# Elementare Wahrscheinlichkeit
# Miniskript zur Vorlesung im HS 2013

Hansruedi Künsch
D-Math, ETH Zürich

Version vom 18. Dezember 2013

## Vorbemerkung

Das Ziel der Vorlesung ist, einen Einblick in ein paar Anwendungen der Wahrscheinlichkeitsrechnung zu geben, die in der Vorlesung des vierten Semesters nicht vorkommen, aber interessant sind und die Vielfalt der Mathematik illustrieren. Dies ist insbesondere keine systematische Einführung in Wahrscheinlichkeitsrechnung. Die nötigen Konzepte werden dann eingeführt, wenn sie gebraucht werden. Der Stoff aus dem ersten Studienjahr wird vorausgesetzt, insbesondere ein Verständnis für exakte Beweise. Das Skript ist rudimentär und soll die Repetition erleichtern, ist aber nicht geeignet als eigenständiger Text zum Selbststudium.

## 1 Zufallsauswahl

Gegeben sind eine endliche Menge $\Omega = \{\omega_1, \dots, \omega_n\}$ und Gewichte $p(\omega_i) \geq 0$ $(i = 1, \dots, n)$, welche normiert sind im Sinne, dass

$$\sum_{i=1}^{n} p(\omega_i) = 1$$

gilt. Dabei soll $p(\omega_i)$ die Wahrscheinlichkeit darstellen, dass bei einer zufälligen Auswahl eines Elements aus $\Omega$ das Element $\omega_i$ gewählt wird. Anschaulich soll das heissen, dass bei $N$ Wiederholungen das Element $\omega_i$ ungefähr $Np(\omega_i)$ mal gewählt wird, bzw. dass $p(\omega_i)$ unsere subjektive Einschätzung quantifiziert, wie sicher wir sind, dass $\omega_i$ gewählt wird. Eine solche Auswahl kann geschehen über die Wahl einer gleichverteilten Zahl in $[0, 1]$, bzw. wenn alle $p(\omega_i) \in \mathbb{Q}$, durch das Ziehen aus einer Urne.

## 2 Kodierung und Entropie

Wir wollen einen Text, der mit dem Alphabet $\Omega = \{1, \dots, n\}$ geschrieben ist, möglichst kurz als binären Text kodieren. Ein Kode ist eine Abbildung

$$\varphi : \Omega \to \cup_{\ell=1}^{\infty} \{0, 1\}^{\ell}, \quad i \mapsto \varphi(i) = (\varphi_1(i), \dots, \varphi_{\ell(i)}(i)).$$

Damit wir den ursprünglichen Text eindeutig zurückgewinnen können, ohne ein weiteres Zeichen zur Trennung von Buchstaben zu benötigen, muss $\varphi$ ein sogenannter *Präfix-Kode* sein. Das bedeutet, dass für alle $i, j$ mit $\ell(i) \leq \ell(j)$

$$(\varphi_1(i), \ldots, \varphi_{\ell(i)}(i)) \neq (\varphi_1(j), \ldots, \varphi_{\ell(i)}(j)).$$

Wenn die Buchstaben zufällig gewählt werden mit Wahrscheinlichkeiten $(p(1), \ldots, p(n))$, dann erscheint in einem Text der Länge $N$ der Buchstabe $i$ ungefähr $Np(i)$-mal. Der kodierte binäre Text hat also ungefähr die Länge $\sum_{i=1}^{n} Np(i)\ell(i)$. Für gegebene Wahrscheinlichkeiten $(p(1), \ldots, p(n))$ suchen wir daher einen Präfix-Kode, der

$$\sum_{i=1}^{n} \ell(i)p(i)$$

minimiert. Diese Summe nennnen wir den Erwartungswert der Zufallsvariable $\ell$ und bezeichnen ihn mit $\mathbb{E}(\ell)$. Wir werden die Lösungen von Huffman, Shannon und Fano für dieses Problem diskutieren. Wir können $\varphi(i)$ als Pfad im vollständigen binären Baum auffassen: Wir starten bei der Wurzel und machen $\ell(i)$ Schritte, wobei $\varphi_k(i)$ angibt, ob wir im $k$-ten Schritt nach rechts oder nach links gehen. Bei einem Präfix-Kode liegt dann kein Endpunkt eines Pfades auf einem anderen Pfad.

Kodierung ist äquivalent zur Identifikation eines Elements aus $\Omega$ mit Hilfe von Fragen der Form "Liegt das gesuchte Element in $\Omega' \subset \Omega$ ?". Eine Fragestrategie wählt sukzessive die Teilmengen $\Omega'$ aus. Wir bezeichnen mit $\varphi(i)$ die Folge der Antworten, die man mit der Strategie erhält, wenn $i$ das gesuchte Element ist. Wenn die Strategie jedes Element eindeutig identifizieren kann, ist $\varphi$ ein Präfix-Kode.

## 2.1 Der Kode von Huffman

Dieses Verfahren bestimmt den optimalen Kode rekursiv, indem die zugehörigen Pfade im binären Baum von unten her konstruiert werden. Wir nehmen o.B.d.A. an dass $p(1) \geq p(2) \geq \ldots \geq p(n)$. Wenn $\varphi$ ein optimaler Kode ist, muss $\ell(i) \leq \ell(j)$ gelten falls $p(i) > p(j)$. Durch Vertauschung von gleichwahrscheinlichen Symbolen können wir daher erreichen, dass $\ell(1) \geq \ell(2) \geq \ldots \geq \ell(n)$. Ausserdem muss für einen optimalen Kode ein $j < n$ existieren, so dass $\ell(j) = \ell(n)$ und $\varphi(j)$ und $\varphi(n)$ sich nur im letzten Bit unterscheiden. Anderenfalls wäre das letzte Bit von $\varphi(n)$ überflüssig. Allenfalls nach Vertauschen von Symbolen gilt dies sogar für $j = n-1$. Wenn wir also das letzte Bit von $\varphi(n-1)$ weglassen, erhalten wir einen Kode $\varphi'$ für $\Omega' = \{1, \ldots, n-1\}$, und wenn wir $p'(n-1) = p(n-1)+p(n)$ setzen, gilt $\mathbb{E}(\ell) = \mathbb{E}(\ell') + p'(n-1)$. Also muss der Kode $\varphi'$ optimal sein für $\Omega'$, und wir können dieses Verfahren iterieren.

## 2.2 Der Kode von Shannon

Shannon konstruiert einen nahezu optimalen Kode in zwei Schritten, indem zuerst die Längen $\ell(i)$ und dann die Kodewörter $\varphi(i)$ bestimmt werden. Für den zweiten Schritt benutzt man das folgende Lemma

**Lemma 1 (Kraft)** *Die folgende Bedingung ist notwendig und hinreichend für die Existenz eines Präfix-Kodes mit Längen* $(\ell(1), \ldots, \ell(n))$*:*

$$\sum_{i=1}^{n} 2^{-\ell(i)} \leq 1$$

**Beweis:** Für den Beweis der Notwendigkeit setzen wir $m = \max \ell(i)$ und bezeichnen mit $C_i$ die Menge der binären $m$-Tuppel, die mit $\varphi(i)$ beginnen. Die Anzahl der Elemente von $C_i$ ist $2^{m-\ell(i)}$, und wegen der Präfix-Eigenschaft sind die $C_i$'s paarweise disjunkt. Weil es $2^m$ binäre $m$-Tuppel gibt, folgt $\sum_{i=1}^n 2^{m-\ell(i)} \leq 2^m$.

Die Umkehrung geht mit Induktion. Wir nehmen an o.B.d.A. an, dass $\ell(1) \geq \ell(2) \geq \ldots \geq \ell(n)$. Nach Induktionsannahme können wir $\varphi(i) \in \{0,1\}^{\ell(i)}$ für $i \leq n-1$ so wählen, dass die Präfix-Eigenschaft erfüllt ist. Weil $\sum_{i=1}^{n-1} 2^{-\ell(i)} < 1$, zeigt der Beweis der Notwendigkeit, dass es dann sogar mindestens ein $\ell(n-1)$-Tuppel gibt, das nicht mit einem $\varphi(i)$ für $i \leq n-1$ beginnt. Also kann man $\varphi(n)$ mit diesem $\ell(n-1)$-Tuppel beginnen lassen und noch $\ell(n) - \ell(n-1)$ Nullen anhängen. □

Für einen optimalen Kode muss man also $\ell(1), \ldots, \ell(n) \in \mathbb{N}$ so wählen, dass $\sum_{i=1}^n \ell(i)p(i)$ minimal wird unter der Nebenbedingung $\sum_{i=1}^n 2^{-\ell(i)} \leq 1$. Das entsprechende vereinfachte Minimierungsproblem, wo alle $\ell(i)$ reell sind statt ganzzahlig, lässt sich leicht lösen mit Hilfe von Lagrange-Multiplikatoren: Man erhält $\ell(i) = -\log_2\ p(i)$. Der Shannon-Kode rundet dann die Längen auf die nächstgrössere ganze Zahl auf oder ab, unter Beachtung der Ungleichung von Kraft.

**Satz 1** *Für jeden Präfix-Kode gilt*

$$\mathbb{E}(\ell) = \sum_{i=1}^n \ell(i)p(i) \geq \sum_{i=1}^n (-log_2\ p(i))p(i) =: H(\mathbb{P})$$

*und der Kode mit Längen* $\ell(i) = \lceil -\log_2\ p(i) \rceil$ *erfüllt*

$$\mathbb{E}(\ell) = \sum_{i=1}^n \ell(i)p(i) < \sum_{i=1}^n (-log_2\ p(i))p(i) + 1 = H(\mathbb{P}) + 1.$$

Die Grösse $H(\mathbb{P}) = \sum_{i=1}^n (-log_2\ p(i))p(i)$ heisst die *Entropie* der Wahrscheinlichkeit $\mathbb{P} = (p(1), \ldots, p(n))$.

## 2.3 Der Kode von Fano

Dieses Verfahren bestimmt den optimalen Kode rekursiv, indem die zugehörigen Pfade im binären Baum von oben her konstruiert werden. Das heisst wir bestimmen zunächst die erste Komponente $\varphi_1$ und wählen nachher zwei Codes $\varphi^0$ und $\varphi^1$ mit Längen $\ell^0$ und $\ell^1$ für die beiden Alphabete $\Omega^0 = \{i; \varphi_1(i) = 0\}$, bzw. $\Omega^1 = \{i; \varphi_1(i) = 1\}$. Das folgende Argument zeigt, wie wir $\Omega^k$ wählen müssen, damit wir einen Kode erhalten, dessen erwartete Länge nahe bei der Shannon-Schranke $H(\mathbb{P})$ ist.

Die erwartete Länge des resultierenden Kodes hat die Form

$$\mathbb{E}(\ell) = \sum_{i \in \Omega^0} p(i)(1 + \ell^0(i)) + \sum_{i \in \Omega^1} p(i)(1 + \ell^1(i)).$$

Für $k = 0, 1$ definieren wir

$$\mathbb{P}(\Omega^k) = \sum_{i \in \Omega^k} p(i), \quad p(i|\Omega^k) = \frac{p(i)}{\mathbb{P}(\Omega^k)}\ (i \in \Omega^k).$$

Weil $\sum_{i\in\Omega^k} p(i|\Omega_k) = 1$, definieren diese Gewichte Wahrscheinlichkeiten $\mathbb{P}(.|\Omega^k)$ auf $\Omega^k$, und wir können schreiben

$$\mathbb{E}(\ell) = 1 + \mathbb{P}(\Omega^0)\mathbb{E}(\ell^0|\Omega^0) + \mathbb{P}(\Omega^1)\mathbb{E}(\ell^1|\Omega^1).$$

Damit ist klar, dass die beiden Kodes $\varphi^k$ optimal sein müssen für die Wahrscheinlichkeiten $\mathbb{P}(.|\Omega^k)$. Ist dies der Fall, dann ist $\mathbb{E}(\ell)$ nur wenig grösser als

$$1 + \mathbb{P}(\Omega^0)H(\mathbb{P}(.|\Omega^0)) + \mathbb{P}(\Omega^1)H(\mathbb{P}(.|\Omega^1)).$$

Andererseits kann man leicht nachrechnen, dass

$$H(\mathbb{P}) = \mathbb{P}(\Omega^0)H(\mathbb{P}(.|\Omega^0)) + \mathbb{P}(\Omega^1)H(\mathbb{P}(.|\Omega^1)) + H((\mathbb{P}(\Omega^0), \mathbb{P}(\Omega^1))). \tag{1}$$

Damit $\mathbb{E}(\ell)$ nahe bei der Shannon-Schranke ist, muss daher $H((\mathbb{P}(\Omega^0), \mathbb{P}(\Omega^1)))$ nahe bei 1 sein. Lemma 2 zeigt, dass dazu $\mathbb{P}(\Omega^0)$ möglichst nahe bei $\frac{1}{2}$ sein muss, d.h. die beiden möglichen Werte von $\varphi_1$ müssen etwa gleich wahrscheinlich sein.

**Lemma 2** *Die Funktion*

$$x \in [0,1] \mapsto -x\log_2(x) - (1-x)\log_2(1-x)$$

*(wobei $0\log 0 = 0$ gesetzt wird) ist symmetrisch bez. $x = \frac{1}{2}$, monoton wachsend auf $[0, \frac{1}{2}]$, konvex und nach unten beschränkt durch* $\min(2x, 2(1-x))$.

Wir nennen $\mathbb{P}(\Omega^k)$ die Wahrscheinlichkeit von $\Omega^k$ und $p(i|\Omega^k)$ die bedingte Wahrscheinlichkeit von $i$, wenn die Auswahl auf ein Element von $\Omega^k$ beschränkt ist. Das ist konsistent mit unserer früheren Interpretation von Wahrscheinlichkeit als approximative relative Häufigkeit bei vielen Wiederholungen. Wenn wir nämlich $N$ mal ein Element aus $\Omega$ wählen, dann kommt ein Element aus $\Omega^k$ ungefähr $N\mathbb{P}(\Omega^k)$ Mal vor, und die relative Häufigkeit von $p(i)$ unter denjenigen Wiederholungen, bei denen ein Element aus $\Omega^k$ gewählt wurde, ist ungefähr gleich $p(i|\Omega^k)$.

Der folgende Satz beschreibt die genauen Eigenschaften eines Kodes, der auf obigen Überlegungen basiert.

**Satz 2 (Fano)** *Sei $p(1) \geq p(2) \geq \ldots \geq p(n)$ und sei $1 \leq m < n$ so bestimmt, dass $|\mathbb{P}(\{1,2,\ldots m\}) - \frac{1}{2}|$ minimal ist. Wenn wir $\varphi_1(i) = 0$ setzen für $i \leq m$ und $\varphi_1(i) = 1$ für $i > m$, und diese Konstruktion auf den beiden entstehenden Teilmengen fortsetzen, wenn sie mehr als 1 Element enthalten, dann gilt*

$$\mathbb{E}(\ell) \leq H(\mathbb{P}) + 1 - 2p(n).$$

Der Beweis geht induktiv mit dem gleichen Argument wie oben. Gemäss Induktionsvoraussetzung gilt

$$\mathbb{E}(\ell) \leq 1 + \mathbb{P}(\Omega^0)H(\mathbb{P}(.|\Omega^0)) + \mathbb{P}(\Omega^1)H(\mathbb{P}(.|\Omega^1)) + 1 - 2p(m) - 2p(n).$$

Aus Lemma 2 und der Definition vom $m$ folgt ferner, dass

$$H((\mathbb{P}(\Omega^0), \mathbb{P}(\Omega^1))) \geq \min(2\mathbb{P}(\Omega^0), 2\mathbb{P}(\Omega^1)) \geq 1 - 2p(m).$$

Zusammen mit der Gleichung (1) ist damit der Induktionsschritt komplett.

## 2.4 Entropierate

Ich gebe in diesem Unterkapitel noch einen Ausblick auf weitere Resultate, die ich nicht mehr im Einzelnen beweisen werde.

Die Länge des Shannon-Kodes ist höchstens um 1 grösser als die Entropie. Da

$$H(\mathbb{P}) \leq H((1/n, \ldots, 1/n)) = \log_2(n),$$

ist die relative Differenz gross für kleines $n$, geht aber für viele Wahrscheinlichkeiten $\mathbb{P}$ gegen null mit wachsendem $n$. Der Shannon-Kode sollte daher auf Blöcken von $k$ aufeinanderfolgenden Symbolen verwendet werden. Das heisst, wir betrachten das neue Alphabet $\Omega_k = \{1, 2, \ldots, n\}^k$. Um herauszufinden, welche Wahrscheinlichkeiten $p_k((i_1, \ldots, i_k))$ angemessen sind, müsste man in einem langen Texten die Häufigkeit der verschiedenen $k$-Tuppel auszählen, was aufwändig ist für grosses $k$. Wahrscheinlichkeiten für $k$-Tuppel implizieren jedoch zwei Wahrscheinlichkeiten für ein $(k-1)$-Tuppel, je nachdem, ob wir das erste oder das letzte Symbol ignorieren. Es ist plausibel, dass diese beiden Wahrscheinlichkeiten gleich sein sollen. Dann gilt der folgende, tiefliegende Satz.

**Satz 3 (Shannon-McMillan-Breiman)** *Sei $(\mathbb{P}_k)$ eine Folge von Wahrscheinlichkeiten auf $\{1, 2, \ldots, n\}^k$, welche konsistent sind im Sinne, dass für alle $k$ und alle $i_1, \ldots, i_k$*

$$\sum_{j=1}^{n} p_k((i_1, \ldots, i_{k-1}, j)) = \sum_{j=1}^{n} p_k((j, i_1, \ldots, i_{k-1})) = p_{k-1}((i_1, \ldots, i_{k-1})).$$

*Unter einer weiteren schwachen Bedingung, der sogenannten Ergodizität, existiert dann die Entropierate $h = \lim_{k\to\infty} k^{-1} H(\mathbb{P}_k)$, und für grosses $k$ ist mit Wahrscheinlichkeit nahe bei 1*

$$-\log_2 p_k((i_1, \ldots, i_k)) \approx k \cdot h.$$

Dies bedeutet, dass der Shannon-Kode einen "typischen" Block mit ungefähr $k \cdot h$ Bits kodiert, und für die seltenen Blöcke ein längeres Kodewort wählt. "Praktisch jeder" Text der Länge $N$ kann daher mit ungefähr $Nh$ Bits kodiert werden. Also treten von den $n^N$ theoretisch möglichen Texten der Länge $N$ praktisch nur $2^{hN}$ Texte auf. Wenn wir diese $2^{hN}$ Texte alle mit dem Shannon-Kode als binäre Folgen der Länge $Nh$ kodiert haben, dann können wir diese Texte auch leicht verschlüsseln: Wir wählen eine Permutation von 32 Elementen (es gibt ungefähr $10^{35}$ davon), die nur der Sender und der Empfänger kennt. Statt der ursprünglichen binären Folge versendet man diejenige Folge, bei der die Permutation auf jeden Block von 5 Bits angewendet wird. Kennt man die Permutation nicht, lässt sich die Verschlüsselung nicht rückgängig machen, denn jede Permutation führt zu einem der praktisch vorkommenden Texte. Selbst der Kontext hilft nicht weiter, weil man nicht zwischen verschiedenen Texten mit dem gleichen Kontext entscheiden kann.

Um den blockweisen Shanon-Kode zu benutzen, müssten wir wie gesagt bestimmen, wie häufig die $n^k$ möglichen Blöcke von $k$ Symbolen sind. In den 70-er Jahren haben Lempel und Ziv jedoch einen Kode gefunden, der ohne die Kenntnis dieser Wahrscheinlichkeiten auskommt und trotzdem asymptotisch einen Text der Länge $N$ mit $h \cdot N$ Bits kodieren kann. Dieser Kode wird von vielen Computerprogrammen benutzt, um Dateien zu komprimieren.

# 3 Mischen von Karten

Wie oft muss man einen Stapel von 52 Karten mischen, damit der Effekt der ursprünglichen Anordnung nicht mehr sichtbar ist ? Bayer und Diaconis haben 1992 die Antwort "Sieben Mal" gegeben, die sogar in der der New York Times erwähnt wurde. In diesem Kapitel werden wir die Begründung dieser Antwort kennenlernen.

## 3.1 Definitionen

Eine Mischung von $n$ Karten ist eine Permutation $\omega$ von $n$ Elementen, d.h. eine bijektive Abbildung von $\{1, 2, \ldots, n\}$ in sich (d.h. wir nehmen an, dass die Karten fortlaufend nummeriert sind). $\omega(i)$ gibt an, an welcher Stelle die Karte, die vor dem Mischen an Stelle $i$ im Stapel war, nach dem Mischen steht. Die Inverse $\omega^{-1}$ beschreibt also, wie ein zu Beginn geordneter Stapel nach dem Mischen angeordnet ist. In diesem Kapitel ist $\Omega$ immer die Menge aller $n!$ Permutationen von $n$ Elementen.

Wir nehmen an, dass die Person, die die Karten mischt, eine Permutation $\omega$ zufällig auswählt gemäss einer Wahrscheinlichkeitsverteilung auf der Menge $\Omega$ aller Permutationen. Idealerweise sollte die Verteilung uniform sein, d.h. $p(\omega) = 1/n!$ für alle $\omega$'s, aber das lässt sich physikalisch mit Karten nicht realisieren. Man begnügt sich damit, durch Wiederholung von einfacheren Mischverfahren möglichst nahe an die uniforme Verteilung zu kommen.

Wenn wir zweimal hintereinander Mischen, zuerst gemäss Permutation $\omega_1$ und dann gemäss Permutation $\omega_2$, dann ist das gleichwertig damit einmal gemäss Permutation $\omega_2 \circ \omega_1$ zu mischen, wobei

$$(\omega_2 \circ \omega_1)(i) = \omega_2(\omega_1(i)).$$

Mit dieser Zusammensetzung wird $\Omega$ zu einer nicht-kommutativen Gruppe. Wir überlegen zuerst, wie wir die Wahrscheinlichkeiten festlegen sollen, wenn wir zweimal hintereinander mischen, d.h. welche Wahrscheinlichkeit soll das Paar $(\omega_1, \omega_2)$ haben, wenn die Wahrscheinlichkeiten $p_1(\omega_1)$ und $p_2(\omega_2)$ gegeben sind ? Es ist plausibel, dass die Wahl der zweiten Permutation nicht von der Wahl der ersten beeinflusst werden soll. Das ist dann der Fall, wenn

$$\frac{Np(\omega_1, \omega_2)}{Np_1(\omega_1)} = p_2(\omega_2) \Leftrightarrow p(\omega_1, \omega_2) = p_1(\omega_1)p_2(\omega_2).$$

Der Quotient links ist nämlich ungefähr gleich der relativen Häufigkeit von $\omega_2$ beim zweiten Mischen unter den Fällen, wo beim ersten Mischen $\omega_1$ gewählt wurde, und dies soll einfach gleich der relativen Häufigkeit von $\omega_2$ beim zweiten Mischen in allen Fällen insgesamt sein. Dies wird als *Unabhängigkeit* der zwei Mischvorgänge bezeichnet.

Damit können wir sofort die Wahrscheinlichkeit berechnen, dass nach zweimaligem Mischen mit Verteilungen $\mathbb{P}_1$ und $\mathbb{P}_2$ die Permutation $\omega$ resultiert

$$p(\omega) = \sum_{\omega_2 \circ \omega_1 = \omega} p_1(\omega_1)p_2(\omega_2) = \sum_{\omega_1} p_1(\omega_1)p_2(\omega \circ \omega_1^{-1}).$$

## 3.2 Riffle shuffle

Wir geben nun eine Wahrscheinlichkeit auf $\Omega$ an, die plausibel erscheint, um einen in der Realität häufig verwendeten Mischvorgang zu beschreiben, der auf Englisch "riffle shuffle"

heisst, auf Deutsch "Bogenmischen". Dabei wird der Kartenstapel in zwei Teile geteilt und dann werden die beiden Teile wieder ineinander geschachtelt (verzahnt), wobei die Ordnung innerhalb der beiden Teile unverändert bleibt. Ein solcher Mischvorgang ist festgelegt durch die Angabe der Grösse $k$ des linken Teils und durch die Angabe der Plätze, wo die Karten des linken Teils nach der Verschachtelung stehen. Diese Plätze bilden eine Teilmenge $L$ von $\{1, 2, \ldots, n\}$ mit $k$ Elementen. Wenn zum Beispiel $n = 8$, $k = 3$ und $L = \{2, 5, 6\}$ ist, dann ist $\omega^{-1} = (4, 1, 5, 6, 2, 3, 7, 8)$ und $\omega = (2, 5, 6, 1, 3, 4, 7.8)$. Welche Wahrscheinlichkeiten für $(k, L)$ das Verhalten der Personen beim Kartenmischen genau beschreiben, ist schwierig zu sagen. Die Teilung erfolgt ungefähr in der Mitte, was durch $p(k) = \binom{n}{k}2^{-n}$ gut wiedergegeben wird (die Binomialkoeffizienten sind in der Mitte am grössten). Ein geschickte Person wird bei der Verschachtelung abwechselnd jeweils wenige Karten von links und von rechts nehmen. Wir nehmen an, dass jede Teilmenge $L$ von $k$ Elementen gleich wahrscheinlich ist: $p(L) = 1/\binom{n}{k}$. Aus der Interpretation von Wahrscheinlichkeit als approximative Häufigkeit bei vielen Wiederholungen des Mischvorgangs erhält man dann

$$p(k, L) = \binom{n}{k}2^{-n} \cdot \frac{1}{\binom{n}{k}} = 2^{-n}$$

(für jedes $L$ mit $k$ Elementen, anderenfalls ist $p(k, L) = 0$). Jede Kombination von Aufteilung und Verschachtelung gleich wahrscheinlich. Das führt jedoch nicht zur Gleichverteilung auf $\Omega$, weil viele Permutationen nicht auf diese Weise erzeugt werden können, und weil andererseits die Identität für beliebiges $k$ und $L = \{1, \ldots, k\}$ entsteht.

Im Prinzip kann man nun durch Iteration die Wahrscheinlichkeiten aller Permutationen berechnen, die durch $k$-fache Wiederholung von riffle shuffles entstehen. Praktisch durchführen lässt sich das aber für $n = 52$ nicht, weil sich eine Summe über $52! \approx 10^{68}$ Terme nicht berechnen lässt. Wir brauchen also noch mehr Einblick, um zu sehen, weshalb 7 riffle shuffle eine Permutation ergeben, die approximativ uniform verteilt ist.

Der Schlüssel dafür ist das Konzept eines aufsteigenden konsekutiven Runs einer Permutation $\omega$. Darunter versteht man eine Folge $(i, i + 1, \ldots, j)$ so dass $\omega(i - 1) > \omega(i) < \omega(i + 1) < \ldots < \omega(j) > \omega(j + 1)$ (für $i = 1$ fällt die erste Ungleichung weg, für $j = n$, die letzte). Zur Vereinfachung der Sprechweise lassen wir auch die Adjektive "aufsteigend" und "konsekutiv" und sprechen einfach von einem Run. Die Runs einer Permutation lassen sich schnell identifizieren, weil jede Inversion, d.h. jede Stelle $j > 1$ mit $\omega(j) > \omega(j - 1)$, einen neuen Run startet. Die Anzahl runs ist also gleich der Anzahl Inversionen plus Eins.

Bei einem riffle shuffle haben offensichtlich nur die Permuationen mit einem oder zwei Runs positive Wahrscheinlichkeit. Nach zwei riffle shuffles kann es dann maximal 4 Runs geben, etc. Man kann aber sogar eine viel präzisere Aussage machen:

**Satz 4** *Nach $m$ unabhängigen riffle shuffles hat jede Permutation $\omega$ mit $r \leq 2^m$ runs die Wahrscheinlichkeit*

$$p_m(\omega) = 2^{-nm}\binom{2^m + n - r}{n}.$$

**Beweis:** Ich gebe den Beweis nur für $m = 2$. Die beiden riffle shuffles sind, wie wir oben gesehen haben, charakterisiert durch 2 Teilmengen $L_1$ und $L_2$ von $\{1, \ldots, n\}$: Die Anzahl Elemente von $L_i$ gibt an, wie viele Karten bei der Teilung im linken Teilstapel sind, und $L_i$ gibt die Position der Karten aus dem linken Teilstapel nach dem Mischen an. Jede Kombination $(L_1, L_2)$ hat die Wahrscheinlichkeit $4^{-n}$. Die Kombination der beiden shuffles kann charakterisiert werden durch eine Partition von $\{1, \ldots, n\}$ in vier (eventuell leere) Teilmengen $(M_{00}, M_{01}, M_{10}, M_{11})$: $M_{00}$ gibt die Endposition der Karten an, die beides

Mal im linken Teilstapel sind; $M_{01}$ gibt die Endposition der Karten an, die beim ersten Mal im linken und beim zweiten Mal im rechten Teilstapel sind, etc. Es gibt $4^n$ solche Partitionen, und wir werden zeigen, dass die Zuordnung $(L_1, L_2) \mapsto (M_{00}, M_{01}, M_{10}, M_{11})$ eine Bijektion ist. Damit haben alle Partitionen die Wahrscheinlichkeit $4^{-n}$.

Aus der Partition $(M_{00}, M_{01}, M_{10}, M_{11})$ können wir die resultierende Permutation $\omega_2 \circ \omega_1$ bestimmen: Die Karten, die zweimal im linken Teilstapel sind, müssen zu Beginn zuoberst sein, und ihre Reihenfolge wird bei der Verteilung auf $M_{00}$ nicht geändert. Als nächstes kommen im ursprünglichen Stapel die Karten, die zuerst im linken und dann im rechten Teilstapel sind, und ihre Reihenfolge bleibt ebenfalls gleich, etc. Jede Inversion in der Permutation $\omega_2 \circ \omega_1$ markiert eine Grenze zwischen zwei Mengen der Partition, weil innerhalb einer Menge die Reihenfolge nicht verändert wird. Wenn die Permutation 4 runs hat, dann ist die Partition eindeutig bestimmt und damit hat diese Permutation die Wahrscheinlichkeit $4^{-n}$. Permutationen mit weniger als 4 runs können jedoch aus verschiedenen Partitionen hervorgehen und haben daher ein Vielfaches von $4^{-n}$ als Wahrscheinlichkeit.

Wir betrachten zuerst ein Beispiel: Sei $n = 8$, $L_1 = \{2, 5, 6\}$, $L_2 = \{1, 3, 5, 8\}$. Der linke Teilstapel besteht also zuerst aus 3 und dann aus 4 Karten. Die erste Karte ist nach dem ersten shuffle auf Platz 2, kommt also bei der zweiten Teilung wieder nach links und endet damit auf Platz 3. Die zweite Karte ist nach dem ersten shuffle auf Platz 5, kommt daher bei der zweiten Teilung nach rechts zuoberst und endet damit auf Platz 2 (dem ersten Element von $L_2^c$). Wenn wir diese Überlegungen fortsetzen, erhalten wir $M_{00} = \{3\}$, $M_{01} = \{2, 4\}$, $M_{10} = \{1, 5, 8\}$ und $M_{11} = \{6, 7\}$. Also muss $\omega_2 \circ \omega_1 = (32415867)$ sein, was sich leicht nachkontrollieren lässt, da $\omega_1 = (25613478)$ und $\omega_2 = (13582467)$. Die Permutation (32415867) hat 4 runs und damit Wahrscheinlichkeit $4^{-n}$.

Als nächstes beweisen wir, dass man aus der Partition $(M_{00}, M_{01}, M_{10}, M_{11})$ $L_1$ und $L_2$ zurückgewinnen kann. Die Karte $i$ ist bei der ersten Teilung genau dann links, wenn $i \leq k_1$, und bei der zweiten Teilung ist sie genau dann links, wenn $\omega_1(i) \leq k_2$. Also ist

$$M_{00} \cup M_{10} = \{\omega_2(\omega_1(i)) \mid \omega_1(i) \leq k_2\} = \{\omega_2(j) \mid j \leq k_2\} = L_2.$$

Aus $L_2$ folgt $\omega_2$ und damit auch $L_1$, weil

$$M_{00} \cup M_{01} = \{\omega_2(\omega_1(i)) \mid i \leq k_1\} \Leftrightarrow L_1 = \{\omega_1(i) \mid i \leq k_1\} = \{\omega_2^{-1}(j) \mid j \in M_{00} \cup M_{01}\}.$$

Für die Permuation (32415867) folgt also $L_2 = \{3, 1, 5, 8\}$ und damit $\omega_2 = (13582467)$ und schliesslich $L_1 = \omega_2^{-1}(\{3, 2, 4\}) = \{2, 5, 6\}$.

Wir machen noch ein Beispiel einer Permutation mit $n = 8$ und 3 runs: $\omega = (23567418)$. Hier sind nur 2 der drei Schnittstellen zwischen den 4 Elementen der Partition durch eine Inversion festgelegt. Die dritte Schnittstelle kann irgendwo zwischen 0 und 8 sein, d.h. es gibt 9 Möglichkeiten. Ist die dritte Schnittstelle zum Beispiel bei 0, erhalten wir $M_{00} = \emptyset$, $M_{01} = \{2, 3, 5, 6, 7\}$, $M_{10} = \{4\}$ $M_{11} = \{1, 8\}$ und $\omega_1 = (34567128)$, $\omega_2 = (41235678)$. Ist die dritte Schnittstelle zum Beispiel bei 2, erhalten wir $M_{00} = \{2, 3\}$, $M_{01} = \{5, 6, 7\}$, $M_{10} = \{4\}$, $M_{11} = \{1, 8\}$ und $\omega_1 = (12567348)$, $\omega_2 = (23415678)$. Dies kann man analog für alle anderen 7 Möglichkeiten durchführen.

Allgemein gilt: wenn $\omega$ $r < 2^m$ runs hat, dann kann man $2^m - r$ Schnittstellen zwischen Mengen der Partition beliebig wählen irgendwo zwischen 0 und $n$. Da Schnittstellen auch zusammenfallen können, ist die Anzahl Möglichkeiten daher durch den Binomialkoeffizienten im Satz gegeben. □

### 3.3 Totaler Variationsabstand

Zunächst definieren wir einen Abstand zwischen zwei Wahrscheinlichkeiten, um zu quantifizieren, was "approximativ uniform" heisst.

**Definition 1** *Der totale Variationsabstand zwischen zwei Wahrscheinlichkeiten $\mathbb{P}$ und $\mathbb{Q}$ auf $\Omega$ ist definiert als*

$$\|\mathbb{P} - \mathbb{Q}\| = \frac{1}{2} \sum_{\omega \in \Omega} |p(\omega) - q(\omega)|.$$

Die Bedeutung dieses Abstands in der Wahrscheinlichkeitsrechnung wird durch das folgende Lemma erklärt.

**Lemma 3** *Es gilt*

$$\|\mathbb{P} - \mathbb{Q}\| = \max_{A \subset \Omega} |\mathbb{P}(A) - \mathbb{Q}(A)| = \sum_{\omega : p(\omega) > q(\omega)} (p(\omega) - q(\omega)) = \sum_{\omega : q(\omega) > p(\omega)} (q(\omega) - p(\omega)).$$

**Beweis:** Für jedes $A \subset \Omega$ gilt

$$\mathbb{P}(A) - \mathbb{Q}(A) = \sum_{\omega \in A} (p(\omega) - q(\omega)).$$

Den grössten Wert rechts erhält man, wenn man über alle $\omega$'s summiert, wo $p(\omega) > q(\omega)$, den kleinsten, wenn man über alle $\omega$'s summiert, wo $p(\omega) < q(\omega)$. Ferner gilt

$$\begin{aligned} 2\,\|\mathbb{P} - \mathbb{Q}\| &= \sum_{p(\omega) > q(\omega)} (p(\omega) - q(\omega)) + \sum_{q(\omega) > p(\omega)} (q(\omega) - p(\omega)), \\ 0 &= \sum_{p(\omega) > q(\omega)} (p(\omega) - q(\omega)) - \sum_{q(\omega) > p(\omega)} (q(\omega) - p(\omega)). \end{aligned}$$

Daher sind die beiden Summen gleich, und die Behauptung folgt. □

Aus Satz 4 folgt sofort, dass der Abstand zwischen der Verteilung $\mathbb{P}_m$ nach $m$ riffle shuffles und der uniformen Verteilung $\mathbb{U}$ gegeben ist durch

$$\|\mathbb{P}_m - \mathbb{U}\| = \frac{1}{2} \sum_{r=1}^{n} A_{n,r} \left| 2^{-nm} \binom{2^m + n - r}{n} - \frac{1}{n!} \right|,$$

wobei $A_{n,r}$ die Anzahl Permutationen von $n$ Elementen mit $r$ Runs bezeichnet und der Binomialkoeffizient gleich Null gesetzt wird, wenn $r > 2^m$. Die Anzahl $A_{n,r}$ kann man rekursiv berechnen mit

$$A_{n,1} = 1, \quad A_{n,r} = r^n - \sum_{j=1}^{r-1} \binom{n + r - j}{n} A_{n,j}.$$

Dies ergibt zwar keine geschlossene Formel, aber man kann $\|\mathbb{P}_m - \mathbb{U}\|$ für $n = 52$ ohne Schwierigkeiten mit dem Computer berechnen, weil man nur $n$ statt $n!$ Summanden hat. Das Resultat ist in Tabelle 1 zu sehen: Nach 7 shuffles ist der Abstand also das erste Mal kleiner als $\frac{1}{2}$.

| $m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $\|\mathbb{P}_m - \mathbb{U}\|$ | 1.000 | 1.000 | 1.000 | 1.000 | 0.924 | 0.614 | 0.334 | 0.167 | 0.085 |

Tabelle 1: Totaler Variationsabstand der Verteilung nach $m$ riffle shuffles von der Gleichverteilung für $n = 52$.

### 3.4 Zusammenhang mit Sortieren

Wenn eine Liste aus $n$ verschiedenen Zahlen gegeben ist, dann will man zum Sortieren diejenige Permutation finden, welche die Zahlen der Liste aufsteigend ordnet. Es gibt viele Algorithmen, welche grosse Listen mit möglichst geringem Aufwand sortieren. Diese werden in der Vorlesung "Algorithmen und Komplexität" besprochen. Oft ist es plausibel anzunehmen, dass die Zahlen in der ursprünglichen Liste zufällig angeordnet sind. Manche Sortieralgorithmen sind unter dieser Annahme besonders effizient. In diesem Fall kann man auf die gegebene Liste auch noch zuerst eine Zufallspermutation anwenden bevor man sortiert. Auf dem Computer sind Zufallspermutationen viel einfacher zu erzeugen als beim physikalischen Kartenmmischen.

Das sogenannte "Radixsort" verwendet Operationen, die riffle shuffles umkehren: Wenn die Zahlen in der Liste im Binärsystem angegeben sind, dann sortiert Radixsort zunächst nach der letzten Ziffer, wobei die Reihenfolge von Zahlen mit gleicher letzter Ziffer unverändert bleibt. Die Zahlen mit letzter Ziffer 0 kommen also auf den linken Stapel, diejenigen mit letzter Ziffer 1 auf den rechten und dann kommt der rechte Stapel unter den linken. Dies ist genau die Inversion eines riffle shuffles, wobei $L$ die Positionen von Zahlen mit letzter Ziffer Null ist. $L$ wird also zufällig, wenn die letzte Ziffer zufällig ist. Im zweiten Schritt wird dann nach der zweitletzten Ziffer sortiert, was wieder eine Inversion eines riffle shuffles ist. Wenn die grösste Zahl $m$ binäre Ziffern hat, dann ist nach spätestens $m$ Iterationen die Liste sortiert. Dies kann man benutzen, um den Abstand zwischen der Verteilung nach $m$ riffle shuffles und der uniformen Verteilung abzuschätzen: Zufällige Wahl der riffle shuffles ist äquivalent zur zufälligen Wahl der zu sortierenden Binärzahlen, und wenn die $n$ Binärzahlen alle verschieden sind, dann hat jede mögliche Anordnung die Wahrscheinlichkeit $1/n!$. Also kommt die Abweichung von der uniformen Verteilung nur dadurch zustande, dass zufällige Binärzahlen gleich sein können. Details zu diesem Argument findet man in Kap. 6 von Aldous (2012).

## 4 Markovketten

### 4.1 Definition und Beispiele

Eine Markovkette beschreibt eine zufällige Bewegung mit kurzem Gedächtnis auf einem Graphen mit gerichteten und gewichteten Kanten. Die Ecken des Graphen nennen wir Zustände. Die gerichteten Kanten beschreiben die möglichen Übergänge und die Gewichte geben die Wahrscheinlichkeiten der Übergänge an. Wenn man auch in $i$ bleiben kann, dann hat der Graph eine Schleife, d.h. eine Kante von $i$ nach $i$.

Wir bezeichnen die Menge der Ecken des Graphen mit $S = \{1, 2, \ldots, n\}$ und die gerichteten Kanten mit $i \to j$. Das Gewicht von $i \to j$ wird mit $\Pi(i, j)$ bezeichnet. Wir setzen $\Pi(i, j) = 0$, wenn es keine Kante von $i$ nach $j$ gibt, so dass $\Pi$ eine $n \times n$ Matrix wird. Aus der Interpretation von $\Pi(i, j)$ als Wahrscheinlichkeit von $i$ nach $j$ zu gehen, folgen sofort

die folgenden Eigenschaften von $\Pi$:

$$\Pi(i,j) \geq 0, \quad \sum_{j=1}^{n} \Pi(i,j) = 1 \quad (i = 1, \ldots, n).$$

Matrizen mit diesen beiden Eigenschaften heissen *stochastisch.*

Das kurze Gedächtnis soll bedeuten, dass es keine Rolle spielt, welche Zustände man vorher besucht hat. Die Übergangswahrscheinlichkeiten hängen nur davon ab, wo man jetzt gerade ist. Wir wählen den Startpunkt $i_0$ ebenfalls zufällig mit Wahrscheinlichkeit $p_0(i_0)$ und machen dann eine beliebige, aber feste Anzahl $T$ von Übergängen. Das heisst, wir betrachten die Menge der Verläufe ("Pfade", "Trajektorien")

$$\Omega = \{(i_0, i_1, \ldots, i_T); 1 \leq i_j \leq n\} = \{1, 2, \ldots, n\}^{T+1}$$

und darauf die Wahrscheinlichkeiten

$$p_{0:T}(i_0, i_1, \ldots, i_T) = p_0(i_0)\Pi(i_0, i_1) \cdots \Pi(i_{T-1}, i_T).$$

Die Rechtfertigung der Produktformel geht wieder über die Interpretation von Wahrscheinlichkeit als approximative relative Häufigkeit bei $N$ Wiederholungen: Der Startpunkt $i_0$ tritt ungefähr bei $Np_0(i_0)$ Wiederholungen auf, und ein Bruchteil $\Pi(i_0, i_1)$ davon hat dann einen Übergang in den Zustand $i_1$, etc.

Man sieht sofort, dass für $t < T$

$$p_{0:t}(i_0, i_1, \ldots, i_t) = \sum_{i_{t+1}=1}^{n} \cdots \sum_{i_T=1}^{n} p_{0:T}(i_0, i_1, \ldots, i_T),$$

d.h. der Anfang einer Markovkette bildet wieder eine Markovkette mit den gleichen Start- und den Übergangswahrscheinlichkeiten. In diesem Sinne ist die Wahl von $T$ nicht wesentlich. Die Konstruktion einer Markovkette kann schrittweise erfolgen: Man wählt zuerst $i_0$ genäss $\mathbb{P}_0$, dann $i_1$ gemäss der $i_0$-ten Zeile von $\Pi$, etc. Man muss also $T$ nicht im voraus festlegen.

Beispiele von Markovketten:

- Irrfahrt: Die Bewegung geht mit Wahrscheinlichkeit $p$ um 1 nach rechts, mit Wahrscheinlichkeit $q$ um 1 nach links und mit Wahrscheinlichkeit $r = 1 - p - q$ bleibt man am gleichen Ort. Der Rand kann absorbierend sein, d.h. $\Pi(1,1) = \Pi(n,n) = 1$, oder reflektierend, d.h. $\Pi(1,1) = q + r$ und $\Pi(n,n) = p + r$. Dies kann den Vermögensverlauf in einem Zwei-Personen Nullsummenspiel beschreiben, und Absorption entspricht dem Ruin von einem der beiden Spieler.

- Soziale Mobilität: Eine soziologische Studie hat die berufliche Mobilität in Grossbritannien nach dem zweiten Weltkrieg untersucht. Dazu wurden die Berufe in 3 Stufen (hoch-, mittel- und wenig qualifiziert) eingeteilt, und es wird angenommen, dass die Qualifikation der Kinder nur von der Qualifikation der Eltern abhängt (und nicht von der der Grosseltern etc.). Die Übergangswahrscheinlichkeiten von einer Generation zur nächsten wurden wie folgt geschätzt:

$$\Pi = \begin{pmatrix} 0.45 & 0.48 & 0.07 \\ 0.05 & 0.70 & 0.25 \\ 0.01 & 0.50 & 0.49 \end{pmatrix}.$$

- Kartenmischen: Der Zustand beschreibt die momentane Anordnung der nummerierten Karten im Stapel, ist also eine Permutation $\omega$ von $n$ Elementen. Es gibt eine gerichtete Kante von $\omega$ nach $\omega'$, wenn $\omega'$ aus $\omega$ durch Anwendung eines riffle shuffles hervorgeht. Gemäss dem vorigen Kapitel gilt $\Pi(\omega, \omega) = (n+1)2^{-n}$ und für $\omega' \neq \omega$ $\Pi(\omega, \omega') = 2^{-n}$, bzw. 0.

- DNA: DNA ist der Träger der Erbinformation und besteht aus einer Kette von Basenpaaren, wobei die 4 möglichen Basen mit $A$, $C$, $G$ und $T$ abgekürzt werden. Da immer $A$ mit $T$ und $C$ mit $G$ gepaart sind, genügt es, eine Sequenz von Basen anzugeben. Das einfachste Modell für eine solche Sequenz ist der Pfad einer Markovkette. Experimentell bestimmte Übergangswahrscheinlichkeiten lauten

  | | $A$ | $C$ | $G$ | $T$ |
  |---|---|---|---|---|
  | $A$ | 0.300 | 0.205 | 0.285 | 0.210 |
  | $C$ | 0.322 | 0.298 | 0.078 | 0.302 |
  | $G$ | 0.248 | 0.246 | 0.298 | 0.208 |
  | $T$ | 0.177 | 0.239 | 0.292 | 0.292 |

  In diesem Beispiel ist es von Interesse, Regionen zu finden, wo die Sequenz sich nicht gemäss diesem Modell verhält.

- Das Ehrenfest-Modell: Ein Gefäss enthält ca. $n = 10^{23}$ Gas-Moleküle die alle in Bewegung sind. Wir beschreiben dieses System durch die Anzahl Moleküle, die zur Zeit $t$ in der linken Hälfte des Gefässes sind. Das Modell von P. und T. Ehrenfest nimmt an, dass in einem Zeitschritt ein Molekül zufällig ausgewählt wird und dieses dann mit Wahrscheinlichkeit $p$ in die andere Hälfte übergeht. Dies entspricht einer Markovkette mit den Zuständen $\{0, 1, \ldots, n\}$ und den Übergangswahrscheinlichkeiten $\Pi(i, i+1) = p(1 - \frac{i}{n})$, $\Pi(i, i-1) = p\frac{i}{n}$, $\Pi(i, i) = 1 - p$ und $\Pi(i, j) = 0$ sonst.

- Surfen auf dem Internet: Angenommen, eine Person surft auf dem Internet und sie wählt auf der gerade besuchten Webseite einen der dort vorhandenen Links zufällig aus, um auf die nächste Seite zu kommen. Das ergibt eine Markovkette auf der Menge aller Webseiten mit den Übergängen $\Pi(i, j) = 1/d(i)$ falls ein Link von Seite $i$ zu Seite $j$ existiert, wobei $d(i)$ die Menge aller Links auf Seite $i$ ist.(Für eine vollständige Definition müssten wir noch festlegen, was im Fall $d(i) = 0$ geschieht).

Wir beschränken uns hier auf Modelle, bei denen die Übergangswahrscheinlichkeiten nicht von der Zeit abhängen, sogenannte homogene Markovketten. Im Beispiel der DNA haben wir gesehen, dass der zeit-inhomogene Fall auch von Interesse ist. Auch im Beispiel der sozialen Mobilität wird es wohl so sein, dass die Übergangswahrscheinlichkeiten nicht während mehrerer Generationen unverändert bleiben.

## 4.2 Übergänge in mehreren Schritten

Oft ist man insbesondere an der Wahrscheinlichkeit interessiert, dass die Markovkette zu einer Zeit $0 < t \leq T$ im Zustand $j$ ist. Wir bezeichnen diese Wahrscheinlichkeit mit $p_t(j)$. Sie lässt sich leicht berechnen unter Benutzung der Additivität der Wahrscheinlichkeit, die wir schon früher benutzt hatten und die sofort aus der entsprechenden Eigenschaft von

relativen Häufigkeiten folgt:

$$\begin{aligned} p_t(j) &= \sum_{i_0=1}^{n} \cdots \sum_{i_{t-1}=1}^{n} \sum_{i_{t+1}=1}^{n}, \cdots \sum_{i_T=1}^{n} p_{0:T}(i_0, \ldots, i_{t-1}, j, i_{t+1}, \ldots, i_T) \\ &= \sum_{i_0=1}^{n} \cdots \sum_{i_{t-1}=1}^{n} p_0(i_0)\Pi(i_0, i_1) \cdots \Pi(i_{t-1}, j) \sum_{i_{t+1}=1}^{n} \Pi(j, i_{t+1}) \cdots \sum_{i_T=1}^{n} \Pi(i_{T-1}, i_T) \\ &= \sum_{i_0=1}^{n} p_0(i_0)\Pi^t(i_0, j), \end{aligned}$$

wobei $\Pi^t$ die $t$-te Matrixpotenz von $\Pi$ bezeichnet. Dies besagt insbesondere, dass $\Pi^t(i,j)$ die Wahrscheinlichkeit angibt, in $t$ Schritten von $i$ nach $j$ zu gelangen (über beliebige Zustände zu Zeiten $s \leq t$). Für das Folgende fassen wir Wahrscheinlichkeiten auf der Menge der Zustände als Zeilenvektoren im $\mathbb{R}^n$ auf. Dann lautet obige Gleichung in Matrixform

$$\mathbb{P}_0\Pi^t = \mathbb{P}_t.$$

Wenn die Anzahl Zustände $n$ klein ist, kann man die Matrixpotenzen numerisch leicht berechnen. Im Beispiel der sozialen Mobilität erhalten wir

$$\Pi^2 = \begin{pmatrix} 0.227 & 0.587 & 0.186 \\ 0.060 & 0.639 & 0.301 \\ 0.034 & 0.600 & 0.366 \end{pmatrix}, \Pi^4 = \begin{pmatrix} 0.093 & 0.620 & 0.287 \\ 0.060 & 0.639 & 0.301 \\ 0.056 & 0.623 & 0.321 \end{pmatrix}, \Pi^8 = \begin{pmatrix} 0.064 & 0.623 & 0.313 \\ 0.062 & 0.624 & 0.314 \\ 0.062 & 0.624 & 0.314 \end{pmatrix}.$$

Die Zeilen von $\Pi^8$ sind fast gleich. Das bedeutet, dass der Einfluss des Anfangszustandes nach 8 Schritten praktisch verschwunden ist. Wir werden dieses Phänomen in Kapitel 4.5 noch genauer untersuchen.

Oft interessiert man sich für den Zeitpunkt, an dem man das erste Mal einen bestimmten Zustand $i$ erreicht, der natürlich auch vom Zufall abhängt. Information darüber erhält man, indem man die Übergänge so modifiziert, dass der Zustand $i$ nicht mehr verlassen wird: $\tilde{\Pi}(i,i) = 1, \tilde{\Pi}(i,j) = 0, \tilde{\Pi}(j,k) = \Pi(j,k)$ für $j \neq i$. Dann ist $\tilde{\Pi}^t(j,i)$ die Wahrscheinlichkeit, dass die ursprüngliche Kette bei Start in $j$ den Zustand $i$ bis zur Zeit $t$ einmal besucht. Im Beispiel der sozialen Mobilität betrachten wir $i = 1$ und $t = 5$. Dann ist

$$\tilde{\Pi}^5 = \begin{pmatrix} 1 & 0 & 0 \\ 0.184 & 0.534 & 0.282 \\ 0.139 & 0.563 & 0.298 \end{pmatrix}.$$

Die Wahrscheinlichkeit, in maximal 5 Generationen von der tiefsten in die höchste Schicht aufzusteigen, ist also etwa 14%.

## 4.3 Klassifikation der Zustände

Um das Verhalten einer Markovkette bei vielen Übergängen zu verstehen, sind die folgenden Definitionen nützlich

**Definition 2** *Zustand $j$ ist* zugänglich *von Zustand $i$, falls $\Pi^t(i,j) > 0$ für ein $t \geq 0$. Die Zustände $i$ und $j$ sind* verbunden*, falls $i$ von $j$ und $j$ von $i$ zugänglich sind. Eine Teilmenge $C \subseteq \{1, \ldots, n\}$ heisst* abgeschlossen*, falls von allen $i \in C$ aus nur Zustände in $C$ zugänglich sind.*

Diese Eigenschaften lassen sich leicht verifizieren durch Betrachtung des zugehörigen gerichteten Graphen. Man sieht sofort, dass Verbundenheit von Zuständen eine Äquivalenzrelation ist und dass der Zustandsraum daher in Äquivalenzklassen von verbundenen Zuständen zerfällt.

Wenn eine Äquivalenzklasse $C$ abgeschlossen ist, dann nennen wir die Zustände in dieser Klasse *rekurrent*. Wenn man in einem Zustand $i \in C$ startet, dann bleibt die Markovkette immer in $C$, die Einschränkung von $\Pi$ auf $C \times C$ ist eine stochastische Matrix.

Zustände in Äquivalenzklassen, die nicht abgeschlossen sind, heissen *transient*. Mit einem Widerspruchsbeweis kann man leicht zeigen, dass es mindestens eine abgeschlossene Äquivalenzklasse gibt. Wenn man die Zustände so umnummeriert, dass die ersten $k$ Zustände transient sind und die Zustände in den $m$ abgeschlossenen Äquivalenzklassen konsekutiv nummeriert sind, dann erhält man für die Übergangsmatrix folgende kanonische Form

$$\Pi = \begin{pmatrix} Q & R_1 & R_2 & \dots & R_m \\ 0 & P_1 & 0 & \dots & 0 \\ 0 & 0 & P_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & P_m \end{pmatrix}.$$

Betrachten wir als Beispiel die Irrfahrt von oben mit Absorption am Rand. Diese Markovkette hat 3 Äquivalenzklassen: $\{1\}$, $\{n\}$ und $\{2, \dots, n-1\}$. Die ersten zwei Klassen sind rekurrent, die dritte transient. Beim Surfen auf dem Internet hängt die Klassifikation davon ab, wie der Übergang von Seiten mit $d(i) = 0$ definiert ist. Die anderen Beispiele von oben haben nur eine Äquivalenzklasse, die demnach rekurrent ist.

Wie verhält sich die Kette, wenn man in einem transienten Zustand $i$ startet ? Wenn man die zugehörige Äquivalenzklasse einnmal verlassen hat, dann folgt sofort aus den Definitionen, dass man nicht mehr in diese Klasse zurückkehren kann. Wir zeigen als nächstes, dass man die zugehörige Äquivalenzklasse früher oder später verlässt und in einer rekurrenten Klasse endet. Ausserdem berechnen wir die Wahrscheinlichkeit dass der erste besuchte rekurrente Zustand gleich $j$ ist. Wenn $\{1, 2, \dots, k\}$ die Menge der transienten Zustände ist, dann definieren wir für $i \leq k$ und $j > k$ das Ereignis

$$U_{ij}^{(T)} = \cup_{t=1}^{T} \{(i, i_1, \dots, i_{t-1}, j, i_{t+1}, \dots, i_T); i_1 \leq k, \dots, i_{t-1} \leq k\}.$$

In Worten, man startet in $i$, der erste rekurrente Zustand ist gleich $j$, und er wird nach höchstens $T$ Schritten besucht. Man sieht sofort, dass $\mathbb{P}_{0:T}(U_{ij}^{(T)})$ monoton wachsend ist in $T$, und damit existiert $u_{ij} = \lim_{T\to\infty} \mathbb{P}_{0:T}(U_{ij}^{(T)})/p_0(i)$. Anschaulich ist $u_{ij}$ die Wahrscheinlichkeit, dass bei Start in $i$ der erste rekurrente Zustand gleich $j$ ist, aber dies würde voraussetzen, dass wir unendlich viele Übergänge betrachten, und dann ist der Raum $\Omega$ nicht mehr endlich, sondern sogar überabzählbar.

**Satz 5** *Wir nehmen an, dass die Übergangsmatrix die kanonische Form von oben hat und bezeichnen die Matrix mit Blöcken $R_1, \dots, R_m$ mit $R$. Dann geht $Q^t$ exponentiell schnell gegen Null, $I - Q$ ist invertierbar mit*

$$(I - Q)^{-1} = \sum_{t=0}^{\infty} Q^t$$

*und für die Matrix* $U = (u_{ij}; 1 \le i \le k, k+1 \le j \le n)$ *gilt*

$$U = (I - Q)^{-1} R = \sum_{t=0}^{\infty} Q^t R.$$

**Beweis:** Zu jedem transienten Zustand $i$ gibt es einen Pfad, der in endlich vielen Schritten einen rekurrenten Zustand erreicht und positive Wahrscheinlichkeit hat. Weil man von einem rekurrenten Zustand nicht zu einem transienten Zustand zurückkehren kann, heisst das: Es gibt ein $\varepsilon > 0$ und ein $s$ so dass

$$\min_{i \le k} \sum_{j=k+1}^{n} \Pi^s(i,j) \ge \varepsilon \Leftrightarrow \max_{i \le k} \sum_{j=1}^{k} Q^s(i,j) \le 1 - \varepsilon.$$

Mit Induktion folgt daraus, dass

$$\min_{i \le k} \sum_{j=1}^{k} Q^t(i,j) \le (1-\varepsilon)^{\lfloor t/s \rfloor},$$

woraus die Konvergenz von $\sum_{t=0}^{\infty} Q^t$ folgt. Ausserdem gilt

$$(I - Q) \sum_{t=0}^{n} Q^t = I - Q^n.$$

Für $n \to \infty$ konvergiert die rechte Seite gegen $I$ und die linke gegen $(I - Q) \sum_{t=0}^{\infty} Q^t$. Damit ist der erste Teil bewiesen. Für den zweiten Teil modifizieren wir die Übergänge so, dass alle rekurrenten Zustände absorbierend sind, d.h.

$$\tilde{\Pi} = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}.$$

Dann ist

$$\tilde{\Pi}^T = \begin{pmatrix} Q^T & (I + Q + \ldots Q^{T-1})R \\ 0 & I \end{pmatrix},$$

und daraus folgt

$$\mathbb{P}_T(U_{ij}^{(T)}) = p_0(i) \tilde{\Pi}^T(i,j) = p_0(i)((I + Q + \ldots + Q^{T-1})R)_{ij}.$$

Wenn wir beide Seiten durch $p_0(i)$ dividieren und $T$ gegen unendlich gehen lassen, erhalten wir somit $U = (I - Q)^{-1} R$. □

Dass $U$ die Gleichung $U = R + QU$ erfüllt, ist auch anschaulich sofort klar: $R$ enthält die Wahrscheinlichkeiten, in einem Schritt in einen rekurrenten Zustand zu kommen, und $QU$ die Wahrscheinlichkeiten, zuerst in einen anderen transienten Zustand und von dort später in einen rekurrenten Zustand zu gelangen. Ausserdem ist $\sum_{j=1}^{k} Q^t(i,j)$ die Wahrscheinlichkeit bei Start in $i$ nach $t$ Schritten immer noch in einem transienten Zustand zu sein. Diese Wahrscheinlichkeit geht also exponentiell schnell gegen 0.

Als Beispiel betrachten wir die Irrfahrt mit Absorption am Rand. Die Wahrscheinlichkeit $u_i := u_{i1}$, dass die Irrfahrt bei Start in $i$ am linken Rand absorbiert wird, erfüllt die Gleichung

$$\begin{aligned} u_i &= p u_{i+1} + q u_{i-1} + (1-p-q) u_i \quad (i = 3, \ldots, n-2) \\ u_2 &= p u_3 + (1-p-q) u_2 + q \\ u_{n-1} &= q u_{n-2} + (1-p-q) u_{n-1}. \end{aligned}$$

Dieses Gleichungssystem lässt sich leicht lösen. Die Wahrscheinlichkeit für Absorption am rechten Rand ist schliesslich gleich $1 - u_i$.

## 4.4 Konvergenz gegen die stationäre Verteilung

Wir betrachten hier den Fall, wo alle Zustände verbunden sind. Wir nennen die Kette in diesem Fall *irreduzibel*. Für die Untersuchung des Verhaltens der Kette bei vielen Übergängen brauchen wir noch das Konzept der *Periode* eines Zustandes.

**Definition 3** *Die Periode $d(i)$ des Zustandes $i$ ist der grösste gemeinsame Teiler von $\{t \geq 1; \Pi^t(i,i) > 0\}$. Der Zustand $i$ heisst* aperiodisch *wenn $d(i) = 1$. Die Kette heisst aperiodisch, wenn alle Zustände aperiodisch sind.*

Das Ehrenfest-Modell ist irreduzibel. Wenn $p = 1$, dann haben alle Zustände die Periode 2. Für $0 < p < 1$ ist die Kette aperiodisch.

Im Beispiel der sozialen Mobilität haben wir empirisch gesehen, dass $\Pi^t$ gegen eine Matrix mit identischen Zeilen zu konvergieren scheint. Dies gilt für jede irreduzible und aperiodische Kette, wie wir nun zeigen werden. Wir benutzen den totalen Variationsabstand aus dem letzten Kapitel und beweisen zuerst eine weitere Eigenschaft dieses Abstandes.

**Lemma 4** *Seien $\mathbb{P}$ und $\mathbb{Q}$ zwei Wahrscheinlichkeiten auf $\Omega$ und $f : \Omega \to \mathbb{R}$. Dann gilt*

$$\left| \sum_{\omega \in \Omega} f(\omega)(p(\omega) - q(\omega)) \right| \leq \|\mathbb{P} - \mathbb{Q}\| \, \Delta f$$

*wobei $\Delta f = \max f(\omega) - \min f(\omega)$.*

**Beweis:** Für beliebiges $c \in \mathbb{R}$ gilt wegen $\sum_\omega p(\omega) = \sum_\omega q(\omega)$

$$\left| \sum_{\omega \in \Omega} f(\omega)(p(\omega) - q(\omega)) \right| = \left| \sum_{\omega \in \Omega} (f(\omega) - c)(p(\omega) - q(\omega)) \right| \leq \max_\omega |f(\omega) - c| \sum_{\omega \in \Omega} |p(\omega) - q(\omega)|.$$

Wählt man $c = \frac{1}{2}(\max f(\omega) + \min f(\omega))$, folgt die Behauptung. □

Damit können wir jetzt eine Kontraktionseigenschaft von einem Markovübergang zeigen.

**Lemma 5** *Seien $\mathbb{P}$ und $\mathbb{Q}$ zwei Wahrscheinlichkeiten auf $\{1, \ldots, n\}$ und sei $\Pi$ eine stochastische Matrix. Dann gilt*

$$\|\mathbb{P}\Pi - \mathbb{Q}\Pi\| \leq \rho(\Pi) \, \|\mathbb{P} - \mathbb{Q}\| \,,$$

*wobei $\rho(\Pi) = \max_{i \neq j} \|\Pi(i, .) - \Pi(j, .)\|$.*

**Beweis:** Für ein beliebiges $A \subset \{1, \ldots, n\}$ setzen wir $f(i) := \sum_{k \in A} \Pi(i, k)$. Nach Lemma 3 ist dann $\Delta f \leq \rho(\Pi)$. Also folgt nach Lemma 4

$$|\mathbb{P}\Pi(A) - \mathbb{Q}\Pi(A)| = \left| \sum_{i=1}^{n} (p(i) - q(i)) \sum_{k \in A} \Pi(i, k) \right| = \left| \sum_{i=1}^{n} f(i)(p(i) - q(i)) \right| \leq \rho(\Pi) \, \|\mathbb{P} - \mathbb{Q}\| \,.$$

Mit einer weiteren Anwendung von Lemma 3 folgt die Behauptung. □

**Satz 6** *Sei $\Pi$ die Übergangsmatrix einer irreduziblen und aperiodischen Markovkette. Dann existiert genau eine Wahrscheinlichkeit $\gamma$ auf $\{1,\ldots,n\}$ mit $\gamma\Pi=\gamma$. Zudem konvergiert für jede Startverteilung $\mathbb{P}_0$ die Folge der Verteilungen $\mathbb{P}_t=\mathbb{P}_0\Pi^t$ bezüglich des totalen Variationsabstandes gegen $\gamma$. Insbesondere konvergiert $\Pi^t$ gegen die Matrix $\Gamma$, deren Zeilen alle gleich $\gamma$ sind.*

Die Eigenschaft $\gamma\Pi=\gamma$ bedeutet, dass bei Start gemäss $\gamma$ die Verteilung des Zustands zu jeder Zeit gleich $\gamma$ ist. Daher nennt man diese Verteilung *stationär*. Aus der Sicht der linearen Algebra besagt $\gamma\Pi=\gamma$, dass $\gamma'$ ein Eigenvektor zum Eigenwert 1 von $\Pi'$ ist. Dass $\Pi'$ den Eigenwert 1 hat, folgt sofort daraus, dass $\sum_j \Pi(i,j)=1$, denn dies heisst nichts anderes als dass $(1,\ldots,1)'$ Eigenvektor von $\Pi$ zum Eigenwert 1 ist.

**Beweis:** Wir nehmen zuerst an, dass $\varepsilon:=\min_{i,j}\Pi(i,j)>0$. Dann ist wegen Lemma 3

$$\|\Pi(i,.)-\Pi(j,.)\|=\sum_{k;\Pi(i,k)>\Pi(j,k)}(\Pi(i,k)-\Pi(j,k))=\sum_{k=1}^{n}(\Pi(i,k)-\min(\Pi(i,k),\Pi(j,k)))\leq 1-n\varepsilon$$

und somit folgt mit Lemma 5 für beliebige Startverteilungen $\mathbb{P}_0$ und $\mathbb{Q}_0$

$$\|\mathbb{P}_t-\mathbb{Q}_t\|\leq(1-n\varepsilon)^t\,\|\mathbb{P}_0-\mathbb{Q}_0\|\to 0.$$

Wir fixieren eine beliebige Startverteilung $\mathbb{P}_0$. Weil die Menge aller Wahrscheinlichkeiten eine kompakte Teilmenge des $\mathbb{R}^n$ ist, existiert eine Wahrscheinlichkeit $\gamma$ und eine Teilfolge $\mathbb{P}_{t_k}$ so dass $\mathbb{P}_{t_k}\to\gamma$. Wenn wir $\mathbb{Q}_0=\mathbb{P}_1$ wählen, dann ist $\mathbb{Q}_{t_k}=\mathbb{P}_{t_k}\Pi$ und damit folgt

$$0=\lim_k\|\mathbb{P}_{t_k}-\mathbb{Q}_{t_k}\|=\lim_k\|\mathbb{P}_{t_k}-\mathbb{P}_{t_k}\Pi\|=\|\gamma-\gamma\Pi\|.$$

Also ist $\gamma\Pi=\gamma$. Wenn wir jetzt $\mathbb{Q}_0=\gamma$ wählen, dann ist $\mathbb{Q}_t=\gamma$ für alle $t$ und damit geht $\|\mathbb{P}_t-\gamma\|$ gegen Null, d.h. man hat Konvergenz auch ohne Bildung einer Teilfolge. Mit dem gleichen Argument folgt auch die Konvergenz für beliebiges $\mathbb{P}_0$. Konzentriert man die Startverteilung auf den Zustand $i$, dann ist $\mathbb{P}_t=\Pi^t(i,.)$, und somit folgt die Konvergenz von $\Pi^t$ gegen $\Gamma$.

Der allgemeine Fall geht analog unter Verwendung von Lemma 6 unten. □

**Lemma 6** *Wenn $\Pi$ die Übergangsmatrix einer irreduziblen und aperiodischen Kette ist, dann existiert ein $t>0$ so, dass alle Elemente von $\Pi^t$ strikt positiv sind.*

**Beweis:** Sei $N_i=\{t;\Pi^t(i,i)>0\}$ mit $i$ beliebig. Dann ist $N_i$ abgeschlossen unter Addition, weil

$$\Pi^{t+s}(i,i)=\sum_{j=1}^{n}\Pi^t(i,j)\Pi^s(j,i)\geq\Pi^t(i,i)\Pi^s(i,i).$$

Nach Voraussetzung existieren $s,t\in N_i$ mit 1 als grössten gemeinsamen Teiler. Gemäss einem Satz aus der elementaren Zahlentheorie existieren dann $a,b\in\mathbb{N}$ mit $as-bt=1$. Dann ist für jedes $j\geq b(t-1)$ und jedes $k$ mit $0\leq k<t$

$$jt+k=jt+k(as-bt)=(ka)s+(j-kb)t\in N_i,$$

d.h. $N_i$ enthält alle natürlichen Zahlen grösser oder gleich $b(t-1)t$. Damit gibt es für jedes $i,j$ ein $t_{ij}$ mit $\Pi^t(i,j)>0$ für alle $t\geq t_{ij}$. Jetzt nimmt man das Maximum aller $t_{ij}$. □

Des Beweis dieses Lemmas zeigt auch, dass eine irreduzible Markovkette aperiodisch ist, wenn ein Zustand aperiodisch ist.

Die Berechnung der stationären Verteilung erfordert also die Lösung des Gleichungssystems $(I - \Pi')x = 0$ unter der zusätzlichen Bedingung $(1, 1, \ldots, 1)x = 1$. Wenn die Anzahl Zustände gross ist, kann das schwierig sein. Im Beispiel des Kartenmischens vermutet man, dass die uniforme Veretilung stationär ist. Dies lässt sich leicht verifizieren, denn bei der Matrix $\Pi(\omega, \omega')$ sind nicht nur die Zeilensummen sondern auch die Spaltensummen gleich eins: $\Pi(\omega, \omega') = 2^{-n}$, wenn $\omega' = \xi \circ \omega$ und $\xi$ eine Permutation mit genau einem run ist, und alle anderen Elemente ausserhalb der Diagonalen sind gleich Null. Weil $\omega' = \xi \circ \omega$ genau dann der Fall ist, wenn $\omega = \xi^{-1} \circ \omega'$, hat man in jeder Zeile und in jeder Spalte gleich viele Elemente, welche gleich $2^{-n}$ sind. Ferner sind alle Elemente von $\Pi^k$ strikt positiv wenn $2^k > n$, vgl. Satz 4, also konvergiert die Verteilung nach $t$ riffle shuffles gegen die uniforme Verteilung für $t \to \infty$. Die Schranken, die man aus Satz 6 erhält, sind jedoch sehr viel schlechter als diejenigen, die wir in Abschnitt 3.3 gefunden haben.

Surfen auf dem Internet und page rank. Wir nummerieren die Webseiten so, dass die Anzahl Links $d(i)$ auf Seite $i > 0$ ist für $i \leq k$ und $= 0$ für $k < i \leq n$. Dann modifizieren wir den oben beschriebenen Übergang und setzen

$$\Pi(i,j) = \begin{cases} (1-r)\frac{1}{n} + \frac{r}{d(i)} & (i \leq k; i \to j) \\ (1-r)\frac{1}{n} & (i \leq k; i \not\to j) \\ \frac{1}{n} & (i > k) \end{cases}$$

Das heisst, wenn man auf einer Seite ohne weiterführende Links ist, beginnt man neu auf einer zufälligen Seite, und wenn man auf einer anderen Seite ist, beginnt man neu mit Wahrscheinlichkeit $1 - r$. Dann erfüllt die stationäre Verteilung $\gamma$ die Gleichung

$$\gamma(j) = \frac{1}{n} \sum_{i=k+1}^{n} \gamma(i) + \frac{1-r}{n} \sum_{i=1}^{k} \gamma(i) + r \sum_{i;i\to j} \frac{\gamma(i)}{d(i)},$$

und der Rang von Seite $i$ ist nun proportional zu $\gamma(i)$. Dieses Gleichungssystem lässt sich vereinfachen: Setzt man $\gamma^*(k+1) = \sum_{i=k+1}^{n} \gamma(i)$ und $e(i) =$ Anzahl Links von Seite $i$ auf Seiten $j > k$, dann kann man schreiben

$$\begin{aligned} \gamma(j) &= \frac{1}{n}\gamma^*(k+1) + \frac{1-r}{n} \sum_{i=1}^{k} \gamma(i) + r \sum_{i\leq k; i\to j} \frac{\gamma(i)}{d(i)} \quad (j \leq k) \\ \gamma^*(k+1) &= \frac{n-k}{n}\gamma^*(k+1) + \sum_{i=1}^{k} \gamma(i) \left( (1-r)\frac{n-k}{n} + r\frac{e(i)}{d(i)} \right). \end{aligned}$$

Das heisst, man kann zuerst ein Gleichungssystem mit $k+1$ statt $n$ Unbekannten lösen, und nachher eines mit $n-k$ Unbekannten. Das obige Gleichungssystem bestimmt die stationäre Verteilung für die reduzierte Kette, bei der alle Zustände mit $d(i) = 0$ zu einem einzigen Zustand zusammengefasst wurden. Dies ist möglich, weil die Zeilen von $\Pi$ für alle diese Zustände gleich sind. Ist das nicht der Fall, dann zerstört eine solche Zusammenfassung von Zuständen die Markoveigenschaft.

Auch das kleinere Gleichungssystem ist jedoch noch sehr gross, und es ist effizienter, rekursiv $\frac{1}{k+1}(1, 1, \ldots, 1)(\Pi^*)^t$ für $t = 1, 2, \ldots, T$ zu berechnen, wobei $\Pi^*$ die Übergangsmatrix für die reduzierte Kette ist. Offenbar ist man nach ca. $T = 100$ Schritten bereits genügend nahe bei der Lösung.

In einigen Anwendungen hat die stationäre Verteilung die zusätzliche Eigenschaft der Reversibilität. Dann kann man die stationäre Verteilung in geschlossener Form angeben.

**Definition 4** *Eine Wahrscheinlichkeit $\gamma$ auf $\{1, 2, \ldots, n\}$ heisst* reversibel *für die Übergangsmatrix $\Pi$, falls für alle $i \neq j$*

$$\gamma(i)\Pi(i,j) = \gamma(j)\Pi(j,i).$$

**Satz 7** *Jede reversible Verteilung ist stationär, aber die Umkehrung gilt nicht. Wenn die Startverteilung reversibel ist, dann gilt für alle $T$ und alle Verläufe mit $T$ Übergängen*

$$p_{0:T}(i_0, i_1, \ldots, i_{T-1}, i_T) = p_{0:T}(i_T, i_{T-1}, \ldots, i_1, i_0).$$

*Eine irreduzible aperiodische Kette hat genau dann eine reversible Verteilung, wenn für alle Zyklen $(i_0, i_1, \ldots, i_{T-1}, i_0)$ gilt*

$$\Pi(i_0, i_1)\Pi(i_1, i_2)\cdots\Pi(i_{T-1}, i_0) = \Pi(i_0, i_{T-1})\Pi(i_{T-1}, i_{T-2})\cdots\Pi(i_1, i_0),$$

*und im Beweis ist eine geschlossene Formel für die reversible Verteilung angegeben.*

**Beweis:** Summation der Gleichungen in der Definition über $i$ ergibt

$$\sum_{i=1}^{n} \gamma(i)\Pi(i,j) = \gamma(j)\sum_{i=1}^{n} \Pi(j,i) = \gamma(j).$$

Reversibilität ergibt ein lineares homogenes Gleichungssystem mit $n(n-1)/2$ Gleichungen und $n$ Unbekannten. Es hat im Allgemeinen für $n > 2$ keine Lösung. Die Konstruktion eines Gegenbeispiels für $n = 3$ ist einfach. Für $T = 1$ ist $p_{0:1}(i_0, i_1) = \gamma(i_0)\Pi(i_0, i_1)$, also ist die zweite Behauptung dann gerade gleich der definierenden Eigenschaft. Für beliebiges $T$ wird die definierende Eigenschaft $T$ mal angewendet. Die letzte Gleichung ist offensichtlich ein Spezialfall der vorangehenden, also ist sie sicher notwendig. Um zu sehen, dass sie auch hinreichend ist, wählen wir $T$ so, dass $\Pi^T(i,j) > 0$ für alle $i, j$ und setzen

$$\gamma(j) = \text{const.}\frac{\Pi(1, j_1)\cdots\Pi(j_{T-1}, j)}{\Pi(j, j_{T-1})\cdots\Pi(j_1, 1)},$$

wobei wir den Pfad von 1 nach $j$ so gewählt haben, dass der Nenner strikt positiv ist. Dann ist $\gamma(i)\Pi(i,j) = \gamma(j)\Pi(j,i)$ äquivalent zu

$$\frac{\Pi(1, i_1)\cdots\Pi(i_{T-1}, i)\Pi(i,j)}{\Pi(i, i_{T-1})\cdots\Pi(i_1, 1)} = \frac{\Pi(1, j_1)\cdots\Pi(j_{T-1}, j)\Pi(j,i)}{\Pi(j, j_{T-1})\cdots\Pi(j_1, 1)},$$

was gemäss Voraussetzung richtig ist. □

Wenn nur Übergänge von $i$ nach $i \pm 1$ möglich sind, dann existiert immer eine reversible Verteilung, weil jeder geschlossene Zyklus entweder keinen oder beide Übergänge $i \to i+1$ und $i+1 \to i$ enthalten muss. Die reversible Verteilung ist in dem Fall gegeben durch

$$\gamma(i) = \text{const.}\frac{\Pi(i-1, i)}{\Pi(i, i-1)}\cdots\frac{\Pi(1,2)}{\Pi(2,1)}.$$

Im Fall der Irrfahrt mit Reflexion am Rand erhalten wir für $p = q$ die uniforme Verteilung, und für $p \neq q$

$$\gamma(i) = \left(\frac{p}{q}\right)^{i-1}\frac{1 - \frac{p}{q}}{1 - (\frac{p}{q})^n}.$$

Im Ehrenfest-Modell erhalten wir

$$\gamma(i) = \text{const.}\frac{n-(i-1)}{i}\cdots\frac{n}{1} = \text{const.}\binom{n}{i} = 2^{-n}\binom{n}{i}.$$

Aus der Stirling-Formel $n! \sim (n/e)^n\sqrt{2\pi n}$ folgt

$$\frac{i}{n} \sim \frac{1}{2} \Rightarrow \frac{\gamma(i)}{\gamma(0)} \sim \frac{2^n}{\sqrt{\pi n}}$$

was für $n = 10^{23}$ ungefähr gleich $10^{3\cdot 10^{22}}$ ist. Es ist daher praktisch sicher, dass sich nie alle Moleküle in der einen Hälfte befinden.

## 4.5 Ausblick

Eine irreduzible und aperiodische Markovkette, welche mit der stationären Verteilung $\gamma$ startet, erfüllt

$$\sum_{j=1}^{n} p_{0:T}(i_0, i_1, \ldots, i_{T-1}, j) = \sum_{j=1}^{n} p_{0:T}(j, i_0, i_1, \ldots, i_{T-1}) = p_{0:T-1}(i_0, i_1, \ldots, i_{T-1}),$$

ist also konsistent im Sinne von Satz 3. Wir können auch die Entropie der Verteilung $\mathbb{P}_{0:T}$ leicht berechnen:

$$\begin{aligned} H(\mathbb{P}_{0:T}) &= \sum_{i_0=1}^{n}\cdots\sum_{i_T=1}^{n} p_{0:T}(i_0, i_1, \ldots, i_T)\,(\log_2\gamma(i_0) + \log_2\Pi(i_0, i_1) + \ldots + \log_2\Pi(i_{T-1}, i_T)) \\ &= H(\gamma) + T\sum_{i=1}^{n}\gamma(i)\sum_{j=1}^{n}\Pi(i,j)\log_2\Pi(i,j). \end{aligned}$$

Daraus folgt

$$(T+1)^{-1}H(\mathbb{P}_{0:T}) \to h := \sum_{i=1}^{n}\gamma(i)H(\Pi(i,.)),$$

das heisst, wir haben die erste Aussage von Satz 3 direkt verifiziert. Die zweite Aussage dieses Satzes lautet, dass mit grosser Wahrscheinlichkeit

$$\frac{1}{T}\gamma(i_0) + \frac{1}{T}\sum_{t=1}^{T}\log_2\Pi(i_{t-1}, i_t) \approx h = \mathbb{E}(\log_2\Pi(i_{t-1}, i_t)).$$

Dies ist ein Spezialfall des allgemeinen Gesetzes der grossen Zahlen, welches besagt, dass das arithmetische Mittel mit grosser Wahrscheinlichkeit nahe beim Erwartungswert liegt, sofern dieser für alle Summanden gleich ist. Der Beweis eines solchen Gesetzes der grossen Zahlen erfordert allerdings mehr Hilfsmittel als wir hier zur Verfügung haben.

Gesetze der grossen Zahlen erlauben auch eine weitere Interpretation der stationären Verteilung $\gamma$ bei irreduziblen und aperiodischen Ketten. Die Anzahl Besuche $N_T(i)$ in einem Zustand $i$ während $T$ Perioden lässt sich schreiben als die Summe von Indikatorfunktionen

$$N_T(i) = \sum_{t=0}^{T} 1_{[i_t=i]}.$$

Mit dem Gesetz der grossen Zahlen ergibt sich also $N_T(i) \approx T\gamma(i)$. Andererseits können wir die Zeiten $\tau_k^{(i)}$ zwischen dem $k$-ten und dem $(k+1)$-ten Besuch in $i$ betrachten. Dann ist

$$T \approx \sum_{k=1}^{N_T(i)-1} \tau_k^{(i)} \approx (N_T(i)-1)\mathbb{E}(\tau^{(i)})$$

(wir vernachlässigen die Zeit bis zum ersten Besuch und die Zeit vom letzten Besuch $\leq T$ bis $T$). Kombiniert man diese beiden Approximationen, so folgt

$$\mathbb{E}(\tau^{(i)}) = \frac{1}{\gamma_i},$$

das heisst, die erwartete Rückkehrzeit in einen Zustand ist gleich dem Kehrwert der stationären Wahrscheinlichkeit dieses Zustands. Im Ehrenfest-Modell sagt dies also insbesondere, wie lange man warten muss, bis sich wieder alle Moleküle in der linken Hälfte befinden, wenn man mit diesen Zustand startet.

# 5 Das Ising Modell

## 5.1 Definition des Modells

Das Ising-Modell beschreibt einen Ferromagneten. Es wurde in den 1920er Jahren von Wilhelm Lenz vorgeschlagen und von seinem Studenten Ernst Ising untersucht. Es ist das bekannteste und best-untersuchte Modell der statistischen Mechanik. Es ist relevant, weil es in zwei und mehr Dimensionen einen Phasenübergang aufweist. Was das genau heisst, werden wir hier erklären. Ising selber hat in seiner Dissertation 1924 ein Argument gegen die Existenz von Phasenübergang gegeben, das aber nicht korrekt ist. Peierls hat dann 1936 die richtige Idee gehabt, allerdings war seine Darstellung auch nicht mathematisch korrekt. Dies wurde dann erst 1964 von Griffiths geleistet.

Wir beschreiben ein magnetisches Material als eine Anordnung von Elementarmagneten auf dem Gitter $L_n = \{-n, -n+1, \dots n\}^d$ deren Spin auf- oder abwärts zeigen kann. Für $d > 1$ benutzen wir die Bezeichnung $\mathbf{i} = (i_1, \dots, i_d)$ für ein Element von $L_n$. Die Menge der möglichen Spinkonfigurationen ist also

$$\Omega_n = \{-1, +1\}^{L_n} = \{\omega = (\omega_{\mathbf{i}}; \mathbf{i} \in L_n); \omega_{\mathbf{i}} = \pm 1\}.$$

Wegen der Interaktion zwischen benachbarten Elementarmagneten ist es energetisch vorteilhaft, wenn alle Spins in die gleiche Richtung zeigen, aber dies wird durch die thermodynaischen Fluktuationen gestört. Die Anordnung der Spins ist daher zufällig, und die Wahrscheinlichkeit für eine Konfiguration $\omega$ ist gegeben durch

$$p_{n,\beta}(\omega) = \frac{1}{Z} \exp\left(\beta \sum_{||\mathbf{i}-\mathbf{j}||=1} \omega_{\mathbf{i}}\omega_{\mathbf{j}}\right) \propto \exp\left(\beta \sum_{||\mathbf{i}-\mathbf{j}||=1} \omega_{\mathbf{i}}\omega_{\mathbf{j}}\right).$$

Dabei ist $\beta = 1/T$ der Kehrwert der absoluten Temperatur $T$, $Z$ ist die Normierungskonstante

$$Z = Z_{n,\beta} = \sum_{\omega \in \Omega_n} \exp\left(\beta \sum_{||\mathbf{i}-\mathbf{j}||=1} \omega_{\mathbf{i}}\omega_{\mathbf{j}}\right),$$

und das Proportionalitätszeichen $\propto$ drückt aus, dass die beiden Ausdrücke sich nur um eine Konstante unterscheiden. Zwei benachbarte gleichgerichtete Spins erhöhen also die Wahrscheinlichkeit um einen Faktor $e^{\beta}$, benachbarte entgegengesetzte Spins reduzieren sie um den Faktor $e^{-\beta}$. Wir bezeichnen die Anzahl benachbarter gleicher Spins von $\omega$ mit $N_g(\omega)$ und die Anzahl benachbarter entgegengesetzter Spins mit $N_u(\omega)$. Da die Gesamtzahl benachbarter Spins konstant ist (gleich $d \cdot 2n \cdot (2n+1)^{d-1}$), ist daher

$$p_{n,\beta}(\omega) \propto \exp(-2\beta N_u(\omega)).$$

Insbesondere ist $p_{n,\beta}(\omega) = p_{n,\beta}(-\omega)$ für alle $\omega$ und damit $\mathbb{P}_{n,\beta}(\{\omega_{\mathbf{i}} = 1\}) = \frac{1}{2}$ für alle $\mathbf{i}$.

Wir können nun untersuchen, wie sich die Wahrscheinlichkeiten ändern, wenn wir die Spins am Rand $\partial L_n = \{\mathbf{i} \in L_n; |i_j| = n \text{ für ein } j\}$ von $L_n$ gleich ausrichten, also nur noch Konfigurationen betrachten, wo für alle $\mathbf{i} \in \partial L_n$ $\omega_{\mathbf{i}} = +1$, bzw. $\omega_{\mathbf{i}} = -1$, ist. Wir bezeichnen die entsprechenden Wahrscheinlichkeiten mit $\mathbb{P}^+_{n,\beta}$ und $\mathbb{P}^-_{n,\beta}$ Für jede Teilmenge $A \subseteq \Omega_n$ ist also

$$\mathbb{P}^+_{n,\beta}(A) = \frac{\mathbb{P}_{n,\beta}(A \cap \{\omega; \omega_{\mathbf{i}} = +1 \ \forall \mathbf{i} \in \partial L\})}{\mathbb{P}_{n,\beta}(\{\omega; \omega_{\mathbf{i}} = +1 \ \forall \mathbf{i} \in \partial L\})}.$$

Es ist unmittelbar plausibel (und nicht schwierig zu beweisen), dass dann für jedes $n$ und jedes $\beta$

$$\mathbb{P}^+_{n,\beta}(\{\omega_0 = 1\}) > \frac{1}{2} > \mathbb{P}^-_{n,\beta}(\{\omega_0 = 1\}).$$

Weniger klar, aber interessant ist es zu untersuchen, ob der Effekt der Randbedingung bei festem $\beta$ verschwindet, wenn $n$ gegen unendlich geht, oder nicht. Es gilt das folgende Resultat.

**Satz 8** *Für $d = 1$ konvergiert $\mathbb{P}^+_{n,\beta}(\{\omega_0 = 1\})$ gegen $\frac{1}{2}$ für $n \to \infty$ und $\beta$ beliebig, aber fest. Für $d \geq 2$ existiert ein $\beta_0 \in (0, \infty)$ so dass*

$$\begin{aligned} \lim_n \mathbb{P}^+_{n,\beta}(\{\omega_0 = 1\}) &= \frac{1}{2} \quad (\beta < \beta_0), \\ \liminf_n \mathbb{P}^+_{n,\beta}(\{\omega_0 = 1\}) &> \frac{1}{2} \quad (\beta > \beta_0). \end{aligned}$$

Man sagt, dass für $d \geq 2$ bei der Temperatur $1/\beta_0$ ein Phasenübergang auftritt: Das Verhalten des Systems ändert sich sprunghaft. Für tiefe Temperaturen hat die lokale Interaktion globale Auswirkungen. Für $d = 2$ lässt sich auch $\beta_0$ berechnen. Man erhält $\beta_0 = \log(1 + \sqrt{2}) = 0.8814$.

## 5.2 Beweis im eindimensionalen Fall

Es ist leicht zu sehen, dass für $d = 1$ das Ising-Modell eine Markovkette auf der Menge $\{+1, -1\}$ ist mit Übergangsmatrix

$$\Pi = \frac{1}{e^{\beta} + e^{-\beta}} \begin{pmatrix} e^{\beta} & e^{-\beta} \\ e^{-\beta} & e^{\beta} \end{pmatrix}$$

und Startverteilung $(\frac{1}{2}, \frac{1}{2})$. Es gilt nämlich

$$\exp\left(\beta \sum_{i=-n+1}^{n} \omega_i \omega_{i-1}\right) = 2\left(e^{\beta} + e^{-\beta}\right)^{2n} \frac{1}{2} \Pi(\omega_{-n}, \omega_{-n+1}) \cdots \Pi(\omega_{n-1}, \omega_n).$$

Also folgt

$$\mathbb{P}^+_{n,\beta}(\{\omega_0 = 1\}) = \frac{\mathbb{P}_{n,\beta}(\{\omega_{-n} = \omega_0 = \omega_n = 1\})}{\mathbb{P}_{n,\beta}(\{\omega_{-n} = \omega_n = 1\})} = \frac{\Pi^n(1,1)^2}{\Pi^{2n}(1,1)}.$$

Wegen Satz 6 konvergiert aber $\Pi^n(1,1)$ gegen $\frac{1}{2}$, denn die stationäre Verteilung ist gleich $(\frac{1}{2}, \frac{1}{2})$. □

### 5.3 Beweis in zwei Dimensionen

Wir beweisen ein Teilresultat, nämlich dass für jedes $0 < c < \frac{1}{2}$ ein $\beta_0$ exisitert, so dass für alle $\beta > \beta_0$ und für alle $n$ $\mathbb{P}^+_{n,\beta}(\{\omega_0 = -1\}) \leq c$. Dazu benutzen wir den Begriff der Konturen: Wir betrachten das duale Gitter mit Ecken in den Punkten $(i_1 \pm \frac{1}{2}, i_2 \pm \frac{1}{2})$ und Kanten zwischen nächsten Nachbarn. Für ein gegebenes $\omega \in \Omega_n$ nennen wir diejenigen Kanten aktiv, welche zwischen zwei verschiedenen Spins liegen. Wenn sich vier aktive Kanten an einer Ecke treffen, runden wir die Ecken so ab, dass die beiden Elementarmagneten mit Spin $-1$ nicht separiert werden. Wenn am Rand alle Spins gleich $+1$ sind, dann bilden die aktiven Kanten geschlossene Konturen $C$, welche Regionen mit Spin $-1$ begrenzen. Die Menge aller Konturen in $L_n$, welche den Ursprung umschliessen, bezeichnen wir mit $\mathcal{C}$, und für jedes $\omega$ mit $\omega_0 = -1$ sei $\Gamma(\omega)$ diejenige Kontur in $\mathcal{C}$, welche den Ursprung umschliesst. Dann gilt

$$\mathbb{P}^+_{n,\beta}(\{\omega_0 = -1\}) = \sum_{C \in \mathcal{C}} \sum_{\omega; \Gamma(\omega) = C} p^+_{n,\beta}(\omega).$$

Der Beweis besteht nun aus 2 Abschätzungen. Wir bezeichnen die Länge einer Kontur $C$ mit $L(C)$. Dann werden wir zeigen, dass

$$\Gamma(\omega) = C \Rightarrow p^+_{n,\beta}(\omega) \leq \exp(-2\beta L(C))$$

und dass die Anzahl Konturen $C \in \mathcal{C}$ mit $L(C) = \ell$ kleiner oder gleich $\frac{\ell}{2} 3^\ell$ ist. Dies impliziert dann

$$\mathbb{P}^+_{n,\beta}(\{\omega_0 = -1\}) \leq \sum_{\ell=4}^{\infty} \frac{\ell}{2} \exp(\ell(\log 3 - 2\beta)).$$

Diese letzte Summe ist offensichtlich beliebig klein, wenn $\beta$ genügend gross ist.

Die Behauptung über die Anzahl Konturen $C \in \mathcal{C}$ mit $L(C) = \ell$ wird wie folgt begründet: Es gibt $\min(\ell/2, n)$ Möglichkeiten für den Schnittpunkt der Kontur mit der positiven $x$-Achse, und an jeder der $\ell$ Ecken von $C$ gibt es maximal 3 Möglichkeiten, die Kontur fortzusetzen.

Es bleibt noch $p^+_{n,\beta}(\omega)$ abzuschätzen, wenn $\omega_0 = -1$ und $\Gamma(\omega) = C$. Für ein solches $\omega$ sei $\bar{\omega}$ die Konfiguration, welche innerhalb von $C$ überall gleich $+1$ ist und ausserhalb von $C$ mit $\omega$ übereinstimmt. Dann gilt

$$p^+_{n,\beta}(\omega) = \frac{\exp(-2\beta N_u(\omega))}{\sum_{\omega'} \exp(-2\beta N_u(\omega'))} \leq \frac{\exp(-2\beta N_u(\omega))}{\exp(-2\beta N_u(\bar{\omega}))} = \exp(-2\beta L(C)).$$

□

## 5.4 Simulation des Ising Modells

Um das Phänomen des Phasenübergangs zu illustrieren, möchte man gerne mit Hilfe des Computers Spinkonfigurationen gemäss den Wahrscheinlichkeiten des Ising-Modells auswählen (simulieren). Wir nehmen dabei an, dass wir gleichverteilte Zahlen in $[0,1]$ simulieren können. Weil die Menge aller Spinkonfigurationen riesig ist, ist das eine schwierige Aufgabe, für die es keine einfache Lösung gibt. Man gibt sich daher zufrieden mit einer Übergangsmatrix $\Pi$ auf der Menge aller Spinkonfigurationen, welche das Ising-Modell als stationäre Verteilung hat und welche einfach genug ist, so dass man für jede vorgegebene Konfiguration $\omega$ eine neue Konfiguration $\omega'$ gemäss den Wahrscheinlichkeiten $\Pi(\omega,\omega')$ auswählen kann. Dann beginnt man mit einer beliebigen Konfiguration $\omega_0$ und wählt dann iterativ für $t = 1,2,\ldots,T$ eine Konfiguration $\omega_t$ gemäss $\Pi(\omega_{t-1},.)$. Wenn $T$ gross genug ist, dann folgt wegen Satz 6

$$\mathbb{P}(\{\omega_T = \omega\}) \approx p_{n,\beta}(\omega).$$

Die einfachste Wahl von $\Pi$ hat die Eigenschaft, dass $\Pi(\omega,\omega')$ meistens gleich Null ist:

$$\Pi(\omega,\omega') = \begin{cases} \frac{1}{(2n+1)^d} \frac{\exp(\beta\omega'_{\mathbf{i}} \sum_{||\mathbf{i}-\mathbf{j}||=1} \omega_{\mathbf{j}})}{\exp(\beta \sum_{||\mathbf{i}-\mathbf{j}||=1} \omega_{\mathbf{j}}) + \exp(-\beta \sum_{||\mathbf{i}-\mathbf{j}||=1} \omega_{\mathbf{j}})} & \exists \mathbf{i}\ \omega'_{\mathbf{j}} = \omega_{\mathbf{j}}\ \forall \mathbf{j} \neq \mathbf{i} \\ 0 & \text{sonst} \end{cases}$$

Dieser Übergang bedeutet, dass man zufällig einen Elementarmagneten wählt und gegeben die benachbarten Spins dessen Spin neu wählt, so wie es das Ising Modell vorschreibt. Es lässt sich leicht nachprüfen, dass das Ising Modell reversibel und damit stationär für diesen Übergang ist. Ferner ist der Übergang irreduzibel und aperiodisch.

Die Idee, Werte gemäss einer komplizierten Verteilung $\gamma$ auf einem grossen Raum $\Omega$ zu wählen, indem man eine einfache Markovkette konstruiert, welche dieses $\gamma$ als stationäre Verteilung hat, hat sich für sehr viele Probleme als nützlich erwiesen. Diese Methode trägt den Namen Markovketten Mone Carlo (in der englischen Abkürzung MCMC).

# Literatur

## Zu Kapitel 2 (Kodierung)


- G. Kersting, A. Wakolbinger. *Elementare Sochastik*, 2. Auflage, Birkhäuser, 2010. Teil VI, Kap. 23 und 24.
- A. Papoulis. S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. 4. Auflage, McGraw Hill, 2002. Kap. 14-5.
- D. Aldous. *On Chance and Unpredictability: 13/20 lectures on the links between mathematical probability and the real world.* Buch-Manuskript, 2012, Kap. 2
  `http://www.stat.berkeley.edu/~aldous/Real-World/draft_book.pdf`


## Zu Kapitel 3 (Mischen von Karten)


- B. Mann. *How Many Times Should You Shuffle a Deck of Cards ?* In J. L. Snell (editor), *Topics in Contemporary Probability and Its Applications*. CRC Press, 1995. 261-289.
- D. Aldous. *On Chance and Unpredictability: 13/20 lectures on the links between mathematical probability and the real world.* Buch-Manuskript, 2012, Kap. 6.
  `http://www.stat.berkeley.edu/~aldous/Real-World/draft_book.pdf`


## Zu Kapitel 4 (Markovketten)


- O. Häggström. Finite Markov Chains and Algorithmic Applications. Cambridge University Press 2002.
- J. G. Kemeny, J. L. Snell. Finite Markov Chains. Springer 1976.


## Zu Kapitel 5 (Ising Modell)


- R. Kindermann, J. L. Snell. Markov random fields and their applications. Contemporary Mathematics, Vol. 1, American Mathematical Society 1980. Kapitel I.