigenschaft erhält man:

$$\begin{aligned} E(U_1) &= E\left(\sum_{n=1}^{T_{11}} f(X_n)\right) \\ &= E\left(\sum_{n=1}^{T_{11}} \sum_{s_j \in S} f(s_j) 1_{\{s_j\}}(X_n)\right) \\ &= \sum_{s_j \in S} f(s_j) E\left(\sum_{n=1}^{T_{11}} 1_{\{s_j\}}(X_n)\right) \\ &= \sum_{s_j \in S} f(s_j) x_j. \end{aligned}$$

Nach Voraussetzung ist  $\sum_{s_j \in S} f(s_j) \gamma_j^1 < \infty$ . Zusammen mit dem Gesetz der großen Zahlen ergibt sich:

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{p=1}^n U_p = \sum_{s_j \in S} f(s_j) \gamma_j^1,$$

und somit

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{l=T_{11}+1}^{t_{n+1}} f(X_l) = \sum_{s_j \in S} f(s_j) \gamma_j^1. \quad (1.31)$$

Da

$$t_{v(n)} \leq n < t_{v(n)+1}$$

gilt, erhält man

$$\frac{\sum_{l=1}^{t_{v(n)}} f(X_l)}{v(n)} \leq \frac{\sum_{l=1}^n f(X_l)}{v(n)} \leq \frac{\sum_{l=1}^{t_{v(n)+1}} f(X_l)}{v(n)}.$$

Wegen der vorausgesetzten Rekurrenz gilt  $\lim_{n \uparrow \infty} v(n) = \infty$ . Zusammen mit (1.31) streben die Terme der obigen Ungleichung mit  $n \rightarrow \infty$  gegen  $\sum_{j \in \{1, \dots, k\}} f(s_j) \gamma_j^1$ , womit (1.30) bewiesen wäre.

Für den Fall, dass nicht nur  $f \geq 0$  gilt, erlangt man die gewünschte Aussage, indem man (1.30) getrennt aufschreibt für  $f^+ = \max(0, f)$  und  $f^- = \max(0, -f)$  und anschließend die Differenz der beiden entstandenen Gleichungen bildet. Wegen (1.29) ist diese Differenz nicht von der Form  $\infty - \infty$ . (vgl. [3, S. 113])

[3]

□

Aus (1.25) geht

$$\sum_{s_j \in S} \gamma_j^1 = E(T_{11})$$

hervor. Startet die Markov Kette in  $s_1$ , so gilt für alle  $i = 2, \dots, k$ :

$$\begin{aligned} \gamma_i^1 &= \sum_{n \geq 1} E(1_{\{s_j\}}(X_n)1_{\{n \leq T_{11}\}}) \\ &= \sum_{n \geq 1} P(X_1 \neq s_1, \dots, X_{n-1} \neq s_1, X_n = s_i) \\ &= \sum_{n \geq 1} P(X_n = s_i, T_{11} \geq n) \\ &= \sum_{n \geq 0} P(X_n = s_i, T_{11} > n). \end{aligned} \tag{1.32}$$

Zum Beweis der Existenz einer invarianten Wahrscheinlichkeitsverteilung (Satz 1.32) wurde bereits gezeigt, dass für die invariante Wahrscheinlichkeitsverteilung  $\pi$  gilt:

$$\pi_i = \frac{\rho_i}{E(T_{11})}$$

mit

$$\rho_i = \sum_{n \geq 0} P(X_n = s_i, T_{11} > n).$$

Mit (1.32) ist  $x_i = \rho_i$  für alle  $i = 2, \dots, k$ .

Dies gilt auch für  $i = 1$ , denn

$$\rho_1 = 1 = \sum_{n \geq 1} E(1_{\{s_1\}}(X_n)1_{\{n \leq T_{11}\}}) = \gamma_1^1.$$

Aus alledem folgt schließlich mit

$$\pi_i = \frac{\rho_i}{E(T_{11})} = \frac{\gamma_i^1}{E(T_{11})} = \frac{\gamma_i^1}{\sum_{s_j \in S} \gamma_j^1}, \tag{1.33}$$

dass  $\pi$  durch Normierung von  $\gamma^1$  hervorgeht.

**Satz 1.53 (Ergodensatz für Markov Ketten).**

Sei  $\{X_t; t \in \mathbb{Z}^+\}$  eine irreduzible und positiv rekurrente Markov Kette mit Zustandsraum  $S$  und invarianter Wahrscheinlichkeitsverteilung  $\pi$ . Weiter sei  $f : S \rightarrow \mathbb{R}$  so, dass gilt:

$$\sum_{s_i \in S} |f(s_i)|\pi_i < \infty. \tag{1.34}$$

Dann gilt für jede Startverteilung  $\mu^{(0)}$

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{j=1}^N f(X_j) = \sum_{s_i \in S} f(s_i)\pi_i \quad \text{fast sicher.}$$

*Beweis.* Man wende Lemma 1.52 auf  $f \equiv 1$  an. Bedingung (1.29) ist erfüllt, da im positiv rekurrenten Fall  $\sum_{s_i \in S} \gamma_i^1 = E(T_{11}) < \infty$  gilt. Aus diesem Grund gilt mit (1.30):

$$\lim_{N \uparrow \infty} \frac{N}{v(N)} = \sum_{s_j \in S} \gamma_j^1.$$

Weil  $f$  (1.34) erfüllt und  $x$  und  $\pi$  wegen (1.33) proportional zueinander sind, erfüllt  $f$  auch (1.29). Aus diesem Grund ist

$$\lim_{N \uparrow \infty} \frac{1}{v(N)} \sum_{j=1}^N f(X_j) = \sum_{s_i \in S} f(s_i) \gamma_i^1.$$

Durch Kombination der beiden obigen Gleichungen ergibt sich:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N f(X_j) = \lim_{N \rightarrow \infty} \frac{v(N)}{N} \frac{1}{v(N)} \sum_{j=1}^N f(X_j) = \frac{\sum_{s_i \in S} f(s_i) \gamma_i^1}{\sum_{s_j \in S} \gamma_j^1}. \quad (1.35)$$

Mit (1.33) folgt schließlich die gewünschte Aussage. [3] □

Diese Version des Ergodensatzes für Markov Ketten ist eine Art Gesetz der großen Zahlen und kann bei einer Simulation benutzt werden. Ist nämlich  $\pi$  unbekannt, kann man mit Hilfe dieses Satzes Größen vom Typ  $E_\pi(f(X_0))$  schätzen. (vgl. [3, S. 113])

# Kapitel 2

## Markov Ketten Monte Carlo Methode

Mit Hilfe der Markov Ketten Monte Carlo (MCMC) Methode wird eine Markov Kette mit Übergangsmatrix  $P$  simuliert. Die Aufgabe besteht darin bei gegebener Wahrscheinlichkeitsverteilung  $\pi$  auf einem sehr großen endlichen Zustandsraum  $S = \{1, \dots, k\}$ <sup>1</sup> eine Stichprobe aus  $S$  unter der Wahrscheinlichkeitsverteilung  $\pi$  zu ziehen. D.h. es wird eine Markov Kette über dem Zustandsraum  $S$  konstruiert, die die invariante Wahrscheinlichkeitsverteilung  $\pi$  besitzt. (vgl. [27])

Die Idee dabei ist folgende:

Angenommen, es gelingt eine aperiodische und irreduzible Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$  mit invarianter Verteilung  $\pi$  zu konstruieren. Dann besagt Satz 1.34, dass bei beliebiger Startverteilung  $\mu^{(0)}$  die Verteilung der Kette zum Zeitpunkt  $n$  mit  $n \rightarrow \infty$  gegen  $\pi$  konvergiert. Das bedeutet, lässt man die Markov Kette eine ausreichend lange Zeit  $n$  laufen, so kommt die Verteilung von  $X_n$  der invarianten Verteilung  $\pi$  sehr nahe.

### 2.1 Hastings Algorithmus

Eine allgemeine Konstruktion einer Markov Kette  $\{X_t; t \in \mathbb{Z}^+\}$  mit Zustandsraum  $S$ , Übergangsmatrix  $P$  und invarianter Wahrscheinlichkeitsverteilung  $\pi$  wurde 1970 durch Hastings [13] geliefert. (vgl. auch [3])

Gesucht werden hierfür Ergebnisse der Form

$$p_{ij} = q_{ij}\alpha_{ij} \quad \text{für } i \neq j$$

mit

$$p_{ii} = 1 - \sum_{i \neq j} p_{ij},$$

wobei  $Q = (q_{ij})$  eine beliebige Übergangsmatrix zum Zustandsraum  $S$ , die sogenannte **Kandidat-erzeugende Matrix** mit den **Vorschlags-Wahrscheinlichkeiten**  $q_{ij}$ , ist.

---

<sup>1</sup>Im Folgenden werden o.B.d.A. die Zustandsräume ausschließlich von der Form  $\{1, 2, \dots, k\}$  sein. Für die Bezeichnung der Übergangswahrscheinlichkeiten ergibt sich entsprechend  $p_{ij} = P(X_1 = j \mid X_0 = i)$ .

**Definition 2.1.**

Die **Akzeptanz-Wahrscheinlichkeit**  $\alpha_{ij}$  des Hastings Algorithmus ist definiert durch

$$\alpha_{ij} = \frac{s_{ij}}{1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}}} \quad \text{für } q_{ji} > 0$$

und

$$\alpha_{ij} = 0 \quad \text{für } q_{ji} = 0.$$

$\Sigma = (s_{ij})$  stellt hierbei eine symmetrische Matrix dar, die so gewählt werden muss, dass

$$\alpha_{ij} \in [0, 1] \quad \text{für alle } i, j \in S$$

gegeben ist.

Der **Hastings Algorithmus** funktioniert wie folgt:

Angenommen, die Markov Kette befindet sich zum gegenwärtigen Zeitpunkt  $t$  im Zustand  $i$ , also  $X_t = i$ . Dann wird der Vorschlag  $j$  für den folgenden Zustand mit der Wahrscheinlichkeit  $q_{ij}$  gewählt. Ist  $i \neq j$ , so wird dieser neue Zustand mit der Wahrscheinlichkeit  $\alpha_{ij}$  angenommen, also gilt dann  $X_{t+1} = j$  und  $p_{ij} = q_{ij}\alpha_{ij}$ , oder andernfalls ist der nächste Zustand  $X_{t+1} = X_t = i$ .

Die aus diesem Algorithmus resultierende Markov Kette erfüllt folgenden Satz:

**Satz 2.2.**

Wird eine Markov Kette mit Hilfe des oben beschriebenen Hastings Algorithmus konstruiert, so ist  $\pi$  deren invariante Wahrscheinlichkeitsverteilung.

*Beweis.* Es seien wie oben beschrieben  $\pi$  eine Wahrscheinlichkeitsverteilung,  $Q = (q_{ij})$  eine Kandidat-erzeugende Matrix und  $\Sigma = (s_{ij})$  eine symmetrische Matrix. Weiter gelte

$$p_{ij} = q_{ij}\alpha_{ij} \quad \text{für } i \neq j$$

mit

$$\alpha_{ij} = \frac{s_{ij}}{1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}}}.$$

Dann erhält man:

$$\begin{aligned} \pi_i p_{ij} &= \pi_i q_{ij} \alpha_{ij} \\ &= \pi_i q_{ij} \frac{s_{ij}}{1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}}} \\ &= \pi_i q_{ij} \frac{s_{ij}}{\frac{\pi_j q_{ji} + \pi_i q_{ij}}{\pi_j q_{ji}}} \\ &= \pi_i q_{ij} s_{ij} \frac{\pi_j q_{ji}}{\pi_j q_{ji} + \pi_i q_{ij}} \end{aligned}$$

$$= \pi_j q_{ji} s_{ji} \frac{\pi_i q_{ij}}{\pi_i q_{ij} + \pi_j q_{ji}} \quad (2.1)$$

$$= \pi_j q_{ji} \frac{s_{ji}}{\frac{\pi_i q_{ij} + \pi_j q_{ji}}{\pi_i q_{ij}}}$$

$$= \pi_j q_{ji} \frac{s_{ji}}{1 + \frac{\pi_j q_{ji}}{\pi_i q_{ij}}}$$

$$= \pi_j q_{ji} \alpha_{ji}$$

$$= \pi_j p_{ji}, \quad (2.2)$$

wobei in (2.1) benutzt wurde, dass die Matrix  $\Sigma$  symmetrisch ist.

Mit (2.2) und Satz 1.37 folgt, dass  $\pi$  die invariante Wahrscheinlichkeitsverteilung der konstruierten Markov Kette ist.  $\square$

Ist somit zusätzlich gegeben, dass die konstruierte Markov Kette aperiodisch und irreduzibel ist, so besagt Satz 1.34, dass die Verteilung der konstruierten Kette zum Zeitpunkt  $n$  mit  $n \rightarrow \infty$  gegen  $\pi$  konvergiert. Ob diese Eigenschaften erfüllt sind, hängt von der Wahl von  $Q$  ab und muss jeweils bei den spezifischen Anwendungen überprüft werden.

### Zusammenfassung:

Angenommen, gegenwärtig ist  $X_t = i$ .

Dann gilt für die Vorgehensweise des Hastings Algorithmus:

- (i) wähle einen Zustand  $j$  mit Hilfe der Verteilung, die durch die  $i$ -te Zeile von  $Q$  gegeben ist
- (ii) nehme  $X_{t+1} = j$  mit der Wahrscheinlichkeit  $\alpha_{ij}$  an oder andernfalls setze  $X_{t+1} = X_t = i$
- (iii) gehe zu (i)

## 2.2 Metropolis Algorithmus

Eine speziellere Form des von Hastings verallgemeinerten Algorithmus wurde bereits 1953 von Metropolis *et al* [17] entwickelt und 1995 noch einmal von Chib und Greenberg [6] beschrieben.

Es handelt sich hierbei um einen Hastings Algorithmus mit einer explizit angegebenen Akzeptanz-Wahrscheinlichkeit  $\alpha_{ij}$ :

### Definition 2.3.

Die **Akzeptanz-Wahrscheinlichkeit**  $\alpha_{ij}$  des Metropolis Algorithmus ist definiert durch

$$\alpha_{ij} = \min \left\{ \frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right\} \quad \text{für } \pi_i q_{ij} > 0$$

und

$$\alpha_{ij} = 1 \quad \text{sonst.}$$

Oft wird der Metropolis Algorithmus als Kombination des Hastings Algorithmus mit der oben definierten Akzeptanz-Wahrscheinlichkeit auch als „**Metropolis-Hastings Algorithmus**“ bezeichnet.

Die Konstruktion von  $\alpha_{ij}$  (siehe [6, S. 329]) erfolgte mit dem Ziel, dass die konstruierte Kette reversibel ist und somit folgender Satz gilt:

### Satz 2.4.

Wird eine Markov Kette mit Hilfe des Metropolis Algorithmus konstruiert, so ist  $\pi$  deren invariante Wahrscheinlichkeitsverteilung.

*Beweis.* Es seien  $\pi$  eine Wahrscheinlichkeitsverteilung,  $Q$  eine Kandidat-erzeugende Matrix,  $\alpha_{ij}$  wie in Definition 2.3 und es gelte

$$p_{ij} = q_{ij} \alpha_{ij} \quad \text{für } i \neq j.$$

Nach Satz 1.37 ist zu zeigen, dass

$$\pi_i p_{ij} = \pi_i q_{ij} \alpha_{ij} = \pi_j q_{ji} \alpha_{ji} = \pi_j p_{ji}$$

erfüllt ist.

Für den Fall  $\pi_i q_{ij} = 0$  ist diese Aussage trivial, da dann  $\alpha_{ji} = 0$  gilt und sich somit auf beiden Seiten 0 ergibt.



Es sei also  $\pi_i q_{ij} > 0$ . Dann sind drei Fälle zu unterscheiden:

1. Fall:

Es gilt  $\alpha_{ij} = \min \left\{ \frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right\} = \frac{\pi_j q_{ji}}{\pi_i q_{ij}}$ .

D.h. es ist  $\pi_j q_{ji} < \pi_i q_{ij}$  und somit gilt  $\alpha_{ji} = \min \left\{ \frac{\pi_i q_{ij}}{\pi_j q_{ji}}, 1 \right\} = 1$ .

Es ergibt sich also:

$$\pi_i p_{ij} = \pi_i q_{ij} \alpha_{ij} = \pi_i q_{ij} \frac{\pi_j q_{ji}}{\pi_i q_{ij}} = \pi_j q_{ji} \cdot 1 = \pi_j q_{ji} \alpha_{ji} = \pi_j p_{ji}.$$

2. Fall:

Es gilt  $\alpha_{ij} = \min \left\{ \frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right\} = 1$ .

D.h. es ist entweder  $\pi_j q_{ji} = \pi_i q_{ij}$  (siehe 3. Fall)

oder es gilt  $\pi_j q_{ji} > \pi_i q_{ij}$  und somit  $\alpha_{ji} = \min \left\{ \frac{\pi_i q_{ij}}{\pi_j q_{ji}}, 1 \right\} = \frac{\pi_i q_{ij}}{\pi_j q_{ji}}$ .

In diesem Fall erhält man:

$$\pi_i p_{ij} = \pi_i q_{ij} \alpha_{ij} = \pi_i q_{ij} \cdot 1 = \pi_j q_{ji} \frac{\pi_i q_{ij}}{\pi_j q_{ji}} = \pi_j q_{ji} \alpha_{ji} = \pi_j p_{ji}.$$

3. Fall:

Es gilt  $\pi_j q_{ji} = \pi_i q_{ij}$ .

Hieraus folgt  $\alpha_{ij} = \alpha_{ji} = 1$  und somit trivialerweise

$$\pi_i p_{ij} = \pi_i q_{ij} \cdot 1 = \pi_j q_{ji} \cdot 1 = \pi_j p_{ji}.$$

□

Ist also gegeben, dass die konstruierte Markov Kette zusätzlich aperiodisch und irreduzibel ist, kann auch hier Satz 1.34 angewandt werden. Somit sind von der Wahl von  $Q$  zwei Dinge abhängig:

- (i) die Aperiodizität und die Irreduzibilität der Kette, sowie
- (ii) die Akzeptanzwahrscheinlichkeit  $\alpha_{ij}$ .

(i) muss auch hier jeweils bei den spezifischen Anwendungen überprüft werden.

(ii) lässt sich jedoch wie folgt vermeiden:

Ein Spezialfall des Algorithmus entsteht nämlich, wenn die Kandidat-erzeugende Matrix  $Q$  symmetrisch gewählt wird. Für die Akzeptanz-Wahrscheinlichkeit ergibt sich in diesem Fall

$$\alpha_{ij} = \min \left\{ 1, \frac{\pi_j}{\pi_i} \right\}, \quad (2.3)$$

was bedeutet, dass diese dann unabhängig von  $Q$  ist.

Insbesondere gilt dies, wenn der Kandidat-erzeugende Vorgang rein zufällig geschieht, d.h. wenn

$$q_{ij} = \frac{1}{|S|} \quad \text{für alle } s_i, s_j \in S$$

gilt. (vgl. [29])

### Zusammenfassung:

Die Konstruktion einer Markov Kette mit der invarianten Verteilung  $\pi$  wird mit dem Metropolis(-Hastings) Algorithmus wie folgt durchgeführt:

Es seien  $i_0, i_1, \dots, i_N \in S$ .

Wähle einen beliebigen Startwert  $i_0$ .

- Wiederhole für  $l = 1, 2, \dots, N$ :
- Wähle einen Zustand  $j$  mit Hilfe der Verteilung, die durch die  $i_l$ -te Zeile von  $Q$  gegeben ist
- Generiere eine Zufallszahl  $u$  aus  $\mathcal{U}(0, 1)$  (d.h. aus unabhängigen, identisch gleichverteilten Zufallszahlen aus  $[0, 1]$ )
- Gilt  $u \leq \alpha_{i_l j}$ , setze  $i_{l+1} = j$
- Andernfalls setze  $i_{l+1} = i_l$ .
- Gib die Werte  $\{i_1, \dots, i_N\}$  aus.

## 2.2.1 Komponentenweiser Metropolis-Hastings Algorithmus

Um den Metropolis-Hastings Algorithmus effizienter zu machen, wird der Zustand  $s \in S$  oft nicht aufeinmal aufdatiert, sondern die Zufallsvariable  $X$  in einzelne Komponenten  $\{X_{.1}, \dots, X_{.n}\}$  von möglicherweise unterschiedlicher Dimension aufgeteilt. (Beschrieben wird diese Methode in [6, S. 332], [13, S. ] und [10, S. 10].) Die Updates werden hier komponentenweise nacheinander durchgeführt, so dass eine Iteration des Komponentweisen Metropolis-Hastings Algorithmus aus  $n$  Update-Schritten besteht.

$$X_{.-i} := \{X_{.1}, \dots, X_{.i-1}, X_{.i+1}, \dots, X_{.n}\}$$

bezeichne dabei alle Komponenten von  $X$  ausschließlich der  $i$ -ten Komponente  $X_{.i}$ .  $X_{t,i}$  sei der Stand von  $X_{.i}$  am Ende der Iteration  $t$ . Es ergibt sich folgende Vorgehensweise:

Für den Schritt  $i$  der Iteration  $t + 1$  wendet man den Metropolis-Hastings Algorithmus auf die Komponente  $X_{.i}$  an.

Der Kandidat  $Y_i$  wird erzeugt mit der Vorschlags-Wahrscheinlichkeit

$$q_i(Y_i | X_{t,i}, X_{t,-i}),$$

wobei

$$X_{t,-i} = \{X_{t+1,1}, \dots, X_{t+1,i-1}, X_{t,i+1}, \dots, X_{t,n}\}$$

den Wert von  $X_{-i}$  nach Beendigung des  $(i-1)$ -ten Schritts der Iteration  $(t+1)$  bezeichnet, bei dem die Komponenten  $1, 2, \dots, i-1$  bereits aufdatiert wurden. Somit erzeugt die  $i$ -te Vorschlags-Verteilung  $q_i(\cdot | \cdot, \cdot)$  ausschließlich Kandidaten für die  $i$ -te Komponente von  $X$  und hängt von allen *aktuellen* Werten jeder einzelnen Komponente von  $X$  ab. Entsprechend dem Metropolis-Hastings Algorithmus wird der vorgeschlagene Kandidat  $Y_i$  mit der Akzeptanz-Wahrscheinlichkeit

$$\alpha(X_{-i}, X_i, Y_i) = \min \left\{ 1, \frac{\pi(Y_i | X_{-i})q_i(X_i | Y_i, X_{-i})}{\pi(X_i | X_{-i})q_i(Y_i | X_i, X_{-i})} \right\} \quad (2.4)$$

angenommen.  $\pi(X_i | X_{-i})$  wird dabei als **vollständig bedingte Verteilung** für  $X$  unter  $\pi$  bezeichnet.

Wird der Kandidat  $Y_i$  akzeptiert, so setzt man  $X_{t+1,i} = Y_i$ , andernfalls ist  $X_{t+1,i} = X_{t,i}$ . Die restlichen Komponenten bleiben im  $i$ -ten Schritt unverändert.

Dass mit diesem Algorithmus wirklich eine Markov Kette mit invarianter Wahrscheinlichkeitsverteilung  $\pi$  konstruiert wird, liegt daran, dass  $\pi$  durch die Menge seiner vollständig bedingten Verteilungen eindeutig bestimmt ist. (Siehe hierzu [10, S. 11])

## 2.2.2 Gibbs Sampler

Der Gibbs Sampler ist eine spezielle Form des komponentenweisen Metropolis-Hastings Algorithmus, wurde 1984 von Geman und Geman eingeführt und soll hier nur kurz erwähnt sein.

Die Vorschlagswahrscheinlichkeit für die  $i$ -te Komponente ist bei diesem Algorithmus gleich der vollständig bedingten Verteilung, d.h. es gilt

$$q_i(Y_i | X_i, X_{-i}) = \pi(Y_i | X_{-i}).$$

Für die Akzeptanz-Wahrscheinlichkeit ergibt sich zusammen mit (2.4):

$$\alpha(X_{-i}, X_i, Y_i) = \min \left\{ 1, \frac{\pi(Y_i | X_{-i})\pi(X_i | Y_i, X_{-i})}{\pi(X_i | X_{-i})\pi(Y_i | X_i, X_{-i})} \right\} = 1,$$

was bedeutet, dass die vorgeschlagene Kandidaten  $Y_i$  immer angenommen werden.

# Kapitel 3

## Unabhängigkeitstest für 2-dimensionale Kontingenztafeln

### 3.1 Bezeichnungen

Im Folgenden sei  $\mathcal{X} = \{(k, l); k = 1, \dots, K, l = 1, \dots, L\}$  ein endlicher Stichprobenraum und  $y = (y_1, \dots, y_N) \in \mathcal{X}^N := \mathcal{Y}$  ein Stichprobenvektor vom Umfang  $N$ .  $f$  bezeichne entsprechend die zugehörige  $K \times L$ -Kontingenztafel<sup>1</sup> mit den Einträgen

$$f_{kl} := f_{kl}(y) = \sum_{i=1}^N 1_{\{(k,l)\}}(y_i),$$

jeweils in Zelle  $(k, l)$ , und dem Stichprobenumfang

$$N = \sum_{k=1}^K \sum_{l=1}^L f_{kl}.$$

Die Randsummen seien bezeichnet mit

$$f_{k+} := f_{k+}(y) = \sum_{l=1}^L f_{kl}(y) \quad \text{und} \quad f_{+l} := f_{+l}(y) = \sum_{k=1}^K f_{kl}(y).$$

Die Wahrscheinlichkeit, dass  $Y_i = (V_i, W_i)$ ,  $i = 1, \dots, N$ , in Zelle  $(k, l)$  fällt, d.h. dass das Merkmal  $V_i$  die Ausprägung  $k$  und das Merkmal  $W_i$  die Ausprägung  $l$  annimmt, sei gegeben durch

$$\pi_{kl} = P(V_i = k, W_i = l) \quad i = 1, \dots, N,$$

womit  $\{\pi_{kl}\}$  die gemeinsame Verteilung von  $V_i$  und  $W_i$  darstellt.

Entsprechend seien die Randverteilungen bezeichnet mit  $\{\pi_{k+}\}$  bzw.  $\{\pi_{+l}\}$ , wobei

$$\pi_{k+} = \sum_{l=1}^L \pi_{kl} \quad \text{bzw.} \quad \pi_{+l} = \sum_{k=1}^K \pi_{kl},$$

---

<sup>1</sup>Kontingenztafeln werden im Folgenden in Form von Matrizen angegeben.

d.h. für  $i = 1, \dots, N$  gilt

$$\pi_{k+} = P(V_i = k) \quad \text{bzw.} \quad \pi_{+l} = P(W_i = l).$$

Die entsprechenden Zufallsvariablen werden mit einem  $Y$  bezeichnet, so ist z.B.

$$f_{kl}(Y) = \sum_{i=1}^N 1_{\{(k,l)\}}(Y_i)$$

die Zufallsvariable für den Zelleneintrag  $f_{kl}$  und

$$F(Y) := (f_{kl}(Y)) = (f_{11}(Y), \dots, f_{1L}(Y), f_{21}(Y), \dots, f_{2L}(Y), \dots, f_{K1}(Y), \dots, f_{KL}(Y))$$

der Zufallsvektor für die Kontingenztafel.

## 3.2 Stochastische Unabhängigkeit und der $\chi^2$ -Test

Oft ist es von Interesse zu testen, ob zwei Merkmale  $V$  und  $W$  unabhängig voneinander auftreten.

Hierfür sei zunächst die folgende grundlegende Definition ins Gedächtnis gerufen:

### Definition 3.1.

$(\Omega, P)$  sei ein diskreter Wahrscheinlichkeitsraum.

Die Zufallsvariablen  $X_1, \dots, X_n$  mit  $X_i : \Omega \longrightarrow \mathcal{X}_i$ ,  $i = 1, \dots, n$ , heißen **(stochastisch) unabhängig**, wenn für beliebige  $(x_1, \dots, x_n) \in \mathcal{X}_1 \times \dots \times \mathcal{X}_n$  gilt:

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i).$$

Bei einem Test auf Unabhängigkeit zweier Zufallsvariablen ergibt sich also die Nullhypothese:

$$H_0 : \pi_{kl} = \pi_{k+}\pi_{+l} \quad \text{für alle } k \text{ und alle } l. \quad (3.1)$$

### 3.2.1 $\chi^2$ -Unabhängigkeitstest

Ein geläufiger Unabhängigkeitstest für  $K \times L$ -Kontingenztafeln ist der  $\chi^2$ -Test mit der (Pearson-) Teststatistik

$$T_{\chi^2} = \sum_{k=1}^K \sum_{l=1}^L \frac{(f_{kl} - \frac{f_{k+}f_{+l}}{N})^2}{\frac{f_{k+}f_{+l}}{N}},$$

wobei  $\tilde{\mu}_{kl} := \frac{f_{k+}f_{+l}}{N}$  die Maximum-Likelihood-Schätzung für  $\mu_{kl} = N\pi_{kl}$  unter  $H_0$  darstellt. Für einen festen Stichprobenumfang  $N$  bewirkt eine größere Differenz zwischen  $f_{kl}$  und  $\tilde{\mu}_{kl}$  größere Werte für  $T_{\chi^2}$  und einen größeren Hinweis gegen  $H_0$ .

Die Teststatistik  $T_{\chi^2}$  besitzt unter der Hypothese  $H_0$  approximativ für  $N \rightarrow \infty$  eine  $\chi^2$ -Verteilung mit  $(K-1)(L-1)$  Freiheitsgraden. Will man jedoch eine befriedigende Approximation erreichen, muss der Stichprobenumfang ausreichend groß sein. Dies ist erfüllt, wenn die erwarteten Häufigkeiten  $\tilde{\mu}_{kl}$  mindestens 5 betragen. Wendet man den  $\chi^2$ -Test an, so sollten also nicht mehr als 20% der erwarteten Häufigkeiten  $\tilde{\mu}_{kl}$  kleiner als 5 sein. (Vgl. [1, S. 28], [4, S.167], [11, S. 103], [23, S. 520] und [26, S. 36])

Der  $\chi^2$ -Test ist somit gut geeignet für Tabellen mit großen Zelleneinträgen. Sind die einzelnen Einträge (und damit die erwarteten Zellhäufigkeiten) jedoch klein, so muss ein alternatives Verfahren gefunden werden.

## 3.3 Die Exponentialfamilie

### Definition 3.2.

Ein Zufallsvektor  $y$  bzw. seine Verteilung gehört zu einer **Exponentialfamilie**, wenn die (stetige oder diskrete) Dichte von der Form

$$P_{\theta}^{Y_i}(\{y_i\}) = P_{\theta}(Y_i = y_i) = c(y_i) \frac{e^{\sum_{j=1}^m \varphi_j(\theta) t_j(y_i)}}{a(\theta)} \quad (3.2)$$

ist. Hierbei ist  $\theta$  ein Parametervektor, die Funktionen  $t_j$  und  $\varphi_j$  sind reellwertig,  $a(\theta) > 0$  und  $c(y_i) \geq 0$  für alle  $i$ .

### Definition 3.3.

Ist in (3.2)  $m$  minimal, so heißt die Exponentialfamilie  **$m$ -parametrig**.

In diesem Fall bezeichnet man  $\theta_1 = \varphi_1(\theta), \dots, \theta_m = \varphi_m(\theta)$  als **natürliche** (oder **kanonische**) **Parameter**.

Mit  $\theta = (\theta_1, \dots, \theta_m)^t$ ,  $b(\theta) = \ln(a(\theta))$  und  $T(y_i) = (t_1(y_i), \dots, t_m(y_i))^t$  mit  $T : \mathcal{X} \rightarrow \mathbb{R}^m$ ,  $y_i \mapsto T(y_i)$ , ergibt sich die Dichte in **natürlicher, minimaler** (oder **kanonischer**) **Form**

$$P_{\theta}^{Y_i}(\{y_i\}) = P_{\theta}(Y_i = y_i) = c(y_i) e^{\theta^t T(y_i) - b(\theta)}.$$

Beispiele für Verteilungen aus der Exponentialfamilie (z.B. die Multinomialverteilung, die Normalverteilung oder die Poisson-Verteilung) werden ausführlich in [8, S. 44f] behandelt.

**Satz 3.4.**

Sind  $Y_1, \dots, Y_N$  unabhängige und identisch verteilte Zufallsvariablen mit den Realisierungen  $y_1, \dots, y_N$  und den Dichten

$$P_\theta(Y_i = y_i) = c(y_i)e^{\theta^t T(y_i) - b(\theta)} \quad \text{für } i = 1, \dots, N,$$

dann ist die gemeinsame Dichte von  $Y = (Y_1, \dots, Y_N)$  gegeben durch

$$P_\theta(Y_1 = y_1, \dots, Y_N = y_N) = e^{\theta^t \sum_{i=1}^N T(y_i) - Nb(\theta)} \prod_{i=1}^N c(y_i).$$

*Beweis.* Da  $Y_1, \dots, Y_N$  unabhängig voneinander sind, gilt:

$$\begin{aligned} P_\theta(Y_1 = y_1, \dots, Y_N = y_N) &= \prod_{i=1}^N P_\theta(y_i) \\ &= \prod_{i=1}^N c(y_i) e^{\theta^t T(y_i) - b(\theta)} \\ &= e^{\theta^t \sum_{i=1}^N T(y_i) - \sum_{i=1}^N b(\theta)} \prod_{i=1}^N c(y_i) \\ &= e^{\theta^t \sum_{i=1}^N T(y_i) - Nb(\theta)} \prod_{i=1}^N c(y_i). \end{aligned}$$

[8]

□

Da viele der für Anwendungen interessanten Verteilungen zur Exponentialfamilie gehören, werde ich mich im Weiteren ausschließlich auf Verteilungen dieser Familie beziehen.

### 3.4 Die Hypergeometrische Verteilung

Ziel dieses Abschnitts ist es einen Weg zu finden, durch den die gemeinsame Verteilung von  $P_\theta$ -verteilten Zufallsvariablen  $Y_1, \dots, Y_N$  unabhängig von dem unbekanntem Parameter (vektor)  $\theta$  ist. Hierfür wird zunächst gezeigt, dass  $T_N(Y_1, \dots, Y_N) = \sum_{i=1}^N T(Y_i)$  eine suffiziente Statistik für  $\theta$  darstellt. Es stellt sich heraus, dass die Verteilung von  $Y = (Y_1, \dots, Y_N)$  gerade dann unabhängig von  $\theta$  ist, wenn man sie mit  $T_N$  bedingt. Die Bezeichnungen sind orientiert an [22] und [7].

#### Definition 3.5.

Es sei  $Y = (Y_1, \dots, Y_N)$  ein Stichprobenvektor mit dem Stichprobenraum  $[\mathbb{R}^N, \mathcal{B}^N]$ . ( $\mathcal{B}^N$  bezeichne die  $\sigma$ -Algebra der Borel-Mengen des  $\mathbb{R}^N$ .) Die Wahrscheinlichkeitsverteilung von  $Y$  gehöre einer parametrisierten Familie  $(P_\theta)_{\theta \in \Theta}$  von Wahrscheinlichkeitsverteilungen auf  $[\mathbb{R}^N, \mathcal{B}^N]$  an. D.h. die Verteilung von  $Y$  sei bis auf einen unbekanntem Parameter (vektor)  $\theta \in \Theta$  bekannt.

Eine Statistik  $T_N = T(Y_1, \dots, Y_N)$  heißt **suffizient** für  $\theta$ , wenn die bedingte Verteilung

$$P_\theta(Y \in A \mid T_N = t)$$

für alle  $A \in \mathcal{B}^N$  unabhängig von  $\theta \in \Theta$  ist.

#### Satz 3.6.

Es sei  $P_\theta^Y$  eine diskrete Verteilung aus der Exponentialfamilie.

Sind  $Y_1, \dots, Y_N$  unabhängige und identisch verteilte  $\mathcal{X}$ -wertige Zufallsvariablen, so ist  $T_N(Y_1, \dots, Y_N) = \sum_{i=1}^N T(Y_i)$  eine suffiziente Statistik für  $\theta$ .

*Beweis.* Es seien  $y = (y_1, \dots, y_N), z = (z_1, \dots, z_N) \in \mathcal{Y}$  und  $P_\theta(T_N(Y) = t) > 0$ . Für alle  $y$  mit  $T_N(y) = t$  ist dann

$$\begin{aligned} P_\theta(Y = y \mid T_N(Y) = t) &= \frac{P_\theta(Y = y, T_N(Y) = t)}{P_\theta(T_N(Y) = t)} \\ &\stackrel{\text{Satz 3.4}}{=} \frac{\prod_{i=1}^N c(y_i) e^{\theta^t \cdot t - N b(\theta)}}{\sum_{\{z \mid T_N(z) = t\}} \prod_{i=1}^N c(z_i) e^{\theta^t T_N(z)}} \\ &= \frac{\prod_{i=1}^N c(y_i) e^{\theta^t \cdot t}}{e^{\theta^t \cdot t} \sum_{\{z \mid T_N(z) = t\}} \prod_{i=1}^N c(z_i)} \\ &= \frac{\prod_{i=1}^N c(y_i)}{\sum_{\{z \mid T_N(z) = t\}} \prod_{i=1}^N c(z_i)} \end{aligned}$$

und für alle  $y$  mit  $T_N(y) \neq t$  gilt

$$P_\theta(Y = y \mid T_N(Y) = t) = 0.$$

$P_\theta(Y = y \mid T_N(Y) = t)$  ist also unabhängig von  $\theta$  und  $T_N(Y)$  somit eine suffiziente Statistik für  $\theta$ .

[7], [22]

□



**Definition 3.7.**

Die Menge der Stichproben mit festem Wert  $t$  der suffizienten Statistik  $T_N$  wird bezeichnet mit

$$\mathcal{Y}_t = \left\{ y = (y_1, \dots, y_N) \in \mathcal{Y} \mid \sum_{i=1}^N T(y_i) = T_N(y) = t \right\}.$$

**Satz 3.8.**

Es sei  $P_\theta^Y$  eine diskrete Verteilung aus der Exponentialfamilie.

Weiter seien  $Y_1, \dots, Y_N$  unabhängige und identisch verteilte  $\mathcal{X}$ -wertige Zufallsvariablen.

Gilt  $c(x) = 1$  für alle  $x \in \mathcal{X}$ , so ist die Verteilung von  $Y = \{Y_1, \dots, Y_N\}$  gegeben  $\{T_N(Y) = t\}$  die Gleichverteilung auf  $\mathcal{Y}_t$ .

*Beweis.* Es sei  $z = (z_1, \dots, z_N) \in \mathcal{Y}$ .

Für alle  $y = (y_1, \dots, y_N) \in \mathcal{Y}$  gilt:

$$\begin{aligned} P_\theta(Y = y \mid T_N(Y) = t) &= \frac{P_\theta(Y = y, T_N(Y) = t)}{P_\theta(T_N(Y) = t)} \\ &\stackrel{\text{Satz 3.4}}{=} \frac{\prod_{i=1}^N c(y_i)}{\sum_{\{z \mid T_N(z) = t\}} \prod_{i=1}^N c(z_i)} \\ &\stackrel{c(x)=1}{=} \frac{1}{\#\{z \mid T_N(z) = t\}} \\ &= \frac{1}{\#\mathcal{Y}_t} \end{aligned}$$

[22]

□

Wie sich herausgestellt hat, ist die Verteilung von  $Y = (Y_1, \dots, Y_N)$  unabhängig von  $\theta$ , wenn man sie mit der suffizienten Statistik  $T_N(Y)$  bedingt. Da diese Tatsache im Folgenden weiter ausgenutzt wird, wird  $P_\theta$  zukünftig der Einfachheit halber nur noch mit  $P$  bezeichnet.

**Definition 3.9.**

Die Menge der Häufigkeitsverteilungen, die man durch die Stichproben mit festem Wert  $t$  der suffizienten Statistik  $T_N$  erhält, wird mit

$$\mathcal{F}_t = \left\{ f := (f(x))_{x \in \mathcal{X}} \in \mathbb{N}_0^{|\mathcal{X}|} \mid \tilde{T}(f) := \sum_{x \in \mathcal{X}} f(x) T(x) = t \right\}$$

bezeichnet.

**Definition 3.10.**

Die Abbildung  $F : \mathcal{Y} \longrightarrow \mathbb{N}_0^{|\mathcal{X}|}$  mit

$$y \longmapsto F(y_1, \dots, y_N) = \left( \sum_{i=1}^N 1_{\{x\}}(y_i) \right)_{x \in \mathcal{X}}$$

verbindet jede Stichprobe mit der dazugehörigen Häufigkeitsverteilung.

Durch die oben definierten Abbildungen ergibt sich folgendes Bild:

$$\begin{array}{ccc} \mathcal{Y} & \xrightarrow{F} & \mathbb{N}_0^{|\mathcal{X}|} \\ & \searrow T_N & \swarrow \tilde{T} \\ & & \mathbb{R}^m \end{array}$$

**Satz 3.11.**

Gilt  $c(x) = 1$  für alle  $x \in \mathcal{X}$ , dann ist das Bildmaß von  $P^{F(Y)}(\cdot | T_N = t)$  unter  $F$  gegeben durch die hypergeometrische Verteilung

$$H_t(f) = \begin{cases} \frac{N!}{\#\mathcal{Y}_t} \prod_{x \in \mathcal{X}} (f(x)!)^{-1} & \text{für } f \in \mathcal{F}_t \\ 0 & \text{sonst} \end{cases}.$$

*Beweis.* Für  $y = (y_1, \dots, y_N) \in \mathcal{Y}$  sei das Bild unter  $F$  gegeben durch

$$F(y) = F(y_1, \dots, y_N) = \left( \sum_{i=1}^N 1_{\{x\}}(y_i) \right)_{x \in \mathcal{X}} := f \in \mathbb{N}_0^{|\mathcal{X}|}.$$

Dann gilt für  $f \in \mathcal{F}_t$ <sup>2</sup>

$$\begin{aligned} P^Y(F^{-1}(f) | T_N(Y) = t) &= P(F(Y) = f | T_N(Y) = t) \\ &= \sum_{\{y | F(Y) = f\}} P(Y = y | T_N(Y) = t) \\ &\stackrel{\text{Satz 3.8}}{=} \frac{\#\{y | F(y) = f\}}{\#\mathcal{Y}_t}, \end{aligned} \tag{3.3}$$

---

<sup>2</sup>Im Falle  $f(y) \notin \mathcal{F}_t$  gilt  $T_N(y) \neq t$ . Hieraus folgt  $P(Y = y | T_N(Y) = t) = 0$ , weswegen gilt:

$$P^Y(F^{-1}(f) | T_N(Y) = t) = \sum_{\{y | F(Y) = f\}} P(Y = y | T_N(Y) = t) = 0.$$

wobei sich  $\#\{y | F(y) = f\}$  auf folgende Weise bestimmen lässt:

Aufgrund der Definition von  $F$  führt das Bild jeder der  $N!$  Umordnungen von  $y$  zum selben Vektor  $f$ .

Es gibt allerdings  $\prod_{x \in \mathcal{X}} f(x)!$  Permutationen, unter denen  $y$  sich nicht ändert. Ist  $c$  dagegen die Anzahl der Permutationen, unter denen  $y$  nicht unverändert bleibt, ergibt sich für die Gesamtzahl der Permutationen:

$$N! = \left( \prod_{x \in \mathcal{X}} f(x)! \right) \cdot c.$$

Es gibt also genau

$$c = \frac{N!}{\prod_{x \in \mathcal{X}} f(x)!}$$

unterschiedliche Stichproben  $y$ , die unter  $F$  das gleiche Bild  $f$  besitzen, d.h.

$$\#\{y | F(y) = f\} = c = \frac{N!}{\prod_{x \in \mathcal{X}} f(x)!}.$$

Setzt man dies in (3.3) ein, erhält man schließlich die gewünschte Aussage.

[22]

□

### 3.5 Bedingter p-Wert

Zur Entscheidung über Annahme oder Ablehnung der Nullhypothese bei einem statistischen Test wird meist der p-Wert herangezogen. In diesem Abschnitt soll ein entsprechender Wert für einen Test mit Hilfe der oben benutzten bedingten Wahrscheinlichkeit eingeführt werden. (Vgl. hierzu [24].)

#### Definition 3.12.

Es sei  $g$  eine Funktion auf  $\mathcal{F}_t$  und  $H_t$  die hypergeometrische Verteilung auf  $\mathcal{F}_t$ .

Der **bedingte p-Wert** zur Teststatistik  $G$  ist definiert als ein Erwartungswert der Form

$$E_{H_t}(g(F(Y))) \quad \text{für } F(y) \in \mathcal{F}_t$$

mit  $g(F(y)) = 1_{\{G(F(y)) \geq G(f^{(0)})\}}$ .

Bei einem statistischen Test zum Niveau  $\alpha$  wird die Nullhypothese abgelehnt, wenn für die Entscheidungsfunktion  $g$  gilt:

$$P(g(F(Y)) = 1) \leq \alpha.$$

Dass dies analog auch für den bedingten p-Wert der Fall ist, zeigt der folgende Satz:

**Satz 3.13.**

Gilt für den bedingten p-Wert

$$E_{H_t}(g(F(Y))) \leq \alpha \quad \text{für alle } t,$$

so ist auch  $P(g(F(Y)) = 1) \leq \alpha$ .

*Beweis.* Es sei  $E_{H_t}(g(F(Y))) \leq \alpha$ .

Dann gilt mit dem Satz von der totalen Wahrscheinlichkeit:

$$\begin{aligned} P(g(F(Y)) = 1) &= \sum_t P(T_N = t) P(g(F(Y)) = 1 \mid T_N = t) \\ &= \sum_t P(T_N = t) \underbrace{E_{H_t}(g(F(Y)))}_{\leq \alpha} \\ &\leq \alpha \cdot \underbrace{\sum_t P(T_N = t)}_{=1} \\ &= \alpha \end{aligned}$$

□

**Definition 3.14.**

Für den  $\chi^2$ -**Unabhängigkeitstest** mit der Teststatistik

$$T_{\chi^2}(f) = \sum_{k=1}^K \sum_{l=1}^L \frac{(f_{kl} - \frac{f_{k+}f_{+l}}{N})^2}{\frac{f_{k+}f_{+l}}{N}}$$

sei  $g(f) := 1_{\{T_{\chi^2}(f) \geq T_{\chi^2}(f^{(0)})\}}(f)$ .

Der Wert

$$\mu = E_{H_t}(1_{\{T_{\chi^2}(f) \geq T_{\chi^2}(f^{(0)})\}}(F(Y)))$$

wird als der **bedingter p-Wert des  $\chi^2$ -Tests** bezeichnet.

Da  $\mathcal{F}_t$  in den meisten Fällen sehr groß ist und deshalb nicht alle Elemente  $f \in \mathcal{F}_t$  aufgezählt werden können, ist es oft nicht möglich den Wert  $E_{H_t}(g(f(Y)))$  exakt zu berechnen.

Mit Hilfe des Ergodensatzes 1.53 kann man stattdessen  $E_{H_t}(g(f(Y)))$  approximieren, indem man Zufallselemente  $f^{(1)}, \dots, f^{(n)}$  aus der hypergeometrischen Verteilung  $H_t$  generiert und die Schätzung

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n g(f^{(i)})$$

für  $\mu = E_{H_t}(g(F(Y)))$  benutzt.

### 3.6 Kontingenztafeln mit gegebenen Zeilen- und Spaltensummen

In Abschnitt 3.4 wurde gezeigt, dass die gemeinsame Verteilung von  $Y_1, \dots, Y_N$ , die suffiziente Statistik  $T_N$  gegeben, unabhängig von  $\theta$  ist. Weiter erhielt man als Verteilung auf  $\mathcal{F}_t$  die von  $\theta$  unabhängige hypergeometrische Verteilung  $H_t$ .

In diesem Abschnitt wird all dies auf die, aus einer Stichprobe der Größe  $N$  resultierende,  $K \times L$ -Kontingenztafel übertragen. Es wird gezeigt, dass unter der Nullhypothese der Unabhängigkeit (3.1) zum Einen die Zeilen- und Spaltensummen eine suffiziente Statistik für  $\theta$  darstellen (Satz 3.16) und zum Anderen auf der Menge  $\mathcal{F}_{(r,c)}$  der Häufigkeitstabellen mit vorgegebenen Zeilen- und Spaltensummen eine von  $\theta$  unabhängige Verteilung  $H_{(r,c)}$  vorliegt (Satz 3.20).

#### Satz 3.15.

Es seien  $Y_1, \dots, Y_N$  voneinander unabhängige Zufallsvariablen mit  $Y_i = (V_i, W_i)$  für  $i = 1, \dots, N$ . Weiter seien  $v_i \in \{1, \dots, K\}$  und  $w_i \in \{1, \dots, L\}$  die Realisierungen von  $V_i$  bzw.  $W_i$ .

Unter der Nullhypothese (3.1) ist die gemeinsame Dichte von  $Y = (Y_1, \dots, Y_N)$  gegeben durch

$$P_\theta(Y_1 = y_1, \dots, Y_N = y_N) = e^{\sum_{k=1}^K f_{k+} \ln(\pi_{k+}) + \sum_{l=1}^L f_{+l} \ln(\pi_{+l})}.$$

Speziell ist sie also eine Dichte aus der Exponentialfamilie mit  $c(y_i) = 1$  für  $i = 1, \dots, N$ .

*Beweis.* Es gilt:

$$\begin{aligned} P(Y_i = y_i) &= P(V_i = v_i, W_i = w_i) \\ &= \prod_{k=1}^K \prod_{l=1}^L \pi_{kl}^{1_{\{k\}}(v_i) 1_{\{l\}}(w_i)} \\ &\stackrel{(3.1)}{=} \prod_{k=1}^K \prod_{l=1}^L \pi_{k+}^{1_{\{k\}}(v_i)} \pi_{+l}^{1_{\{l\}}(w_i)} \\ &= \prod_{k=1}^K \pi_{k+}^{1_{\{k\}}(v_i)} \prod_{l=1}^L \pi_{+l}^{1_{\{l\}}(w_i)} \\ &= \prod_{k=1}^K e^{\ln(\pi_{k+}) 1_{\{k\}}(v_i)} \prod_{l=1}^L e^{\ln(\pi_{+l}) 1_{\{l\}}(w_i)} \\ &= e^{\sum_{k=1}^K \ln(\pi_{k+}) 1_{\{k\}}(v_i) + \sum_{l=1}^L \ln(\pi_{+l}) 1_{\{l\}}(w_i)} \end{aligned}$$

Man setze

$$\theta = (\ln(\pi_{1+}), \dots, \ln(\pi_{K+}), \ln(\pi_{+1}), \dots, \ln(\pi_{+L}))$$

und

$$\begin{aligned} T(y_i) &= (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0) \\ &= (e_{v_i}^1, e_{w_i}^2), \end{aligned}$$

wobei die 1 jeweils an der Stelle  $v_i$  und der Stelle  $K + w_i$  steht und  $e_{v_i}^1 \in \mathbb{R}^K$  und  $e_{w_i}^2 \in \mathbb{R}^L$  den jeweiligen kanonischen Basisvektor darstellen.

D.h. es sei

$$T(y_i) = \left( \sum_{l=1}^L 1_{\{(1,l)\}}(y_i), \dots, \sum_{l=1}^L 1_{\{(K,l)\}}(y_i), \sum_{k=1}^K 1_{\{(k,1)\}}(y_i), \dots, \sum_{k=1}^K 1_{\{(k,L)\}}(y_i) \right)$$

Mit  $c(y_i) = 1$  für  $i = 1, \dots, N$  und  $b(\theta) = 0$  ergibt sich somit:

$$P_\theta(Y_i = y_i) = 1 \cdot e^{\theta^t T(y_i) - 0}.$$

Weiter gilt:

$$\sum_{i=1}^N T(y_i) = (f_{1+}, \dots, f_{K+}, f_{+1}, \dots, f_{+L})$$

Mit Hilfe von Satz 3.4 erhält man schließlich

$$\begin{aligned} P_\theta(Y_1 = y_1, \dots, Y_N = y_N) &= e^{\theta^t \sum_{i=1}^N T(y_i) - Nb(\theta)} \prod_{i=1}^N c(y_i) \\ &= e^{\theta^t \sum_{i=1}^N T(y_i)} \\ &= e^{(\ln(\pi_{1+}), \dots, \ln(\pi_{K+}), \ln(\pi_{+1}), \dots, \ln(\pi_{+L}))^t (f_{1+}, \dots, f_{K+}, f_{+1}, \dots, f_{+L})} \\ &= e^{\sum_{k=1}^K f_{k+} \ln(\pi_{k+}) + \sum_{l=1}^L f_{+l} \ln(\pi_{+l})}. \end{aligned}$$

□

### Satz 3.16.

Es seien  $Y = (Y_1, \dots, Y_N)$  ein Beobachtungsvektor und  $F(Y) = (f_{kl}(Y))$  die zugehörige  $K \times L$ -Kontingenztafel.

Unter der Nullhypothese (3.1) ist

$$T_N(Y) = (f_{1+}(Y), \dots, f_{K+}(Y), f_{+1}(Y), \dots, f_{+L}(Y))$$

eine suffiziente Statistik für  $\theta$ .

*Beweis.*  $c(y_i)$  und  $\theta$  seien wie in Satz 3.15 definiert.

Weiter seien

$$r = (r_1, \dots, r_K) \in \mathbb{N}_0^K \quad \text{und} \quad c = (c_1, \dots, c_L) \in \mathbb{N}_0^L$$

gegeben. Dann gilt:

$$\begin{aligned}
P(F(Y) = f \mid T_N(Y) = (r, c)) &= \sum_{\{y \mid F(y)=f\}} P(Y = y \mid T_N(Y) = (r, c)) \\
&= \sum_{\{y \mid F(y)=f\}} \frac{P(Y = y, T_N(Y) = (r, c))}{P(T_N(Y) = (r, c))} \\
&\stackrel{\text{Satz 3.15}}{=} \sum_{\{y \mid F(y)=f\}} \frac{e^{\sum_{k=1}^K r_k \ln(\pi_{k+}) + \sum_{l=1}^L c_l \ln(\pi_{+l})}}{\sum_{\{z \mid T_N(z)=(r,c)\}} e^{\sum_{k=1}^K f_{k+}(z) \ln(\pi_{k+}) + \sum_{l=1}^L f_{+l}(z) \ln(\pi_{+l})}} \\
&= \sum_{\{y \mid F(y)=f\}} \frac{e^{\sum_{k=1}^K r_k \ln(\pi_{k+}) + \sum_{l=1}^L c_l \ln(\pi_{+l})}}{e^{\sum_{k=1}^K r_k \ln(\pi_{k+}) + \sum_{l=1}^L c_l \ln(\pi_{+l})} \sum_{\{z \mid T_N(z)=(r,c)\}} 1} \\
&= \sum_{\{y \mid F(y)=f\}} \frac{1}{\#\{z \mid T_N(z) = (r, c)\}} \\
&= \frac{\#\{y \mid F(y) = f\}}{\#\{z \mid T_N(z) = (r, c)\}}. \tag{3.4}
\end{aligned}$$

$P(F(Y) = f \mid T_N(Y) = (r, c))$  ist also unabhängig von  $\theta$  und  $T_N$  somit eine suffiziente Statistik für  $\theta$ .  $\square$

Entsprechend der Definitionen von  $\mathcal{Y}_t$  und  $\mathcal{F}_t$  (Abschnitt 3.4) werden nun folgende Bezeichnungen eingeführt:

**Definition 3.17.**

$\mathcal{Y}_{(r,c)}$  bezeichne die Menge der Stichproben mit festen Werten  $(r, c)$  der suffizienten Statistik  $T_N$ , d.h.

$$\mathcal{Y}_{(r,c)} = \left\{ y = (y_1, \dots, y_N) \in \mathcal{Y} \mid \sum_{i=1}^N T(y_i) = T_N(y) = (r, c) \right\}.$$

**Definition 3.18.**

Es sei

$$\mathcal{F}_{(r,c)} := \left\{ f = (f_{kl}) \in \mathbb{N}_0^{|\mathcal{X}|} \mid \tilde{T}(f) = \sum_{k=1}^K \sum_{l=1}^L f_{kl} T((k, l)) = (r, c) \right\}$$

die Menge der  $K \times L$ -Häufigkeitstabellen mit nichtnegativen Einträgen, gegebenen Zeilensummen  $r = (r_1, \dots, r_K)$  und gegebenen Spaltensummen  $c = (c_1, \dots, c_L)$ .

Analog zu Definition 3.10 erhält man:

**Definition 3.19.**

Die Abbildung  $F : \mathcal{Y} \longrightarrow \mathbb{N}_0^{|\mathcal{X}|}$  mit

$$y \longmapsto F(y_1, \dots, y_N) = \left( \sum_{i=1}^N 1_{\{(k,l)\}}(y_i) \right)_{\substack{k=1, \dots, K \\ l=1, \dots, L}} := (f_{kl}(y))_{\substack{k=1, \dots, K \\ l=1, \dots, L}}$$

verbindet jede Stichprobe  $y = (y_1, \dots, y_N)$  mit der dazugehörigen  $K \times L$ -Kontingenztafel.

**Satz 3.20.**

Unter der Nullhypothese (3.1) ist die Verteilung auf  $\mathcal{F}_{(r,c)}$  gegeben durch

$$H_{(r,c)}(f) = \begin{cases} \frac{\prod_{l=1}^L \binom{c_l}{f_{1l} \dots f_{Kl}}}{\binom{N}{r_1 \dots r_K}} & \text{für } f \in \mathcal{F}_{(r,c)} \\ 0 & \text{sonst} \end{cases}.$$

*Beweis.* Mit (3.4) gilt analog zum Beweis von Satz 3.11:

$$P(F(Y) = (f_{kl})_{\substack{k=1, \dots, K \\ l=1, \dots, L}} \mid T_N(Y) = (r, c)) = \frac{N!}{\#\mathcal{Y}_{(r,c)}} \prod_{l=1}^L \prod_{k=1}^K (f_{kl})^{-1}. \quad (3.5)$$

$\#\mathcal{Y}_{(r,c)}$  ergibt sich hierfür wie folgt:

Es existieren

$$a = \binom{N}{r_1 \dots r_K}$$

unterschiedliche Stichproben  $y$ , die unter  $F$  eine Häufigkeitstabelle mit den Zeilensummen  $r_1, \dots, r_K$  erzeugen.

Entsprechend ergeben sich

$$b = \binom{N}{c_1 \dots c_L}$$

verschiedene Stichproben  $y$ , für die die Tabellen  $F(y) = f$  die Spaltensummen  $c_1, \dots, c_L$  aufweisen.

Somit gibt es genau

$$\#\mathcal{Y}_{(r,c)} = a \cdot b = \binom{N}{r_1 \dots r_K} \cdot \binom{N}{c_1 \dots c_L}$$

verschiedene Stichproben  $y = (y_1, \dots, y_N)$ , die als Bild unter  $F$  eine Häufigkeitstabelle mit den Zeilensummen  $r$  und den Spaltensummen  $c$  besitzen.



Setzt man dies in (3.5) ein, erhält man

$$\begin{aligned}
H_{(r,c)}(f) &= \frac{N!}{\#\mathcal{Y}_{(r,c)}} \prod_{l=1}^L \prod_{k=1}^K (f_{kl})^{-1} \\
&= \frac{N!}{\binom{N}{r_1 \dots r_K} \cdot \binom{N}{c_1 \dots c_L}} \prod_{l=1}^L \prod_{k=1}^K (f_{kl})^{-1} \\
&= \frac{N!}{\binom{N}{r_1 \dots r_K} \prod_{l=1}^L \prod_{k=1}^K f_{kl}} \cdot \frac{\prod_{l=1}^L c_l!}{N!} \\
&= \prod_{l=1}^L \frac{c_l!}{\prod_{k=1}^K f_{kl}} \cdot \frac{1}{\binom{N}{r_1 \dots r_K}} \\
&= \frac{\prod_{l=1}^L \binom{c_l}{f_{1l} \dots f_{kl}}}{\binom{N}{r_1 \dots r_K}}.
\end{aligned}$$

□

### 3.7 Modifizierter Metropolis Algorithmus

Um einen bedingten p-Wert des  $\chi^2$ -Tests (siehe Abschnitt 3.5) approximativ mit Hilfe des Ergodensatzes (Satz 1.53) zu berechnen, muss eine irreduzible, aperiodische Markov Kette  $\{f^{(i)}; i \in \mathbb{Z}^+\}$  mit der invarianten Wahrscheinlichkeitsverteilung  $H_t$  simuliert und der Wert  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n g(f^{(i)})$  berechnet werden.

Im letzten Abschnitt wurde gezeigt, dass unter der Annahme der Unabhängigkeit auf der Menge  $\mathcal{F}_{(r,c)}$  die hypergeometrische Verteilung  $H_{(r,c)}$  vorliegt (Satz 3.20). Entsprechend muss für einen Test auf Unabhängigkeit also eine irreduzible, aperiodische Markov Kette mit Zustandsraum  $\mathcal{F}_{(r,c)}$  simuliert werden.

**Definition 3.21.**

Eine  $K \times L$ -Matrix des  $\begin{matrix} + & - \\ - & + \end{matrix}$ -Typs ist eine Matrix der Form

$$M_{k_1 k_2 l_1 l_2} := \begin{pmatrix} & & l_1 & & l_2 & & \\ 0 & \dots & 0 & & \dots & 0 & \\ \vdots & 0 & +1 & \vdots & -1 & 0 & \vdots \\ 0 & \dots & 0 & & \dots & 0 & \\ \vdots & 0 & -1 & \vdots & +1 & 0 & \vdots \\ 0 & \dots & 0 & & \dots & 0 & \end{pmatrix} \begin{matrix} k_1 \\ k_2 \end{matrix} \quad \begin{matrix} k_1, k_2 \in \{1, \dots, K\}, \\ l_1, l_2 \in \{1, \dots, L\}. \end{matrix}$$

D.h. es ist eine  $K \times L$ -Matrix  $M_{k_1 k_2 l_1 l_2} = (m_{kl})$ , bei der alle Einträge 0 sind, ausgenommen der Einträge  $(k_1, l_1)$  und  $(k_2, l_2)$ , die jeweils 1 betragen, und der Einträge  $(k_1, l_2)$  und  $(k_2, l_1)$ , die beide gleich  $-1$  sind.

Angenommen, es sind  $f \in \mathcal{F}_{(r,c)}$  und eine Matrix  $M_{k_1 k_2 l_1 l_2}$  des  $\begin{matrix} + & - \\ - & + \end{matrix}$ -Typs,

$k_1, k_2 \in \{1, \dots, K\}$  und  $l_1, l_2 \in \{1, \dots, L\}$  beliebig aber fest, gegeben.

Addiert oder subtrahiert man  $M_{k_1 k_2 l_1 l_2}$  zu bzw. von  $f$ , so bleiben die Zeilen- und Spaltensummen unverändert. D.h. sind alle Einträge der Matrix

$$\tilde{f} := f + M_{k_1 k_2 l_1 l_2}$$

bzw.

$$\bar{f} := f - M_{k_1 k_2 l_1 l_2}$$

nichtnegativ, gilt wieder  $\tilde{f} \in \mathcal{F}_{(r,c)}$  bzw.  $\bar{f} \in \mathcal{F}_{(r,c)}$ .

Um also ausgehend von einem  $f^{(0)} \in \mathcal{F}_{(r,c)}$  eine Markov Kette auf  $\mathcal{F}_{(r,c)}$  zu simulieren, geht man wie folgt vor:

- Wähle zufällig zwei Zeilen  $k_1$  und  $k_2$  und zwei Spalten  $l_1$  und  $l_2$  aus
- Wähle  $\epsilon = \pm 1$  jeweils mit der Wahrscheinlichkeit  $\frac{1}{2}$
- Berechne  $\tilde{f} = f^{(i)} + \epsilon \cdot M_{k_1 k_2 l_1 l_2}$
- Gilt  $\tilde{f}_{kl} \geq 0$  für alle  $k = 1, \dots, K$  und alle  $l = 1, \dots, L$ ,  
setze  $f^{(i+1)} = \tilde{f}$  mit der Akzeptanzwahrscheinlichkeit  $\alpha_{f\tilde{f}}$
- Setze  $f^{(i+1)} = f^{(i)}$  sonst.

Wendet man den Metropolis Algorithmus an, ergibt sich für die Akzeptanzwahrscheinlichkeit:

**Satz 3.22.**

Es seien  $\epsilon \in \{-1, 1\}$ ,  $f \in \mathcal{F}_{(r,c)}$  und  $M_{k_1 k_2 l_1 l_2} = (m_{kl})$  eine Matrix vom  $\begin{matrix} + & - \\ - & + \end{matrix}$ -Typ.

Weiter sei  $\tilde{f} = f + \epsilon \cdot M_{k_1 k_2 l_1 l_2}$  mit  $\tilde{f}_{kl} \geq 0$ .

Dann ergibt sich für die Akzeptanz-Wahrscheinlichkeit

$$\alpha_{f\tilde{f}} = \min \left\{ 1, \prod_{\substack{l=1 \\ m_{kl} \neq 0}}^L \prod_{\substack{k=1 \\ m_{kl} \neq 0}}^K \frac{f_{kl}!}{(f_{kl} + \epsilon \cdot m_{kl})!} \right\}.$$

*Beweis.* Es sei  $r = (r_1, \dots, r_K)$  der Vektor mit den vorgegebenen Zeilensummen, d.h. es seien  $f_{K+} = r_K$  für alle  $k = 1, \dots, K$ .  
 $c = (c_1, \dots, c_L)$  bezeichne entsprechend den Vektor mit den vorgegebenen Spaltensummen, d.h.  $f_{+L} = c_l$  für alle  $l = 1, \dots, L$ .

Es gilt

$$\begin{aligned}
\frac{H_{(r,c)}(\tilde{f})}{H_{(r,c)}(f)} &= \frac{H_{(r,c)}(f + \epsilon \cdot M_{k_1 k_2 l_1 l_2})}{H_{(r,c)}(f)} \\
&= \frac{\prod_{l=1}^L \binom{c_l}{(f_{1l} + \epsilon \cdot m_{1l}) \dots (f_{kl} + \epsilon \cdot m_{kl})}}{\binom{N}{r_1 \dots r_K}} \cdot \frac{\binom{N}{r_1 \dots r_K}}{\prod_{l=1}^L \binom{c_l}{f_{1l} \dots f_{kl}}} \\
&= \prod_{l=1}^L \frac{c_l!}{\prod_{k=1}^K (f_{kl} + \epsilon \cdot m_{kl})!} \cdot \frac{\prod_{k=1}^K f_{kl}!}{c_l!} \\
&= \prod_{l=1}^L \prod_{k=1}^K \frac{f_{kl}!}{(f_{kl} + \epsilon \cdot m_{kl})!} \\
&= \prod_{\substack{l=1 \\ f_{kl} \neq 0}}^L \prod_{\substack{k=1 \\ m_{kl} \neq 0}}^K \frac{f_{kl}!}{(f_{kl} + \epsilon \cdot m_{kl})!}
\end{aligned}$$

und mit (2.3) ergibt sich schließlich

$$\alpha_{f\tilde{f}} = \min \left\{ 1, \prod_{\substack{l=1 \\ m_{kl} \neq 0}}^L \prod_{\substack{k=1 \\ m_{kl} \neq 0}}^K \frac{f_{kl}!}{(f_{kl} + \epsilon \cdot m_{kl})!} \right\}.$$

[22]

□

Da die Matrix  $M_{k_1 k_2 l_1 l_2}$  lediglich vier Einträge  $\neq 0$  besitzt, fließen auch in die Berechnung der Akzeptanz-Wahrscheinlichkeit nur vier Einträge von  $f$  ein, nämlich  $f_{k_1 l_1}$ ,  $f_{k_1 l_2}$ ,  $f_{k_2 l_1}$  und  $f_{k_2 l_2}$ .

Es muss nun nur noch gegeben sein, dass die mit den oben beschriebenen Algorithmus simulierte Markov Kette irreduzibel und aperiodisch ist. Hierfür wurde in [7] folgende Definition eingeführt:

**Definition 3.23.**

Eine **Markov Basis** von  $\mathcal{F}_{(r,c)}$  ist eine Menge von Funktionen  $m_1, m_2, \dots, m_M : \mathcal{X} \rightarrow \mathbb{Z}$  so, dass für jedes  $1 \leq i \leq M$

$$\sum_{x \in \mathcal{X}} m_i(x) T(x) = 0$$

gilt, wobei  $T = (r, c)$  die suffiziente Statistik ist.

Für alle  $m, m' \in \mathcal{F}_{(r,c)}$  gibt es weiter  $(\epsilon_1, m_{i_1}), \dots, (\epsilon_A, m_{i_A})$  mit  $\epsilon_i = \pm 1$  so, dass

$$m' = m + \sum_{j=1}^A \epsilon_j m_{i_j}$$

und

$$m + \sum_{j=1}^a \epsilon_j m_{i_j} \geq 0 \quad \text{für } 1 \leq a \leq A.$$

In [7] und [22] wird gezeigt, dass die mit dem Metropolis Algorithmus simulierte Markov Kette irreduzibel und aperiodisch ist, wenn man die Vorschläge  $y$  gleichmäßig aus der Markov Basis zieht. Betrachtet man zusätzlich folgenden Satz aus [22, S.13], ist damit auch gewährleistet, dass durch den oben beschriebenen Algorithmus eine irreduzible und aperiodische Kette entsteht.

**Satz 3.24.**

Im Falle der Unabhängigkeit besteht die Markov Basis aus allen Matrizen des

$$\begin{array}{cc} + & - \\ - & + \end{array} \text{-Typs.}$$

Zusammenfassend ergibt sich also Folgendes:

**Algorithmus:**

Es sei  $f^{(0)}$  eine  $K \times L$ -Tabelle mit  $f_{kl}^{(0)} \geq 0$  für alle  $k = 1, \dots, K$  und alle  $l = 1, \dots, L$ , sowie mit  $f_{K+}^{(0)} = r_K$  für  $k = 1, \dots, K$  und  $f_{+L}^{(0)} = c_l$  für  $l = 1, \dots, L$ .

Möchte man eine Markov Kette mit Zustandsraum  $\mathcal{F}_{(r,c)}$  und invarianter Wahrscheinlichkeitsverteilung  $H_{(r,c)}$  simulieren, geht man wie folgt vor:

- Wiederhole für  $i = 1, \dots, n$ :
- (\*) Wähle zufällig zwei Zeilen  $k_1$  und  $k_2$  und zwei Spalten  $l_1$  und  $l_2$  aus
- Wähle  $\epsilon = \pm 1$  jeweils mit der Wahrscheinlichkeit  $\frac{1}{2}$
- Berechne  $\tilde{f} = f^{(i)} + \epsilon \cdot M_{k_1 k_2 l_1 l_2}$
- Existiert ein  $\tilde{f}_{kl} < 0$ , setze  $f^{(i+1)} = f^{(i)}$  und gehe zu (\*)
- Generiere ein  $u$  aus  $\mathcal{U}(0, 1)$
- Gilt  $\tilde{f}_{kl} \geq 0$  für alle  $k = 1, \dots, K$  und alle  $l = 1, \dots, L$  und ist  $u < \alpha_{f\tilde{f}}$ , setze  $f^{(i+1)} = \tilde{f}$
- Setze  $f^{(i+1)} = f^{(i)}$  sonst.

# Kapitel 4

## Anwendungsbeispiele

Der in den vorherigen Kapiteln theoretisch erarbeitete Algorithmus zum Testen stochastischer Unabhängigkeit mit Hilfe der Markov Ketten Monte Carlo Methode soll nun anhand von Beispielen praktisch angewandt werden. Hierfür dient im Folgenden die Funktion `unab(x0,n,d,schritt)`, die in Anhang A.2 näher beschrieben wird. Der Schätzwert für den bedingten p-Wert des  $\chi^2$ -Tests sei auch hier mit

$$\hat{\mu} := \frac{1}{n+1} \sum_{i=0}^n \mathbf{1}_{\{T_{\chi^2}(f^{(i)}) \geq T_{\chi^2}(f^{(0)})\}}(f^{(i)})$$

bezeichnet.

Zum Vergleich soll zusätzlich der  $\chi^2$ -Unabhängigkeitstest durchgeführt werden, der bei der freien Software R Version 2.6.2 mit dem Befehl `chisq.test()` aufrufbar ist.

### 4.1 Automarkt

Für einen Autohändler soll getestet werden, ob es beim Kauf eines Autos einer bestimmten Marke darauf ankommt, aus welchem Viertel der Stadt der Kunde kommt. Durch Umfragen sei Tabelle 4.1 entstanden, die die Beziehung zwischen dem Wohnviertel einer befragten Person und der bevorzugten Automarke für den in Kürze geplanten Autokauf angibt.

Testet man nun mit Hilfe des  $\chi^2$ -Unabhängigkeitstests, ob die beiden Variablen (Wohnviertel und bevorzugte Automarke) unabhängig voneinander sind, ergibt sich mit R ein p-Wert von 0,01375:

```
Pearson's Chi-squared test
```

```
data: y  
X-squared = 16, df = 6, p-value = 0.01375
```

Da bei dieser Tabelle alle erwarteten Häufigkeiten  $> 5$  sind, wird mit der  $\chi^2$ -Teststatistik eine zufriedenstellende Approximation der  $\chi^2$ -Verteilung erreicht.

Wohn- viertel	Automarke			Summe
	A	B	C	
I	60	50	40	150
II	40	50	60	150
III	60	50	40	150
IV	40	50	60	150
Summe	200	200	200	600

Tabelle 4.1: Beziehung zwischen Wohnviertel und bevorzugter Automarke

Wendet man zum Vergleich die Funktion `unab(x0,n,d,schritt)` an, ergibt sich der bedingte p-Wert:

```
> set.seed(000)
> unab(y1,50000,0,1)
```

```
      [,1] [,2] [,3]
[1,]   60   50   40
[2,]   40   50   60
[3,]   60   50   40
[4,]   40   50   60
```

approximierter p-Wert= 0.01343973

In Abbildung 4.1<sup>1</sup> wird der Verlauf von  $\hat{\mu}$  aufgezeigt. Offensichtlich ist bei 50000 Iterationen ein stabiler Wert für  $\hat{\mu}$  erreicht, so dass man den zu diesem Zeitpunkt berechneten Wert als approximierten p-Wert ansehen kann.

Insgesamt ist also zu erkennen, dass bei dieser Tabelle mit großen Einträgen (d.h. erwartete Häufigkeiten  $> 5$ ) ein vernachlässigbaren Unterschied (Differenz 0,00031027) zwischen dem mit dem  $\chi^2$ -Test berechneten p-Wert und dem mit Hilfe der Markov Ketten Monte Carlo Methode berechneten approximierten bedingten p-Wert vorliegt. Die Nullhypothese der Unabhängigkeit von Wohnviertel und bevorzugter Automarke der befragten Personen würde man mit Hilfe beider Verfahren auf dem Niveau 0,05 ablehnen.

---

<sup>1</sup>Der Verlauf von  $\hat{\mu}$  wurde hier und wird im Folgenden ohne Einschwingphase ( $d = 0$ ) und unter Einbeziehung jedes einzelnen Schrittes ( $\text{schritt} = 1$ ) erzeugt.

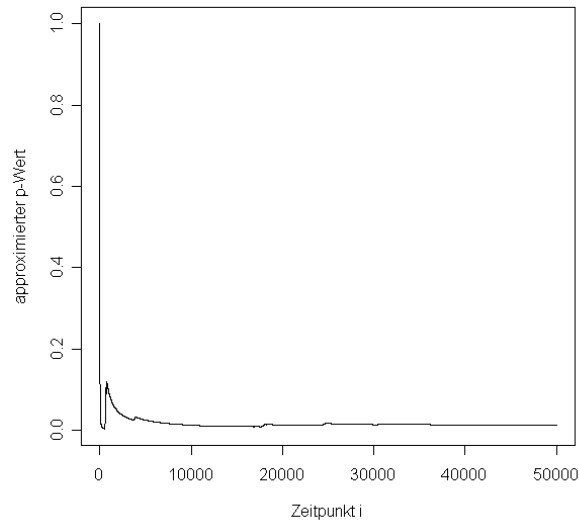


Abbildung 4.1: Verlauf des approximierten bedingten p-Werts  $\hat{\mu}$  ( $n = 50000$ , Tabelle 4.1)

## 4.2 Scheidungsrecht

Mit Hilfe von Tabelle 4.2 soll getestet werden, ob die Wahl einer Partei mit der Einstellung des Wählers zu einem neuen Scheidungsrecht zusammenhängt. (Beispiel orientiert an [21, S.78])

Einstellung zum Scheidungsrecht	Partei				Summe
	A	B	C	Sonstige	
negativ	60	40	20	10	130
positiv	20	15	10	5	50
Summe	80	55	30	15	180

Tabelle 4.2: Beziehung zwischen gewählter Partei und Einstellung zum neuen Scheidungsrecht

Man teste hierfür die beiden Variablen (Partei und Einstellung) auf Unabhängigkeit. Der  $\chi^2$ -Unabhängigkeitstest ergibt hier einen p-Wert von 0,7996:

```
> chisq.test(y)
```

```
      Pearson's Chi-squared test
```

```
data:  y
```

```
X-squared = 1.007, df = 3, p-value = 0.7996
```

```
Warning message:
```

```
In chisq.test(y) : Chi-squared approximation may be incorrect
```

Da

$$\tilde{\mu}_{24} = \frac{f_{2+}f_{+4}}{N} = \frac{50 \cdot 15}{180} = 4,166667 < 5$$

gilt (vgl. Abschnitt 3.2.1), wird in R eine Warnung angezeigt, die auf eine möglicherweise nicht zufriedenstellende Approximation der  $\chi^2$ -Verteilung durch die  $\chi^2$ -Teststatistik hinweist.

Um dieses Problem zu umgehen kann der modifizierte Metropolis Algorithmus angewandt werden. Abbildung 4.2 zeigt, dass bei 40000 – 50000 Iterationen ein stabiler Wert für  $\hat{\mu}$  erreicht wird und zudem die ersten ca. 9000 Iterationen wegen der großen Ausschläge vernachlässigt werden können. Bei 50000 Iterationen, mit einer Einschwingphase von 9000 und unter Einbeziehung jedes 10. Schrittes werden somit insgesamt 4101 Iterationen für die Berechnung des bedingten p-Werts benutzt:

```
> set.seed(123456)
```

```
> unab(y,50000,9000,10)
```

```
      [,1] [,2] [,3] [,4]
[1,]   60   40   20   10
[2,]   20   15   10    5
```

```
approximierter p-Wert= 0.7983419
```

Mit beiden Verfahren wird also die Nullhypothese der Unabhängigkeit auf dem Niveau 0,05 angenommen, so dass die Wahl der Partei und die Einstellung des Wählers zu dem neuen Scheidungsrecht hier unabhängig voneinander sind.

Auch in diesem Fall ist die Differenz zwischen dem approximierten bedingten p-Wert und dem p-Wert des  $\chi^2$ -Tests gering (0,0012581). Da in Tabelle 4.2 weniger als 20% der erwarteten Zelhäufigkeiten  $< 5$  sind, erhält man durch die  $\chi^2$ -Teststatistik eine noch befriedigende Approximation der  $\chi^2$ -Verteilung, weswegen das Ergebnis des  $\chi^2$ -Unabhängigkeitstests akzeptabel ist.

Was aber geschieht, wenn eine Kontingenztafel ausschließlich geringe Einträge vorweist?



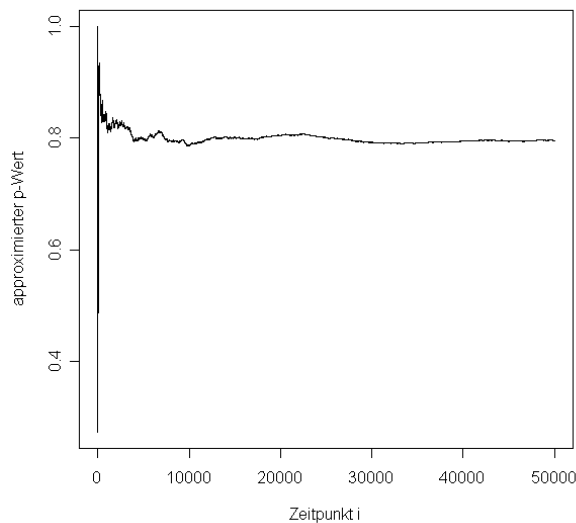


Abbildung 4.2: Verlauf von  $\hat{\mu}$  ( $n = 50000$ , Tabelle 4.2)

### 4.3 Tod durch Pferdetritte

Tabelle 4.4 auf Seite 90 berichtet über die Anzahl der Tode durch Pferdetritte in der Preussischen Armee (vgl. [5, S.326]). Von Interesse sei zu testen, ob es einen Zusammenhang zwischen Todesjahr und Regiment der Toten gab.

Wendet man dafür den  $\chi^2$ -Unabhängigkeitstests an, so ergibt sich für diese Tabelle ein p-Wert von 0,6173:

```
> chisq.test(y1)
```

```
Pearson's Chi-squared test
```

```
data: y1
```

```
X-squared = 239.7686, df = 247, p-value = 0.6173
```

```
Warning message:
```

```
In chisq.test(y1) : Chi-squared approximation may be incorrect
```

Da hier jedoch alle erwarteten Zellhäufigkeiten  $< 5$  sind, wird mit der  $\chi^2$ -Teststatistik nur eine unbefriedigende Approximation der  $\chi^2$ -Verteilung erreicht, was in R einen Warnhinweis verursacht. Wendet man deshalb stattdessen den modifizierten Metropolis Algorithmus an, ergibt sich für  $\hat{\mu}$  der in Abbildung 4.3 gezeigte Verlauf.

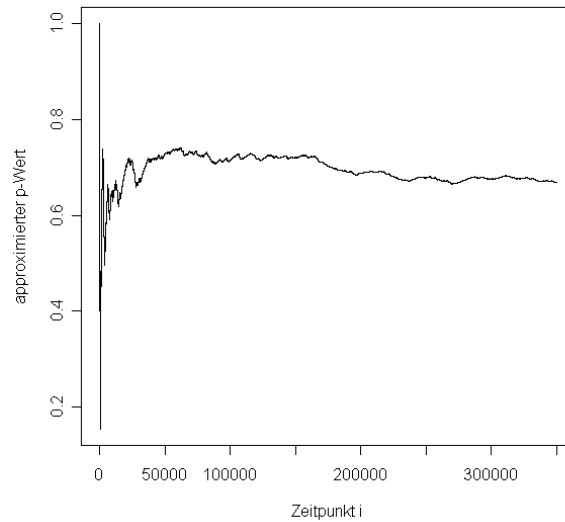


Abbildung 4.3: Verlauf von  $\hat{\mu}$  ( $n = 350000$ , Tabelle 4.4)

Hier ist zu erkennen, dass bei etwa 350000 Iterationen ein relativ stabiler Wert von  $\hat{\mu}$  erreicht ist und die ersten 20000 Schritte für die Berechnung des bedingten p-Werts vernachlässigt werden können. Bezieht man nun nur jeden 100. Schritt in die Rechnung mit ein, so erhält man mit Hilfe der verbleibenden 3301 Iterationen einen approximierten bedingten p-Wert von 0,6422296:

```
> set.seed(1234)
> unab(y,350000,20000,100)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
[1,]    0    2    2    1    0    0    1    1    0    3    0    2    1
[2,]    0    0    0    2    0    3    0    2    0    0    0    1    1
[3,]    0    0    0    2    0    2    0    0    1    1    0    0    2
[4,]    0    0    0    1    1    1    2    0    2    0    0    0    1
[5,]    0    1    0    1    1    1    1    0    0    0    0    1    0
[6,]    0    0    0    0    2    1    0    0    1    0    0    1    0
...
...
```

approximierter p-Wert= 0.6422296

Die Unabhängigkeitshypothese wird also mit Hilfe beider Verfahren auf dem Niveau 0,05 angenommen, was bedeutet, dass es keinen Zusammenhang zwischen dem Jahr und dem Wohnort der durch Pferdetritte verstorbenen Bauern gab. Obwohl beide Verfahren zur Annahme der Nullhypothese führen, ist die Differenz von bedingtem p-Wert und p-Wert

des  $\chi^2$ -Tests (0,0349327) doch relativ groß.

Die Frage ist nun, wie groß die Differenz ausfällt, wenn man die Nullhypothese ablehnen müsste.

### 4.3.1 Beispiel „Tod durch Pferdetritte“ mit Ablehnung der Nullhypothese

Tabelle 4.4 soll nun so abgewandelt werden, dass man die Nullhypothese der Unabhängigkeit ablehnen muss. Ein Beispiel hierfür ist Tabelle 4.5 auf Seite 91. Wie schon bei Tabelle 4.4 sind auch hier alle erwarteten Zelhäufigkeiten kleiner als fünf. Wendet man trotzdem den  $\chi^2$ -Unabhängigkeitstest an, erhält man einen p-Wert von 0,01771:

```
> chisq.test(y2)
```

```
Pearson's Chi-squared test
```

```
data: y2
```

```
X-squared = 296.0065, df = 247, p-value = 0.01771
```

```
Warning message:
```

```
In chisq.test(y2) : Chi-squared approximation may be incorrect
```

Um auch hier die nicht zufriedenstellende Approximation der  $\chi^2$ -Verteilung zu umgehen, soll der modifizierte Metropolis Algorithmus angewandt werden. Abbildung 4.4 zeigt, dass auch hier bei 350000 Iterationen ein stabiler Wert für  $\hat{\mu}$  erreicht wird. Geht man entsprechend wie oben vor, ergibt sich mit der Markov Ketten Monte Carlo Methode ein approximierter bedingter p-Wert von 0,01242048:

```
> set.seed(1234)
```

```
> unab(y2,350000,20000,100)
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
[1,]    0    1    1    0    0    0    1    1    0    3    0    2    0
[2,]    0    0    0    2    0    3    0    2    0    0    0    1    1
[3,]    0    0    0    2    0    2    0    0    1    1    0    0    2
[4,]    0    0    0    1    1    1    2    0    2    0    0    0    1
[5,]    0    1    0    1    0    2    1    0    0    0    0    1    0
[6,]    0    0    0    0    2    0    0    0    1    0    3    1    0
...
...
```

```
approximierter p-Wert= 0.01242048
```

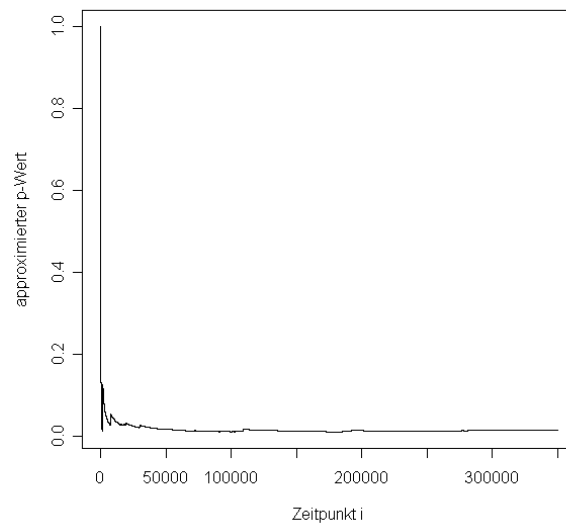


Abbildung 4.4: Verlauf von  $\hat{\mu}$ , ( $n = 350000$ , Tabelle 4.5)

In diesem Beispiel müsste man mit Hilfe beider Methoden die Unabhängigkeitshypothese auf dem Niveau 0,05 ablehnen. Die Differenz der beiden Werte (0,00528952) ist hier im Gegensatz zu obigem Beispiel relativ klein.

Aufgrund dieses Unterschiedes stellt sich die Frage, wie groß die Differenz zwischen p-Wert und bedingtem p-Wert des  $\chi^2$ -Tests ausfällt, wenn diese im Bereich des Signifikanzniveaus 0,05 liegen.

### 4.3.2 Beispiel „Tod durch Pferdetritte“ mit p-Wert im Bereich des Signifikanzniveaus

Zu obigem Zweck soll wieder Tabelle 4.4 abgewandelt werden. Dieses Mal mit dem Ziel, einen (bedingten) p-Wert im Bereich des Signifikanzniveaus 0,05 zu erreichen. Tabelle 4.6 auf Seite 92 ist ein Beispiel dafür.

Der  $\chi^2$ -Unabhängigkeitstests ergibt einen p-Wert von 0,04719:

```
> chisq.test(y3)
```

```
Pearson's Chi-squared test
```

```
data: y3
```

```
X-squared = 285.3412, df = 247, p-value = 0.04719
```

```
Warning message:
```

```
In chisq.test(y3) : Chi-squared approximation may be incorrect
```

Aufgrund der kleinen Einträge wird wie bei den obigen Beispielen auch hier durch die  $\chi^2$ -Teststatistik nur eine unbefriedigende Approximation der  $\chi^2$ -Verteilung erreicht. Es bietet sich also wieder an, den modifizierten Metropolis Algorithmus zu verwenden:

Der Verlauf von  $\hat{\mu}$  für Tabelle 4.6 wird in Abbildung 4.5 gezeigt. Offensichtlich kommt man bei diesem Beispiel ebenfalls zu den oben bereits beschriebenen Erkenntnissen, so dass sich der bedingten p-Wert auf folgende Weise approximativ bestimmen lässt:

```
> set.seed(1234)
```

```
> unab(y3,350000,20000,100)
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
[1,]    0    3    0    0    2    0    1    1    0    3    0    2    1
[2,]    0    0    0    2    0    3    0    2    0    0    0    1    1
[3,]    0    0    0    2    0    2    0    0    1    1    0    0    2
[4,]    0    0    0    3    0    0    1    0    2    0    0    0    1
[5,]    0    1    0    1    0    2    1    0    0    0    0    1    0
[6,]    0    0    0    0    2    1    0    2    0    0    0    1    0
...
...
```

```
approximierter p-Wert= 0.03423205
```

Beide Verfahren führen also zu der Ablehnung der Nullhypothese auf dem Niveau 0,05. Während jedoch der p-Wert des  $\chi^2$ -Unabhängigkeitstests sehr nahe bei 0,05 liegt, ist der Abstand des bedingten p-Werts zu 0,05 schon deutlicher. Die Differenz zwischen p-Wert und bedingtem p-Wert beträgt hier 0,01295795. D.h. auch im Bereich des Signifikanzniveaus 0,05 weist sie einen relativ hohen Wert auf, der hier möglicherweise in einigen Fällen sogar zwischen Ablehnung und Annahme der Unabhängigkeitshypothese entscheiden kann.

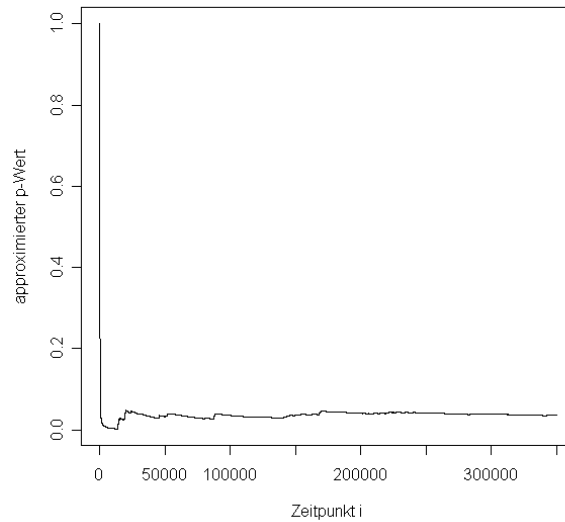


Abbildung 4.5: Verlauf von  $\hat{\mu}$ , ( $n = 350000$ , Tabelle 4.6)

## 4.4 Ähnlichkeit zweier Gedichte

In diesem Beispiel (vgl. hierzu [5, S.340 ff]) soll die Ähnlichkeit zweier Gedichte, beide namens „Winter Retrospect“, von J.A. Lindon und H.W. Bergerson untersucht werden. Hierfür schrieb zunächst Bergerson sein Gedicht und schickte eine alphabetisch geordnete Liste der Wörter, die er für dieses Gedicht verwendete, zusammen mit einer Aufstellung der benutzten Strophengröße und Zeilenlänge an Lindon. Dieser schrieb anschließend ohne Kenntnis des Originals ein Gedicht mit eben diesen Wörtern, wobei er die gleiche Strophengröße und Zeilenlänge benutzte. Beide Gedichte bestehen aus 24 Zeilen mit insgesamt 478 Wörtern, unter denen 162 Wörter jeweils nur einmal auftauchen.

Es soll nun untersucht werden, ob sich die beiden Gedichte im Hinblick auf die Positionierung dieser 162 Wörter in den entsprechenden Zeilen ähneln. In Tabelle 4.7 auf Seite 93 wird die Platzierung der entsprechenden Wörter in dem jeweiligen Gedicht angegeben. Zu testen ist, ob die Platzierung der 162 Wörter in den beiden Gedichten unabhängig voneinander sind. Da hier alle Zellhäufigkeiten kleiner als 5 sind, ist der  $\chi^2$ -Unabhängigkeitstest für diesen Fall ungeeignet, was die in R angezeigte Warnung bestätigt:

```
> chisq.test(x1)
```

```
Pearson's Chi-squared test
```

```
data: x1
```

```
X-squared = 589.1646, df = 529, p-value = 0.0356
```

```
Warning message:
```

```
In chisq.test(x1) : Chi-squared approximation may be incorrect
```

Stattdessen soll wiederum der modifizierte Metropolis Algorithmus angewandt werden. Abbildung 4.6 zeigt, dass bei ca. 700000 Iterationen ein relativ stabiler Wert für  $\hat{\mu}$  erreicht ist und man die ersten 100000 Berechnungen vernachlässigen kann. Auf diese Weise erhält man einen approximierten bedingten p-Wert von:

```
> set.seed(12345)
```

```
> unab(x1,700000,100000,100)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
[1,]	2	0	1	0	0	1	0	0	0	1	1	1	0
[2,]	1	1	1	2	0	0	0	0	0	0	0	1	0
[3,]	0	0	1	0	0	1	0	0	1	0	0	0	0
[4,]	1	1	1	2	1	0	0	0	0	0	0	0	0
[5,]	1	0	0	1	0	0	0	0	0	0	0	1	0
[6,]	0	0	1	0	0	0	1	0	0	0	0	0	0
...													
...													

```
approximierter p-Wert= 0.03582736
```

Die Hypothese der Unabhängigkeit der Wortplatzierungen in den beiden Gedichten muss also auf dem Niveau 0,05 verworfen werden, was bedeutet, dass die Platzierungen der Wörter in beiden Gedichten nicht unabhängig voneinander sind und beide somit eine Ähnlichkeit zueinander aufweisen.

Die Differenz zwischen bedingten p-Wert und p-Wert des  $\chi^2$ -Tests ist bei diesem Beispiel sehr gering (0,00022736). Was aber passiert, wenn die Nullhypothese angenommen werden würde?

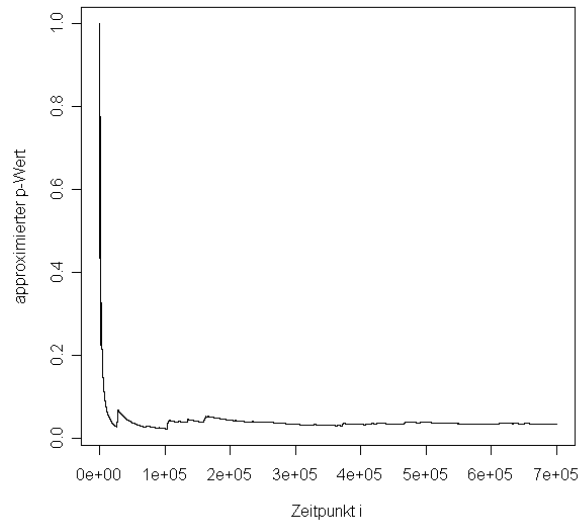


Abbildung 4.6: Verlauf von  $\hat{\mu}$ , ( $n = 700000$ , Tabelle 4.7)

#### 4.4.1 Beispiel „Ähnlichkeit zweier Gedichte“ mit Annahme der Nullhypothese

Nach dem letzten Beispiel besteht auch hier die Frage, was bezüglich des Unterschieds zwischen dem p-Wert und dem bedingten p-Wert des  $\chi^2$ -Unabhängigkeitstests passiert, wenn Tabelle 4.7 so abgewandelt wird, dass man die Unabhängigkeitshypothese annehmen müsste. Hierfür sei Tabelle 4.8 auf Seite 94 gegeben.

Da auch hier alle erwarteten Zellhäufigkeiten  $< 5$  sind, wird mit Hilfe des  $\chi^2$ -Tests nur eine unbefriedigende Approximation der  $\chi^2$ -Verteilung erreicht. Mit diesem Test ergäbe sich ein p-Wert von 0,8092:

```
> chisq.test(x2)
```

```
Pearson's Chi-squared test
```

```
data: x2
```

```
X-squared = 500.4074, df = 529, p-value = 0.8092
```

```
Warning message:
```

```
In chisq.test(x2) : Chi-squared approximation may be incorrect
```

Der Verlauf von  $\hat{\mu}$  für Tabelle 4.8 (Abbildung 4.7) zeigt, dass auch hier 600000-700000 Iterationen ausreichen um einen relativ stabilen Wert für  $\hat{\mu}$  zu erreichen. Bei 700000



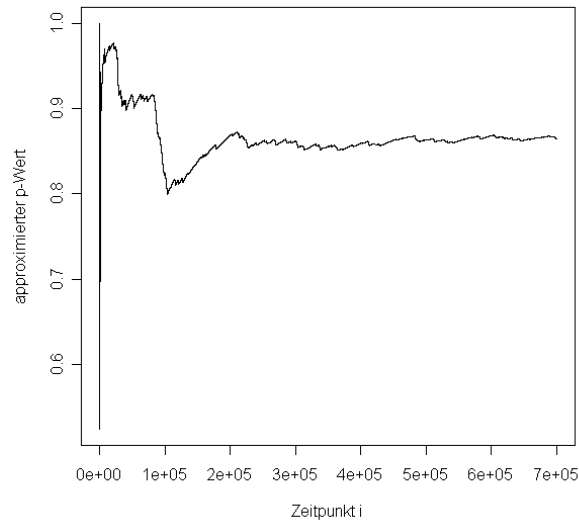


Abbildung 4.7: Verlauf von  $\hat{\mu}$ , ( $n = 700000$ , Tabelle 4.8)

Iterationen mit einer Einschwingphase von 100000 und unter Einbeziehung jedes 100. Schrittes ergibt sich hier ein approximierter bedingter p-Wert von 0,8693551:

```
> set.seed(12345)
> unab(x2,700000,100000,100)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
[1,]	2	0	1	0	0	1	0	0	0	1	1	1	0
[2,]	1	1	1	2	0	0	0	0	0	0	0	1	0
[3,]	0	0	1	0	0	1	0	0	1	0	0	0	0
[4,]	1	1	1	2	1	0	0	0	0	0	0	0	0
[5,]	1	0	0	1	0	0	0	0	0	0	0	1	0
[6,]	0	0	1	0	0	0	1	0	0	0	0	0	0
...													
...													

```
approximierter p-Wert= 0.8693551
```

Als Differenz zwischen den beiden Werten erhält man hier 0,0601551, was in besonderen Fällen einen Unterschied zwischen Annahme und Ablehnung der Nullhypothese bedeuten könnte. Es stellt sich nun also wieder die Frage, was im Bereich des Signifikanzniveaus  $\alpha = 0,05$  geschieht.

## 4.4.2 Beispiel „Ähnlichkeit zweier Gedichte“ mit p-Wert im Bereich des Signifikanzniveaus

Um einen (bedingten) p-Wert im Bereich des Signifikanzniveaus  $\alpha = 0,05$  zu erreichen, wurde wieder Tabelle 4.7 abgewandelt. (Siehe Tabelle 4.9 auf Seite 95.) Wie schon oben wird auch hier durch die  $\chi^2$ -Teststatistik keine befriedigende Approximation der  $\chi^2$ -Verteilung erreicht. Mit dem  $\chi^2$ -Unabhängigkeitstest erhält man bei Tabelle 4.9 einen p-Wert von 0,05107:

```
> chisq.test(x3)
```

```
Pearson's Chi-squared test
```

```
data: x3
```

```
X-squared = 583.2569, df = 529, p-value = 0.05107
```

```
Warning message:
```

```
In chisq.test(x3) : Chi-squared approximation may be incorrect
```

Man würde also mit diesem Test die Unabhängigkeitshypothese annehmen.

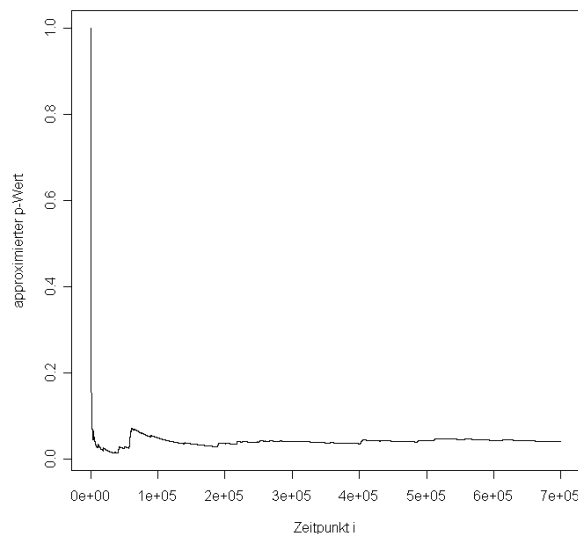


Abbildung 4.8: Verlauf von  $\hat{\mu}$ , ( $n = 700000$ , Tabelle 4.9)

Es soll nun getestet werden, ob dies auch mit Hilfe des modifizierten Metropolis Algorithmus der Fall ist. Abbildung 4.8 zeigt den Verlauf von  $\hat{\mu}$  für Tabelle 4.9. Wie oben ist auch hier zu erkennen, dass es ausreicht 600000 - 700000 Iterationen durchzuführen um einen stabilen Wert für  $\hat{\mu}$  zu erhalten. Für den bedingten p-Wert des  $\chi^2$ -Tests erhält man analog zu den obigen beiden Beispielen dieses Abschnitts:

```
> set.seed(12345)
> unab(x3,700000,100000,100)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
[1,]	0	0	1	0	0	1	0	0	0	0	0	0	0
[2,]	1	0	0	2	0	0	0	0	0	0	0	1	0
[3,]	0	0	1	0	0	1	0	0	1	0	0	0	0
[4,]	1	1	0	3	1	0	0	0	0	0	0	0	0
[5,]	1	0	1	1	0	0	0	0	0	0	0	0	0
[6,]	0	0	1	0	0	0	1	0	0	1	0	0	0
...													
...													

```
approximierter p-Wert= 0.03799367
```

Die Nullhypothese der Unabhängigkeit muss auf diese Weise also, im Gegensatz zu dem Ergebnis des  $\chi^2$ -Unabhängigkeitstests, abgelehnt werden. Die Differenz zwischen p-Wert und bedingten p-Wert beträgt hier 0,01307633 und ist somit ebenfalls relativ groß. In diesem Fall hat dieser große Unterschied sogar zwischen Annahme und Ablehnung der Nullhypothese entschieden.

## 4.5 Zusammenfassung der Ergebnisse

Tabelle 4.3 zeigt noch einmal zusammenfassend alle Ergebnisse der obigen Beispiele. Hier ist zu erkennen, dass bei den Beispielen (4.1 und 4.2), bei denen keine oder nur eine der erwarteten Zellhäufigkeiten  $< 5$  sind, es nur einen sehr geringen Unterschied zwischen dem p-Wert und dem approximativ berechneten bedingten p-Wert des  $\chi^2$ -Tests gibt. Egal, ob man schließlich die Nullhypothese auf dem Niveau 0,05 annehmen oder ablehnen muss, beide Werte gleichen sich jeweils mindestens bis zur zweiten Nachkommastelle. Der Grund dafür ist die zufriedenstellende Approximation der  $\chi^2$ -Verteilung, die beim  $\chi^2$ -Unabhängigkeitstest erreicht wird.

Sind jedoch alle erwarteten Zellhäufigkeiten  $< 5$ , kommt man zu unterschiedlichen Ergebnissen:

Bei den Beispielen 4.3.1 und 4.4, bei denen die Nullhypothese auf dem Niveau 0,05 abgelehnt werden musste, stimmen die beiden berechneten p-Werte ebenfalls mindestens bis zur zweiten Nachkommastelle überein.

Bei Annahme der Nullhypothese (Beispiele 4.3 und 4.4.1) ergibt sich jedoch ein großer

### Ergebnisse der obigen Beispiele

Beispiel	p-Wert des $\chi^2$ -Tests	bedingter p-Wert	Betrag der Differenz
4.1	0,01375	0,01343973	0,00031027
4.2	0,7996	0,7983419	0,0012581
4.3	0,6173	0,6422296	0,0249296
4.3.1	0,01771	0,01242048	0,00528952
4.3.2	0,04719	0,03423205	0,01295795
4.4	0,0356	0,03582736	0,00022736
4.4.1	0,8092	0,8693551	0,0601551
4.4.2	0,05107	0,03799367	0,01307633

Tabelle 4.3: Zusammenfassung der obigen Ergebnisse

Unterschied zwischen p-Wert und bedingten p-Wert. Als Differenzen erhält man hier 0,0249296 und sogar 0,0601551.

Es stellt sich somit die Frage, ob ein solcher großer Unterschied auch auftritt, wenn der (bedingte) p-Wert im Bereich des Signifikanzniveaus von 0,05 liegt, da er in diesem Bereich in manchen Fällen möglicherweise einen Unterschied in Annahme und Ablehnung der Nullhypothese bewirken kann. Nach entsprechenden Umwandlungen der Ausgangstabellen (Beispiele 4.3.2 und 4.4.2) ergab sich auch in diesem Bereich noch ein relativ großer Unterschied zwischen dem p-Wert und dem approximativ berechneten bedingten p-Wert. Die Differenzen liegen hier betragsmäßig bei rund 0,01, also in einer Größenordnung, die in bestimmten Fällen tatsächlich zu einem Unterschied in Annahme und Ablehnung der Nullhypothese führen können, wie es bei Beispiel 4.4.2 tatsächlich der Fall ist.

Es ist zu beachten, dass hier nur einige wenige Beispiele untersucht wurden, weswegen sich aufgrund der Beobachtungen keine allgemeine Aussage über den Unterschied von  $\chi^2$ -Unabhängigkeitstest und modifiziertem Metropolis Algorithmus machen lässt. Es zeigt sich jedoch allein bei diesen wenigen Beispielen, dass es bei Kontingenztafeln mit kleinen Einträgen (bzw. erwarteten Zellhäufigkeiten  $< 5$ ) doch auffallende Unterschiede zwischen den beiden Verfahren gibt.

Regiment	Jahr (1800)														Summe						
	'75	'76	'77	'78	'79	'80	'81	'82	'83	'84	'85	'86	'87	'88		'89	'90	'91	'92	'93	'94
I	0	2	2	1	0	0	1	1	0	3	0	2	1	0	0	1	0	1	0	1	16
II	0	0	0	2	0	3	0	2	0	0	0	1	1	1	0	2	0	3	1	0	16
III	0	0	0	2	0	2	0	0	1	1	0	0	2	1	1	0	0	2	0	0	12
IV	0	0	0	1	1	1	2	0	2	0	0	0	1	0	1	2	1	0	0	0	12
V	0	1	0	1	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	0	8
VI	0	0	0	0	2	1	0	0	1	0	0	1	0	1	1	1	1	1	1	0	11
VII	0	0	1	0	2	0	0	1	2	0	1	1	3	1	1	1	0	3	0	0	17
VIII	1	0	1	0	0	0	1	0	1	1	0	0	2	0	0	2	1	0	2	0	12
IX	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	1	0	1	7
X	0	0	0	0	0	2	1	1	1	0	2	1	1	0	1	2	0	1	0	0	13
XI	0	0	1	1	0	1	0	2	0	2	0	0	0	0	2	1	3	0	1	1	15
XII	0	0	0	0	2	4	0	1	3	0	1	1	1	1	2	1	3	1	3	1	25
XIII	1	1	2	1	1	3	0	4	0	1	0	3	2	1	0	2	1	1	0	0	24
XIV	0	1	0	0	0	0	0	1	0	1	1	0	0	0	2	2	0	0	0	0	8
Summe	3	5	7	9	10	18	6	14	11	9	5	11	15	6	11	17	12	15	8	4	196

Tabelle 4.4: Anzahl der Toten durch Pferdetritte (vgl. [5, S.326])

Regiment	Jahr (1800)														Summe						
	'75	'76	'77	'78	'79	'80	'81	'82	'83	'84	'85	'86	'87	'88		'89	'90	'91	'92	'93	'94
I	0	1	1	0	0	0	1	1	0	3	0	2	0	0	0	0	0	1	0	1	11
II	0	0	0	2	0	3	0	2	0	0	0	1	1	1	0	2	0	3	1	0	16
III	0	0	0	2	0	2	0	0	1	1	0	0	2	1	1	0	0	2	0	0	12
IV	0	0	0	1	1	1	2	0	2	0	0	0	1	0	1	2	1	0	0	0	12
V	0	1	0	1	0	2	1	0	0	0	0	1	0	0	0	0	1	1	0	0	8
VI	0	0	0	0	2	0	0	0	1	0	3	1	0	1	0	1	4	1	1	0	15
VII	0	0	0	0	2	0	0	2	3	0	0	0	3	1	1	1	0	3	0	0	16
VIII	3	0	0	0	0	0	1	0	1	0	0	0	2	0	0	2	1	0	2	0	12
IX	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	1	0	1	7
X	0	0	0	0	0	2	1	1	1	0	2	1	1	0	1	2	0	1	0	0	13
XI	0	0	1	1	0	1	0	2	0	2	0	0	0	0	2	1	3	0	1	1	15
XII	0	0	0	0	2	4	0	1	3	0	1	4	0	0	3	1	3	1	3	1	27
XIII	1	1	2	1	1	3	0	4	0	1	0	3	2	1	0	2	1	1	0	0	24
XIV	0	1	0	0	0	0	0	1	0	1	1	0	0	0	2	2	0	0	0	0	8
Summe	5	4	4	8	9	18	6	15	12	8	7	13	13	5	11	16	15	15	8	4	196

Tabelle 4.5: Abwandlung von Tabelle 4.4, durch die eine Ablehnung der Nullhypothese erreicht wird

Regiment	Jahr (1800)														Summe						
	'75	'76	'77	'78	'79	'80	'81	'82	'83	'84	'85	'86	'87	'88		'89	'90	'91	'92	'93	'94
I	0	3	0	0	2	0	1	1	0	3	0	2	1	0	0	1	0	1	0	1	16
II	0	0	0	2	0	3	0	2	0	0	0	1	1	1	0	2	0	3	1	0	16
III	0	0	0	2	0	2	0	0	1	1	0	0	2	2	0	0	0	2	0	0	12
IV	0	0	0	3	0	0	1	0	2	0	0	0	1	0	1	2	1	0	0	0	11
V	0	1	0	1	0	2	1	0	0	0	0	1	0	0	0	1	1	1	0	0	8
VI	0	0	0	0	2	1	0	2	0	0	0	1	0	3	0	0	1	1	1	0	12
VII	0	0	1	0	2	0	0	1	2	0	1	1	3	1	1	1	0	3	0	0	17
VIII	1	0	1	0	0	0	0	2	1	1	0	0	2	0	0	2	1	0	2	0	13
IX	1	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	1	1	0	1	8
X	0	0	0	0	1	1	1	1	1	0	2	1	1	0	1	2	0	1	0	0	13
XI	0	0	2	0	0	1	0	2	0	2	0	0	0	0	2	1	3	0	1	1	15
XII	0	0	0	0	2	2	1	2	1	0	2	2	0	0	2	1	3	1	3	1	23
XIII	2	0	2	2	0	3	0	4	0	1	0	3	2	1	0	2	1	1	0	0	24
XIV	0	1	0	0	0	0	0	0	0	2	1	0	0	0	2	2	0	0	0	0	8
Summe	4	5	6	10	10	15	5	18	8	11	6	12	14	8	9	16	12	15	8	4	196

Tabelle 4.6: Abwandlung von Tabelle 4.4, durch die ein (bedingter) p-Wert im Bereich von 0,05 erreicht wird

		Reihe bei Lindon																								Summe	
Reihe bei Bergerson		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
1	2	0	1	1	0	0	1	0	0	0	1	1	1	0	0	0	0	0	2	1	0	0	0	0	0	10	
2	1	1	1	2	0	0	0	0	0	0	0	0	1	0	0	2	1	0	0	0	0	0	1	1	0	11	
3	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	4	
4	1	1	1	2	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	7	
5	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0	0	1	0	1	0	8	
6	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	1	0	6	
7	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	4	1	0	0	9	
8	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	3	1	0	0	1	8	
9	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	1	0	0	0	0	1	0	1	0	7	
10	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	6	
11	0	0	1	0	0	0	4	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	8	
12	0	0	0	0	0	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	5	
13	0	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	6	
14	0	0	0	0	0	0	0	1	2	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	7	
15	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	4	
16	0	0	0	0	0	0	0	1	0	2	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	5	
17	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	6	
18	1	0	0	0	0	2	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	7	
19	0	2	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	7	
20	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	5	
21	0	1	0	0	1	0	0	0	0	0	1	0	0	0	2	0	0	0	1	0	0	0	1	0	1	8	
22	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	6	
23	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	5	
24	0	0	1	0	0	0	0	0	1	1	0	1	1	1	0	1	0	0	0	0	1	0	0	0	0	7	
Summe		6	6	8	10	6	9	7	6	6	8	10	6	4	7	12	7	3	3	3	3	7	10	8	6	4	162

Tabelle 4.7: Platzierung der 162 einzeln auftauchenden Wörter in den Gedichten „Winter Retrospect“ von J.A. Lindon und H.W. Bergerson (vgl.[5, S.341])



		Reihe bei Lindon																								Summe
Reihe bei Bergerson		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Summe
1	2	0	1	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	2	1	0	0	0	0	0	10
2	1	1	1	2	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	1	0	10
3	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	4
4	1	1	1	2	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	7
5	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	8
6	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	5
7	0	1	0	0	0	0	0	0	0	0	1	1	0	2	1	0	0	0	0	0	0	2	1	0	0	9
8	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	1	6
9	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	0	0	1	0	1	0	7
10	1	0	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	6
11	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	4
12	1	0	0	0	0	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	6
13	0	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	6
14	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	8
15	0	1	0	1	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	6
16	0	0	0	0	0	0	0	0	1	0	2	0	0	0	1	0	0	0	0	1	0	0	1	0	0	6
17	1	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	7
18	1	0	0	1	0	1	0	0	1	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	8
19	0	2	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	7
20	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	5
21	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	2	0	0	1	0	0	0	1	0	1	8
22	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	1	0	7
23	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	5
24	0	0	1	0	0	0	0	0	1	1	0	1	1	1	0	1	0	0	0	0	1	0	0	0	0	7
Summe	10	8	8	11	6	9	3	5	8	8	10	6	4	7	11	8	8	3	3	4	4	8	8	6	4	162

Tabelle 4.8: Abwandlung von Tabelle 4.7, durch die eine Annahme der Nullhypothese erreicht wird

		Reihe bei Lindon																								Summe	
Reihe bei Bergerson		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	1	0	0	1	7	
2	1	0	0	2	0	0	0	0	0	0	0	1	0	0	2	1	0	0	0	0	0	0	1	1	0	9	
3	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	4	
4	1	1	0	3	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	7	
5	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	7	
6	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	6	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2	3	1	0	0	8	
8	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	2	3	0	0	1	8	
9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	4	
10	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	4	
11	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	3	
12	0	2	0	0	1	2	0	0	0	0	0	0	1	1	0	0	0	1	0	0	4	1	0	0	0	13	
13	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	6	
14	0	0	0	0	0	0	1	1	2	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	8	
15	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	4	
16	0	0	3	0	0	0	0	1	0	2	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	8	
17	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	6	
18	1	0	0	0	0	2	0	0	0	0	0	0	0	2	0	2	0	0	0	2	0	0	0	0	0	9	
19	0	2	0	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	7	
20	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	3	0	8	
21	0	1	0	0	1	0	0	0	0	0	1	0	0	0	2	0	0	0	1	0	0	0	1	0	1	8	
22	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	6	
23	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	5	
24	0	0	1	0	0	0	0	0	0	1	1	0	1	1	0	1	0	0	0	0	1	0	0	0	0	7	
Summe	4	6	9	10	7	8	4	4	7	6	7	4	5	5	5	13	7	7	3	3	4	12	12	8	9	5	162

Tabelle 4.9: Abwandlung von Tabelle 4.7, durch die ein (bedingter) p-Wert im Bereich von 0,05 erreicht wird

# Anhang A

## Programme

Im Folgenden werden die Quellcodes für die in dieser Arbeit aufgerufenen Funktionen aufgeführt und erläutert. Für das Programmieren und Aufrufen dieser Funktionen wurde die freie Software R Version 2.6.2 vom 08.02.2008 benutzt.

### A.1 Computer Simulation von Markov Ketten

Mit dem folgenden Programm wird eine Markov Kette simuliert, sowie der Verlauf von

$$Y(n) := \frac{1}{n+1} \sum_{i=0}^n 1_{\{s\}}(X_i) \quad , s \in S \text{ beliebig aber fest,}$$

ausgegeben. Theoretisch wurde die Simulation bereits in Abschnitt 1.2 beschrieben und in 1.2.1 angewandt.

Man ruft die Funktion durch den Befehl `mk(n, s, mu, p, d)` auf, wobei als Eingabe folgende Parameter notwendig sind:

- (i) `n` : gewünschte Anzahl der zu simulierenden Kettenglieder,
- (ii) `s` : Zustandsraum  $S$ ,
- (iii) `mu` : Anfangsverteilung  $\mu^{(0)}$ ,
- (iv) `p` : Übergangsmatrix  $P$  und
- (v) `d` : gewünschter Zustand  $s_j$ , für den die Kurve von  $Y$  ausgegeben werden soll.

Dem Programm werden zunächst die Einführungsfunktion `psi` gemäß (1.5), sowie die Update-Funktion `phi` gemäß (1.7) übergeben. Anschließend werden, wie in (1.10) beschrieben, die einzelnen  $X_i$  für  $i = 1, \dots, n$  erzeugt, sowie die entsprechenden Werte für  $Y$  berechnet und ausgegeben.

```

> mk<-function(n,s,mu,p,d){
+ #Einfuehrungsfunktion:
+ psi<-function(x,s,mu){
+     y<-cumsum(mu)
+     i<-2
+     if(x>=0 & x<mu[1]){i<-1}
+     else{
+         repeat{
+             if(x>=y[i-1] & x< y[i]){i&break}
+             else{i<-i+1}
+         }
+     }
+     return(i)
+ }
+ #Update-Funktion:
+ phi<-function(x,t,p){
+     z<-cumsum(p[t, ])
+     j<-2
+     if(x>=0 & x<p[t,1]){j<-1}
+     else{
+         repeat{
+             if(x>=z[j-1] & x< z[j]){j&break}
+             else{j<-j+1}
+         }
+     }
+     return(j)
+ }
+ #Erzeugen der X[i]
+ u<-runif(n+1,min=0,max=1)
+ X<-rep(3,n+1)
+ Y<-rep(0,n+1)
+ j<-rep(0,n+1)
+ j[1]<-psi(u[1],s,mu)
+ X[1]<-s[j[1]]
+ z<-0
+ if(X[1]==d){Y[1]<-1}
+ if(X[1]==d){z<-1}
+ i<-1
+ repeat{
+     i<-i+1
+     j[i]<-phi(u[i],j[i-1],p)
+     X[i]<-s[j[i]]
+     if(X[i]==d){z<-z+1} #Summe 1_[Xi=z]
+     Y[i]<-(1/i)*z #Berechnung von Y
+     if(i==n+1)break

```

```

+   }
+   print(X)
+   plot(Y,type="l",xlab="Zeitpunkt n")
+ }

```

## A.2 Modifizierter Metropolis Algorithmus

Ausgehend von einer  $K \times L$ -Tabelle  $f^{(0)}$  mit nichtnegativen Einträgen, den Zeilensummen  $r = (r_1, \dots, r_I)$  und den Spaltensummen  $c = (c_1, \dots, c_J)$ , dient die folgende Funktion dazu,  $K \times L$ -Tabellen  $f^{(i)} \in \mathcal{F}_{(r,c)}$  mit den gleichen Zeilensummen  $r$  und den gleichen Spaltensummen  $c$  zu simulieren.<sup>1</sup> Die Grundlage bildet der modifizierte Metropolis Algorithmus, der in Abschnitt 3.7 beschrieben wurde.

Zusätzlich wird der Wert

$$\hat{\mu} = \frac{1}{n+1} \sum_{i=0}^n 1_{\{T_{\chi^2}(f^{(i)}) \geq T_{\chi^2}(f^{(0)})\}}(f^{(i)})$$

berechnet, der sich für  $n \rightarrow \infty$  dem bedingten p-Wert des  $\chi^2$ -Tests annähert. (vgl. Abschnitt 3.5)

Aufgerufen wird diese Funktion mit dem Befehl `unab(x0,n,d,schritt)`, wobei folgende Parameter gegeben sein müssen:

- (i) `x0` :  
die Startmatrix  $f^{(0)}$  mit nichtnegativen Einträgen, den Zeilensummen  $r = (r_1, \dots, r_I)$  mit  $r_i \neq 0$  für  $i = 1, \dots, I$  und den Spaltensummen  $c = (c_1, \dots, c_J)$  mit  $c_j \neq 0$  für  $j = 1, \dots, J$ ,
- (ii) `n` : mit  $n > 0$ ,  
die gewünschte Anzahl der zu simulierenden Tabellen,
- (iii) `d` : mit  $0 \leq d < n$ ,  
die Länge der Einschwingphase (burn-in), d.h. die gewünschte Anzahl der anfangs simulierten Tabellen, die nicht in die Rechnung von  $\hat{\mu}$  einbezogen werden sollen,
- (iv) `schritt` : mit  $1 \leq \text{schritt} \leq n$ ,  
die Schrittlänge nach der Einschwingphase, d.h. nach der Einschwingphase wird zunächst  $f^{(d+1)}$  und anschließend jede `schritt`-te Tabelle (also  $f^{(d+1+\text{schritt})}$ ,  $f^{(d+1+2 \cdot \text{schritt})}$ , ...) in die Rechnung von  $\hat{\mu}$  einbezogen.

Als Ausgabe erhält man hier die vorgegebene Tabelle  $f^{(0)}$ , den zuletzt berechneten Wert für  $\hat{\mu}$  (`approximierter p-Wert`), sowie den Verlauf von  $\hat{\mu}$  in einer Grafik.

---

<sup>1</sup>Diese Funktion steht lediglich für Matrizen zur Verfügung, bei der alle Randsummen  $\neq 0$  sind.

```

> unab<-function(x0,n,d,schritt){
+ x<-x0
+ print(x)
+ l<-0
+ m<-0
+ f<-0
+ s<-seq(1,(n+1)*schritt,by=schritt)
+ mu<-s+1
+ k<-1
+
+ #p-Wert:
+ #Funktion fuer die Teststatistik:
+ Tchi<-function(x0){
+
+ #ML-Schaetzer:
+ ML<-function(x0,i,j){rowSums(x0)[i]*colSums(x0)[j]/sum(colSums(x0))}
+
+ #T(Chi^2)-Wert, i und j gegeben:
+ Chi<-function(x0,i,j){(x0[i,j]-ML(x0,i,j))^2/ML(x0,i,j)}
+
+ a<-rep(0,nrow(x0))
+
+ #Summen der T(Chi^2)-Werte:
+ for(i in 1:nrow(x0)){a[i]<-sum(Chi(x0,i,1:ncol(x0)))}
+ chiwert<-sum(a)
+ return(chiwert)
+ }
+
+ TCHI<-Tchi(x0)
+
+ repeat{
+
+ l<-l+1
+
+ if(l==d+s[k]){
+
+ #mu erst ab (einschl.) l=d+1 berechnen,
+ #in "schritt"-Schritten
+
+ f<-f+1
+
+ if(Tchi(x) >= TCHI){m<-m+1}
+
+ #Summe 1_{Tchi(x) >= Tchi(x0)}
+
+ mu[f]<- m/f
+
+ k<-k+1
+
+ }
+
+
+
+ if(l==n+1)break
+

```

```

+
+
+ j<-c(sample(1:ncol(x),2))      #zufaellig zwei Spalten aussuchen
+ if(j[1]>j[2]){j<-c(j[2],j[1])} #richtige Reihenfolge der
+                               #ausgewaehlten Spalten
+
+ i<-c(sample(1:nrow(x),2))      #zufaellig zwei Zeilen aussuchen
+ if(i[1]>i[2]){i<-c(i[2],i[1])} #richtige Reihenfolge der
+                               #ausgewaehlten Zeilen
+
+ e<-runif(1,min=0,max=1)
+ if(e<=0.5){epsilon<-1}else{epsilon<-(-1)} #epsilon berechnen
+
+ #Matrix M addieren
+ y<-x
+ y[i[1],j[1]]<- x[i[1],j[1]]+epsilon
+ y[i[2],j[2]]<- x[i[2],j[2]]+epsilon
+ y[i[1],j[2]]<- x[i[1],j[2]]-epsilon
+ y[i[2],j[1]]<- x[i[2],j[1]]-epsilon
+
+ if(y[i[1],j[1]]<0){x & next}   #y verwerfen, wenn ein Eintrag <0
+ if(y[i[1],j[2]]<0){x & next}
+ if(y[i[2],j[1]]<0){x & next}
+ if(y[i[2],j[2]]<0){x & next}
+
+ #alpha berechnen
+ A<-(factorial(x[i[1],j[1]])*factorial(x[i[2],j[2]])*factorial(x[i[1],j[2]])
+ *factorial(x[i[2],j[1]]))/(factorial(y[i[1],j[1]])*factorial(y[i[2],j[2]])
+ *factorial(y[i[1],j[2]])*factorial(y[i[2],j[1]]))
+ alpha<-min(A,1)
+ u<-runif(1,min=0,max=1)
+ if(u<=alpha){x<-y}           #y mit Wahrscheinlichkeit alpha annehmen
+
+ x
+ }
+
+ cat("approximierter p-Wert=",mu[f],"\n")
+ plot(mu,type="l",xlab="Zeitpunkt i",ylab="approximierter p-Wert")
+ }

```

# Literaturverzeichnis

- [1] A. Agresti, *An Introduction to Categorical Data Analysis*, Wiley Series in Probability and Statistics, Wiley Interscience (1996)
- [2] J. Besag, *Markov Chain Monte Carlo Methods for Statistical Inference*, Department of Statistics University of Washington, Seattle, USA (2004)
- [3] P. Brémaud, *Markov Chains–Gibbs Fields, Monte Carlo Simulation an Queues*, Springer (1999)
- [4] J. Bortz, *Statistik für Sozialwissenschaftler*, 5. Auflage, Springer (1999)
- [5] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press (1975)
- [6] S. Chib, E. Greenberg, *Understanding the Metropolis-Hastings Algorithm*, The American Statistician, **49** (4), S. 327-335 (1995)
- [7] P. Diaconis, B. Sturmfels, *Algebraic algorithms for sampling from conditional distributions*, The Annals of Statistics, **26** (1), S. 363-397 (1998)
- [8] L. Farhmeir, A. Hamerle, G. Tutz, *Multivariate statistische Verfahren*, Walter de Gruyter (1996)
- [9] G.S. Fishman, *Monte Carlo*, Springer (1996)
- [10] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall (1996)
- [11] S.J. Haberman, *Analysis of Qualitative Data - Volume 1*, Academic Press (1978)
- [12] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press (2002)
- [13] W.K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika, **57**, S. 97-109 (1970)
- [14] W. König, *Stochastische Prozesse I: Markovketten in diskreter und stetiger Zeit*, [www.math.uni-leipzig.de/~koenig/www/StPrI.ps](http://www.math.uni-leipzig.de/~koenig/www/StPrI.ps) (2005)
- [15] U. Krenzel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Vieweg (2000)



- [16] K. Krickeberg, H. Ziezold, *Stochastische Methoden*, 4. Auflage, Springer (1995)
- [17] N. Metropolis, A.W. Rosenbluth, M. Rosenbluth, A.H. Teller, E. Teller, *Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics, **21** (6), S. 1087-1092 (1953)
- [18] C. Müller, *Skript zu Einführung in die Stochastik* (2006)
- [19] C. Müller, *Skript zu Stochastik für Ingenieure* (2006)
- [20] J.R. Norris, *Markov Chains*, Cambridge University Press (1997)
- [21] J. Puhani, *Statistik*, Lexika Verlag (2005)
- [22] F. Rapallo, *Algebraic Markov Bases and MCMC for Two-Way Contingency Tables*, Scand. J. Stat., **30** (2), S. 385-397 (2003)
- [23] L. Sachs, J. Hedderich, *Angewandte Statistik - Methodensammlung mit R*, 12. Auflage, Springer (2006)
- [24] S. Sullivant, *Log-Linear Models, Toric Varieties, and Their Markov Bases*, [www.math.harvard.edu/~seths/lecture2.pdf](http://www.math.harvard.edu/~seths/lecture2.pdf)
- [25] A. Vadeby, *On Gibbs Sampler and Metropolis-Hastings Applied to Pairwise Poisson and Car Crash Data*, [www.mai.liu.se/~öleri/trafikgrupp/pdffil/paperE.pdf](http://www.mai.liu.se/~öleri/trafikgrupp/pdffil/paperE.pdf) (2003)
- [26] G. Walz, *Lexikon der Statistik*, Spektrum (2004)
- [27] E. Wiebe, *Markov-Ketten Monte-Carlo Methode*, [wwwcs.uni-paderborn.de/cs/ag-madh/vorl/Perlen/Wiebe.ps.gz](http://wwwcs.uni-paderborn.de/cs/ag-madh/vorl/Perlen/Wiebe.ps.gz) (2002)
- [28] *Metropolis-Hastings algorithm II*, [www.angelfire.com/falcon/isinotes/statcomp04/mcmc/mh2.html](http://www.angelfire.com/falcon/isinotes/statcomp04/mcmc/mh2.html)
- [29] *Metropolis-Hastings-Algorithmus*, [www.mathematik.uni-ulm.de/stochastik/lehre/ss03/markov/skript/skript.html](http://www.mathematik.uni-ulm.de/stochastik/lehre/ss03/markov/skript/skript.html)
- [30] *Kapitel 5: Markov Chain Monte Carlo (MCMC) Methoden*, [www.stochastik.jku.at/LEHRE/Simulation/Skript07T4.pdf](http://www.stochastik.jku.at/LEHRE/Simulation/Skript07T4.pdf)

## **Erklärung**

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Kassel, den 3. Juni 2008