

Kapitel 2

Diskrete Markov-Ketten

Nachdem wir im vorherigen Kapitel die allgemeinen theoretischen Grundlagen über MK auf beliebigen Zustandsräumen bereitgestellt haben, konzentrieren wir uns jetzt auf den abzählbaren Fall, wobei immer zeitliche Homogenität unterstellt wird. Unser besonderes Augenmerk gilt der Frage nach dem asymptotischen Verhalten diskreter und endlicher MK für gegen unendlich strebende Zeit. Beginnen wollen wir aber mit einer Reihe von Beispielen, die gleichsam als Motivation für unsere Untersuchungen dienen.

2.1 Beispiele diskreter Markov-Ketten

Mit interessanten Beispielen zeitlich homogener DMK ließe sich mühelos ein eigenes Buch füllen, und die folgende Auswahl soll lediglich einen ersten Einblick vermitteln.

2.1.1 Markov-Ketten mit zwei Zuständen

Betrachten wir einen Telefonanschluss, dessen Leitung zum Zeitpunkt n frei bzw. besetzt ist, was wir durch $M_n = 0$ bzw. $M_n = 1$ codieren. Wir nehmen an, dass die Wahrscheinlichkeit für einen Anruf während eines Zeitintervalls $p > 0$ beträgt und dass ferner höchstens ein Anruf pro Zeitintervall eingeht. Ist die Leitung besetzt, geht der Anruf nicht durch und wird folglich nicht registriert. Wir nehmen weiter an, dass sie in diesem Fall mit einer Wahrscheinlichkeit $q > 0$ im nächsten Intervall wieder frei ist. Wir erhalten so eine EMK $(M_n)_{n \geq 0}$ mit Zustandsraum $\mathcal{S} = \{0, 1\}$ und Übergangsmatrix

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

Offenkundig hat jede Übergangsmatrix einer EMK mit zwei Zuständen diese Form, und es bedarf zu ihrer Spezifikation lediglich der Angabe der Werte p und q .

2.1.2 Ein einfaches Bedienungssystem

Wir greifen auf das vorherige Beispiel zurück und nehmen nun an, dass der Anschluss einen Anrufer auf Abruf halten kann. Die Anzahl der Anrufer im System ist folglich ein Element der Menge $\mathcal{S} = \{0, 1, 2\}$. Wie bisher betrage die Wahrscheinlichkeit q dafür, dass ein Anruf während eines Zeitintervalls beendet wird, und p dafür, dass ein neuer Anruf eingeht, sofern das System nicht bereits voll ist. Zur Modellierung setzen wir

$$p_{00} = 1 - p, \quad p_{01} = p \quad \text{und} \quad p_{02} = 0,$$

denn ein neuer Anruf kommt mit Wahrscheinlichkeit p an (wobei wir weiterhin höchstens einen eingehenden Anruf pro Zeitintervall annehmen). Analog erhalten wir

$$p_{20} = 0, \quad p_{21} = q \quad \text{und} \quad p_{22} = 1 - q,$$

denn kein neuer Anruf wird registriert, wenn bereits zwei in der Leitung sind, und

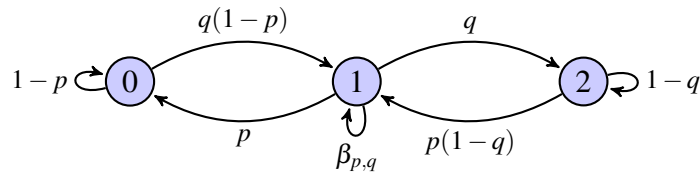


Abb. 2.1 Markov-Kette mit 3 Zuständen und den obigen Übergangswahrscheinlichkeiten, wobei $\beta_{p,q} = 1 - q(1 - p) - p(1 - q)$.

höchstens einer der beiden Anrufe kann während des Intervalls beendet werden. Befindet sich genau ein Gespräch in der Leitung, ist die Situation etwas komplizierter. Das System geht von 1 in 0 über, wenn das aktuelle Gespräch während des Intervalls beendet wird und kein neuer Anruf eingeht, also $p_{10} = q(1 - p)$ (unter der vernünftigen Annahme, dass sich Gesprächsdauern und Gesprächseingänge unabhängig voneinander verhalten). Ähnlich ergibt sich $p_{12} = p(1 - q)$, weil das System von 1 in 2 übergeht, wenn das aktuelle Gespräch während des Intervalls nicht beendet wird und zugleich ein neuer Anruf eingeht, der auf Abruf gestellt wird. Da sich Zeilen einer Übergangsmatrix stets zu 1 addieren, folgt $p_{11} = 1 - q(1 - p) - p(1 - q)$ und somit insgesamt

$$P = \begin{pmatrix} 1-p & p & 0 \\ q(1-p) & 1-q(1-p)-p(1-q) & p(1-q) \\ 0 & q & 1-q \end{pmatrix}.$$

Übergangswahrscheinlichkeiten werden oft mittels gerichteter Graphen dargestellt, wobei die Knoten die Zustände und die gerichteten Kanten (Pfeile) die Übergänge beschreiben. Für die obige Matrix zeigt dies Abb. 2.1.

2.1.3 Irrfahrten mit reflektierenden Barrieren

Stellen wir uns vor, ein Teilchen springt auf den Gitterpunkten $\{0, 1, \dots, N\}$. Zu jedem Zeitpunkt springt es eine Einheit nach rechts bzw. links mit den Wahrscheinlichkeiten $p \in (0, 1)$ bzw. $1 - p$. Befindet es sich in einem der Randpunkte (Barrieren) 0 oder N , wird es mit Wahrscheinlichkeit 1 reflektiert, wandert also wieder in Richtung des Intervalls. Die Übergangsmatrix dieser EMK lautet

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

Im Fall $p = \frac{1}{2}$ sprechen wir von einer *symmetrischen*, andernfalls von einer *asymmetrischen Irrfahrt auf $\{0, 1, \dots, N\}$ mit reflektierenden Barrieren*. Es ist manchmal sinnvoll, lediglich teilweise reflektierende Barrieren zu betrachten, was bedeuten soll, dass das Teilchen mit der Wahrscheinlichkeit p eine Zeiteinheit in den Randpunkten 0 und N verharren kann. Es gilt dann also

$$p_{00} = 1 - p_{01} = p \quad \text{und} \quad p_{NN} = 1 - p_{N,N-1} = p.$$

2.1.4 Irrfahrten mit absorbierenden Barrieren

Diese EMK verhält sich genauso wie die in 2.1.3, außer für den Fall, dass das Teilchen einen der Randpunkte 0 oder N erreicht, wo es nun absorbiert wird. Die Übergangsmatrix lautet dann

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

2.1.5 Einfache Irrfahrten auf einem Graphen

Betrachten wir einen (endlichen, einfachen und ungerichteten) Graphen $G = (V, E)$, V die Menge der Knoten und $\mathbb{E} \subset \mathfrak{P}(V)$ die Menge der Kanten. Dann verbindet jede Kante zwei *verschiedene* Knoten, und je zwei Knoten sind durch höchstens eine Kante verbunden. Wir schreiben $v \sim w$, wenn zwei Knoten v, w benachbart, d.h. durch eine Kante verbunden sind.

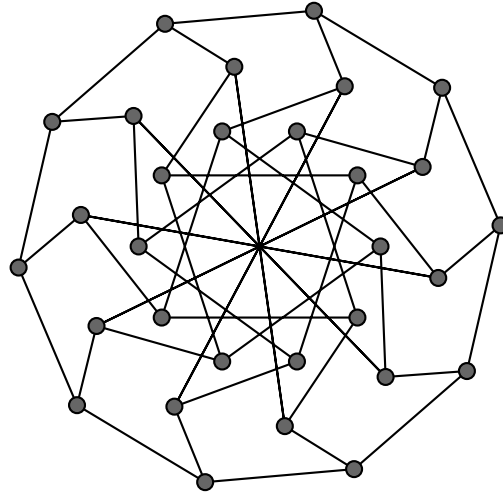


Abb. 2.2 Der Tutte-Coxeter-Graph, auch Tutte-8-Käfig genannt, als Beispiel eines endlichen, einfachen und ungerichteten Graphen. Dieser Graph hat 30 Knoten, 45 Kanten und ist ferner regulär von der Ordnung 3, d.h., jeder Knoten hat genau 3 Nachbarn.

Eine einfache Irrfahrt auf G ist eine EMK mit Zustandsraum V , die zu jedem Zeitpunkt vom augenblicklichen Aufenthaltsknoten mit gleicher Wahrscheinlichkeit in einen der Nachbarknoten springt. Sie hat demnach die Übergangswahrscheinlichkeiten $p_{vw} = \frac{1}{d(v)}$, $v \sim w$, wobei $d(v)$ die Anzahl benachbarter Knoten von v angibt (im Fall $d(v) = 0$ setzen wir $p_{vv} = 1$). Die in 2.1.3 vorgestellte symmetrische Irrfahrt mit reflektierenden Barrieren bildet ein spezielles Beispiel einer solchen einfachen Irrfahrt. Für den in Abb. 2.2 dargestellten Tutte-Coxeter-Graphen beträgt die Übergangswahrscheinlichkeit stets $\frac{1}{3}$, in einen der Nachbarknoten zu springen.

2.1.6 Das Ehrenfest-Modell für Wärmeaustausch

Das Ehrenfest-Modell bildet eine klassische mathematische Beschreibung sowohl für den *Wärmeaustausch zwischen zwei nach außen isolierten, sich berührenden Körpern* als auch für die *Diffusion durch eine Membran*. Stellen wir uns dazu zwei

Urnen A und B vor, die insgesamt $2N$ Kugeln enthalten, von denen sich zu Beginn k in Urne A und $2N - k$ in Urne B befinden. Es wird nun immer eine Kugel zufällig ausgewählt (jede also mit der gleichen Wahrscheinlichkeit) und von der Urne, in der sie sich gerade befindet, in die jeweils andere gelegt. Im Zeitablauf wandern die Kugeln also zwischen beiden Urnen hin und her, wobei stets eine mittlere Drift in Richtung derjenigen mit der geringeren Zahl von Kugeln besteht. Die Anzahl der Kugeln in den beiden Urnen interpretiert man beim Wärmeaustausch als Temperatur der sie repräsentierenden Körper. Sei X_n die Anzahl der Kugeln in Urne A nach der n -ten Ziehung und $M_n = X_n - N$. Dann bildet $(M_n)_{n \geq 0}$ eine EMK mit Zustandsraum $\mathcal{S} = \{-N, -N+1, \dots, -1, 0, 1, \dots, N-1, N\}$ und Übergangswahrscheinlichkeiten

$$p_{ij} = \begin{cases} \frac{N-i}{2N}, & \text{falls } j = i + 1, \\ \frac{N+i}{2N}, & \text{falls } j = i - 1, \\ 0, & \text{sonst.} \end{cases}$$

2.1.7 Markov-Ketten in der Genetik: Die Modelle von Wright-Fisher und Moran

Das folgende idealisierte genetische Modell wurde von S. WRIGHT [62] vorgeschlagen, um die Fluktuationen von Genfrequenzen unter dem Einfluss von Mutation und Selektion zu untersuchen.

Das Wright-Fisher-Modell

Wir beginnen mit der Beschreibung eines *Grundmodells zufälliger Reproduktion ohne Mutation und Selektion* und betrachten eine endliche Population von N Individuen (Zellen), die jeweils zwei gleichartige (homologe) Chromosomensätze besitzen (*Diploidie*). Auf jedem Chromosomensatz befindet sich jeweils eine Kopie eines bestimmten Gens, das entweder vom Typ a oder A ist. Diese Typen, genannt *Allele*, bilden die beiden möglichen Zustandsformen des Gens und befinden sich in homologen Chromosomen an gleichen Loci (Genorten). Wir richten nun unser Augenmerk ausschließlich auf die $2N$ Gene der N Individuen, die wir als gegebene Genpopulation auffassen, wobei die nachfolgenden Überlegungen gleichermaßen gelten, wenn diese $2N$ Gene von $2N$ Individuen einer haploiden Population stammen. Man spricht von *Haploidie* bei Populationen (z.B. von Keimzellen), deren Individuen nur einen Chromosomensatz und daher das Gen immer nur in einer seiner Varianten besitzen.

Ausgehend von einer Elternpopulation (Generation 0) mit i Typ- a -Allelen und $2N - i$ Typ- A -Allelen, ergibt sich die nächste Generation wie durch $2N$ -maliges Ziehen mit Zurücklegen aus einer Urne mit $2N$ Kugeln, von denen i rot und $2N - i$ schwarz sind. Liefert die k -te Ziehung eine rote Kugel, ist das k -te Gen der Toch-

terpopulation vom Typ a , andernfalls vom Typ A (Bernoulli-Experiment). Die Ziehungen sind offenbar unabhängig und die Wahrscheinlichkeiten für ein Typ- a bzw. Typ- A -Allel gegeben durch

$$\alpha_i = \frac{i}{2N} \quad \text{bzw.} \quad \beta_i = 1 - \frac{i}{2N}.$$

Die Evolution der Population unter diesem ad infinitum fortgesetzten Mechanismus lässt sich durch

$$M_n = \sum_{k=1}^{2N} X_{n,k}, \quad n \geq 1 \quad (2.1)$$

beschreiben, wobei M_n die Anzahl der Typ- a -Allele der n -ten Generation bezeichnet und $X_{n,k}$ das Ergebnis des k -ten Bernoulli-Experiments zur Erzeugung dieser Generation angibt. Die $X_{n,k}$, $1 \leq k \leq 2N$, sind bedingt unter M_0, \dots, M_{n-1} stochastisch unabhängig und identisch $Bern(\alpha_{M_{n-1}})$ -verteilt, d.h.

$$\mathbb{P}(X_{n,1}, \dots, X_{n,2N} | M_0 = i_0, \dots, M_{n-1} = i_{n-1}) = \mathbb{P}(X_{n,1}, \dots, X_{n,2N} | M_{n-1} = i_{n-1}) = Bern(\alpha_{i_{n-1}})^{2N}. \quad (2.2)$$

Die $X_{n,k}$ hängen also von M_0, \dots, M_{n-1} nur über M_{n-1} ab, und vermöge (2.2) folgt weiter

$$\mathbb{P}^{M_n} | M_0 = i_0, \dots, M_{n-1} = i_{n-1} = \mathbb{P}^{M_n} | M_{n-1} = i_{n-1} = Bin(2N, \alpha_{i_{n-1}}).$$

$(M_n)_{n \geq 0}$ bildet demnach eine EMK mit Zustandsraum $\{0, \dots, 2N\}$ und Übergangswahrscheinlichkeiten

$$p_{ij} = \binom{2N}{j} \alpha_i^j \beta_i^{2N-j}, \quad i, j = 0, \dots, 2N. \quad (2.3)$$

Auf eine Diskussion der biologischen Rechtfertigung der hier gemachten Voraussetzungen verzichten wir und verweisen auf FISHER [22]. Beachte, dass die Zustände 0 und $2N$, in denen die Population nur noch Gene vom Typ A bzw. a enthält, absorbierend sind. Man spricht in diesem Fall von *Fixierung*. Eine Frage von beträchtlichem Interesse lautet: Mit welcher Wahrscheinlichkeit tritt bedingt unter $M_0 = k$, $1 \leq k \leq 2N - 1$, Fixierung ein? Darüber hinaus stellt sich im positiven Fall die Frage nach der Rate, mit der dieses geschieht.

Dasselbe Modell mit Mutationseffekten

Ein verallgemeinertes Modell berücksichtigt Mutationseffekte, etwa durch folgende Modellierung: Vor Bildung einer neuen Generation hat jedes Allel die Chance zu mutieren, d.h. hier, sich in ein Allel der anderen Art zu verwandeln. Wir nehmen an, dass eine Mutation $a \rightarrow A$ mit Wahrscheinlichkeit γ_1 und eine Mutation $A \rightarrow a$ mit Wahrscheinlichkeit γ_2 geschieht. Es gelten dann weiter (2.1) – (2.3), jedoch mit den neuen Ziehungswahrscheinlichkeiten

$$\begin{aligned}\alpha_i &= \frac{i}{2N}(1 - \gamma_1) + \left(1 - \frac{i}{2N}\right)\gamma_2, \\ \beta_i &= \frac{i}{2N}\gamma_1 + \left(1 - \frac{i}{2N}\right)(1 - \gamma_2).\end{aligned}\tag{2.4}$$

Zur genaueren Erläuterung schlüsseln wir den Mechanismus weiter auf: Wir nehmen an, dass Mutation der Ziehung nachgeschaltet ist. Sei $Y_{n,k} = 1$ bzw. $= 0$, falls das k -te aus der $(n-1)$ -ten Generation selektierte Gen vor Auftreten einer möglichen Mutation vom Typ a bzw. A ist. Die $Y_{n,1}, \dots, Y_{n,2N}$ erfüllen dann (2.2), sind also bedingt unter M_{n-1} unabhängig und jeweils $Bern(M_{n-1}/2N)$ -verteilt. Seien weiter $I_{n,k}, J_{n,k}$ unabhängige (auch von M_{n-1} und den $Y_{n,k}$) Bernoulli-Variablen, $I_{n,k} \sim Bern(\gamma_1)$, $J_{n,k} \sim Bern(\gamma_2)$, und setze

$$X_{n,k} = Y_{n,k}(1 - I_{n,k}) + (1 - Y_{n,k})J_{n,k}.$$

$I_{n,k} = 1$ bedeutet demnach eine Mutation $a \rightarrow A$ des k -ten gezogenen Gens der $(n-1)$ -ten Generation und $J_{n,k} = 1$ eine Mutation $A \rightarrow a$. Wie man sofort einsieht, erfüllen auch hier die $X_{n,k}$ (2.2), jedoch mit den α_i aus (2.4). Es gilt nämlich unter Beachtung der Unabhängigkeitsannahmen

$$\begin{aligned}\mathbb{P}(X_{n,k} = 1 | M_{n-1} = i) &= \mathbb{P}(I_{n,k} = 0, Y_{n,k} = 1 | M_{n-1} = i) + \mathbb{P}(J_{n,k} = 1, Y_{n,k} = 0 | M_{n-1} = i) \\ &= \mathbb{P}(Y_{n,k} = 1 | M_{n-1} = i) \mathbb{P}(I_{n,k} = 0) + \mathbb{P}(Y_{n,k} = 0 | M_{n-1} = i) \mathbb{P}(J_{n,k} = 1) \\ &= \frac{i}{2N}(1 - \gamma_1) + \left(1 - \frac{i}{2N}\right)\gamma_2 = \alpha_i\end{aligned}$$

für alle $k = 1, \dots, 2N$.

Sofern $\gamma_1 \gamma_2 > 0$, tritt offenbar in keinem Zustand Fixierung ein. Stattdessen strebt M_n in diesem Fall für $n \rightarrow \infty$ in Verteilung gegen einen stationären Limes π , den wir als *Genfrequenz im Gleichgewicht* bezeichnen.

Dasselbe Modell mit Selektionsdruck

Wir kehren zurück zum Grundmodell und wollen für dieses als weitere Variante das Konzept eines Selektionsdrucks zugunsten von, sagen wir, Typ- a -Allelen diskutieren. Es sei zunächst bemerkt, dass im Grundmodell (neutrale Selektion) unter Benutzung von (2.3)

$$\mathbb{E}(M_n | M_{n-1} = i) = 2N \cdot \frac{i}{2N} = i\tag{2.5}$$

für alle $i = 0, \dots, 2N$ folgt, die mittleren Reproduktionsraten $r_n = \frac{\mathbb{E}(M_n | M_{n-1})}{M_{n-1}}$ und $R_n = \frac{\mathbb{E}(2N - M_n | M_{n-1})}{2N - M_{n-1}}$ für beide Alleltypen also stets 1 betragen. Stellen wir uns nun vor, dass der Ziehungsmechanismus Allelen vom Typ a gegenüber denen vom Typ A

einen mittleren selektiven Vorteil gibt, präzisiert durch $r_n = (1+s)R_n$ für alle $n \geq 1$ und ein $s > 0$ (klein). Gesucht sind also Selektionswahrscheinlichkeiten α_k, β_k , die dieses gewährleisten. Da weiterhin $\mathbb{E}(M_n | M_{n-1} = i) = 2N\alpha_i$ für alle $i = 0, \dots, 2N$ gilt, ergeben sich $\alpha_i, \beta_k = 1 - \alpha_i$ vermöge

$$r_n = \frac{2N\alpha_{M_{n-1}}}{M_{n-1}} = (1+s) \frac{2N(1 - \alpha_{M_{n-1}})}{2N - M_{n-1}} = (1+s)R_n, \quad n \geq 1,$$

eindeutig zu

$$\alpha_i = \frac{(1+s)i}{2N+si} \quad \text{und} \quad \beta_i = \frac{2N-k}{2N+si}. \quad (2.6)$$

Der Quotient der erwarteten Populationsgrößen von Typ- a - und Typ- A -Allelen in der n -ten Generation (bedingt unter M_{n-1}) ergibt sich zu

$$\begin{aligned} \frac{\mathbb{E}(M_n | M_{n-1})}{\mathbb{E}(2N - M_n | M_{n-1})} &= \frac{\alpha_{M_{n-1}}}{\beta_{M_{n-1}}} = \frac{(1+s)M_{n-1}}{2N - M_{n-1}} \\ &= \left(\frac{1+s}{1} \right) \left(\frac{\text{Anzahl von Typ-}a\text{-Genen in der } (n-1)\text{-ten Generation}}{\text{Anzahl von Typ-}A\text{-Genen in der } (n-1)\text{-ten Generation}} \right) \end{aligned}$$

und verdeutlicht auf alternative Weise die Bedeutung von Selektion. Beachte, dass Zustände 0 und $2N$ auch unter Selektionsdruck absorbierend sind. Eine wichtige Frage lautet demnach auch hier, mit welcher Wahrscheinlichkeit bedingt unter $M_0 = k$ Fixierung eintritt.

Das Moran-Modell

Wir betrachten wieder eine Population von $2N$ Genen, die entweder vom Typ a oder A seien, wobei sich jedoch Generationen anders als beim Wright-Fisher-Modell hier überlappen. Der im Anschluss beschriebene Reproduktionsmechanismus, der in jedem Zeitschritt immer nur ein Element der Population betrifft, wurde von MORAN [48] vorgeschlagen. Wir beschränken uns auf die einfache Modellvariante ohne Mutations- und Selektionseffekte, weisen aber darauf hin, dass auch diese problemlos bei der Modellierung berücksichtigt werden können. Stellen wir uns die $2N$ Gene wieder als rote (Typ a) und schwarze (Typ A) Kugeln einer Urne vor. Die Zusammensetzung der Urne nach n -facher Anwendung des folgenden zwei-stufigen Mechanismus' beschreibt die Population nach n Zeitschritten, wobei wir der Einfachheit wiederum von der n -ten Generation sprechen. Wir ziehen zufällig eine Kugel mit Zurücklegen und "duplizieren" diese (Reproduktionsschritt), wobei das Duplikat zunächst beiseite gelegt wird. Anschließend ziehen wir nochmals eine Kugel aus der Urne, legen diese aber nicht zurück, sondern ersetzen sie durch das beiseite gelegte Duplikat (Austauschschritt). Seien wieder M_n die Anzahl der Typ- a -Allele (roter Kugeln) der n -ten Generation (nach n Doppelziehungen) und $\alpha_k = \frac{k}{2N}, \beta_k = 1 - \alpha_k = \frac{2N-k}{2N}$, d.h., $\alpha_{M_n}, \beta_{M_n}$ bezeichnen die relativen Häufigkeiten der beiden Allele in der n -ten Generation. Nach der obigen Beschreibung bildet

$(M_n)_{n \geq 0}$ offenkundig eine EMK mit Zustandsraum $\{0, \dots, 2N\}$ und Übergangswahrscheinlichkeiten

$$p_{ij} = \begin{cases} \alpha_i \beta_i, & \text{falls } j = i \pm 1 \\ \alpha_i^2 + \beta_i^2 & \text{falls } i = j \\ 0, & \text{sonst} \end{cases}.$$

Die Zustände 0 und $2N$ sind erneut absorbierend, und es gilt wieder (2.5), d.h. $\mathbb{E}(M_n | M_{n-1} = i) = i$ für alle $i = 0, \dots, 2N$.

Für eine ausführlichere Diskussion der zuvor beschriebenen Modelle lese man etwa in der Monographie von GALE [23].

War der Zustandsraum bisher stets endlich, so ist oder kann dieser in den nachfolgenden Beispielen abzählbar unendlich sein.

2.1.8 Irrfahrten auf \mathbb{Z}^d

Wir betrachten das Gitter \mathbb{Z}^d , versehen mit der gewöhnlichen Nachbarschaftsrelation, die durch die ℓ_1 -Norm $|i|_1 := \sum_{k=1}^d |i_k|$ determiniert ist, d.h., zwei Punkte in \mathbb{Z}^d sind genau dann durch eine Kante verbunden, wenn ihr ℓ_1 -Abstand 1 beträgt. Stellen wir uns vor, ein Teilchen auf diesem Gitter springt pro Zeiteinheit unabhängig vom gegenwärtigen Aufenthaltsort i mit Wahrscheinlichkeit je $1/2d$ in einen der $2d$ Nachbarpunkte $i \pm e_k$, $k = 1, \dots, d$, wobei e_k den k -ten kanonischen Einheitsvektor im \mathbb{R}^d bezeichnet. Abb. 6.3 veranschaulicht die Situation für den Fall “ $d = 2$ ”. Bezeichnet M_0 den Startpunkt und M_n die Position des Teilchens zum Zeitpunkt n , so gilt

$$\mathbb{P}(M_{n+1} = i_n \pm e_k | M_n = i_n, \dots, M_0 = i_0) = \mathbb{P}(M_{n+1} = i_n \pm e_k | M_n = i_n) = \frac{1}{2d}$$

für alle $n \geq 0$, $k = 1, \dots, d$ und $(i_0, \dots, i_n) \in \mathbb{Z}^{(n+1)d}$ mit $\mathbb{P}(M_0 = i_0, \dots, M_n = i_n) > 0$. Es liegt also eine DMK mit Zustandsraum \mathbb{Z}^d vor, wobei es sich im Rückblick auf 2.1.5 offenbar um eine einfache Irrfahrt auf dem *unendlichen* Graphen G mit \mathbb{Z}^d als Knotenmenge und der ℓ_1 -Nachbarschaftsrelation handelt. Wir sprechen in diesem Fall auch von einer *symmetrischen Irrfahrt auf $(\mathbb{Z}^d, |\cdot|_1)$* . Allgemeiner bezeichnen wir $(M_n)_{n \geq 0}$ als *Irrfahrt auf $(\mathbb{Z}^d, |\cdot|_1)$ mit Parametern p_{-d}, \dots, p_d* , falls

$$\mathbb{P}(M_{n+1} = i | M_n = i) = p_0 \quad \text{und} \quad \mathbb{P}(M_{n+1} = i \pm e_k | M_n = i) = p_{\pm k}$$

für alle $i \in \mathbb{Z}^d$ und $k = 1, \dots, d$ gilt.

Eine andere Nachbarschaftsrelation auf \mathbb{Z}^d , für $d = 2$ in Abb. 2.4 dargestellt, ergibt sich bei Zugrundelegung der Maximums(ℓ_∞ -)norm $|i|_\infty := \max_{1 \leq k \leq d} |i_k|$. In diesem Fall sind zwei Punkte genau dann durch eine Kante verbunden, wenn ihr ℓ_∞ -Abstand 1 beträgt. Eine DMK $(M_n)_{n \geq 0}$ auf \mathbb{Z}^d heißt *Irrfahrt auf $(\mathbb{Z}^d, |\cdot|_\infty)$ mit Parametern p_α , $\alpha \in \{-1, 0, 1\}^d$* , wenn

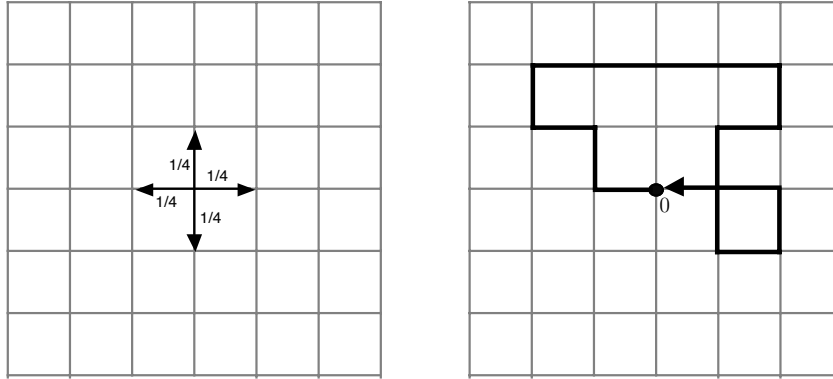


Abb. 2.3 Symmetrische Irrfahrt auf \mathbb{Z}^2 .

$$\mathbb{P}(M_{n+1} = i + \alpha | M_n = i) = p_\alpha$$

für alle $\alpha \in \{-1, 0, 1\}^d$. Sie heißt ferner *symmetrisch*, wenn $p_0 = 0$ und $p_\alpha = 1/(3^d - 1)$ für alle anderen α . Jede Irrfahrt auf $(\mathbb{Z}^d, |\cdot|_1)$ ist offenkundig ebenfalls eine solche auf $(\mathbb{Z}^d, |\cdot|_\infty)$.

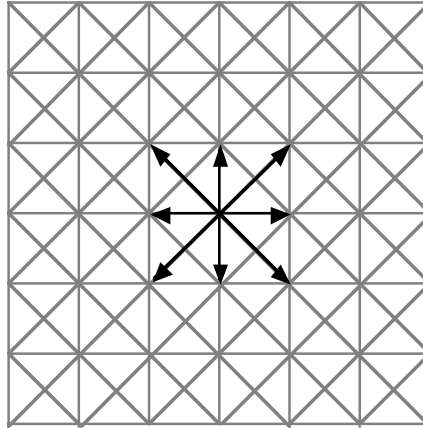


Abb. 2.4 Nachbarschaftsstruktur von \mathbb{Z}^2 unter der Maximumsnorm $|\cdot|_\infty$.

Eine interessante Frage für Irrfahrten auf \mathbb{Z}^d lautet: Für welche Parameterkombinationen kehrt diese, ausgehend von einem Gitterpunkt i , mit Wahrscheinlichkeit 1 in endlicher Zeit nach i zurück? Man nennt den Zustand i dann rekurrent. Abb. 2.3 (rechts) zeigt einen rekurrenten Pfad für den Zustand $i = 0$. Die Frage wird in Abschnitt 2.3 eingehend untersucht.

2.1.9 Eine Variante: Reflektierende Irrfahrten auf \mathbb{Z}

Das folgende Beispiel ist nicht zuletzt deswegen interessant, weil bei diesem das Vorliegen der Markov-Eigenschaft nicht unbedingt sofort klar ist. Sei $(S_n)_{n \geq 0}$ eine einfache, im Ursprung startende Irrfahrt auf \mathbb{Z} mit Parametern $p, q \in (0, 1)$, $p + q = 1$, d.h.

$$p = \mathbb{P}(S_{n+1} = i + 1 | S_n = i) \quad \text{und} \quad q = \mathbb{P}(S_{n+1} = i - 1 | S_n = i)$$

für alle $n \in \mathbb{N}_0$ und $i \in \mathbb{Z}$. Dann heißt

$$M_n := M_n, \quad n \geq 0$$

reflektierende Irrfahrt auf \mathbb{Z} und bildet eine zeitlich homogene MK auf \mathbb{N}_0 , wie wir im Anschluss zeigen werden.

Wir berechnen zunächst die bedingte Wahrscheinlichkeit

$$\mathbb{P}(S_n = i | M_n = i, M_{n-1} = i_{n-1}, \dots, M_1 = i_1)$$

für beliebige $i_1, \dots, i_{n-1}, i \in \mathbb{N}_0$ mit $\mathbb{P}(M_n = i, M_{n-1} = i_{n-1}, \dots, M_1 = i_1) > 0$. Sei dazu $m := \max\{0 \leq k \leq n : i_k = 0\}$. Dann sieht man leicht ein, dass

$$\begin{aligned} & \mathbb{P}(S_n = i | M_n = i, M_{n-1} = i_{n-1}, \dots, M_1 = i_1) \\ &= \mathbb{P}(S_n = i | M_n = i, M_{n-1} = i_{n-1}, \dots, M_{m+1} = i_{m+1}, M_m = 0) \end{aligned}$$

und dass es nur zwei Pfade für (S_m, \dots, S_n) gibt, die zu $M_n = i, M_{n-1} = i_{n-1}, \dots, M_m = 0$ führen, nämlich $(0, i_1, \dots, i_{n-1}, i)$ sowie $(0, -i_1, \dots, -i_{n-1}, -i)$ mit den Wahrscheinlichkeiten

$$p^{(n-m+i)/2} q^{(n-m-i)/2} \quad \text{bzw.} \quad p^{(n-m-i)/2} q^{(n-m+i)/2}.$$

Dies liefert

$$\begin{aligned} & \mathbb{P}(S_n = i | M_n = i, M_{n-1} = i_{n-1}, \dots, M_{m+1} = i_{m+1}, M_m = 0) \\ &= \frac{\mathbb{P}(S_n = i, S_{n-1} = i_{n-1}, \dots, S_{m+1} = i_{m+1}, S_m = 0)}{\mathbb{P}(S_n = i, M_{n-1} = i_{n-1}, \dots, M_{m+1} = i_{m+1}, M_m = 0)} \\ &= \frac{p^{(n-m+i)/2} q^{(n-m-i)/2}}{p^{(n-m+i)/2} q^{(n-m-i)/2} + p^{(n-m-i)/2} q^{(n-m+i)/2}} \\ &= \frac{p^i}{p^i + q^i}. \end{aligned}$$

Für $i \geq 1$ erhalten wir nun

$$\begin{aligned} & \mathbb{P}(M_{n+1} = i + 1 | M_n = i, M_{n-1} = i_{n-1}, \dots, M_1 = i_1) \\ &= \mathbb{P}(S_{n+1} = i + 1, S_n = i | M_n = i, M_{n-1} = i_{n-1}, \dots, M_1 = i_1) \\ &+ \mathbb{P}(S_{n+1} = -i - 1, S_n = -i | M_n = i, M_{n-1} = i_{n-1}, \dots, M_1 = i_1) \end{aligned}$$

$$\begin{aligned}
&= \mathbb{P}(S_{n+1} = i+1 | S_n = i) \mathbb{P}(S_n = i | M_n = i, M_{n-1} = i_{n-1}, \dots, M_1 = i_1) \\
&\quad + \mathbb{P}(S_{n+1} = -i-1 | S_n = -i) \mathbb{P}(S_n = -i | M_n = i, M_{n-1} = i_{n-1}, \dots, M_1 = i_1) \\
&= p \cdot \frac{p^i}{p^i + q^i} + q \cdot \frac{q^i}{p^i + q^i} \\
&= \frac{p^{i+1} + q^{i+1}}{p^i + q^i},
\end{aligned}$$

und da diese Übergangswahrscheinlichkeiten nicht von i_{n-1}, \dots, i_1 abhängen, folgt

$$p_{i,i+1} = \frac{p^{i+1} + q^{i+1}}{p^i + q^i} = 1 - p_{i,i-1}.$$

Für den verbleibenden Fall $i = 0$ genügt der Hinweis, dass offenbar

$$\mathbb{P}(M_{n+1} = 1 | M_n = 0, M_{n-1} = i_{n-1}, \dots, M_0 = i_0) = \mathbb{P}(M_{n+1} = 1 | M_n = 0) = 1$$

gilt, also $p_{01} = 1$.

Unterstellt man statt $S_0 = 0$ lediglich, dass S_0 symmetrisch verteilt ist, so bildet $(M_n)_{n \geq 0}$ weiterhin eine zeitlich homogene MK mit denselben Übergangswahrscheinlichkeiten. Verzichtet man jedoch auch auf die Symmetrie von \mathbb{P}^{S_0} , so ist die Aussage im Fall $0 < p, q < 1$ mit $p \neq q$ falsch, wie der Leser zur Übung überprüfen mag.

2.1.10 Diskrete Random Walks in \mathbb{Z}^d

M_0, X_1, X_2, \dots seien stochastisch unabhängige Zufallsvariablen mit Werten in \mathbb{Z}^d , wobei die X_n ferner alle dieselbe Verteilung $(p_k)_{k \in \mathbb{Z}^d}$ besitzen, d.h.

$$\mathbb{P}(X_n = k) = p_k$$

für alle $k \in \mathbb{Z}^d$. Dann heißt

$$M_n := M_0 + \sum_{k=1}^n X_k$$

diskreter Random Walk auf \mathbb{Z}^d mit Zuwachsverteilung $(p_k)_{k \in \mathbb{Z}^d}$ und definiert eine DMK mit Zustandsraum \mathbb{Z}^d und Übergangswahrscheinlichkeiten

$$\begin{aligned}
p_{ij} &= \mathbb{P}(M_{n+1} = j | M_n = i) \\
&= \mathbb{P}(X_{n+1} = j - i | M_n = i) = \mathbb{P}(X_{n+1} = j - i) = p_{j-i}
\end{aligned}$$

für alle $i, j \in \mathbb{Z}^d$. Einen Spezialfall bilden offensichtlich die zuvor vorgestellten Irrfahrten auf \mathbb{Z}^d , für die die Verschiebung pro Zeiteinheit, gemessen in $|\cdot|_1$ oder $|\cdot|_\infty$,

höchstens 1 beträgt. Die formale Bedingung hierfür lautet bezüglich der $|\cdot|_\infty$ -Norm

$$p_j = 0 \quad \text{für alle } j \notin \{-1, 0, 1\}^d. \quad (2.7)$$

Man spricht dann auch von einem *sprungfreien* (engl. “skip-free”) oder auch *Nächste-Nachbarn-* (engl. “nearest neighbour”) Random Walk.

2.1.11 Ein Bedienungssystem mit konstanten Bedienungszeiten

Kunden betreten ein Bedienungssystem und reihen sich in die Warteschlange ein, wenn der Bedienungsschalter besetzt ist. In jeder Bedienungsperiode (eine Zeiteinheit) wird genau ein Kunde bedient, sofern sich ein solcher überhaupt im System befindet; andernfalls wird in der Periode niemand bedient. Der Arbeitsmodus (Schlangendisziplin) sei FIFO (“first in first out”), d.h. die Kunden werden in der Reihenfolge ihres Erscheinens abgefertigt. Für $n \geq 1$ bezeichne M_n die Anzahl wartender Kunden am Ende und X_n die Anzahl ankommender Kunden während der n -ten Bedienungsperiode. Wir nehmen an, dass $(X_n)_{n \geq 1}$ eine unabhängige Folge identisch gemäß $(p_k)_{k \geq 0}$ verteilter Zufallsgrößen bildet und außerdem unabhängig ist von M_0 , dem Anfangszustand des Systems. Aufgrund dieser Beschreibung besitzt die Folge $(M_n)_{n \geq 0}$ die folgende rekursive Struktur: Für jedes $n \geq 0$ ist

$$M_{n+1} = \begin{cases} X_{n+1}, & \text{falls } M_n = 0, \\ (M_n - 1) + X_{n+1}, & \text{falls } M_n \geq 1 \end{cases} = (M_n - 1)^+ + X_{n+1}. \quad (2.8)$$

Induktiv folgt die Unabhängigkeit von (M_0, \dots, M_n) und X_{n+1} und daraus schließlich, dass $(M_n)_{n \geq 0}$ eine DMK bildet mit Zustandsraum \mathbb{N}_0 und Übergangswahrscheinlichkeiten

$$p_{ij} = \mathbb{P}(M_{n+1} = j | M_n = i) = \mathbb{P}(X_{n+1} = j - (i - 1)^+) = p_{j - (i-1)^+}$$

für alle $j + 1 \geq i \geq 0$. Die Übergangsmatrix hat demnach die Form

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & p_3 & p_4 & \dots \\ p_0 & p_1 & p_2 & p_3 & p_4 & \dots \\ 0 & p_0 & p_1 & p_2 & p_3 & \dots \\ 0 & 0 & p_0 & p_1 & p_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Es ist intuitiv klar, dass die Warteschlange über alle Schranken wächst, wenn die mittlere Anzahl der pro Bedienungsperiode im System erscheinenden Kunden $\sum_{k \geq 1} k p_k$ größer als 1 ist, der Server also langsamer arbeitet als die Ankunftsrate des Systems gebietet. Gilt dagegen $\sum_{k \geq 1} k p_k < 1$, strebt die Schlängellänge M_n für $n \rightarrow \infty$ gegen eine stationäre Verteilung, d.h.

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_n = j | M_0 = i) = \pi_j$$

für alle $j \geq 0$ (unabhängig von i), wobei $\sum_{j \geq 0} \pi_j = 1$. Von besonderem Interesse sind dann Größen wie der mittlere Zeiteanteil, den der Server unbeschäftigt ist, gegeben durch π_0 , oder auch die mittlere Wartezeit eines Kunden im Gleichgewicht, gegeben durch $\sum_{j \geq 0} (j+1)\pi_j$. Nähere Erläuterungen bedürfen allerdings noch weiterer Überlegungen, für die wir auf 2.9.4 verweisen.

2.1.12 Ein Lagerhaltungsmodell

Betrachten wir als nächstes die Situation eines Auslieferungslagers, in dem ein bestimmtes, laufend nachgefragtes Gut gelagert wird. Wir nehmen an, dass der Lagerbestand immer am Ende der Lagerperioden, numeriert mit $n = 0, 1, 2, \dots$, überprüft und gegebenenfalls aufgefüllt wird. Die Gesamtnachfrage in den einzelnen Perioden sei eine Folge X_1, X_2, \dots identisch verteilter Zufallsgrößen mit Verteilung $(p_k)_{k \geq 0}$. Die Lagerhaltungspolitik ist durch zwei Parameter $s < S$ determiniert: Fällt der Lagerbestand bis zum Ende einer Periode auf einen Wert echt kleiner als s , wird dieser sofort auf den Bestand S aufgestockt. Liegt der Bestand dagegen zwischen s und S , geschieht nichts bis zur nächsten Überprüfung [Abb. 2.5]. Sei M_n der Lagerbestand am Ende der n -ten Periode unmittelbar vor einer eventuellen Aufstockung. Dann hat diese Folge den Zustandsraum

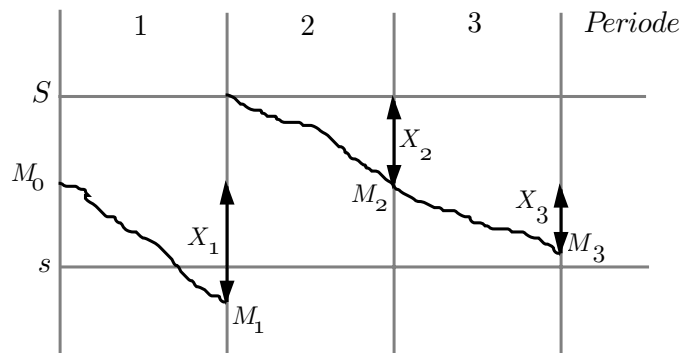


Abb. 2.5 Der Lagerbestandsprozess.

$$\mathcal{S} = \{S, S-1, \dots, 1, 0, -1, -2, \dots\},$$

wobei ein negativer Wert einen Nachfrageüberschuss bedeutet, der mittels Aufstockung sofort befriedigt wird. Aufgrund der beschriebenen Lagerhaltungspolitik besteht folgende Beziehung zwischen M_n, M_{n+1} und X_{n+1} :

$$M_{n+1} = \begin{cases} M_n - X_{n+1}, & \text{falls } s \leq M_n \leq S, \\ S - X_{n+1}, & \text{falls } M_n < s. \end{cases}$$

Nehmen wir weiter an, dass M_0, X_1, X_2, \dots stochastisch unabhängig sind, so bildet $(M_n)_{n \geq 0}$ eine DMK mit den Übergangswahrscheinlichkeiten

$$p_{ij} = \begin{cases} \mathbb{P}(X_{n+1} = i - j) = p_{i-j}, & \text{falls } s \leq M_n \leq S, \\ \mathbb{P}(X_{n+1} = S - j) = p_{S-j}, & \text{falls } M_n < s \end{cases}$$

für $i, j \in \mathcal{S}$. Wichtige Fragen für Modelle dieses Typs bilden der mittlere Anteil der Perioden n mit Nachfrageüberhang ($M_n < 0$) und auch der durchschnittliche Lagerbestand im Zeitablauf. Diese ergeben sich, wie wir noch sehen werden (ÜB 2.9.6), zu $\lim_{n \rightarrow \infty} \sum_{j < 0} p_j^{(n)}$ bzw. $\lim_{n \rightarrow \infty} \sum_{j > 0} j p_j^{(n)}$, wobei $p_j^{(n)} := \mathbb{P}(M_n = j)$.

2.1.13 Der Galton-Watson-Verzweigungsprozess

Im Folgenden stellen wir ein einfaches Modell für Populationswachstum vor. Zu Beginn (0-te Generation) bestehe die Population aus einem Mitglied, genannt *Urahne*. M_n bezeichne die Anzahl der Individuen der n -ten Generation, d.h. $M_0 = 1$. Jedes Individuum habe eine Lebenszeit von einer Zeiteinheit und produziere am Lebensende eine zufällige Anzahl von Nachkommen. Wir machen zwei weitere Annahmen über den Reproduktionsprozess der Population:

- (1) Die Anzahl der Nachkommen sei für jedes Individuum identisch verteilt gemäß $(p_k)_{k \geq 0}$, genannt *Reproduktionsverteilung*.
- (2) Individuen reproduzieren unabhängig voneinander und von der Anzahl der Mitglieder der eigenen und aller vorhergehenden Generationen.

Denken wir uns die Mitglieder der n -ten Generation durchnummeriert mit $1, \dots, M_n$, und bezeichne $X_{n,k}$ dann die Anzahl der Nachkommen des k -ten Mitglieds, so gilt offenbar

$$M_{n+1} = \sum_{k=1}^{M_n} X_{n,k}, \quad n \geq 0,$$

und aufgrund der obigen Voraussetzungen bildet $(M_n)_{n \geq 0}$ eine DMK mit Zustandsraum \mathbb{N}_0 und Übergangsverteilungen

$$\mathbb{P}^{M_{n+1}|M_n=i} = (p_k^{*(i)})_{k \geq 0}, \quad i \geq 0,$$

wobei $(p_k^{*(i)})_{k \geq 0}$ die i -fache Faltung der Reproduktionsverteilung $(p_k)_{k \geq 0}$ bezeichnet, $(p_k^{*(0)})_{k \geq 0} := (\delta_{0k})_{k \geq 0}$ (das Dirac-Maß in 0). $(M_n)_{n \geq 0}$ heißt (*einfacher*) *Galton-Watson-(Verzweigungs-)Prozess*. Der Zustand 0 ist offensichtlich absorbierend und bedeutet, dass die Population ausstirbt. Zwei Fragen drängen sich damit unmittelbar

auf: Wie groß ist die *Aussterbewahrscheinlichkeit* q in Abhängigkeit von $(p_k)_{k \geq 0}$? Wie verhält sich der Prozess auf dem Ereignis $E = \{M_n \neq 0 \text{ für alle } n \geq 0\}$? Es zeigt sich, dass die Population, sieht man von dem uninteressanten Fall “ $p_1 = 1$ ” ($\Rightarrow M_n = 1$ für alle $n \geq 0$) einmal ab, nur aussterben oder explodieren kann. Letzteres bedeutet natürlich $\lim_{n \rightarrow \infty} M_n = \infty$ auf E . Wir werden dies in 14.7 eingehender untersuchen. Dem interessierten Leser sei außerdem die historische Einführung über Verzweigungsprozesse in der Monographie von JAGERS [36] ans Herz gelegt, wo er insbesondere Informationen, teilweise amüsanter Art, zur Entstehungsgeschichte des Galton-Watson-Prozesses einschließlich seiner Namensgebung findet.

2.2 Zustandseigenschaften und Irreduzibilität

2.2.1 Irreduzibilität

Gegeben sei fortan stets eine DMK $M = (M_n)_{n \geq 0}$ mit Zustandsraum \mathcal{S} und Übergangsmatrix $P = (p_{ij})_{i,j \in \mathcal{S}}$ in einem Standardmodell $(\Omega, \mathfrak{A}, M, (\mathbb{P}_\lambda)_{\lambda \in \mathcal{P}(\mathcal{S})})$. Für $i \in \mathcal{S}$ und $A \subset \mathcal{S}$ seien ferner $\tau^0(i)$ und $\tau^0(A)$ die *Ersteintrittszeit* in den Zustand i bzw. die Menge A , und zwar

$$\tau^0(i) = \inf\{n \geq 0 : M_n = i\} \quad \text{und} \quad \tau^0(A) = \inf\{n \geq 0 : M_n \in A\}, \quad (2.9)$$

wobei $\inf \emptyset := \infty$. Diese unterscheiden sich von den in (1.23) bzw. (1.24) definierten Rückkehrzeiten $\tau(i)$ bzw. $\tau(A)$ nur dann, wenn sie den Wert 0 haben.

Um die zeitliche Evolution einer DMK zu analysieren, muss man sich zunächst klarmachen, welche Pfade (Realisierungen) durch den Zustandsraum überhaupt möglich sind, was möglicherweise zu einer Zerlegung des Zustandsraums führt. Grundlegend für die Beantwortung der hiermit verknüpften Frage, welche Zustände von einem beliebigen Ausgangszustand $i \in \mathcal{S}$ erreicht werden können, ist die folgende

Definition 2.1. Gegeben $i, j \in \mathcal{S}$, heißt j *erreichbar von* i , kurz $i \rightarrow j$, wenn

$$\mathbb{P}_i(\tau^0(j) < \infty) > 0.$$

Mit anderen Worten: j ist erreichbar von i , wenn die Kette, in i startend, den Zustand j mit positiver Wahrscheinlichkeit irgendwann erreicht.

Da $n = 0$ in (2.9) zugelassen ist, folgt $i \rightarrow i$ für alle $i \in \mathcal{S}$, denn

$$\mathbb{P}_i(\tau^0(i) < \infty) \geq \mathbb{P}_i(\tau^0(i) = 0) = 1.$$

Das wohl nützlichste Kriterium für Erreichbarkeit lautet:

Lemma 2.2. Für alle $i, j \in \mathcal{S}$ gilt:

$$i \rightarrow j \Leftrightarrow p_{ij}^{(n)} > 0 \text{ für ein } n \geq 0.$$

Beweis. “ \Rightarrow ” Aus der Inklusion

$$\{\tau^0(j) < \infty\} = \sum_{n \geq 0} \{\tau^0(j) = n\} \subset \bigcup_{n \geq 0} \{M_n = j\}$$

folgt in Kombination mit der Voraussetzung

$$0 < \mathbb{P}_i(\tau^0(j) < \infty) \leq \mathbb{P}_i\left(\bigcup_{n \geq 0} \{M_n = j\}\right) \leq \sum_{n \geq 0} p_{ij}^{(n)}$$

und somit $p_{ij}^{(n)} > 0$ für mindestens ein $n \geq 0$.

“ \Leftarrow ” Umgekehrt liefert $p_{ij}^{(n)} > 0$ zusammen mit der Inklusion

$$\{M_n = j\} \subset \{\tau^0(j) \leq n\} \subset \{\tau^0(j) < \infty\}$$

offenkundig

$$0 < p_{ij}^{(n)} \leq \mathbb{P}_i(\tau^0(j) < \infty). \quad \square$$

Während Erreichbarkeit für jedes $i \in \mathcal{S}$ all diejenigen Zustände $j \in \mathcal{S}$ beschreibt, die von i aus irgendwann mit positiver Wahrscheinlichkeit erreicht werden, richtet sich der nachfolgende Begriff an die Frage, für welche Zustände j auch ein Rückkehrpfad positiver Wahrscheinlichkeit nach i existiert.

Definition 2.3. Zwei Zustände $i, j \in \mathcal{S}$ heißen *kommunizierend* oder *verbunden*, kurz $i \leftrightarrow j$, wenn $i \rightarrow j$ und $j \rightarrow i$ gilt.

Mit Hilfe der Verbundenheit erhalten wir eine Zerlegung des Zustandsraums, denn:

Lemma 2.4. Verbundenheit (\leftrightarrow) bildet eine Äquivalenzrelation.

Beweis. Reflexivität ($i \leftrightarrow i$) und Symmetrie ($i \leftrightarrow j \Leftrightarrow j \leftrightarrow i$) sind offensichtlich. Für den Nachweis der Transitivität seien $i \leftrightarrow j$ und $j \leftrightarrow k$ für $i, j, k \in \mathcal{S}$ angenommen. Nach Lemma 2.2 existieren $m, n \geq 0$, so dass $p_{ij}^{(m)} > 0$ und $p_{jk}^{(n)} > 0$. Vermöge der Kolmogorov-Chapman-Gleichungen (Korollar 1.7) folgt dann aber

$$p_{ik}^{(m+n)} = \sum_{l \in \mathcal{S}} p_{il}^{(m)} p_{lk}^{(n)} \geq p_{ij}^{(m)} p_{jk}^{(n)} > 0$$

und somit $i \rightarrow k$. Analog zeigt man $k \rightarrow i$, so dass insgesamt $i \leftrightarrow k$ gilt. \square

Die Äquivalenzklassen bezüglich “ \leftrightarrow ”, im Folgenden nur *Klassen* genannt, zerlegen den Zustandsraum \mathcal{S} in disjunkte Teilmengen kommunizierender Zustände, bilden also eine Partition von \mathcal{S} . Die zu $i \in \mathcal{S}$ gehörende Klasse bezeichnen wir mit \mathcal{C}_i , d.h.

$$\mathcal{C}_i := \{j \in \mathcal{S} : i \leftrightarrow j\}.$$

Im einfachsten Fall gibt es nur eine Klasse, was zu folgender Definition führt:

Definition 2.5. Eine DMK M heißt *irreduzibel*, wenn alle Zustände kommunizieren, d.h., wenn $i \leftrightarrow j$ oder, was äquivalent ist, $\mathcal{C}_i = \mathcal{C}_j$ für alle $i, j \in \mathcal{S}$ gilt.

Da Irreduzibilität im Grunde genommen eine Eigenschaft der Übergangsmatrix P darstellt, nämlich $\sup_{n \geq 0} p_{ij}^{(n)} > 0$ für alle $i, j \in \mathcal{S}$, spricht man auch von einem irreduziblen P . Dies ist offensichtlich insbesondere dann der Fall, wenn P^n für ein $n \geq 1$ eine *positive Matrix* bildet ($P^n > 0$), d.h. aus lauter positiven Komponenten besteht.

Beispiel 2.6. [☞ 2.1.1] Aus $0 < p, q < 1$ folgt hier offenkundig $P > 0$ und damit Irreduzibilität. Falls $p = 0$ oder $q = 0$, ist der Zustand 0 oder 1 absorbierend und somit nur mit sich selbst verbunden. Es folgt in beiden Fällen $\mathcal{C}_0 = \{0\}$ und $\mathcal{C}_1 = \{1\}$. Es bleiben die Fälle $p = 1, q > 0$ und $q = 1, p > 0$, wobei es aus Symmetriegründen reicht, den ersten zu betrachten: Falls $p = 1$ und $q > 0$, so springt die Kette stets von 0 sofort nach 1 verharret dort eine geometrisch verteilte Zeit T ($\mathbb{P}_1(T = n) = q(1 - q)^n$ für alle $n \geq 0$) und springt dann zurück nach 0. Es liegt somit erneut Irreduzibilität vor.

Beispiel 2.7. [☞ 2.1.3] Irrfahrten mit reflektierenden Barrieren 0 und N sind stets irreduzibel, denn unter der Voraussetzung $p \in (0, 1)$ folgt für alle $j > i \geq 0$

$$p_{ij}^{(j-i)} \geq \prod_{k=1}^{j-i} p_{i+k-1, i+k} = \begin{cases} p^{j-i}, & \text{falls } i \geq 1 \\ p^{j-i-1}, & \text{falls } i = 0 \end{cases} > 0 \quad (2.10)$$

und analog

$$p_{ji}^{(j-i)} \geq \prod_{k=1}^{j-i} p_{i+k, i+k-1} = \begin{cases} (1-p)^{j-i}, & \text{falls } j \leq N-1 \\ (1-p)^{j-i-1}, & \text{falls } j = N \end{cases} > 0, \quad (2.11)$$

d.h. $i \leftrightarrow j$. Dasselbe gilt auch für partiell reflektierende Barrieren (p_{00} oder $p_{NN} \in (0, 1)$).

Beispiel 2.8. [☞ 2.1.4] Liegen absorbierende Barrieren vor, so sind diese nur mit sich selbst verbunden, während alle Zustände dazwischen, also $1, \dots, N-1$, weiter kommunizieren, da für diese (2.10) und (2.11) gültig bleiben. Die Klassen lauten demnach $\{0\}, \{1, \dots, N-1\}$ und $\{N\}$.

Beispiel 2.9. [☞ 2.1.8] Symmetrische Irrfahrten auf $(\mathbb{Z}^d, \|\cdot\|_r)$, $r = 1$ oder $r = \infty$, sind stets irreduzibel, denn jeder Pfad auf dem Gitter von irgendeinem Punkt x zu einem anderen Punkt y hat die positive Wahrscheinlichkeit $(2d)^{-n}$ bzw. $(3^d - 1)^{-n}$, wenn n die Anzahl der Kanten des Pfades (Pfadlänge in der ℓ_1 - bzw. ℓ_∞ -Norm) bezeichnet.

Beispiel 2.10. [☞ 2.1.10] Für allgemeine diskrete Random Walks auf \mathbb{Z}^d mit Zuwachsverteilung $(p_k)_{k \in \mathbb{Z}^d}$ ist die Frage der Irreduzibilität schwerer, und vor allem nicht einheitlich zu beantworten. Wir machen hier keinen Versuch einer vollständigen Klassifikation und betrachten zudem nur den eindimensionalen Fall ($d = 1$).

Falls $p_k = 0$ für alle $k < 0$, so hat der Random Walk $M_n = M_0 + \sum_{j=1}^n X_j$ nicht-negative Zuwächse und ist entweder \mathbb{P}_i -f.s. konstant ($p_0 = 1$) oder driftet \mathbb{P}_i -f.s. monoton nach $+\infty$ ($p_0 < 1$) für alle $i \in \mathbb{Z}$. Jeder Zustand ist folglich nur mit sich selbst verbunden, d.h. $\mathcal{C}_i = \{i\}$ für alle $i \in \mathbb{Z}$. Dasselbe gilt aus Symmetriegründen, wenn $p_k = 0$ für alle $k > 0$.

Aufgrund der additiven Struktur von M_n folgt weiter, dass, wenn die X_j f.s. nur Werte in einer Untergruppe $m\mathbb{Z}$, $m > 1$, annehmen, wenn also $p_k = 0$ für alle $k \notin m\mathbb{Z}$ gilt, M_n unter \mathbb{P}_i f.s. nur Werte in $i + m\mathbb{Z}$ annimmt für alle $n \geq 0$. Die Kette ist also wiederum reduzibel mit $\mathcal{C}_{r+mn} \subset r + m\mathbb{Z}$ für alle $0 \leq r < m$ und $n \in \mathbb{Z}$. Die tatsächliche Form der einzelnen Klassen hängt von der weiteren Struktur der Zuwachsverteilung $(p_k)_{k \in \mathbb{Z}}$ ab.

Betrachten wir abschließend die Situation $p_1 > 0$ und $p_{k_0} > 0$ für ein $k_0 \leq -1$. In diesem Fall ist $(M_n)_{n \geq 0}$ irreduzibel, denn für beliebige $i, j \in \mathbb{Z}$, $i < j$, folgt

$$p_{ij}^{(j-i)} = \mathbb{P}_i \left(\sum_{k=1}^{j-i} X_k = j-i \right) \geq \mathbb{P}_i(X_1 = 1, \dots, X_{j-i} = 1) = p_1^{j-i} > 0,$$

d.h. $i \rightarrow j$, sowie umgekehrt, falls $j + lk_0 \leq i < j + (l-1)k_0$,

$$\begin{aligned} p_{ji}^{(l(1-k_0)+i-j)} &\geq \mathbb{P}_j(X_1 = \dots = X_l = k_0, X_{l+1} = \dots = X_{l(1-k_0)+i-j} = 1) \\ &= p_{k_0}^l p_1^{-lk_0+i-j} > 0, \end{aligned}$$

also $j \rightarrow i$. Entsprechend ergibt sich die Irreduzibilität im Fall $p_{-1} > 0$ und $p_{k_0} > 0$ für ein $k_0 \geq 1$.

Als nächstes wollen wir uns anschauen, welche Teilmengen von Zuständen von anderen Teilmengen des Zustandsraums aus erreichbar sind.

Definition 2.11. Eine Teilmenge $\mathcal{C} \subset \mathcal{S}$ heißt *abgeschlossen*, wenn

$$\mathbb{P}_i(\tau^0(\mathcal{C}^c) = \infty) = 1$$

für alle $i \in \mathcal{C}$. Ist speziell $\mathcal{C} = \{i\}$ abgeschlossen, nennen wir i (wie bereits geschehen) *absorbierend*.

Eine abgeschlossene Teilmenge von Zuständen wird demnach bei Erreichen niemals mehr verlassen. Als nützliches Kriterium notieren wir:

Lemma 2.12.

- (a) $\mathcal{C} \subset \mathcal{S}$ ist genau dann abgeschlossen, wenn $p_{ij} = 0$ für alle $i \in \mathcal{C}$ und $j \in \mathcal{C}^c$ gilt.
- (b) Ein Zustand $i \in \mathcal{S}$ ist genau dann absorbierend, wenn $p_{ii} = 1$.

Beweis. Wir brauchen nur Teil (a) zu zeigen, da dieser (b) als Spezialfall enthält.

“ \Rightarrow ” Ist \mathcal{C} abgeschlossen, folgt wegen $\tau^0(j) \geq \tau^0(\mathcal{C}^c)$ für alle $j \in \mathcal{C}^c$

$$0 = \mathbb{P}_i(\tau^0(\mathcal{C}^c) < \infty) \geq \mathbb{P}_i(\tau^0(j) < \infty) \geq \mathbb{P}_i(\tau^0(j) \leq n) \geq \mathbb{P}_i(M_n = j) = p_{ij}^{(n)}$$

für alle $n \geq 0$ und $i \in \mathcal{C}$, insbesondere $p_{ij} = 0$.

“ \Leftarrow ” Ist $p_{ij} = 0$ für alle $i \in \mathcal{C}$ und $j \in \mathcal{C}^c$, folgt zuerst

$$\mathbb{P}_i(\tau^0(\mathcal{C}^c) = 1) = \sum_{j \in \mathcal{C}^c} p_{ij} = 0$$

für alle $i \in \mathcal{C}$. Analog ergibt sich für $n \geq 2$

$$\begin{aligned} \mathbb{P}_i(\tau^0(\mathcal{C}^c) = n) &= \mathbb{P}_i(M_1 \in \mathcal{C}, \dots, M_{n-1} \in \mathcal{C}, M_n \in \mathcal{C}^c) \\ &= \sum_{i_1 \in \mathcal{C}} \dots \sum_{i_{n-1} \in \mathcal{C}} \sum_{j \in \mathcal{C}^c} p_{ii_1} p_{i_1 i_2} \dots p_{i_{n-2} i_{n-1}} \underbrace{p_{i_{n-1} j}}_{=0} = 0 \end{aligned}$$

und deshalb insgesamt $\mathbb{P}_i(\tau^0(\mathcal{C}^c) < \infty) = 0$. □

Die Abschätzung unter “ \Rightarrow ” hat gezeigt, dass für eine abgeschlossene Menge \mathcal{C} sogar $p_{ij}^{(n)} = 0$ für alle $i \in \mathcal{C}$, $j \in \mathcal{C}^c$ und alle $n \geq 0$ gilt. Ist \mathcal{S} endlich, so ergibt sich aus Lemma 2.12(a) für die Übergangsmatrix P der DMK die Blockgestalt

$$P = \begin{pmatrix} P_{\mathcal{C}} & 0 \\ * & * \end{pmatrix}, \quad P_{\mathcal{C}} := (p_{ij})_{i,j \in \mathcal{C}},$$

sofern man die Elemente der abgeschlossenen Menge \mathcal{C} mit $1, \dots, |\mathcal{C}|$ und die übrigen mit $|\mathcal{C}| + 1, \dots, |\mathcal{S}|$ durchnummeriert. Zerfällt \mathcal{S} gar in lauter abgeschlossene Mengen $\mathcal{C}_1, \dots, \mathcal{C}_k$, folgt bei entsprechender Numerierung der Zustände

$$P = \begin{pmatrix} P_{\mathcal{C}_1} & 0 & \dots & 0 \\ 0 & P_{\mathcal{C}_2} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & P_{\mathcal{C}_k} \end{pmatrix}.$$

Generell kann man eine abgeschlossene Menge \mathcal{C} als *absorbierenden Makrozustand* interpretieren, der zwar nie mehr verlassen, i.A. aber von Zuständen außerhalb von \mathcal{C} erreicht werden kann. Beachte, dass die zuvor eingeführten Irreduzibilitätsklassen (bzgl. \leftrightarrow) nicht abgeschlossen zu sein brauchen. Man denke etwa an die Klasse $\{1, \dots, N-1\}$ einer Irrfahrt auf $\{0, \dots, N\}$ mit absorbierenden Barrieren.

2.2.2 Periodizität

Um das evolutionäre Verhalten einer DMK zu verstehen, benötigt man den Begriff der Periodizität eines Zustands. Es ist nämlich durchaus möglich, dass gewisse Zustände nur zu bestimmten Zeitpunkten erreicht werden können. Um die richtige Anschauung zu bekommen, bemühen wir das Beispiel der symmetrischen Irrfahrt auf $(\mathbb{Z}^2, |\cdot|_1)$. Startet diese im Ursprung, so kann der Zustand 0 offenbar nur zu geraden Zeitpunkten $2, 4, 6, \dots$ wieder erreicht werden, weil jeder Vorwärtsschritt in die x - oder y -Richtung eines kompensierenden Rückwärtsschritts bedarf.

Die anschließende Definition präzisiert, wie sämtliche Zustände des Zustandsraums als periodisch oder aperiodisch klassifiziert werden können:

Definition 2.13. Ein Zustand $i \in \mathcal{S}$ heißt *periodisch mit Periode d* oder kurz *d -periodisch*, falls

$$d = d(i) := \text{ggT}\{n \geq 1 : p_{ii}^{(n)} > 0\}, \quad (2.12)$$

wobei “ggT” den größten gemeinsamen Teiler bezeichnet. Ein Zustand mit Periode 1 heißt auch *aperiodisch*.

Für das Beispiel der symmetrischen Irrfahrt auf $(\mathbb{Z}^2, |\cdot|_1)$ hat der Zustand 0 somit die Periode 2, denn $p_{00}^{(2n+1)} = 0$ für alle $n \geq 0$ impliziert zunächst $d(0) \geq 2$, und die Gleichheit folgt dann aus $p_{00}^{(2)} = \frac{1}{4} > 0$.

Dieselbe Argumentation liefert allgemein: Gilt $p_{ii}^{(n)} = 0$ für alle $n \notin d\mathbb{N}$ sowie $p_{ii}^{(d)} > 0$, so ist i d -periodisch. Insbesondere ist jeder Zustand i mit $p_{ii} > 0$ aperi-

odisch. Für ein weiteres Kriterium benötigen wir das folgende elementare Ergebnis aus der Zahlentheorie:

Lemma. Sei $H \subset \mathbb{N}$ eine Halbgruppe bezüglich der Addition mit $\text{ggT } d$. Dann ist $d\mathbb{N} - H$ endlich, d.h., H enthält bis auf endlich viele Ausnahmen alle Zahlen $d, 2d, 3d, \dots$

Beweis. Zuerst überlegt man sich leicht, dass ein $k \geq 2$ und teilerfremde $n_1, \dots, n_k \in \mathbb{N}$ mit $n_1d, \dots, n_kd \in H$ existieren. Aus der Teilerfremdheit folgt weiter $n_1\mathbb{Z} \oplus \dots \oplus n_k\mathbb{Z} = \mathbb{Z}$ (andernfalls ergäbe sich $c\mathbb{Z}$ für ein $c \geq 2$ und dann mit $n_1 + \dots + n_k = mc$ für ein $m \in \mathbb{N}$ ein Widerspruch) und folglich die Existenz ganzer Zahlen m_1, \dots, m_k mit $\sum_{j=1}^k m_j n_j = 1$. Setze $n = \sum_{j=1}^k n_j$, und wähle $m \in \mathbb{N}$ so groß, dass $m + (n-1)m_j \geq 0$ für $j = 1, \dots, k$ und $mn \in H$. Aus der Halbgruppeneigenschaft von H ergibt sich dann

$$\begin{aligned} mn, (mn+1)d &= \sum_{j=1}^k (m+m_j)n_jd, \dots, (mn+n-1)d \\ &= \sum_{j=1}^k (m+(n-1)m_j)n_jd \in H \end{aligned}$$

und daraus bei nochmaliger Verwendung $ld \in H$ für alle $l \geq mn$. \square

Lemma 2.14. Ein Zustand $i \in \mathcal{S}$ ist genau dann d -periodisch, wenn $p_{ii}^{(n)} = 0$ für alle $n \notin d\mathbb{N}$ und $p_{ii}^{(md)} > 0$ für alle hinreichend großen $m \geq m_0$ gilt.

Beweis. Offensichtlich reicht es, die Notwendigkeit der Charakterisierung nachzuweisen. Da aber die Menge $H = \{n \in \mathbb{N} : p_{ii}^{(n)} > 0\}$ eine Halbgruppe bildet (wegen $p_{ii}^{(m)} > 0, p_{ii}^{(n)} > 0 \Rightarrow p_{ii}^{(m+n)} \geq p_{ii}^{(m)} p_{ii}^{(n)} > 0$), ergibt sich das Gewünschte unmittelbar als Konsequenz des obigen zahlentheoretischen Lemmas. \square

Zum Abschluss zeigen wir, dass die Periode eines Zustands i auch über die Verteilung der Rückkehrzeit $\tau(i)$ gemäß (1.23) [nicht $\tau^0(i)$] unter \mathbb{P}_i charakterisiert werden kann, wobei an $f_{ij}^{(n)} = \mathbb{P}_i(\tau(j) = n)$ für $n \geq 0$ erinnert sei.

Lemma 2.15. Ist $i \in \mathcal{S}$ d -periodisch und $d' := \text{ggT}\{n \geq 1 : f_{ii}^{(n)} > 0\}$, so gilt $d = d'$.

Beweis. Aus $f_{ii}^{(n)} \leq p_{ii}^{(n)}$ für alle $n \geq 1$ folgt $d' \geq d$. Für die umgekehrte Ungleichung notieren wir als erstes, dass $p_{ii}^{(md)} = \sum_{k=1}^m f_{ii}^{(kd)} p_{ii}^{((m-k)d)}$ für alle $m \geq 1$ gilt (1.30). Aus $p_{ii}^{(d)} > 0$ folgt nun $f_{ii}^{(d)} > 0$ und somit $d' = d$. Nehmen wir also $p_{ii}^{(d)} = 0$ und $d' > d$ an. Nach Lemma 2.14 existiert ein $m \geq 2$, so dass $md \notin d'\mathbb{N}$ und $p_{ii}^{(md)} > 0$. Die obige Formel sichert dann $f_{ii}^{(k_1d)} > 0$ und $p_{ii}^{((m-k_1)d)} > 0$ für ein $1 \leq k_1 < m$, was $(m-k_1)d \notin d'\mathbb{N}$ wegen $k_1d \in d'\mathbb{N}$ nach sich zieht. Im Fall

$m - k_1 = 1$, d.h. $p_{ii}^{(d)} > 0$, steht dies im Widerspruch zur Annahme. Falls $m - k_1 \geq 2$, finden wir aber durch Anwendung desselben Schlusses wie zuvor auf $p_{ii}^{((m-k_1)d)}$ ein $k_1 < k_2 < m$, so dass $f_{ii}^{(k_2d)} > 0$ und $p_{ii}^{((m-k_2)d)} > 0$, also $(m - k_2)d \notin d'\mathbb{N}$. Nach einer endlichen Wiederholung dieses Vorgehens gelangen wir schließlich zu einem k_l mit $m - k_l = 1$ und $p_{ii}^{((m-k_l)d)} = p_{ii}^{(d)} > 0$, d.h. ebenfalls zu einem Widerspruch zur Annahme. \square

2.2.3 Zyklische Zerlegung einer DMK

Betrachten wir eine DMK, für die ein Zustand i existiert, den sie unter \mathbb{P}_i mit Wahrscheinlichkeit 1 unendlich oft aufsucht. Es ist aufgrund der (starken) Markov-Eigenschaft und der zeitlichen Homogenität intuitiv klar, dass die Kette mit jedem Besuch des Zustands i einen Neuanfang macht, d.h., danach unabhängig von ihrer Vergangenheit weiterläuft, und zwar genauso, als wäre sie gerade in diesem Zustand gestartet. Mit anderen Worten: Die Kette zerfällt in unabhängige, identisch verteilte *Zyklen* zufälliger Länge, an deren Anfang und Ende ein Besuch in i steht. Diese anschauliche Begründung gilt es im Folgenden mathematisch zu beweisen, wobei wir auch die Möglichkeit einschließen wollen, dass der betrachtete Zustand nur endlich oft aufgesucht wird.

Es bezeichne $(\sigma_n(i))_{n \geq 1}$ die Folge der sukzessiven Rückkehrzeiten in den Zustand i , d.h.

$$\sigma_n(i) := \inf\{k \geq \sigma_{n-1}(i) + 1 : M_k = i\},$$

für $n \geq 1$, wobei $\sigma_0(i) \equiv 0$. Beachte $\sigma_1(i) = \tau(i)$ gemäß (1.23). Aus $\sigma_n(i) = \infty$ für ein n folgt natürlich $\sigma_m(i) = \infty$ für alle $m > n$. Die Zuwächse dieser Folge bezeichnen wir mit $\tau_n(i)$, d.h.

$$\tau_n(i) = \sigma_n(i) - \sigma_{n-1}(i), \quad n \geq 1,$$

wobei $\tau_n(i) := \infty$ auf $\{\sigma_n(i) = \infty\}$. Offensichtlich gilt dann

$$\{\sigma_n(i) < \infty\} = \{\tau_1(i) < \infty, \dots, \tau_n(i) < \infty\} \quad (2.13)$$

für alle $n \geq 1$. Schließlich setzen wir für $n \geq 0$

$$Z_n := \begin{cases} (\tau_{n+1}(i), M_{\sigma_n(i)}, \dots, M_{\sigma_{n+1}(i)-1}), & \text{falls } \sigma_n(i) < \infty, \\ (\infty, \Delta, \Delta, \dots), & \text{falls } \sigma_n(i) = \infty, \end{cases}$$

wobei Δ irgendeinen Friedhof der Kette bezeichnet (\mathbb{E}^{\otimes} Abschnitt 1.4). Die Z_n bilden die bereits erwähnten *Zyklen* (*Segmente*, *Exkursionen*) der Kette zwischen den Aufenthalten im Zustand i , wobei die Festlegung auf $\{\sigma_n(i) = \infty\}$ lediglich der Definiertheit auf dem gesamten zugrundeliegenden W-Raum dient. Wir interessieren uns im folgenden für die gemeinsame Verteilung von Z_0, \dots, Z_{n-1} unter der Bedingung, dass M den Zustand i überhaupt n -mal aufsucht, d.h., bedingt unter dem Ereignis $\{\sigma_n(i) < \infty\}$. Dazu bedarf es allerdings der Voraussetzung $\mathbb{P}_i(\sigma_n(i) < \infty) > 0$.

Zum besseren Verständnis der anschließenden Rechnungen notieren wir, dass $\tau(i) = \tau_1(i) = \rho(M)$ und $Z_0 = \theta(M)$ für geeignete meßbare Abbildungen ρ, θ gilt und allgemein für $n \geq 0$

$$\tau_{n+1}(i) = \rho(M^{(\sigma_n(i))}), \quad Z_n = \theta(M^{(\sigma_n(i))}) \quad (2.14)$$

wobei wie bisher $M^{(n)} = (M_k)_{k \geq n}$. Aus (2.13), (2.14) und der starken Markov-Eigenschaft folgt

$$\begin{aligned} \mathbb{P}_i(\sigma_n(i) < \infty) &= \mathbb{P}_i(\tau_1(i) < \infty, \dots, \tau_n(i) < \infty) \\ &= \int_{\{\tau_1(i) < \infty, \dots, \tau_{n-1}(i) < \infty\}} \mathbb{P}(\rho(M^{(\sigma_{n-1}(i))}) < \infty | M_{\sigma_{n-1}(i)}) d\mathbb{P}_i \\ &= \int_{\{\tau_1(i) < \infty, \dots, \tau_{n-1}(i) < \infty\}} \mathbb{P}_{M_{\sigma_{n-1}(i)}}(\rho(M) < \infty) d\mathbb{P}_i \\ &= \mathbb{P}_i(\tau_1(i) < \infty, \dots, \tau_{n-1}(i) < \infty) \mathbb{P}_i(\tau(i) < \infty) \end{aligned}$$

und dann induktiv

$$\mathbb{P}_i(\sigma_n(i) < \infty) = \mathbb{P}_i(\tau_1(i) < \infty, \dots, \tau_n(i) < \infty) = \mathbb{P}_i(\tau(i) < \infty)^n. \quad (2.15)$$

Keht die Kette also mindestens einmal mit positiver Wahrscheinlichkeit bzw. fast sicher nach i zurück, so auch n -mal für jedes $n \geq 2$. Dies führt nun zu dem wichtigen

Satz 2.16. *Gegeben ein $i \in \mathcal{S}$ mit $\mathbb{P}_i(\tau(i) < \infty) > 0$, gilt für jedes $n \geq 1$ unter*

$$\widehat{\mathbb{P}}_i^{(n)} := \mathbb{P}_i(\cdot | \sigma_n(i) < \infty):$$

Z_0, \dots, Z_{n-1} sind stochastisch unabhängig und identisch verteilt mit

$$\widehat{\mathbb{P}}_i^{(n)}(Z_0 \in \cdot) = \mathbb{P}_i(Z_0 \in \cdot | \tau(i) < \infty).$$

Ist $\tau(i)$ sogar \mathbb{P}_i -f.s. endlich, d.h. $\widehat{\mathbb{P}}_i^{(n)} = \mathbb{P}_i$ für jedes $n \geq 1$, so bildet $(Z_n)_{n \geq 0}$ unter \mathbb{P}_i eine unabhängige Folge identisch verteilter Zufallsvariablen.

Beweis. Zur Abkürzung schreiben wir im Folgenden σ_n, τ_n für $\sigma_n(i), \tau_n(i)$. Es genügt,

$$\mathbb{P}_i(Z_0 \in A_0, \dots, Z_{n-1} \in A_{n-1}, \tau_1 < \infty, \dots, \tau_n < \infty) = \prod_{k=0}^{n-1} \mathbb{P}_i(Z_0 \in A_k, \tau(i) < \infty) \quad (2.16)$$

für alle $n \geq 1$ und messbaren A_0, \dots, A_{n-1} zu zeigen, weil dann mit (2.12) und (2.14)

$$\widehat{\mathbb{P}}_i^{(n)}(Z_0 \in A_0, \dots, Z_{n-1} \in A_{n-1}) = \frac{\mathbb{P}_i(Z_0 \in A_0, \dots, Z_{n-1} \in A_{n-1}, \tau_1 < \infty, \dots, \tau_n < \infty)}{\mathbb{P}_i(\tau_1 < \infty, \dots, \tau_n < \infty)}$$

$$\begin{aligned}
&= \prod_{k=0}^{n-1} \frac{\mathbb{P}_i(Z_0 \in A_k, \tau(i) < \infty)}{\mathbb{P}_i(\tau(i) < \infty)} \\
&= \prod_{k=0}^{n-1} \mathbb{P}_i(Z_0 \in A_k | \tau(i) < \infty).
\end{aligned}$$

Für den Beweis von (2.16) führt man eine Induktion über n durch, wobei im Fall $n = 1$ nichts zu zeigen ist. Für den Induktionsschritt $n - 1 \rightarrow n$ seien A_0, \dots, A_n messbare Mengen aus der auf dem Bildraum der Z_k zu wählenden σ -Algebra (\mathfrak{E} hierzu den Beweis von Satz 1.18(f)). Dann folgt unter Hinweis auf (2.14)

$$\begin{aligned}
&\mathbb{P}_i(Z_0 \in A_0, \dots, Z_n \in A_n, \tau_1 < \infty, \dots, \tau_{n+1} < \infty) \\
&= \int_{\{Z_k \in A_k, \tau_{k+1} < \infty, 0 \leq k < n\}} \mathbb{P}(\theta(M^{(\sigma_n)}) \in A_n, \rho(M^{(\sigma_n)}) < \infty | M_{\sigma_n}) d\mathbb{P}_i \\
&= \mathbb{P}_i(Z_k \in A_k, \tau_{k+1} < \infty, 0 \leq k < n) \mathbb{P}_i(\theta(M) \in A_n, \rho(M) < \infty), \\
&= \mathbb{P}_i(Z_k \in A_k, \tau_{k+1} < \infty, 0 \leq k < n) \mathbb{P}_i(Z_0 \in A_n, \tau(i) < \infty),
\end{aligned}$$

wobei die starke Markov-Eigenschaft für die zweite sowie die zeitliche Homogenität in Verbindung mit $M_{\sigma_n} = i$ auf $\{\sigma_n < \infty\}$ für die dritte Zeile verwendet wurden. \square

Als direkte Folgerung, die insbesondere (2.15) verallgemeinert, notieren wir:

Korollar 2.17. *In der Situation von Satz 2.16 gilt für jedes $n \geq 1$: Unter $\widehat{\mathbb{P}}_i^{(n)}$ sind $\tau_1(i), \dots, \tau_n(i)$ unabhängige, identisch verteilte Zufallsgrößen mit Werten in \mathbb{N} , also*

$$\widehat{\mathbb{P}}_i^{(n)}(\tau_1(i) = k_1, \dots, \tau_n(i) = k_n) = \prod_{j=1}^n \widehat{\mathbb{P}}_i^{(n)}(\tau(i) = k_j) = \prod_{j=1}^n \frac{\mathbb{P}_i(\tau(i) = k_j)}{\mathbb{P}_i(\tau(i) < \infty)}$$

für alle $(k_1, \dots, k_n) \in \mathbb{N}^n$. Falls $\mathbb{P}_i(\tau(i) < \infty) = 1$, so bildet $(\tau_n(i))_{n \geq 1}$ unter \mathbb{P}_i eine unabhängige Folge identisch verteilter Zufallsgrößen.

Es gibt eine ziemlich offensichtliche Verallgemeinerung von Satz 2.16 für den Fall, wenn die Kette nicht im Zustand i startet. Wir beschränken uns auf die Angabe des Resultats, überlassen den Beweis jedoch dem Leser zur Übung:

Satz 2.18. *Seien $i \in \mathcal{S}$, $\lambda \in \mathcal{P}(\mathcal{S})$ mit $\mathbb{P}_i(\tau(i) < \infty) > 0$ und $\mathbb{P}_\lambda(\tau(i) < \infty) > 0$. Dann gilt für jedes $n \geq 1$: Unter $\widehat{\mathbb{P}}_{\lambda i}^{(n)} := \mathbb{P}_\lambda(\cdot | \sigma_n(i) < \infty)$ sind Z_0, \dots, Z_{n-1} stochastisch unabhängig und Z_1, \dots, Z_{n-1} ferner identisch verteilt mit*

$$\widehat{\mathbb{P}}_{\lambda i}^{(n)}(Z_1 \in \cdot) = \mathbb{P}_i(Z_0 \in \cdot | \tau(i) < \infty). \quad (2.17)$$

Ist $\tau(i)$ sogar sowohl unter \mathbb{P}_i als auch unter \mathbb{P}_λ f.s. endlich, d.h. $\widehat{\mathbb{P}}_{\lambda i}^{(n)} = \mathbb{P}_\lambda$ für jedes $n \geq 1$, so bildet $(Z_n)_{n \geq 0}$ unter \mathbb{P}_λ eine unabhängige Folge von Zufallsvariablen, die für $n \geq 1$ ferner identisch verteilt sind.

2.2.4 Rekurrenz und Transienz

Wir kommen als nächstes zu wichtigen Klassifikationen der Zustände im Hinblick auf die Frage, wie häufig diese von der Kette aufgesucht werden. Für $i \in \mathcal{S}$ seien $(\sigma_n(i))_{n \geq 0}$ und $\tau(i) = \sigma_1(i)$ wie im vorherigen Teilabschnitt definiert. Wir erinnern außerdem an $f_{ij}^{(n)} = \mathbb{P}_i(\tau(j) = n)$ und setzen $f_{ij}^* := \sum_{n \geq 1} f_{ij}^{(n)} = \mathbb{P}_i(\tau(j) < \infty)$ und $\mu_{ij} = \mathbb{E}_i \tau(j)$ für $i, j \in \mathcal{S}$.

Definition 2.19. Ein Zustand $i \in \mathcal{S}$ heißt

- *rekurrent* bzw. *transient*, falls $f_{ii}^* = 1$ bzw. < 1 .
- *positiv rekurrent*, falls $f_{ii}^* = 1$ und $\mu_{ii} < \infty$.
- *null-rekurrent*, falls $f_{ii}^* = 1$ und $\mu_{ii} = \infty$.

Ausgehend von einem rekurrenten Zustand i , kehrt die Kette also fast sicher irgendwann in diesen zurück, während dies für einen transienten Zustand gerade nicht der Fall ist. Unter Verwendung von (2.15) folgt sofort

Lemma 2.20. Ein Zustand $i \in \mathcal{S}$ ist genau dann rekurrent, wenn

$$\mathbb{P}_i(M_n = i \text{ u.o.}) = 1. \quad (2.18)$$

Beweis. Unter Hinweis auf (2.15) gilt $f_{ii}^* = 1$ genau dann, wenn alle $\sigma_n(i)$ \mathbb{P}_i -f.s. endlich sind, was wiederum zu (2.18) äquivalent ist, weil offenkundig

$$\mathbb{P}_i(M_n = i \text{ u.o.}) = \mathbb{P}_i(\sigma_n(i) < \infty \text{ für alle } n \geq 1). \quad \square$$

Wir können also festhalten: Ein rekurrenter Zustand i wird unter \mathbb{P}_i immer schon fast sicher unendlich oft aufgesucht, ein transienter Zustand dagegen höchstens endlich oft, wiederum unter Hinweis auf (2.15).

Die weitere Unterscheidung in positive und Null-Rekurrenz erfolgt für einen Zustand i anhand seiner *mittleren Rekurrenzzeit* μ_{ii} , die angibt, wie lange die Kette im Mittel braucht, um nach i zurückzukehren. Sie spielt für die Stabilität der MK

(Langzeitverhalten) eine entscheidende Rolle, wie wir schon bald sehen werden (☞ Abschnitt 2.5).

Es liegt auf der Hand, dass wir uns ein möglichst einfaches Kriterium wünschen, mit dem sich Zustände hinsichtlich Rekurrenz oder Transienz klassifizieren lassen. Auf dem Weg dorthin (Satz 2.22) benötigen wir zunächst ein anderes, auch sonst nützliches Ergebnis: Wir führen erzeugende Funktionen ein, indem wir für $i, j \in \mathcal{S}$

$$P_{ij}(s) := \sum_{n \geq 0} p_{ij}^{(n)} s^n \quad \text{und} \quad F_{ij}(s) := \sum_{n \geq 1} f_{ij}^{(n)} s^n, \quad s \in (-1, 1),$$

setzen, wobei an $p_{ij}^{(0)} = \delta_{ij}$ erinnert sei.

Lemma 2.21. *Für alle $i, j \in \mathcal{S}$ gilt*

$$P_{ij}(s) = \delta_{ij} + F_{ij}(s)P_{jj}(s), \quad s \in (-1, 1), \quad (2.19)$$

und speziell für $i = j$

$$P_{ii}(s) = \frac{1}{1 - F_{ii}(s)}, \quad s \in (-1, 1). \quad (2.20)$$

Beweis. Gemäß (1.30) in Beispiel 1.30 gilt $p_{ij}^{(n)} = \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}$ für alle $n \geq 1$, was bei Einsetzen in $P_{ij}(s) = \delta_{ij} + \sum_{n \geq 1} p_{ij}^{(n)} s^n$ und Verwendung der Cauchyschen Produktformel für Reihen (2.19) liefert. \square

Das angekündigte Rekurrenzkriterium ergibt sich nun als einfache Folgerung:

Satz 2.22. *Ein Zustand $i \in \mathcal{S}$ ist genau dann rekurrent, wenn*

$$\sum_{n \geq 0} p_{ii}^{(n)} = \infty.$$

Er ist somit transient genau dann, wenn $\sum_{n \geq 0} p_{ii}^{(n)} < \infty$, wobei in diesem Fall sogar $\sum_{n \geq 0} p_{ji}^{(n)} < \infty$ für alle $j \in \mathcal{S}$ gilt.

Beweis. Aufgrund monotoner Konvergenz und (2.20) gilt

$$\sum_{n \geq 0} p_{ii}^{(n)} = \lim_{s \uparrow 1} P_{ii}(s) = \lim_{s \uparrow 1} \frac{1}{1 - F_{ii}(s)} = \frac{1}{1 - f_{ii}^*},$$

woraus offenkundig die behauptete Äquivalenz folgt. Ist i transient, ergibt sich in (2.19) unter Verwendung des zuvor Gezeigten

$$\sum_{n \geq 0} p_{ji}^{(n)} = \lim_{s \uparrow 1} P_{ji}(s) = \lim_{s \uparrow 1} F_{ji}(s) P_{ii}(s) = f_{ji}^* \sum_{n \geq 0} p_{ii}^{(n)} < \infty$$

für alle $j \neq i$. □

Das erhaltene Kriterium besitzt die folgende, sehr einleuchtende Interpretation: Definiere für $i \in \mathcal{S}$ die Zählvariable

$$N(i) = \sum_{n \geq 1} \mathbf{1}_{\{M_n=i\}},$$

die die Gesamtzahl der Aufenthalte der Kette in i nach dem Zeitpunkt 0 angibt. Es folgt für alle $i, j \in \mathcal{S}$

$$\mathbb{E}_j N(i) = \mathbb{E}_j \left(\sum_{n \geq 1} \mathbf{1}_{\{M_n=i\}} \right) = \sum_{n \geq 1} \mathbb{P}_j(M_n = i) = \sum_{n \geq 1} p_{ji}^{(n)}.$$

Wählen wir $i = j$, besagt das obige Kriterium gerade, dass ein Zustand i genau dann rekurrent ist, wenn, in i startend, die erwartete Anzahl von Aufenthalten in i unendlich beträgt. Kombinieren wir dies mit (2.18), können wir also festhalten:

$$i \text{ rekurrent} \Leftrightarrow N(i) = \infty \text{ } \mathbb{P}_i\text{-f.s.} \Leftrightarrow \mathbb{E}_i N(i) = \infty.$$

Für die Zählvariable $N(i)$ haben wir demnach die (i.A. natürlich ungültige!) Aussage, dass sie genau dann unendlichen Erwartungswert besitzt, wenn sie selbst schon f.s. unendlich ist.

Zum Abschluss notieren wir noch einen Satz, der die Beziehung zwischen Rekurrenz/Transienz und der Anzahl von Aufenthalten eines Zustands weiter beleuchtet. Seinen Beweis, der die Ergebnisse des vorherigen Teilabschnitts über die zyklische Zerlegung einer DMK verwendet, überlassen wir dem Leser zur Übung:

Satz 2.23. Für alle $i, j \in \mathcal{S}$ und $k \in \mathbb{N}_0$ gilt

$$\mathbb{P}_j(N(i) = k) = \begin{cases} 1 - f_{ji}^*, & \text{falls } k = 0 \\ f_{ji}^* f_{ii}^{*k-1} (1 - f_{ii}^*) & \text{falls } k \geq 1 \end{cases}.$$

Für transientes i folgen deshalb $\mathbb{P}_j(N(i) < \infty) = 1$ und

$$\mathbb{E}_j N(i) = \frac{f_{ji}^*}{1 - f_{ii}^*} = \sum_{n \geq 1} p_{ji}^{(n)} < \infty$$

für alle $j \in \mathcal{S}$. Außerdem besitzt $N(i)$ unter \mathbb{P}_i eine geometrische Verteilung mit Parameter $1 - f_{ii}^*$, d.h. $\mathbb{P}_i(N(i) = k) = (1 - f_{ii}^*) f_{ii}^{*k}$ für alle $k \in \mathbb{N}_0$. Ist i rekurrent, gilt dagegen $\mathbb{P}_i(N(i) = \infty) = 1$ und $\mathbb{P}_j(N(i) = \infty) = f_{ji}^*$ für alle $j \in \mathcal{S}$.

2.3 Rekurrenz/Transienz von Irrfahrten auf \mathbb{Z}^d

Wir verlassen einen Abschnitt lang die allgemeine Theorie diskreter Markov-Ketten und beantworten unter Benutzung des zuvor entwickelten Rekurrenzkriteriums 2.22 die spannende Frage, wann eine Irrfahrt auf \mathbb{Z}^d mit Wahrscheinlichkeit 1 in ihren Anfangspunkt M_0 zurückkehrt, wobei wir aus Symmetriegründen $M_0 = 0$ wählen dürfen. Es gilt nämlich

$$M_n = M_0 + \sum_{k=1}^n X_k$$

mit unter jedem $\mathbb{P}_i, i \in \mathbb{Z}^d$, unabhängigen (auch von M_0), identisch verteilten X_k , deren Verteilung unter \mathbb{P}_i ferner nicht von i abhängt. Damit folgt aber für jedes $i \in \mathbb{Z}^d$

$$\begin{aligned} \mathbb{P}_i(M_n = i \text{ u.o.}) &= \mathbb{P}_i\left(i + \sum_{k=1}^n X_k = i \text{ u.o.}\right) = \mathbb{P}_i\left(\sum_{k=1}^n X_k = 0 \text{ u.o.}\right) \\ &= \mathbb{P}_0\left(\sum_{k=1}^n X_k = 0 \text{ u.o.}\right) = \mathbb{P}_0(M_n = 0 \text{ u.o.}). \end{aligned}$$

Wir sehen also: Entweder sind alle Zustände $i \in \mathbb{Z}^d$ rekurrent oder gar keiner. Der nächste Abschnitt wird zeigen, dass diese *Solidarität* hinsichtlich Rekurrenz ganz allgemein für Zustände derselben Irreduzibilitätsklasse gilt.

2.3.1 Der eindimensionale Fall

Sei also $M_n = \sum_{k=1}^n X_k$ mit (unter \mathbb{P}_0) unabhängigen, identisch verteilten X_k , wobei

$$\mathbb{P}_0(X_1 = 1) = 1 - \mathbb{P}_0(X_1 = -1) = p \in (0, 1).$$

Der Zustand 0 hat die Periode 2, wie bereits früher bemerkt, d.h. $p_{00}^{(2n+1)} = 0$ für alle $n \geq 0$. Da sämtliche Pfade der Länge $2n$ mit Anfangs- und Endpunkt 0 dieselbe Wahrscheinlichkeit $p^n(1-p)^n$ besitzen (n Schritte nach links und n Schritte nach rechts) und es $\binom{2n}{n}$ solche Pfade gibt, folgt

$$p_{00}^{(2n)} = \binom{2n}{n} p^n (1-p)^n \quad (2.21)$$

für alle $n \geq 0$. Falls $p \neq \frac{1}{2}$, liefert das starke Gesetz der großen Zahlen $n^{-1}M_n = n^{-1} \sum_{k=1}^n X_k \rightarrow \mathbb{E}_0 X_1 = 2p - 1 \neq 0$ \mathbb{P}_0 -f.s. und somit $|M_n| \rightarrow \infty$ \mathbb{P}_0 -f.s. Der Zustand 0 wird demnach fast sicher nur endlich oft aufgesucht und ist deshalb transient.

Im symmetrischen Fall $p = \frac{1}{2}$ benutzen wir die Stirlingsche Formel

$$n! \simeq \sqrt{2\pi} e^{-n} n^{n+1/2} \quad (n \rightarrow \infty).$$

Mit ihrer Hilfe ergibt sich

$$\binom{2n}{n} \simeq \frac{4^n}{\sqrt{n\pi}} \quad (n \rightarrow \infty)$$

und damit

$$p_{00}^{(2n)} \simeq \frac{1}{\sqrt{n\pi}} \quad (n \rightarrow \infty), \quad (2.22)$$

was schließlich $\sum_{n \geq 0} p_{00}^{(n)} = \infty$, d.h. die Rekurrenz von 0 impliziert.

Statt die Stirlingsche Formel zu benutzen, kann man $\mathbb{P}_{00}(s) = \sum_{n \geq 0} p_{00}^{(n)} s^n$ aber auch in geschlossener Form berechnen und anschließend einen Grenzübergang $s \uparrow 1$ durchführen. Unterstellen wir hierfür zunächst ein beliebiges $p \in (0, 1)$. Dann gilt aufgrund von (2.21), der Beziehung $4^n = \sum_{k=0}^n \binom{2k}{k} \binom{2n-2k}{n-k}$ und der Cauchyschen Produktformel

$$\begin{aligned} \mathbb{P}_{00}(s)^2 &= \left(\sum_{n \geq 0} \binom{2n}{n} (p(1-p)s^2)^n \right)^2 \\ &= \sum_{n \geq 0} (p(1-p)s^2)^n \sum_{k=0}^n \binom{2k}{k} \binom{2n-2k}{n-k} \\ &= \sum_{n \geq 0} (4p(1-p)s^2)^n = \frac{1}{1-4p(1-p)s^2}, \end{aligned}$$

d.h.

$$\mathbb{P}_{00}(s) = \frac{1}{\sqrt{1-4p(1-p)s^2}}. \quad (2.23)$$

Da $4p(1-p) = 1$ für $p = \frac{1}{2}$ und < 1 für $p \neq \frac{1}{2}$, erhalten wir schließlich

$$\sum_{n \geq 0} p_{00}^{(n)} = \lim_{s \uparrow 1} \mathbb{P}_{00}(s) = \begin{cases} \infty, & \text{falls } p = \frac{1}{2}, \\ \left(\sqrt{1-4p(1-p)} \right)^{-1} < \infty, & \text{falls } p \neq \frac{1}{2}, \end{cases}$$

also die Bestätigung der Rekurrenz des Zustands 0 im symmetrischen Fall und die Transienz sonst.

Um die Klasse der Irrfahrten auf \mathbb{Z} vollständig zu klassifizieren, müssen wir noch den Fall

$$\mathbb{P}_0(X_1 = 1) = p, \quad \mathbb{P}_0(X_1 = -1) = q \quad \text{und} \quad \mathbb{P}_0(X_1 = 0) = 1 - p - q$$

für $p, q \in (0, 1)$ mit $p + q < 1$ untersuchen. Der Unterschied besteht hier darin, dass die Irrfahrt mit der positiven Wahrscheinlichkeit $1 - p - q$ in einem Zustand verweilen kann. Falls $p \neq q$, liefert wiederum das starke Gesetz der großen Zahlen wegen $\mathbb{E}_0 X_1 = p - q \neq 0$ die Transienz des Zustands 0 (und damit aller $i \in \mathbb{Z}$). Im symme-

trischen Fall “ $p = q$ ” dürfen wir dagegen wiederum Rekurrenz erwarten, was sich vermöge eines einfachen Einbettungsarguments zeigen lässt: Betrachte, zunächst für beliebige p, q , die Folge $(\tau_n)_{n \geq 1}$ der sukzessiven Sprungzeiten der Irrfahrt, rekursiv gegeben durch

$$\tau_n = \inf\{k > \tau_{n-1} : M_k \neq M_{\tau_{n-1}}\}, \quad (2.24)$$

wobei $\tau_0 = 0$. Wir überlassen es dem Leser zu zeigen, dass die $\tau_n - \tau_{n-1}$ unabhängige, jeweils geometrisch verteilte Stopzeiten bilden ($\mathbb{P}_0(\tau_n - \tau_{n-1} = k) = (p+q)(1-p-q)^k$ für $k \in \mathbb{N}_0$) und dass $M_{\tau_n} = \sum_{k=0}^n \widehat{X}_k$ gilt mit ebenfalls unabhängigen, identisch verteilten Zufallsgrößen \widehat{X}_k , wobei

$$\mathbb{P}_0(\widehat{X}_1 = 1) = 1 - \mathbb{P}_0(\widehat{X}_1 = -1) = \frac{p}{p+q}.$$

$(M_{\tau_n})_{n \geq 0}$ definiert somit ebenfalls eine Irrfahrt auf \mathbb{Z} , und zwar der zuvor betrachteten Form. Falls $p = q$, d.h. $\frac{p}{p+q} = \frac{1}{2}$, erhalten für diese folglich Rekurrenz des Zustands 0 und damit natürlich auch für $(M_n)_{n \geq 0}$ selbst.

Zum Abschluss wollen wir kurz skizzieren, dass die $p_{00}^{(2n)}$ auch im zuletzt betrachteten symmetrischen Fall mit Verharrung mindestens von der Größenordnung $n^{-1/2}$ für $n \rightarrow \infty$ sind. Dies wollen wir uns nämlich im mehrdimensionalen Fall zunutze machen. Bezeichnet $\hat{p}_{00}^{(2n)}$ die entsprechende Wahrscheinlichkeit ohne Verharrungsmöglichkeit, gemäß (2.21) gegeben durch $\binom{2n}{n} 4^{-n}$, so gilt die Beziehung

$$p_{00}^{(2n)} = \sum_{k=0}^n \binom{2n}{2k} (1-2p)^{2k} (2p)^{2n-2k} \hat{p}_{00}^{(2n-2k)}, \quad (2.25)$$

die man erhält, wenn man die Menge der Pfade der Länge $2n$ mit Anfangs- und Endpunkt 0 unterscheidet nach der Anzahl der Verharrungsschritte, notwendig eine gerade Zahl $2k$, und deren Wahrscheinlichkeiten (konstant in k) dann jeweils unter den möglichen Verharrungszeitpunkten bedingt. Die Details kann sich der Leser leicht selbst überlegen. Durch Berechnung der Quotienten $\hat{p}_{00}^{(2n+2)}/\hat{p}_{00}^{(2n)}$ erhält man ferner die strenge Monotonie der $\hat{p}_{00}^{(2n)}$, so dass in (2.25)

$$p_{00}^{(2n)} > \hat{p}_{00}^{(2n)} \sum_{k=0}^n \binom{2n}{2k} (1-2p)^{2k} (2p)^{2n-2k} = \hat{p}_{00}^{(2n)} \text{Bin}(2n, 1-2p)(2\mathbb{N}_0).$$

Da außerdem $c(p) := \inf_{n \geq 0} \text{Bin}(2n, 1-2p)(2\mathbb{N}_0) > 0$, ergibt sich schließlich mit (2.22)

$$p_{00}^{(2n)} > c(p) \hat{p}_{00}^{(2n)} \simeq \frac{c(p)}{\sqrt{n\pi}} \quad (n \rightarrow \infty). \quad (2.26)$$

2.3.2 Der zweidimensionale Fall

Wir schreiben $M_n = (M_{n,1}, M_{n,2})$ und $X_n = (X_{n,1}, X_{n,2})$, so dass $M_{n,k} = \sum_{j=1}^n X_{j,k}$ unter \mathbb{P}_0 ($k = 1, 2$). Beachte, dass $(M_{n,k})_{n \geq 0}$, $k = 1, 2$, eine eindimensionale Irrfahrt bildet.

Falls $\mathbb{E}_0 X_n = (\mathbb{E}_0 X_{n,1}, \mathbb{E}_0 X_{n,2}) \neq 0$, liefert einmal mehr das starke Gesetz der großen Zahlen $|M_{n,1}| \rightarrow \infty$ oder $|M_{n,2}| \rightarrow \infty$ \mathbb{P}_0 -f.s. und folglich die Transienz des Zustands 0. Zu untersuchen bleibt also lediglich der Fall " $\mathbb{E}_0 X_1 = 0$ ". Wir beschränken uns dabei auf die symmetrischen Fälle, weil eine vollständige Behandlung mit den uns zur Verfügung stehenden Mitteln zu aufwendig wäre.

Der einfachste Fall liegt vor, wenn

$$\mathbb{P}_0(X_1 = (\pm 1, \pm 1)) = \frac{1}{4},$$

weil die X_n dann offensichtlich unabhängige, jeweils auf $\{-1, 1\}$ gleichverteilte Komponenten $X_{n,1}$ und $X_{n,2}$ besitzen, $(M_{n,1})_{n \geq 0}$ und $(M_{n,2})_{n \geq 0}$ also unabhängige, symmetrische Irrfahrten auf \mathbb{Z} bilden. Für $(M_n)_{n \geq 0}$ ergibt dies

$$\begin{aligned} p_{00}^{(2n)} &= \mathbb{P}_0(M_{2n} = 0) = \mathbb{P}_0(M_{2n,1} = 0)^2 \\ &= \binom{2n}{n}^2 \frac{1}{4^{2n}} \simeq \frac{1}{n\pi} \quad (n \rightarrow \infty) \end{aligned} \quad (2.27)$$

unter Hinweis auf (2.22). Da weiterhin $\sum_{n \geq 0} p_{00}^{(2n)} = \infty$ gilt, ist $0=(0,0)$ rekurrent.

Als nächstes wenden wir uns der symmetrischen Irrfahrt auf $(\mathbb{Z}^2, |\cdot|_1)$ mit

$$\mathbb{P}_0(X_1 = (1, 0)) = \mathbb{P}_0(X_1 = (0, 1)) = \mathbb{P}_0(X_1 = (-1, 0)) = \mathbb{P}_0(X_1 = (0, -1)) = \frac{1}{4}.$$

zu (Abb. 6.3). Ein einfacher Trick, der leider nur für die Dimension 2 funktioniert, liefert uns hier ohne weitere Rechnung die Antwort. Durch Drehen des Gitters in 0 um 45° wird diese Irrfahrt nämlich offenkundig in die zuvor betrachtete überführt (Abb. 2.6), so dass weiterhin (2.27) und folglich Rekurrenz des Zustands 0 gilt.

Unabhängige und identisch verteilte Komponentenfolgen $(M_{n,1})_{n \geq 0}$ und $(M_{n,2})_{n \geq 0}$ besitzt auch die Irrfahrt $(M_n)_{n \geq 0}$ auf $(\mathbb{Z}^2, |\cdot|_\infty)$ mit

$$\mathbb{P}_0(X_1 = x) = \frac{1}{9}$$

für alle $x \in \{-1, 0, 1\}^2$. $(M_{n,k})_{n \geq 0}$ bildet in diesem Fall eine symmetrische Irrfahrt auf \mathbb{Z} mit Verharrung, wobei $\mathbb{P}_0(X_{1,k} = i) = \frac{1}{3}$ für $i = -1, 0, 1$. Wir erhalten deshalb vermöge (2.26) und derselben Rechnung wie in (2.27), dass $p_{00}^{(2n)}$ für $n \rightarrow \infty$ mindestens von der Größenordnung $c(\frac{1}{3})^2 (n\pi)^{-1}$ ist, was erneut die Rekurrenz des Zustands 0 zeigt.

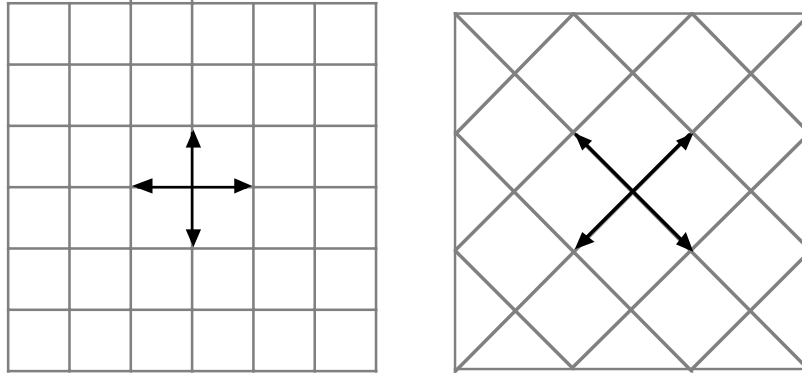


Abb. 2.6 Drehung des Gitters $(\mathbb{Z}^2, |\cdot|_1)$ um 45° .

Betrachten wir zuletzt die symmetrische Irrfahrt auf $(\mathbb{Z}^2, |\cdot|_\infty)$ ohne Verharrung, d.h.

$$\mathbb{P}_0(X_1 = x) = \frac{1}{8}$$

für alle $x \in \{-1, 0, 1\}^2 \setminus \{(0, 0)\}$. Sie ergibt sich gerade aus der vorherigen als Teilfolge zu den Sprungzeiten τ_n gemäß (2.24). Da jene den Zustand 0 mit Wahrscheinlichkeit 1 nicht nur unendlich oft aufsucht, sondern auch unendlich oft verlässt (Verharrungszeiten sind geometrisch verteilt, also f.s. endlich), ist 0 auch rekurrent für die eingebettete Irrfahrt ohne Verharrung.

2.3.3 Der drei- und mehrdimensionale Fall

Hier wollen wir uns kürzer fassen. Seien $M_{n,k}, X_{n,k}$, $k = 1, \dots, d$, die Komponenten von M_n bzw. X_n . Dasselbe Argument wie im zweidimensionalen Fall ergibt die Transienz des Zustands 0, falls $\mathbb{E}_0 X_{n,k} \neq 0$ für mindestens ein $k = 1, \dots, d$. Wir dürfen also gleich wieder $\mathbb{E} X_n = 0$ voraussetzen. Unabhängige, identisch verteilte Komponenten $(M_{n,k})_{n \geq 0}$ besitzt $(M_n)_{n \geq 0}$, falls

$$\mathbb{P}_0(X_1 = x) = \frac{1}{2^d} \quad \text{oder} \quad = \frac{1}{3^d}$$

für alle $x \in \{-1, 1\}^d$ bzw. $\{-1, 0, 1\}^d$. Jedes $(M_{n,k})_{n \geq 0}$ definiert eine symmetrische Irrfahrt auf \mathbb{Z} ohne bzw. mit Verharrung, woraus vermöge derselben Rechnung wie in (2.27) unter Benutzung von (2.22) bzw. (2.25) nun

$$p_{00}^{(2n)} = O(n^{-d/2}) \quad (n \rightarrow \infty) \tag{2.28}$$

folgt und daraus $\sum_{n \geq 0} p_{00}^{(2n)} < \infty$ wegen $d \geq 3$. Bei positiver Verharrungswahrscheinlichkeit ist $p_{00}^{(2n+1)}$ zwar nicht 0, aber immer noch beschränkt durch eine Konstante mal $n^{-d/2}$, wie man sich weiter überlegen kann, so dass in jedem Fall $\sum_{n \geq 0} p_{00}^{(n)} < \infty$ und damit die Transienz des Zustands 0 folgt.

Werfen wir abschließend einen Blick auf die symmetrische Irrfahrt in $(\mathbb{Z}^d, |\cdot|_1)$ (ohne Verharrung), d.h.

$$\mathbb{P}_0(X_1 = \pm e_k) = \frac{1}{2d}$$

für alle $k = 1, \dots, d$, e_k der k -te kanonische Einheitsvektor des \mathbb{R}^d . Dann gilt wieder $p_{00}^{(2n+1)} = 0$ für alle $n \in \mathbb{N}_0$, während

$$p_{00}^{(2n)} = \frac{1}{(2d)^{2n}} \sum_{k_1=0}^n \sum_{k_2=0}^{n-k_1} \dots \sum_{k_d=0}^{n-k_1-\dots-k_{d-1}} \frac{(2n)!}{k_1!k_1! \cdot \dots \cdot k_d!k_d!},$$

denn es gibt genau $\frac{(2n)!}{k_1!k_1! \cdot \dots \cdot k_d!k_d!}$ Pfade der Länge $2n$ mit Anfangs- und Endpunkt 0, die aus je k_i Vorwärts- und Rückwärtsschritten in die i -te Koordinatenrichtung bestehen. Eine einfache Umformung liefert unter Benutzung von Multinomialkoeffizienten

$$p_{00}^{(2n)} = \frac{1}{4^n} \binom{2n}{n} \sum_{k_1, \dots, k_d} \left(\binom{n}{k_1 k_2 \dots k_d} \frac{1}{d^n} \right)^2.$$

Aus $\sum_{k_1, \dots, k_d} \binom{n}{k_1 k_2 \dots k_d} \frac{1}{d^n} = 1$ (Multinomialwahrscheinlichkeiten) und (2.22) folgt dann

$$p_{00}^{(2n)} \leq \frac{1}{4^n} \binom{2n}{n} \max_{k_1, \dots, k_d} \binom{n}{k_1 k_2 \dots k_d} \frac{1}{d^n} \simeq \frac{1}{\sqrt{n\pi}} \max_{k_1, \dots, k_d} \binom{n}{k_1 k_2 \dots k_d} \frac{1}{d^n},$$

wobei das auftretende Maximum für $k_1 \approx \dots \approx k_d \approx \frac{n}{d}$ angenommen wird und von der Größenordnung $n^{-(d-1)/2}$ ist, wie eine Anwendung der Stirlingschen Formel zeigt. Es folgt schließlich erneut (2.28) und damit die Transienz des Zustands 0.

Der allgemeine Satz hinter den zuvor dargestellten Ergebnissen, den wir nur für den eindimensionalen Fall vollständig gezeigt haben, lautet:

Satz 2.24. Eine Irrfahrt $M_n = M_0 + \sum_{k=1}^n X_k$ in \mathbb{Z}^d ist genau dann rekurrent (d.h. alle Zustände $i \in \mathbb{Z}^d$ sind rekurrent), wenn $d \leq 2$ und $\mathbb{E}_0 X_1 = 0$.

Der Satz bleibt übrigens auch dann noch richtig, wenn $(M_n)_{n \geq 0}$ einen diskreten Random Walk in \mathbb{Z}^d (siehe Beispiel 2.9) bildet.

Zum Abschluss sei eine scherzhafte Bemerkung von KAKUTANI während eines Kolloquiumsvortrags an der U.C.L.A. zitiert, die das Ergebnis in einprägsamer Weise zusammenfasst:

“A drunk man will find his way home but a drunk bird may get lost forever.”

2.4 Solidaritätseigenschaften

Eine Zustandseigenschaft, die, wenn gültig für ein $i \in \mathcal{S}$, immer schon für jeden Zustand aus $\mathcal{C}_i = \{j \in \mathcal{S} : j \leftrightarrow i\}$ gilt, nennt man *Solidaritäts-* oder auch *Klassen-*eigenschaft. Die gute Nachricht diesbetreffend lautet:

Satz 2.25. *Rekurrenz, Transienz, positive Rekurrenz, Null-Rekurrenz sowie die Periode eines Zustands bilden Solidaritätseigenschaften.*

Die praktische Konsequenz des Satzes besteht darin, dass die genannten Eigenschaften immer nur für einen Vertreter statt für jedes Element einer Klasse überprüft werden müssen. So gilt beispielsweise: Aus $i \leftrightarrow j$ folgt $d(i) = d(j)$. Im Fall einer irreduziblen MK, für die $\mathcal{C}_i = \mathcal{S}$ für alle $i \in \mathcal{S}$ gilt, bedeutet dies, dass jede der im Satz genannten Solidaritätseigenschaften entweder für alle oder gar keinen Zustand vorliegt, was folgende Definition rechtfertigt:

Definition 2.26. Eine DMK mit Zustandsraum $(\mathcal{S}, \mathfrak{S})$ heißt
rekurrent/transient,
positiv rekurrent/null-rekurrent,
aperiodisch/ d -periodisch,
wenn sie irreduzibel ist und die jeweilige Eigenschaft für einen und somit alle $i \in \mathcal{S}$ gilt.

Beweis (von Satz 2.25). Seien $i, j \in \mathcal{S}$ zwei verschiedene kommunizierende Zustände und i rekurrent. Dann existieren $m, m' \geq 1$, so dass $p_{ij}^{(m)} > 0$ und $p_{ji}^{(m')} > 0$. Wir geben im Folgenden zwei Beweise dafür, dass auch j rekurrent sein muss. Der erste ist kurz und elegant, aber weniger intuitiv als der zweite.

1. *Beweis:* Nach Satz 2.22 ist $\sum_{n \geq 1} p_{ii}^{(n)} = \infty$. Da ferner $p_{jj}^{(n)} \geq p_{ji}^{(m')} p_{ii}^{(n-m-m')} p_{ij}^{(m)}$ gilt für alle $n > m + m'$, liefert dies

$$\sum_{n \geq 1} p_{jj}^{(n)} \geq p_{ji}^{(m')} \left(\sum_{n \geq 1} p_{ii}^{(n)} \right) p_{ij}^{(m)} = \infty$$

und somit wiederum gemäß Satz 2.22 die Rekurrenz von j .

2. *Beweis:* Wir zeigen zuerst $f_{ij}^* = \mathbb{P}_i(\tau(j) < \infty) = 1$. Betrachte die Rekurrenzzeiten $\sigma_n = \sigma_n(i)$, $n \geq 1$, die gemäß Satz 2.16 unter \mathbb{P}_i f.s. endliche, unabhängige und identisch verteilte Zuwächse τ_n besitzen, wobei $\tau_1 = \sigma_1 = \tau(i)$ \mathbb{P}_i -f.s. Wie schon früher bezeichne $(\mathcal{G}_n)_{n \geq 0}$ die kanonische Filtration von M . Unter Benutzung der starken Markov-Eigenschaft folgt für jedes $n \geq 1$

$$\begin{aligned}
\mathbb{P}_i(\tau(j) > \sigma_n) &= \int_{\{\tau(j) > \sigma_{n-1}\}} \mathbb{P}(\tau(j) - \sigma_{n-1} > \tau_n | \mathcal{G}_{\sigma_{n-1}}) d\mathbb{P}_i \\
&= \int_{\{\tau(j) > \sigma_{n-1}\}} \mathbb{P}(\tau(j) - \sigma_{n-1} > \tau_n | M_{\sigma_{n-1}}) d\mathbb{P}_i \\
&= \int_{\{\tau(j) > \sigma_{n-1}\}} \mathbb{P}_{M_{\sigma_{n-1}}}(\tau(j) > \tau(i)) d\mathbb{P}_i \\
&= \mathbb{P}_i(\tau(j) > \sigma_{n-1}) \mathbb{P}_i(\tau(j) > \tau(i))
\end{aligned}$$

wenn man beachtet, dass auf dem Ereignis $\{\tau(j) > \sigma_{n-1}\}$ offenbar

$$\mathbb{P}^{\tau(j) - \sigma_{n-1} | \mathcal{G}_{\sigma_{n-1}}} = \mathbb{P}^{\tau(j) - \sigma_{n-1} | M_{\sigma_{n-1}}} = \mathbb{P}_i^{\tau(j)} \quad \mathbb{P}_i\text{-f.s.}$$

gilt. Induktiv erhalten wir damit

$$\mathbb{P}_i(\tau(j) > \sigma_n) = \mathbb{P}_i(\tau(j) > \tau(i))^n \quad (2.29)$$

und folglich wie behauptet

$$1 - f_{ij}^* = \lim_{n \rightarrow \infty} \mathbb{P}_i(\tau(j) > \sigma_n) = \lim_{n \rightarrow \infty} \mathbb{P}_i(\tau(j) > \tau(i))^n = 0,$$

denn

$$\lim_{n \rightarrow \infty} \mathbb{P}_i(\tau(j) > \sigma_n) = \mathbb{P}_i(\tau(j) = \infty) \leq 1 - p_{ij}^{(m)} < 1$$

impliziert $\mathbb{P}_i(\tau(j) > \tau(i)) < 1$. Wir haben folglich nicht nur $f_{ij}^* = 1$ nachgewiesen, sondern auch, dass j in jedem der durch die σ_n markierten Zyklen mit positiver Wahrscheinlichkeit besucht wird. Da die Zyklen unabhängig und identisch verteilt sind (Satz 2.16), folgern wir weiter, dass die Anzahl ν von Zyklen vor demjenigen, in dem j tatsächlich erstmals besucht wird, unter \mathbb{P}_i geometrisch verteilt ist. Dasselbe Argument, genannt “*geometric trials argument*”, liefert dann aber auch, dass j unendlich oft aufgesucht wird und somit rekurrent ist. Ein formaleres Argument ist das folgende: Sei $\widehat{\tau}(i) := \inf\{n \geq 1 : M_{\tau(j)+n} = i\}$ und $\widehat{\tau}(j)$ analog definiert mit vertauschten Rollen für i, j . Dann ergibt sich unter erneuter Verwendung der starken Markov-Eigenschaft und Beachtung von $\sigma_{\nu+1} = \tau(j) + \widehat{\tau}(i)$ \mathbb{P}_i -f.s.

$$\begin{aligned}
1 &= \mathbb{P}_i(\sigma_{\nu+1} < \infty) = \mathbb{P}_i(\tau(j) < \infty, \widehat{\tau}(i) < \infty) \\
&= \int_{\{\tau(j) < \infty\}} \mathbb{P}(\widehat{\tau}(i) < \infty | M_{\tau(j)}) d\mathbb{P}_i = \mathbb{P}_i(\tau(j) < \infty) \mathbb{P}_j(\tau(i) < \infty) = f_{ij}^* f_{ji}^*
\end{aligned}$$

sowie anschließend in ähnlicher Weise

$$\begin{aligned}
f_{jj}^* &= \mathbb{P}_j(\tau(j) < \infty) \geq \mathbb{P}_j(\tau(i) < \infty, \widehat{\tau}(j) < \infty) \\
&= \mathbb{P}_j(\tau(i) < \infty) \mathbb{P}_i(\tau(j) < \infty) = f_{ji}^* f_{ij}^* = 1.
\end{aligned}$$

Mit i ist also auch j rekurrent (bzw. transient als Komplementäreigenschaft).

Nehmen wir als nächstes an, dass i positiv rekurrent ist, also $\mu_{ii} = \mathbb{E}_i \tau(i) < \infty$ gilt, so impliziert zunächst $\tau(j) \leq \tau(i) + \widehat{\tau}(j)$

$$\mu_{jj} = \mathbb{E}_j \tau(j) \leq \mathbb{E}_j \tau(i) + \mathbb{E}_i \widehat{\tau}(j) = \mu_{ji} + \mu_{ij},$$

denn $\mathbb{E}_j \widehat{\tau}(j) = \mathbb{E}_j \mathbb{E}(\widehat{\tau}(j) | \mathcal{M}_{\tau(i)}) = \mathbb{E}_i \tau(j) = \mu_{ij}$. Es genügt also $\mu_{ij}, \mu_{ji} < \infty$ nachzuweisen. Sei ν die, wie oben erläutert, unter \mathbb{P}_i geometrisch verteilte Anzahl von Zyklen (markiert durch die σ_n) vor dem ersten Zyklus, in dem die Kette j besucht. Es folgt $\mu_{ij} < \infty$ vermöge der folgenden Rechnung unter Beachtung von $\tau(j) \leq \sigma_{\nu+1} = \sum_{k=1}^{\nu+1} \tau_k$, $\mathbb{E}_i(\nu+1) = \mathbb{P}_i(\tau(j) < \tau(i))^{-1}$ sowie der Unabhängigkeit von τ_k und $\{\nu \geq k-1\} \in \mathcal{G}_{\sigma_{k-1}}$ für alle $k \geq 1$:

$$\begin{aligned} \mu_{ij} &\leq \mathbb{E}_i \sigma_{\nu+1} = \mathbb{E}_i \left(\sum_{k=1}^{\nu+1} \tau_k \right) \\ &= \mathbb{E}_i \left(\sum_{k \geq 1} \tau_k \mathbf{1}_{\{\nu+1 \geq k\}} \right) = \sum_{k \geq 1} \mathbb{E}_i \tau_k \mathbb{P}_i(\nu \geq k-1) \\ &= \mathbb{E}_i \tau(i) \mathbb{E}_i(\nu+1) = \frac{\mu_{ii}}{\mathbb{P}_i(\tau(j) < \tau(i))} < \infty. \end{aligned}$$

Beachtet man nun noch, dass $\sigma_{\nu+1}$ die erste Rückkehr nach i nach $\tau(j)$ bezeichnet und dass folglich

$$\mathbb{P}_i^{\sigma_{\nu+1} - \tau(j)} = \mathbb{P}_j^{\tau(i)}$$

vermöge der starken Markov-Eigenschaft gilt, so erhalten wir auch $\mu_{ji} < \infty$, denn

$$\mu_{ji} = \mathbb{E}_i(\sigma_{\nu+1} - \tau(j)) = \mathbb{E}_i \sigma_{\nu+1} - \mu_{ij} < \infty.$$

Rekurrente kommunizierende Zustände sind somit stets vom gleichen Typ: positiv oder null-rekurrent.

Es bleibt noch zu zeigen, dass kommunizierende Zustände immer dieselbe Periode besitzen. Gegeben $i \leftrightarrow j$, $i \neq j$ und $m, n \in \mathbb{N}$ derart, dass $p_{ij}^{(m)} > 0$ und $p_{ji}^{(n)} > 0$, gilt

$$p_{ii}^{(m+n+k)} \geq p_{ij}^{(m)} p_{jj}^{(k)} p_{ji}^{(n)} > 0$$

für alle $k \in \mathcal{D}(j) := \{l \geq 1 : p_{jj}^{(l)} > 0\}$ und $k = 0$. $p_{ii}^{(m+n)} > 0$ impliziert aber $m+n = \nu_0 d(i)$ für ein $\nu_0 \geq 1$, und $p_{ii}^{(m+n+k)} > 0$ für alle $k \in \mathcal{D}(j)$ liefert die Existenz von $\nu_k \in \mathbb{N}$ mit $m+n+k = \nu_k d(i)$. Es folgt $k = (\nu_k - \nu_0) d(i)$, was $\mathcal{D}(j) \subset d(i)\mathbb{N}$, d.h. $d(j) \geq d(i)$ beweist. Die umgekehrte Ungleichung ergibt sich analog durch Vertauschen der Rollen von i und j . \square

Der obige Beweis hat zur Rekurrenz sogar etwas mehr hervorgebracht als im Satz behauptet wird.

Satz 2.27. Gegeben $i, j \in \mathcal{S}$, $i \neq j$, i rekurrent und $i \rightarrow j$ folgt:

- (a) $i \leftrightarrow j$,
- (b) j ist rekurrent,
- (c) $f_{ij}^* = f_{ji}^* = 1$.

Jeder von einem rekurrenten Zustand i erreichbare Zustand ist also bereits mit diesem verbunden und aus Solidarität ebenfalls rekurrent. Dies hat zur Konsequenz, dass die zugehörige Klasse \mathcal{C}_i abgeschlossen ist, nach Erreichen also nie mehr verlassen wird.

Beweis. Offenbar reicht es, Aussage (a) und hierfür $j \rightarrow i$ zu zeigen. Wie im vorherigen Beweis gesehen, impliziert $i \rightarrow j$, dass nach einer geometrisch verteilten Anzahl v von durch Besuchen in i markierten Zyklen der Zustand j aufgesucht wird, d.h. $\sigma_v < \tau(j) \leq \sigma_{v+1}$ in den dortigen Bezeichnungen. Wählt man $m, n \in \mathbb{N}$ mit $\mathbb{P}_i(\tau(j) = m, \sigma_{v+1} = n) > 0$, liefert dies aber

$$0 < \mathbb{P}_i(\tau(j) = m, \sigma_{v+1} = n) = \mathbb{P}_i(\tau(j) = m) \mathbb{P}_j(\tau(i) = n - m) = f_{ij}^{(m)} f_{ji}^{(n-m)}$$

vermöge der starken Markov-Eigenschaft und daher $p_{ji}^{(n-m)} \geq f_{ji}^{(n-m)} > 0$. \square

Mit Hilfe von Satz 2.27 (☞ auch Lemma 2.12 und danach) erhalten wir schließlich ohne weitere Mühe den folgenden Zerlegungssatz:

Satz 2.28. Der Zustandsraum \mathcal{S} einer DMK besitzt eine eindeutig bestimmte disjunkte Zerlegung

$$\mathcal{S} = \mathcal{T} + \sum_{\alpha} \mathcal{R}_{\alpha} \quad (2.30)$$

in eine Menge \mathcal{T} transienter Zustände (nicht notwendig eine Klasse) und endlich oder abzählbar unendlich viele abgeschlossene Klassen \mathcal{R}_{α} rekurrenter Zustände (Rekurrenzklassen), wobei, falls $i \in \mathcal{R}_{\alpha}$,

$$f_{ij}^* = \mathbb{P}_i(\tau(j) < \infty) = \begin{cases} 1, & \text{falls } j \in \mathcal{R}_{\alpha}, \\ 0, & \text{falls } j \notin \mathcal{R}_{\alpha}. \end{cases}$$

Bei geeigneter Numerierung der Zustände besitzt die Übergangsmatrix P außerdem die Form

$$P = \begin{pmatrix} Q_0 & Q_1 & Q_2 & Q_3 & Q_4 & \dots \\ 0 & P_1 & 0 & 0 & 0 & \dots \\ 0 & 0 & P_2 & 0 & 0 & \dots \\ 0 & 0 & 0 & P_3 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix},$$

wobei P_α und Q_0 die Übergangsmatrizen der Kette bei Einschränkung auf \mathcal{R}_α bzw. \mathcal{T} bilden und Q_α die Übergangswahrscheinlichkeiten p_{ij} von $i \in \mathcal{T}$ nach $j \in \mathcal{R}_\alpha$ enthält.

Eine DMK mit abzählbar unendlichem Zustandsraum kann durchaus überhaupt keine rekurrenten Zustände besitzen, wie das triviale Beispiel der deterministischen Kette $0 \rightarrow 1 \rightarrow 2 \rightarrow \dots$ verdeutlicht. Der Anteil $\sum_\alpha \mathcal{R}_\alpha$ von \mathcal{S} in (2.30) tritt demnach möglicherweise gar nicht auf. Für eine EMK ergibt sich dagegen unter Vorgriff auf ein Ergebnis im nächsten Abschnitt:

Satz 2.29. *Jede EMK besitzt mindestens einen positiv rekurrenten Zustand.*

Beweis. Sind alle Zustände transient oder null-rekurrent, folgt $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)} = 0$ für alle $i, j \in \mathcal{S}$ aus Satz 2.22 bzw. Korollar 2.39. Dies führt aber vermöge

$$0 = \sum_{j \in \mathcal{S}} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \sum_{j \in \mathcal{S}} p_{ij}^{(k)} = 1$$

zu einem Widerspruch. □

2.5 Stationäre Maße und Zeitmittel

Wir sind nun hinreichend präpariert, um die zentrale Frage nach dem Langzeitverhalten diskreter MK anzugehen, wobei zwei unterschiedlich starke Konvergenzarten betrachtet werden:

(1) *Gleichmäßige Konvergenz im Zeitmittel (Césaro-Mittel):*

$$\lim_{n \rightarrow \infty} \sup_{A \subset \mathcal{S}} \left| \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda(M_k \in A) - \pi(A) \right| = 0.$$

(2) *Gleichmäßige Verteilungskonvergenz (Konvergenz in Totalvariation):*

$$\lim_{n \rightarrow \infty} \sup_{A \subset \mathcal{S}} |\mathbb{P}_\lambda(M_n \in A) - \pi(A)| = 0.$$

Es ist klar, dass aus (2) stets (1) folgt, da jede konvergente Folge auch im Césaro-Limes konvergiert. Bezeichnet $\|\cdot\|$ die *Totalvariation* auf dem Raum signierter Maße $\lambda - \mu$ auf $(\mathcal{S}, \mathfrak{S})$ (λ, μ endliche Maße), d.h.

$$\|\lambda - \mu\| = \sup_{A \in \mathcal{S}} |\lambda(A) - \mu(A)|,$$

so bildet diese eine Norm, und es gilt offenkundig

$$(1) \Leftrightarrow \lim_{n \rightarrow \infty} \left\| \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda^{M_k} - \pi \right\| = 0,$$

$$(2) \Leftrightarrow \lim_{n \rightarrow \infty} \|\mathbb{P}_\lambda^{M_n} - \xi\| = 0.$$

Der Variationsabstand wurde in [2, Abschnitt 29] über Poisson-Approximation kurz eingeführt, und wir hatten dort gezeigt (Lemma 34.5), dass für abzählbares \mathcal{S} (wie hier der Fall) und beliebige *Wahrscheinlichkeitsmaße* λ, μ auf $(\mathcal{S}, \mathfrak{G})$

$$\|\lambda - \mu\| = \frac{1}{2} \sum_{i \in \mathcal{S}} |\lambda_i - \mu_i| \quad (2.31)$$

gilt.

In diesem Abschnitt beschäftigen wir uns mit der schwächeren Konvergenzart (1) und weisen als erstes darauf hin, dass der Limes π stets eine stationäre Verteilung von $(M_n)_{n \geq 0}$ bildet. Dies wurde für die stärkere Konvergenzart (2) bereits in Lemma 1.34 gezeigt und kann vom Leser mühelos auf die Konvergenz im Zeitmittel ausgedehnt werden. Stationäre Verteilungen spielen also bei den avisierten Konvergenzbetrachtungen eine wichtige Rolle und stehen deshalb im ersten Teilabschnitt im Vordergrund.

2.5.1 Stationäre Maße via zyklischer Zerlegungen

Der kanonische Weg, stationäre Maße und Verteilungen einer DMK $M = (M_n)_{n \geq 0}$ zu bestimmen, besteht gemäß Satz 1.35 in der Lösung des Gleichungssystems $\pi_j = \sum_{i \in \mathcal{S}} \pi_i p_{ij}$, $j \in \mathcal{S}$, kurz $\pi = \pi P$, was auf eine Analyse der Übergangsmatrix P hinsichtlich des Eigenwerts 1 hinausläuft. Im Folgenden bevorzugen wir aber einen anderen, probabilistischen Zugang gegenüber dieser matrixanalytischen Betrachtungsweise, weil er im Hinblick auf das Langzeitverhalten der Kette zum einen ein besseres Verständnis vermittelt und zum anderen für die späteren Konvergenzbeweise nützlicher ist. Wir betonen jedoch, dass nichtsdestotrotz der kanonische Weg in vielen Beispielen, in denen eine eindeutig bestimmte stationäre Verteilung existiert, für deren explizite Berechnung der sinnvollste ist und daher als Alternative immer in der Hinterhand bleibt.

Mit Hilfe der in Teilabschnitt 2.2.3 kennengelernten zyklischen Zerlegung einer DMK mittels der Rückkehrzeiten in einen rekurrenten Zustand werden wir nun zeigen, wie sich in kanonischer Weise ein stationäres Maß für M ergibt: Seien dazu $i \in \mathcal{S}$ rekurrent und $\sigma_n = \sigma_n(i)$ die zugehörigen \mathbb{P}_i -f.s. endlichen sukzessiven Rückkehrzeiten. Setzen wir

$$Z_n = (M_{\sigma_n}, \dots, M_{\sigma_{n+1}-1}), \quad n \geq 0,$$

so bilden diese gemäß Satz 2.16 unter \mathbb{P}_i eine unabhängige Folge identisch verteilter Zufallsvariablen, und es folgt wegen $M = (Z_0, Z_1, \dots)$

$$\mathbb{P}_i^M = \bigotimes_{n \geq 0} \mathbb{P}_i^{Z_n} = (\mathbb{P}_i^{Z_0})^\infty.$$

Die Verteilung der Kette unter \mathbb{P}_i ist somit vollständig durch die Verteilung des ersten Zyklus' Z_0 determiniert. Mit anderen Worten: Alle Information über die Verteilung von M steckt bereits in der von Z_0 (unter \mathbb{P}_i). Für jeden weiteren Zustand $j \in \mathcal{S}$ mit $i \leftrightarrow j$ sind die $\sigma_n(i)$ unter \mathbb{P}_j ebenfalls f.s. endlich (Satz 2.27) und die Z_n deshalb nach Satz 2.18 weiterhin unabhängig und für $n \geq 1$ identisch gemäß $\mathbb{P}_i^{Z_0}$ verteilt, was

$$\mathbb{P}_j^M = \mathbb{P}_j^{Z_0} \otimes (\mathbb{P}_i^{Z_0})^\infty \quad (2.32)$$

impliziert. Schreiben wir nun (1) mit $\lambda = \delta_j$ in der Form

$$\lim_{n \rightarrow \infty} \sup_{A \subset \mathcal{S}} \left| \mathbb{E}_j \left(\frac{1}{n+1} \sum_{k=0}^n \mathbf{1}_A(M_k) \right) - \pi^*(A) \right| = 0, \quad (2.33)$$

so wird deutlich, dass es sich um eine gleichmäßige Konvergenz von mittleren relativen Häufigkeiten handelt. Andererseits ist mit Blick auf (2.32) klar, dass die mittlere Anzahl von Aufenthalten in einer Menge A innerhalb eines Zyklus', also

$$\mathbb{E}_j \left(\sum_{k=\sigma_n}^{\sigma_{n+1}-1} \mathbf{1}_A(M_k) \right)$$

für $n \geq 1$ konstant ist, und zwar gleich

$${}^{(i)}\pi(A) := \mathbb{E}_i \left(\sum_{k=0}^{\tau(i)-1} \mathbf{1}_A(M_k) \right). \quad (2.34)$$

Somit bildet ${}^{(i)}\pi^*(A)$, definiert durch

$${}^{(i)}\pi^*(A) := \frac{{}^{(i)}\pi(A)}{\mu_{ii}} = \frac{1}{\mathbb{E}_i \tau(i)} \mathbb{E}_i \left(\sum_{k=0}^{\tau(i)-1} \mathbf{1}_A(M_k) \right), \quad (2.35)$$

die in *jedem* (bis auf den ersten) Zyklus (eventuell $\equiv 0$) mittlere Anzahl von Aufenthalten in A relativ zur mittleren Zykluslänge. ${}^{(i)}\pi^*(A)$ ist deshalb auch ein natürlicher Kandidat für den Limes $\pi^*(A)$ in (2.33), der gerade die asymptotische relative Häufigkeit von Aufenthalten über die ganze Zeitachse angibt. Dass ${}^{(i)}\pi$ und, sofern i positiv rekurrent ist, ${}^{(i)}\pi^*$ ein Maß bzw. eine Verteilung auf \mathcal{S} definieren, ist offensichtlich. Die Stationarität jedoch, die bei Gültigkeit von (2.33) ja notwendig gelten muss, bedarf einiger Arbeit:

Satz 2.30. Gegeben einen rekurrenten Zustand i , bildet das Okkupationsmaß ${}^{(i)}\pi$ gemäß (2.34) ein stationäres Maß der DMK M und folglich ${}^{(i)}\pi^*$ gemäß (2.35) eine stationäre Verteilung, sofern i positiv rekurrent ist.

Beweis. Es bezeichne $(\mathcal{G}_n)_{n \geq 0}$ wieder die zu M gehörende kanonische Filtration. Da i im Folgenden nicht variiert, schreiben wir kurz π für ${}^{(i)}\pi$. Wir zeigen zunächst die σ -Endlichkeit von π , d.h. $\pi_j = \pi(\{j\}) < \infty$ für alle $j \in \mathcal{S}$. Da $\pi_i = 1$, sei gleich $j \neq i$ vorausgesetzt. Offensichtlich gilt dann

$$\pi_j = \sum_{n \geq 1} n \mathbb{P}_i(\sigma_n(j) < \tau(i) < \sigma_{n+1}(j)) \leq \sum_{n \geq 1} n \mathbb{P}_i(\sigma_n(j) < \tau(i))$$

Mittels der starken Markov-Eigenschaft sowie (2.29) erhalten wir weiter

$$\begin{aligned} \mathbb{P}_i(\sigma_n(j) < \tau(i)) &= \mathbb{P}_i(\tau(j) < \tau(i), \sigma_n(j) < \tau(i)) \\ &= \int_{\{\tau(j) < \tau(i)\}} \mathbb{P}(\sigma_n(j) < \tau(i) | M_{\tau(j)}) d\mathbb{P}_i \\ &= \mathbb{P}_i(\tau(j) < \tau(i)) \mathbb{P}_j(\sigma_{n-1}(j) < \tau(i)) \\ &= \mathbb{P}_i(\tau(j) < \tau(i)) \mathbb{P}_j(\tau(j) < \tau(i))^{n-1} \end{aligned}$$

für alle $n \geq 1$. Wegen $\mathbb{P}_i(\sigma_n(j) < \tau(i)) \rightarrow 0$ muss $\mathbb{P}_i(\tau(j) < \tau(i)) = 0$ oder $\mathbb{P}_j(\tau(j) < \tau(i)) < 1$ sein, was schließlich

$$\begin{aligned} \pi_j &\leq \sum_{n \geq 1} n \mathbb{P}_i(\tau(j) < \tau(i)) \mathbb{P}_j(\tau(j) < \tau(i))^{n-1} \\ &= \frac{\mathbb{P}_i(\tau(j) < \tau(i))}{\mathbb{P}_j(\tau(j) \geq \tau(i))^2} < \infty \end{aligned} \tag{2.36}$$

liefert. Zu zeigen bleibt die Invarianz, d.h. $\pi_j = \sum_{k \in \mathcal{S}} \pi_k p_{kj}$ für alle $j \in \mathcal{S}$. Unter Beachtung von $M_0 = M_{\tau(i)} = i$ \mathbb{P}_i -f.s. gilt

$$\begin{aligned} \pi_j &= \mathbb{E}_i \left(\sum_{n=0}^{\tau(i)-1} \mathbf{1}_{\{M_n=j\}} \right) \\ &= \mathbb{E}_i \left(\sum_{n=1}^{\tau(i)} \mathbf{1}_{\{M_n=j\}} \right) \\ &= \mathbb{E}_i \left(\sum_{n=0}^{\tau(i)-1} \mathbf{1}_{\{M_{n+1}=j\}} \right) \\ &= \sum_{k \in \mathcal{S}} \mathbb{E}_i \left(\sum_{n=0}^{\tau(i)-1} \mathbf{1}_{\{M_n=k, M_{n+1}=j\}} \right) \\ &= \sum_{k \in \mathcal{S}} \sum_{n \geq 0} \mathbb{P}_i(\tau(i) > n, M_n = k, M_{n+1} = j) \end{aligned}$$

$$\begin{aligned}
&= \sum_{k \in \mathcal{S}} \sum_{n \geq 0} \mathbb{P}_i(\tau(i) > n, M_n = k) p_{kj} \\
&= \sum_{k \in \mathcal{S}} \mathbb{E}_i \left(\sum_{n=0}^{\tau(i)-1} \mathbf{1}_{\{M_n=k\}} \right) p_{kj} \\
&= \sum_{k \in \mathcal{S}} \pi_k p_{kj},
\end{aligned}$$

wobei in der vorletzten Zeile einmal mehr die starke Markov-Eigenschaft benutzt wurde. \square

Satz 2.30 beschert uns also bei Vorliegen mindestens eines (positiv) rekurrenten Zustands i – für eine EMK gemäß Satz 2.29 immer der Fall – automatisch die Existenz eines stationären Maßes (einer stationären Verteilung), das außerdem auf die zugehörige Klasse \mathcal{C}_i kommunizierender Zustände konzentriert ist, wie das folgende Korollar lehrt:

Korollar 2.31. *In der Situation von Satz 2.10 gilt ferner*

$$\pi(\mathcal{C}_i^c) = 0.$$

Beweis. Da i rekurrent ist, gilt $\mathbb{P}_i(\tau(i) < \infty) = 1$. Wir hatten mittels Satz 2.27 festgestellt, dass \mathcal{C}_i abgeschlossen ist, also nach Erreichen nicht mehr verlassen wird. Dies bedeutet aber $f_{ij}^{(n)} = 0$ für $j \in \mathcal{C}_i^c$ und $n \geq 0$, d.h. $\mathbb{P}_i(\tau(j) = \infty) = 1$. Insgesamt folgt $\mathbb{P}_i(\tau(j) < \tau(i)) = 0$ und daraus $\pi_j = 0$ für alle $j \in \mathcal{C}_i^c$ vermöge (2.36). \square

Eine DMK besitzt somit auf jeder Klasse rekurrenter Zustände mindestens ein stationäres Maß. Dagegen werden wir schon bald sehen, dass für rekurrente Zustände i, j derselben Klasse die zugehörigen stationären Maße ${}^{(i)}\pi$ bzw. ${}^{(j)}\pi$ bis auf ein skalares Vielfaches gleich sind.

2.5.2 Zeitmittel

Wir kommen nun zur angekündigten Untersuchung von Zeitmitteln diskreter Markov-Ketten, und zwar sowohl *pfadweise*:

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \mathbf{1}_A(M_k) = \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \delta_{M_k}(A), \quad A \subset \mathcal{S}$$

als auch *in Erwartung unter beliebigem* \mathbb{P}_λ :

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{E}_\lambda \left(\sum_{k=0}^n \mathbf{1}_A(M_k) \right) = \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda(M_k \in A), \quad A \subset \mathcal{S}.$$

Zur Abkürzung schreiben wir für derartige Césaro-Limiten im Folgenden auch $C\text{-}\lim_{n \rightarrow \infty}$, d.h.

$$C\text{-}\lim_{n \rightarrow \infty} a_n := \frac{1}{n+1} \sum_{k=0}^n a_k.$$

Zum Beweis des Hauptsatzes 2.33 weiter unten benötigen wir ein Ergebnis, nämlich die erste Aussage in (2.38), das eigentlich in Kapitel 1 des 2. Bands dieses Skriptums [4] über Erneuerungstheorie gehört. Für einen Moment verlassen wir die Welt diskreter Markov-Ketten und begeben uns in die schon recht vertraute Welt der Random Walks, d.h. Summen unabhängiger identisch verteilter Zufallsgrößen.

Lemma 2.32. *Gegeben eine unabhängige Folge $(X_n)_{n \geq 1}$ identisch verteilter Zufallsgrößen mit positivem Erwartungswert μ , seien*

$$S_0 = 0 \quad \text{und} \quad S_n = \sum_{k=1}^n X_k \quad \text{für } n \geq 1$$

der zugehörige Random Walk und $(\mathcal{F}_n)_{n \geq 0}$ eine Filtration, so dass

- (F1) $\sigma(S_0, \dots, S_n) \subset \mathcal{F}_n$ für alle $n \geq 0$,
(F2) \mathcal{F}_n und $(X_k)_{k > n}$ sind unabhängig.

Dann gelten folgende Aussagen:

- (a) Für jede $(\mathcal{F}_n)_{n \geq 0}$ -Zeit τ mit $\mathbb{E}\tau < \infty$ gilt die sogenannte **(erste) Waldsche Gleichung**

$$\mathbb{E}S_\tau = \mu \mathbb{E}\tau. \quad (2.37)$$

- (b) Für die Erstaustrittszeiten $\tau(t) := \inf\{n \geq 0 : S_n > t\}$, $t \geq 0$, gelten

$$\frac{\tau(t)}{t} \rightarrow \frac{1}{\mu} \quad \mathbb{P}\text{-f.s.} \quad \text{und} \quad \frac{\mathbb{E}\tau(t)}{t} \rightarrow \frac{1}{\mu} \quad (2.38)$$

sowie im Fall $\mu < \infty$

$$\frac{S_{\tau(t)} - t}{t} \rightarrow 0 \quad \mathbb{P}\text{-f.s.} \quad \text{und} \quad \frac{\mathbb{E}(S_{\tau(t)} - t)}{t} \rightarrow 0, \quad (2.39)$$

falls $t \rightarrow \infty$.

Beweis. (a) Wir setzen zunächst $X_1 \geq 0$ voraus. Der entscheidende Trick zum Nachweis der Waldschen Gleichung besteht in der Darstellung von S_τ als

$$S_\tau = \sum_{n \geq 1} \mathbf{1}_{\{\tau \geq n\}} X_n \quad \mathbb{P}\text{-f.s.}$$

wie sich durch Vertauschung der Summation aus $S_\tau = \sum_{n \geq 1} \mathbf{1}_{\{\tau \geq n\}} \sum_{k=1}^n X_k$ ergibt. Dabei beachte man, dass τ wegen $\mathbb{E}\tau < \infty$ insbesondere f.s. endlich ist. Da $\{\tau \geq n\} = \{\tau \leq n-1\}^c \in \mathcal{F}_{n-1}$ und daher von X_n unabhängig ist, folgt nun (alle Summanden nichtnegativ)

$$\mathbb{E}S\tau = \mathbb{E}\left(\sum_{n \geq 1} \mathbf{1}_{\{\tau \geq n\}} X_n\right) = \sum_{n \geq 1} \mathbb{P}(\tau \geq n) \mathbb{E}X_n = \mu \sum_{n \geq 1} \mathbb{P}(\tau \geq n) = \mu \mathbb{E}\tau.$$

Für beliebige X_n ergibt sich die Behauptung wegen $\mathbb{E}X_1^- < \infty$ aus dem zuvor Bewiesenen durch Zerlegung in Positiv- und Negativteil:

$$\mathbb{E}S\tau = \mathbb{E}\left(\sum_{n=1}^{\tau} X_n^+\right) - \mathbb{E}\left(\sum_{n=1}^{\tau} X_n^-\right) = (\mathbb{E}X_1^+ - \mathbb{E}X_1^-) \mathbb{E}\tau = \mu \mathbb{E}\tau.$$

(b) Wegen $\mu > 0$ gilt $S_n \rightarrow \infty$ und damit $\tau(t) < \infty$ \mathbb{P} -f.s. für alle $t \geq 0$. Da außerdem $\tau(t) \uparrow \infty$ \mathbb{P} -f.s., folgt $\frac{S_{\tau(t)}}{\tau(t)} \rightarrow \mu$ \mathbb{P} -f.s. aus dem starken Gesetz der großen Zahlen. Zusammen mit der Ungleichung $S_{\tau(t)-1} \leq t < S_{\tau(t)}$ liefert dies

$$\limsup_{t \rightarrow \infty} \frac{t}{\tau(t)} \leq \lim_{t \rightarrow \infty} \frac{S_{\tau(t)}}{\tau(t)} = \mu \quad \mathbb{P}\text{-f.s.}$$

sowie umgekehrt

$$\liminf_{t \rightarrow \infty} \frac{t}{\tau(t)} \geq \lim_{t \rightarrow \infty} \frac{S_{\tau(t)-1}}{\tau(t)} = \mu \quad \mathbb{P}\text{-f.s.}$$

also insgesamt die erste Behauptung in (2.38). Selbige in (2.39), falls $\mu < \infty$, ergibt sich dann aber aus

$$\frac{S_{\tau(t)} - t}{t} = \frac{S_{\tau(t)}}{\tau(t)} \cdot \frac{\tau(t)}{t} - 1.$$

Etwas mehr Arbeit bereiten die zweiten Konvergenzaussagen in (2.38) und (2.39): Wir zeigen als erstes $\mathbb{E}\tau(t) < \infty$ für alle $t \geq 0$ und setzen dazu $Y_n := X_n \wedge c$, $W_n = Y_1 + \dots + Y_n$ sowie $\tau_c(t) := \inf\{n : W_n > t\}$ für beliebiges hinreichend großes $c > 0$, so dass $\mu_c := \mathbb{E}Y_1 > 0$. Natürlich bildet auch $(W_n)_{n \geq 0}$ einen Random Walk, der \mathbb{P} -f.s. gegen ∞ strebt ($\Rightarrow \tau_c(t) < \infty$ \mathbb{P} -f.s. für alle $t \geq 0$). Aus $W_n \leq S_n$ folgt $\tau(t) \leq \tau_c(t)$ für alle $t \geq 0$ und aus $Y_{\tau_c(t)} \leq c$ die Ungleichung

$$t < W_{\tau_c(t)} = W_{\tau_c(t)-1} + Y_{\tau_c(t)} \leq t + c.$$

Vermöge der in (a) gezeigten Waldschen Gleichung ergibt sich nun

$$\mu_c \mathbb{E}(\tau_c(t) \wedge n) = \mathbb{E}W_{\tau_c(t) \wedge n} = \mathbb{E}W_{\tau_c(t)} \mathbf{1}_{\{\tau_c(t) \leq n\}} + \mathbb{E}W_n \mathbf{1}_{\{\tau_c(t) > n\}} \leq c + t$$

und daraus mit dem Satz von der monotonen Konvergenz

$$\mathbb{E}\tau(t) \leq \mathbb{E}\tau_c(t) = \lim_{n \rightarrow \infty} \mathbb{E}(\tau_c(t) \wedge n) \leq \frac{c+t}{\mu_c} < \infty \quad (2.40)$$

für alle $t \geq 0$. Eine erneute Anwendung der Waldschen Gleichung liefert als nächstes $\mu \mathbb{E}\tau(t) = \mathbb{E}S\tau(t) \geq t$ und damit

$$\liminf_{t \rightarrow \infty} \frac{\mathbb{E}\tau(t)}{t} \geq \frac{1}{\mu}.$$

(Dies kann übrigens auch mit dem Lemma von Fatou aus dem ersten Teil von (2.38) gefolgert werden.) Für die umgekehrte Ungleichung

$$\limsup_{t \rightarrow \infty} \frac{\mathbb{E}\tau(t)}{t} \leq \frac{1}{\mu} \quad (2.41)$$

bemühen wir Gleichung (2.40), aus der sich

$$\limsup_{t \rightarrow \infty} \frac{\mathbb{E}\tau(t)}{t} \leq \lim_{t \rightarrow \infty} \frac{c+t}{\mu_c t} = \frac{1}{\mu_c}$$

für alle hinreichend großen c ergibt und daraus (2.41) per Grenzübergang $c \uparrow \infty$, weil $\mu_c \uparrow \mu$. Damit ist die zweite Behauptung in (2.38) gezeigt, die aber zusammen mit der Waldschen Gleichung auch sofort die zweite Aussage in (2.39) impliziert, sofern $\mu < \infty$, denn

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}(S_{\tau(t)} - t)}{t} = \lim_{t \rightarrow \infty} \frac{\mu \mathbb{E}\tau(t)}{t} - 1 = 0. \quad \square$$

Zur Formulierung der Hauptergebnisse dieses Abschnitts sei zuvor folgende Erläuterung gegeben. Wir hatten in Satz 2.28 festgehalten, dass der Zustandsraum einer DMK in lauter abgeschlossene Rekurrenzklassen sowie eine weitere Menge transienter Zustände zerfällt. Sehen wir für einen Moment von der Möglichkeit transienter Zustände ab, so ist klar, dass wir das Verhalten der Kette auf jeder Rekurrenzklasse getrennt beschreiben können, da es Übergänge zwischen diesen Klassen nicht gibt. Mit anderen Worten: Es reicht, (irreduzible) rekurrente DMK zu betrachten. Die Berücksichtigung transienter Zustände wird sich anschließend relativ mühelos unter Verwendung der zuvor erzielten Resultate bewerkstelligen lassen.

Satz 2.33. *Für eine rekurrente DMK $M = (M_n)_{n \geq 0}$ gelten folgende Aussagen:*

- (a) *M besitzt ein bis auf skalares Vielfaches eindeutig bestimmtes stationäres Maß $\pi = (\pi_i)_{i \in \mathcal{S}}$, $0 < \pi_i < \infty$ für alle $i \in \mathcal{S}$, das genau dann endlich ist, wenn M positiv rekurrent ist. In diesem Fall bezeichnet $\pi^* := \pi/\pi(\mathcal{S})$ die eindeutig bestimmte stationäre Verteilung von M .*
- (b) *Für $\pi^* = (\pi_i^*)_{i \in \mathcal{S}}$ gilt außerdem*

$$\pi_i^* = \frac{1}{\mu_{ii}}, \quad i \in \mathcal{S}, \quad (2.42)$$

sofern die rechte Seite im null-rekurrenten Fall ($\mu_{ii} = \infty$ für alle $i \in \mathcal{S}$) als 0 interpretiert wird, d.h. $\pi^* \equiv 0$.

- (c) Die empirischen Verteilungen $\frac{1}{n+1} \sum_{k=0}^n \delta_{M_k}$ konvergieren auf $\mathfrak{S}_\pi := \{A : \pi(A) < \infty\}$ \mathbb{P}_λ -f.s. punktweise gegen π^* , d.h.

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \delta_{M_k}(A) = \pi^*(A) \quad \mathbb{P}_\lambda\text{-f.s.} \quad (2.43)$$

für jedes $A \in \mathfrak{S}_\pi$ und jede Anfangsverteilung λ .

- (d) Die Césaro-Mittel $\frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda^{M_k}$ konvergieren auf \mathfrak{S}_π ebenfalls punktweise gegen π^* , also

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda^{M_k}(A) = \pi^*(A) \quad (2.44)$$

für jedes $A \in \mathfrak{S}_\pi$ und jede Anfangsverteilung λ , wobei im positiv rekurrenten Fall ($\Rightarrow \mathfrak{S}_\pi = \mathfrak{P}(\mathcal{S})$) sogar

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda^{M_k} - \pi^* \right\| = 0, \quad (2.45)$$

d.h. in (2.44) gleichmäßige Konvergenz in $A \subset \mathcal{S}$ vorliegt.

Anmerkung 2.34. (2.44) bildet nichts anderes als die L^1 -Version von (2.43), denn

$$\frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda^{M_k}(A) = \mathbb{E}_\lambda \left(\frac{1}{n+1} \sum_{k=0}^n \delta_{M_k}(A) \right). \quad (2.46)$$

für jedes $A \in \mathfrak{S}_\pi$. Aufgrund majorisierter Konvergenz folgt dann, dass (2.44) äquivalent ist zu

$$\lim_{n \rightarrow \infty} \mathbb{E}_\lambda \left| \frac{1}{n+1} \sum_{k=0}^n \mathbf{1}_A(M_k) - \pi^*(A) \right| = 0,$$

d.h. $\frac{1}{n+1} \sum_{k=0}^n \mathbf{1}_A(M_k) \xrightarrow{L^1} \pi^*(A)$ für alle $A \in \mathfrak{S}_\pi$ und $\lambda \in \mathcal{P}(\mathcal{S})$. Dagegen besagt (2.43) gerade $\frac{1}{n+1} \sum_{k=0}^n \mathbf{1}_A(M_k) \rightarrow \pi^*(A)$ \mathbb{P}_λ -f.s.

Anmerkung 2.35. Per Funktions-Erweiterungsargument ergeben sich sofort folgende Äquivalenzen zu (2.43)–(2.45): Bezeichnet $b\mathfrak{S}$ den Raum aller beschränkten Funktionen $f : (\mathcal{S}, \mathfrak{S}) \rightarrow (\mathbb{R}, \mathfrak{B})$, $\|\cdot\|_\infty$ die Supremumsnorm auf diesem und $b\mathfrak{S}_\pi$ den Teilraum aller f mit Träger in \mathfrak{S}_π , d.h. $f|_{A^c} \equiv 0$ für ein $A \in \mathfrak{S}_\pi$, so implizieren (2.43) und (2.44)

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n f(M_k) = \int_{\mathcal{S}} f(s) \pi^*(ds) \quad \mathbb{P}_\lambda \text{ f.s.} \quad (2.47)$$

$$\text{bzw.} \quad \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \mathbb{E}_\lambda f(M_k) = \int_{\mathcal{S}} f(s) \pi^*(ds) \quad (2.48)$$

für alle $f \in b\mathfrak{S}_\pi$ und $\lambda \in \mathcal{P}(\mathcal{S})$, während (2.45) äquivalent ist zu

$$\lim_{n \rightarrow \infty} \sup_{f \in b\mathfrak{S}, \|f\|_\infty \leq 1} \left| \frac{1}{n+1} \sum_{k=0}^n \mathbb{E}_\lambda f(M_k) - \int_{\mathcal{S}} f(s) \pi^*(ds) \right| = 0 \quad (2.49)$$

für alle $\lambda \in \mathcal{P}(\mathcal{S})$. Dabei steht im null-rekurrenten Fall auf der rechten Seite in (2.47) und (2.48) natürlich 0, und es gilt $b\mathfrak{S}_\pi = b\mathfrak{S}$, falls M positiv rekurrent ist. (2.47) und (2.48) bleiben ferner auch für nichtnegative f mit Träger in \mathfrak{S}_π gültig, wie man nach Lesen des Beweises von Satz 2.33 leicht einsieht.

Anmerkung 2.36. Aus der für alle $n \geq 0$, $A \in \mathfrak{S}^\infty$ und $\lambda \in \mathcal{P}(\mathcal{S})$ gültigen Beziehung $\mathbb{P}_\lambda(M^{(n)} \in A) = \int_{\mathcal{S}} \mathbb{P}_s(M \in A) \mathbb{P}_\lambda^{M_n}(ds)$ in Verbindung mit (2.49) ergibt sich im positiv rekurrenten Fall leicht

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda^{M^{(k)}} - \mathbb{P}_{\pi^*}^M \right\| = 0. \quad (2.50)$$

als weitere Äquivalenz zu (2.45) für alle $\lambda \in \mathcal{P}(\mathcal{S})$. Auch existieren entsprechende Verallgemeinerungen der Aussagen (2.43) und (2.44) für die Post- n -Folgen $M^{(n)}$, sind aber von geringerer Bedeutung. Wir verzichten deshalb auf weitere Details.

Anmerkung 2.37. Ergebnisse der Form (2.43) – (2.45), die die Konvergenz irgendeiner Art (f.s., in L^1 , in Totalvariation, etc.) eines stochastischen Prozesses, hier einer DMK M , gegen einen stationären Limes zum Inhalt haben, bezeichnet man als *Ergodensätze*, wobei der Begriff “*Ergodizität*” aus der statistischen Physik stammt. Wir verzichten an dieser Stelle auf nähere Erläuterungen, weisen allerdings noch auf eine Interpretation der Ergebnisse in diesem Zusammenhang hin, wobei wir (2.43) im positiv rekurrenten Fall herausgreifen: Auf der linken Seite steht dort, gegeben $A \subset \mathcal{S}$, das *zeitliche Mittel* der Aufenthalte der Kette M in A , auf der rechten Seite dagegen das *räumliche Mittel* $\pi^*(A) = \int_{\mathcal{S}} \mathbf{1}_A(s) \pi^*(ds) = \mathbb{P}_{\pi^*}(M_n \in A)$ im Gleichgewicht. Unter \mathbb{P}_{π^*} stimmen demnach für jedes $A \subset \mathcal{S}$ zeitliches und räumliches Mittel f.s. überein.

Anmerkung 2.38. Wählt man $A = \{j\}$ und $\lambda = \delta_i$ in (2.44), so ergibt sich die Césaro-Konvergenz der Übergangswahrscheinlichkeiten, nämlich

$$C\text{-}\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j^* \quad (2.51)$$

für alle $i, j \in \mathcal{S}$, wobei im null-rekurrenten Fall auf der rechten Seite stets 0 steht. Kombiniert man dies mit dem Rekurrenz Kriterium in Satz 2.22, so ergibt sich nun direkt

Korollar 2.39. Ein Zustand $i \in \mathcal{S}$ ist genau dann null-rekurrent, wenn

$$\sum_{n \geq 0} p_{ii}^{(n)} = \infty \quad \text{und} \quad C\text{-}\lim_{n \rightarrow \infty} p_{ii}^{(n)} = 0$$

gilt. In diesem Fall gilt weiter $C\text{-}\lim_{n \rightarrow \infty} p_{ji}^{(n)} = 0$ für alle $j \in \mathcal{S}$.

Beweis. Wir dürfen gleich $i \in \mathcal{S}$ als rekurrent voraussetzen. Da die zu i gehörende Klasse dann abgeschlossen ist, bildet M unter \mathbb{P}_i eine irreduzible rekurrente DMK mit Zustandsraum \mathcal{C}_i . Gemäß (2.51) liegt somit Null-Rekurrenz genau dann vor, wenn neben $\sum_{n \geq 0} p_{ii}^{(n)} = \infty$ auch $C\text{-}\lim_{n \rightarrow \infty} p_{ii}^{(n)} = 0$ gilt. Mittels (1.30) folgt dann leicht auch $C\text{-}\lim_{n \rightarrow \infty} p_{ji}^{(n)} = 0$ für alle $j \in \mathcal{S}$. Die Details überlassen wir dem Leser zur Übung. \square

Da im Fall eines endlichen Zustandsraums jedes stationäre Maß notwendigerweise endlich ist, erhalten wir als weitere triviale Konsequenz:

Korollar 2.40. Jede irreduzible EMK M ist positiv rekurrent und besitzt folglich eine eindeutig bestimmte stationäre Verteilung, nämlich π^* gemäß (2.42).

Für beliebige EMK ergibt sich aus Korollar 2.40 ferner leicht die folgende Ergänzung zu Satz 2.29:

Korollar 2.41. Jeder rekurrente Zustand einer EMK ist bereits positiv rekurrent.

Beweis (von Satz 2.33). (a) Sei $\pi = (\pi_i)_{i \in \mathcal{S}}$ irgendein stationäres Maß der DMK M , d.h. $\pi P = \pi$, $\pi_i < \infty$ für alle $i \in \mathcal{S}$ und $\pi_i > 0$ für mindestens ein $i \in \mathcal{S}$. Wir zeigen zuerst, dass aus der Irreduzibilität von M sogar $\pi_i > 0$ für alle $i \in \mathcal{S}$ folgt. Wählen wir dazu ein $i_0 \in \mathcal{S}$ mit $\pi_{i_0} > 0$ und dann ein beliebiges $i \in \mathcal{S}$, $i \neq i_0$. Wegen $i_0 \rightarrow i$ existiert ein $n \geq 1$ mit $p_{i_0 i}^{(n)} > 0$. Daher ergibt sich wie behauptet unter Beachtung von $\pi = \pi P^n$

$$\pi_i = \sum_{j \in \mathcal{S}} \pi_j p_{ji}^{(n)} \geq \pi_{i_0} p_{i_0 i}^{(n)} > 0,$$

Als nächstes wählen wir ein beliebiges $i \in \mathcal{S}$ und nehmen o.B.d.A. $\pi_i = 1$ an. Andernfalls gehe man einfach zu dem wiederum stationären Maß $\pi' = \pi/\pi_i$ über. Die Invarianzgleichungen haben dann die Form

$$\pi_i = 1 \quad \text{und} \quad \pi_j = \sum_{k \in \mathcal{S}} \pi_k p_{kj}, \quad j \neq i,$$

was zusammengefasst

$$\pi_j = \delta_{ij} + \sum_{k \in \mathcal{S}} \pi_k q_{kj}, \quad \text{d.h.} \quad \pi = \delta_{i\bullet} + \pi Q \quad (2.52)$$

ergibt, wobei $Q = (q_{jk})_{j,k \in \mathcal{S}}$ aus der Matrix P entsteht, indem man in deren i -ten Spalte alle Komponenten durch Nullen ersetzt. Wir zeigen nun

$$\pi \geq {}^{(i)}\pi, \quad \text{d.h.} \quad \pi_j \geq {}^{(i)}\pi_j \quad \text{für alle } j \in \mathcal{S}, \quad (2.53)$$

und notieren zu diesem Zweck, dass für die Komponenten $q_{jk}^{(n)}$ von Q^n

$$\begin{aligned} q_{jk}^{(n)} &= \sum_{j_1 \neq i} \dots \sum_{j_{n-1} \neq i} P_{jj_1} P_{j_1 j_2} \dots P_{j_{n-1} k} \\ &= \mathbb{P}_j(M_1 \neq i, \dots, M_{n-1} \neq i, M_n = k) = \mathbb{P}_j(M_n = k, \tau(i) > n) \end{aligned}$$

im Fall $k \neq i$ gilt und dass außerdem $(Q^0 := I = (\delta_{ij})_{i,j \in \mathcal{S}})$

$$\begin{aligned} {}^{(i)}\pi_j &= \mathbb{E}_i \left(\sum_{n=0}^{\tau(i)-1} \mathbf{1}_{\{M_n=j\}} \right) \\ &= \sum_{m \geq 1} \sum_{n=0}^{m-1} \mathbb{P}_i(M_n = j, \tau(i) = m) \\ &= \sum_{n \geq 0} \sum_{m > n} \mathbb{P}_i(M_n = j, \tau(i) = m) \\ &= \sum_{n \geq 0} \mathbb{P}_i(M_n = j, \tau(i) > n) \\ &= \sum_{n \geq 0} q_{ij}^{(n)} \end{aligned} \quad (2.54)$$

für alle $j \neq i$. Damit ergibt sich in (2.52) per Iteration und unter Beachtung von $\delta_{i\bullet} Q^n = q_{i\bullet}^{(n)}$

$$\begin{aligned} \pi &= \delta_{i\bullet} + (\delta_{i\bullet} + \pi Q)Q = \delta_{i\bullet} + \delta_{i\bullet} Q + \pi Q^2 \\ &= \dots = \sum_{k=0}^n \delta_{i\bullet} Q^k + \pi Q^{n+1} \geq \sum_{k=0}^n \delta_{i\bullet} Q^k = \sum_{k=0}^n q_{i\bullet}^{(k)} \end{aligned}$$

für alle $n \geq 1$ und schließlich bei Grenzübergang $n \rightarrow \infty$ in Kombination mit (2.54)

$$\pi \geq \sum_{k \geq 0} q_{i\bullet}^{(k)} = {}^{(i)}\pi,$$

also (2.53). Setzen wir nun $\Delta := \pi - {}^{(i)}\pi \geq 0$, so bildet Δ ein σ -endliches Maß, das offenkundig wieder der Invarianzgleichung $\Delta P = \Delta$ genügt. Wäre es ein stationäres Maß, also nicht identisch 0, müsste außerdem $0 < \Delta_j < \infty$ für alle $j \in \mathcal{S}$ gelten,

wie weiter oben gezeigt wurde. Aus ${}^{(i)}\pi_i = 1$ und $\pi_i = 1$ folgt aber $\Delta_i = 0$ und somit $\Delta \equiv 0$. Jedes stationäre Maß π stimmt somit bis auf skalares Vielfaches mit ${}^{(i)}\pi$ überein, was die Eindeutigkeit beweist.

Nun ergibt sich aber weiter $\pi = c_i {}^{(i)}\pi$ für alle $i \in \mathcal{S}$ und Konstanten $c_i \in (0, \infty)$ und somit genau dann die Existenz und Eindeutigkeit einer stationären Verteilung, wenn ${}^{(i)}\pi(\mathcal{S}) = \mu_{ii}$ für alle $i \in \mathcal{S}$ endlich, d.h. M positiv rekurrent ist.

(b) Im null-rekurrenten Fall ist für (2.42) offenbar nichts zu zeigen, während dies im positiv rekurrenten Fall aus $\pi_i^* = {}^{(i)}\pi_i/\mu_{ii} = 1/\mu_{ii}$ folgt.

(c) Sei $i \in \mathcal{S}$ ein beliebig gewählter Zustand und $(\sigma_n)_{n \geq 1}$ einmal mehr die Folge der sukzessiven Rekurrenzzeiten in diesen Zustand, insbesondere also $\sigma_1 = \tau(i)$. Da alle Zustände rekurrent und verbunden sind, folgt $\sigma_n < \infty$ \mathbb{P}_j -f.s. für alle $n \geq 1$ und $j \in \mathcal{S}$. Unter \mathbb{P}_i sind die Zyklen

$$Z_n = (\tau_{n+1}, M_{\sigma_n}, \dots, M_{\sigma_{n+1}-1}), \quad n \geq 0 \quad [\sigma_0 := 0]$$

unabhängig und identisch verteilt (\mathbb{P}_i Satz 2.16), wobei insbesondere σ_n aus den unabhängigen, identisch verteilten Summanden $\tau_1 = \tau(i), \tau_2, \dots, \tau_n$ besteht. Wir definieren nun

$$\begin{aligned} N_n(A) &:= \sum_{k=0}^{\sigma_n-1} \mathbf{1}_A(M_k) = \sum_{k=1}^n f_A(Z_{k-1}), \\ f_A(n, s_0, \dots, s_{n-1}) &:= \sum_{k=0}^{n-1} \mathbf{1}_A(s_k) \\ T_m &:= \inf\{n \geq 1 : \sigma_n > m\} \end{aligned}$$

für $A \subset \mathcal{S}$, $m \in \mathbb{N}_0$, $n \in \mathbb{N}$ und $(s_0, \dots, s_{n-1}) \in \mathcal{S}^n$. Dann gilt offensichtlich

$$N_{T_n-1}(A) \leq \sum_{k=0}^n \mathbf{1}_A(M_k) \leq N_{T_n}(A)$$

und somit auch

$$\left(\frac{N_{T_n-1}(A)}{T_n-1} \right) \left(\frac{T_n-1}{n+1} \right) \leq \frac{1}{n+1} \sum_{k=0}^n \mathbf{1}_A(M_k) \leq \left(\frac{N_{T_n}(A)}{T_n} \right) \left(\frac{T_n}{n+1} \right). \quad (2.55)$$

Mit Hilfe des starken Gesetzes der großen Zahlen sowie von Lemma 2.32 (dessen Anwendbarkeit auf T_n sollte der Leser selbst überprüfen) erhalten wir

$$\lim_{n \rightarrow \infty} \frac{N_{T_n}(A)}{T_n} = \mathbb{E}_i f_A(Z_0) \quad \text{und} \quad \lim_{n \rightarrow \infty} \frac{T_n}{n} = \frac{1}{\mu_{ii}} \quad \mathbb{P}_i\text{-f.s.}, \quad (2.56)$$

wobei außerdem

$$\mathbb{E}_i f_A(Z_0) = \mathbb{E}_i \left(\sum_{k=0}^{\tau(i)-1} \mathbf{1}_A(M_k) \right) = {}^{(i)}\pi(A).$$

Folglich ergibt sich in (2.55) für jedes $A \subset \mathcal{S}$ mit ${}^{(i)}\pi(A) < \infty$ offenbar (2.43) für $\lambda = \delta_i$ und $\pi^* = {}^{(i)}\pi/\mu_{ii}$. Da i beliebig vorgegeben war, gilt also

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \mathbf{1}_A(M_k) = \frac{{}^{(i)}\pi(A)}{\mu_{ii}} = {}^{(i)}\pi^*(A) \quad \mathbb{P}_i\text{-f.s.} \quad (2.57)$$

für alle ${}^{(i)}\pi$ -endlichen A und alle $i \in \mathcal{S}$. Andererseits behaupten wir, dass (2.56) und folglich (2.57) auch \mathbb{P}_λ -f.s. für alle $\lambda \in \mathcal{P}(\mathcal{S})$ gilt, was zum ersten zeigt, dass $\pi^*(A) := {}^{(i)}\pi^*(A)$ nicht von i abhängt, und zum zweiten dann (2.43) in voller Allgemeinheit.

Die noch offene Behauptung betreffend notieren wir als erstes, dass unter \mathbb{P}_λ die Z_n sowie speziell die τ_{n+1} weiterhin für $n \geq 0$ stochastisch unabhängig und für $n \geq 1$ auch identisch wie unter \mathbb{P}_i verteilt sind (Satz 2.18), was die \mathbb{P}_λ -f.s. Gültigkeit der ersten Aussage in (2.56) liefert. Setzen wir $\widehat{T}_m = \inf\{n \geq 0 : \sigma_{n+1} - \sigma_1 > m\}$, folgt

$$\mathbb{P}_\lambda((\widehat{T}_n)_{n \geq 0} \in \cdot) = \mathbb{P}_i((T_n)_{n \geq 0} \in \cdot)$$

aus $\mathbb{P}_\lambda((\sigma_{n+1} - \sigma_1)_{n \geq 0} \in \cdot) = \mathbb{P}_i((\sigma_n)_{n \geq 0} \in \cdot)$ und daher $n^{-1}\widehat{T}_n \rightarrow \mu_{ii}^{-1}$ \mathbb{P}_λ -f.s. per erneuter Anwendung von Lemma 2.32. Aufgrund der Darstellung

$$T_n = 1 + \mathbf{1}_{\{\tau(i) \leq n\}} \widehat{T}_{n-\tau(i)}$$

erhalten wir schließlich auch die zweite Aussage in (2.56) \mathbb{P}_λ -f.s., denn $\tau(i) < \infty$ \mathbb{P}_λ -f.s.

(d) (2.44) folgt unter Hinweis auf (2.46) und dem Satz von der majorisierten Konvergenz direkt aus (2.43). Für (2.45) bleibt somit nur noch die Gleichmäßigkeit der Konvergenz in A im positiv rekurrenten Fall nachzuweisen. Diese ergibt sich aber als Folgerung aus Satz 12.3 in [2], wenn wir dort $p = 1$, μ als Zählmaß auf \mathcal{S} sowie

$$f_n(i) := \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda(M_k = i) \quad \text{und} \quad f(i) := \pi_i^*$$

für $n \geq 0$ und $i \in \mathcal{S}$ wählen. Wegen $f_n(i) \rightarrow f(i)$ für alle $i \in \mathcal{S}$ (gemäß (2.44)) und $\int f_n d\mu = \int f d\mu = 1$ für alle $n \geq 0$ liefert dieser Satz nämlich

$$\lim_{n \rightarrow \infty} \int |f_n - f| d\mu = \lim_{n \rightarrow \infty} \sum_{i \in \mathcal{S}} |f_n(i) - f(i)| = 0,$$

woraus sich das Gewünschte vermöge der Identität (2.31)

$$\left\| \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_\lambda^{M_k} - \pi^* \right\| = \frac{1}{2} \sum_{i \in \mathcal{S}} |f_n(i) - f(i)|$$

ergibt. □

Anmerkung 2.42. Satz 2.33(a) betreffend, sei darauf hingewiesen, dass die Eindeutigkeit der stationären Verteilung π^* einer positiv rekurrenten DMK M auch ohne die Eindeutigkeit bis auf skalares Vielfaches des stationären Maßes folgt, und zwar aus (2.44). Setzen wir dort nämlich $\lambda = \zeta$ für irgendeine weitere stationäre Verteilung ζ , folgt zum einen $\frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_{\zeta}^{M^k}(A) \rightarrow \pi^*(A)$ für alle $A \subset \mathcal{S}$, zum anderen $\frac{1}{n+1} \sum_{k=0}^n \mathbb{P}_{\zeta}^{M^k} = \zeta$ aus der Stationarität von ζ (Lemma 1.32), was $\zeta = \pi^*$ zeigt.

Anmerkung 2.43. Nach Satz 2.33(a) gilt ${}^{(i)}\pi = c_{ij} {}^{(j)}\pi$ für geeignetes $c_{ij} \in (0, \infty)$, was zusammen mit ${}^{(i)}\pi_i = 1$ sofort $c_{ij} = 1/{}^{(j)}\pi_i$ ergibt, also

$${}^{(i)}\pi = {}^{(j)}\pi / {}^{(j)}\pi_i$$

für alle $i, j \in \mathcal{S}$. Ist M positiv rekurrent, liefert ${}^{(i)}\pi = \mu_{ii} \pi^*$ in Verbindung mit (2.42) außerdem

$${}^{(i)}\pi_j = \frac{\mu_{ii}}{\mu_{jj}}$$

für alle $i, j \in \mathcal{S}$.

Satz 2.33 besitzt eine kanonische Verallgemeinerung auf den Fall, in dem neben einer Klasse rekurrenter Zustände auch noch eine Menge transienter Zustände existiert, von der aus man f.s. irgendwann einen rekurrenten Zustand erreicht:

Satz 2.44. Sei $M = (M_n)_{n \geq 0}$ eine DMK mit Zustandsraum \mathcal{S} , der in eine Menge (nicht notwendig Klasse) \mathcal{T} transienter und eine Klasse \mathcal{R} rekurrenter Zustände zerfällt, wobei ferner $\mathbb{P}_i(\tau(\mathcal{R}) < \infty) = 1$ für alle $i \in \mathcal{T}$. Ersetzt man dann in Satz 2.33 “ M (null-/positiv) rekurrent” durch “ \mathcal{R} (null-/positiv) rekurrent”, so gelten dort weiterhin sämtliche Aussagen außer, dass $\pi_i = 0$ für alle $i \in \mathcal{T}$.

Beweis. Wir geben nur den Hinweis, dass $(M_{\tau(\mathcal{R})+n})_{n \geq 0}$ eine irreduzible rekurrente DMK mit Zustandsraum \mathcal{R} bildet. Die weiteren Details bleiben dem Leser überlassen. \square

Als letzte Folgerung aus den bisherigen Ergebnissen wollen wir zeigen, dass, wenn immer eine DMK eine stationäre Verteilung π^* besitzt, die Menge ihrer nullrekurrenten und transienten Zustände eine π^* -Nullmenge darstellt. Da keine Irreduzibilität vorausgesetzt wird, kann die Menge \mathcal{R} der positiv rekurrenten Zustände natürlich in mehrere Klassen zerfallen.

Satz 2.45. Sei $M = (M_n)_{n \geq 0}$ eine DMK, die eine stationäre Verteilung π besitzt. Dann zerfällt \mathcal{S} in eine (möglicherweise leere) Menge \mathcal{N} nullrekurrenter oder transienter sowie eine Menge $\mathcal{R} \neq \emptyset$ lauter positiv rekurrenter Zustände, wobei $\pi_i = 0$ für alle $i \in \mathcal{N}$ gilt.

Beweis. Sei $\mathcal{N} \neq \emptyset$ und $j \in \mathcal{N}$ beliebig. Aus Satz 2.22 und Korollar 2.39 folgern wir $C\text{-}\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$ für alle $i \in \mathcal{S}$. Damit ergibt sich unter Benutzung des Satzes von der majorisierten Konvergenz

$$\pi_j = C\text{-}\lim_{n \rightarrow \infty} \sum_{i \in \mathcal{S}} \pi_i p_{ij}^{(n)} = \sum_{i \in \mathcal{S}} \pi_i C\text{-}\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0.$$

Also gilt für jeden null-rekurrenten oder transienten Zustand j notwendig $\pi_j = 0$. Da andererseits $\sum_{i \in \mathcal{S}} \pi_i = 1$, muss es mindestens einen rekurrenten Zustand geben, d.h. $\mathcal{R} = \mathcal{N}^c$ ist nicht leer. \square

Als Ergänzung notieren wir, dass \mathcal{R} nach Satz 2.28 in disjunkte (abgeschlossene) Rekurrenzklassen zerfällt, auf denen M irreduzibel und positiv rekurrent ist. Bezeichnet $\pi^{*(\alpha)}$ die eindeutige stationäre Verteilung von M bei Einschränkung auf \mathfrak{R}_α und setzt man $\pi_j^{*(\alpha)} = 0$ für $j \notin \mathfrak{R}_\alpha$, so ergibt sich jede stationäre Verteilung der Kette auf ganz \mathcal{S} als konvexe Kombination der $\pi^{*(\alpha)}$, wie der Leser leicht nachweisen kann ($\mathfrak{E}^{\mathfrak{S}}$ auch Diskussion vor Lemma 1.34).

2.5.3 Null-Rekurrenz unter der Lupe

Nicht zuletzt zum besseren Verständnis des Verhaltens null-rekurrenter DMK wollen wir nun auf diese einen genaueren Blick werfen und insbesondere die Frage klären, welche Bedeutung ein *unendliches* stationäres Maß aus probabilistischer Sicht für die jeweilige Kette besitzt. Den Schlüssel hierfür bildet der anschließende Satz, der in Kürze besagt, dass jede DMK nach *Ausdünnung* hinsichtlich ihrer Aufenthalte außerhalb einer beliebigen *Rekurrenzmenge* \mathcal{R} des Zustandsraums immer noch eine zeitlich homogene DMK bildet. Dabei heißt \mathcal{R} Rekurrenzmenge (für M), wenn $\mathbb{P}_i(\tau(\mathcal{R}) < \infty) = 1$ für alle $i \in \mathcal{R}$. Sie kann also durchaus transiente Zustände enthalten, ja sogar nur aus transienten Zuständen bestehen ($\mathcal{R} = \mathcal{S}$ ist schließlich immer rekurrent) und sollte daher nicht mit einer Rekurrenzklasse oder einer Menge rekurrenter Zustände verwechselt werden!

Satz 2.46. Sei $M = (M_n)_{n \geq 0}$ eine DMK mit kanonischer Filtration $(\mathcal{G}_n)_{n \geq 0}$ und \mathcal{R} eine Rekurrenzmenge. Bezeichnet $(\sigma_n(\mathcal{R}))_{n \geq 1}$ die zugehörige Folge der sukzessiven Eintritte in \mathcal{R} , d.h.

$$\sigma_n(\mathcal{R}) := \inf\{k > \sigma_{n-1}(\mathcal{R}) : M_k \in \mathcal{R}\}, \quad n \geq 1, \quad [\sigma_0(\mathcal{R}) := 0]$$

und $M_n^{\mathcal{R}} := M_{\sigma_n(\mathcal{R})}$, so bildet $M^{\mathcal{R}} = (M_n^{\mathcal{R}})_{n \geq 0}$ unter jedem \mathbb{P}_λ mit $\lambda(\mathcal{R}^c) = 0$ eine DMK bezüglich $(\mathcal{G}_{\sigma_n(\mathcal{R})})_{n \geq 0}$ mit Zustandsraum \mathcal{R} , die genau dann irreduzibel ist, wenn alle Zustände in \mathcal{R} miteinander kommunizieren. Besitzt

M einen rekurrenten Zustand $i \in \mathcal{R}$, so bildet ${}^{(i)}\pi^{\mathcal{R}} = {}^{(i)}\pi(\cdot \cap \mathcal{R})$, also die Einschränkung von ${}^{(i)}\pi$ auf \mathcal{R} , ein stationäres Maß für $M^{\mathcal{R}}$.

Im Englischen nennt man $M^{\mathcal{R}}$ auch die zu \mathcal{R} gehörende “hit-chain”.

Beweis. Zur Abkürzung schreiben wir σ_n anstelle von $\sigma_n(\mathcal{R})$ und setzen wieder $\tau_n = \sigma_n - \sigma_{n-1}$ für $n \geq 1$. Aus der Rekurrenz von \mathcal{R} folgt die f.s. Endlichkeit der σ_n unter jedem $\lambda \in \mathcal{P}(\mathcal{R})$. Dass $M^{\mathcal{R}}$ bezüglich $(\mathcal{G}_{\sigma_n})_{n \geq 0}$ adaptiert ist, bedarf keines Beweises. Die starke Markov-Eigenschaft für M impliziert dann sofort die gewöhnliche Markov-Eigenschaft für $M^{\mathcal{R}}$ sowie für alle $i, j \in \mathcal{R}$ und $n \geq 0$

$$\begin{aligned} \mathbb{P}(M_{n+1}^{\mathcal{R}} = j | M_n^{\mathcal{R}} = i) &= \mathbb{P}(M_{\sigma_{n+1}} = j | M_{\sigma_n} = i) \\ &= \mathbb{P}(M_{\sigma_n + \tau_{n+1}} = j | M_{\sigma_n} = i) \\ &= \mathbb{P}_i(M_{\tau(\mathcal{R})} = j). \end{aligned}$$

$M^{\mathcal{R}}$ hat also die Übergangsmatrix $P^{\mathcal{R}} = (p_{ij}^{\mathcal{R}})_{i, j \in \mathcal{R}}$ mit

$$p_{ij}^{\mathcal{R}} = \mathbb{P}_i(M_{\tau(\mathcal{R})} = j) = \sum_{k \geq 1} \mathbb{P}_i(M_k = j, M_l \notin \mathcal{R}, 1 \leq l < k).$$

Die Irreduzibilitätsbehauptung kann der Leser leicht selbst verifizieren, so dass wir uns gleich der letzten Behauptung zuwenden und annehmen, dass \mathcal{R} einen rekurrenten Zustand i enthält. Sei $\tau^{\mathcal{R}}(j) = \inf\{n \geq 1 : M_n^{\mathcal{R}} = j\}$ für $j \in \mathcal{R}$. Dann definiert ${}^{(i)}\pi^{\mathcal{R}} := \mathbb{E}_i(\sum_{n=0}^{\tau^{\mathcal{R}}(i)-1} \mathbf{1}_{\{M_n^{\mathcal{R}} \in \cdot\}})$, $i \in \mathcal{R}$, gemäß Satz 2.30 ein stationäres Maß für $M^{\mathcal{R}}$. Außerdem gilt aber

$$\sum_{n=0}^{\tau^{(i)}-1} \mathbf{1}_{\{M_n=j\}} = \sum_{n=0}^{\tau^{\mathcal{R}}(i)-1} \mathbf{1}_{\{M_n^{\mathcal{R}}=j\}}$$

für alle $j \in \mathcal{R}$, denn M und $M^{\mathcal{R}}$ unterscheiden sich bis zum ersten Erreichen von i höchstens durch zusätzliche Aufenthalte von M in \mathcal{R}^c . Durch Übergang zum Erwartungswert unter \mathbb{P}_i folgt ${}^{(i)}\pi = {}^{(i)}\pi^{\mathcal{R}}$ auf \mathcal{R} , also das Gewünschte. \square

Für rekurrente DMK halten wir als einfache Folgerung des vorherigen Satzes fest:

Korollar 2.47. Gegeben eine rekurrente DMK M mit stationärem Maß π , bildet $M^{\mathcal{R}}$ für jede nichtleere Teilmenge \mathcal{R} des Zustandsraums ebenfalls eine rekurrente DMK mit dem bis auf skalares Vielfaches eindeutig bestimmten stationären Maß $\pi^{\mathcal{R}} = \pi(\cdot \cap \mathcal{R})$.

Beweis. Klar mit Satz 2.46 und Satz 2.33(a). \square

Mit Hilfe der vorherigen Ergebnisse lässt sich nun leicht die Bedeutung des stationären Maßes für null-rekurrente DMK erklären:

Satz 2.48. *Gegeben eine null-rekurrente DMK M mit stationärem Maß π , gilt für jedes nichtleere $\mathcal{R} \in \mathfrak{S}_\pi$: $M^{\mathcal{R}}$ bildet eine positiv rekurrente DMK mit eindeutig bestimmter stationärer Verteilung $\pi(\cdot \cap \mathcal{R})/\pi(\mathcal{R})$.*

Beweis. Da $\pi_i \in (0, \infty)$ für alle $i \in \mathcal{S}$ (Satz 2.33(a)), folgt $0 < \pi(\mathcal{R}) < \infty$ für jedes nichtleere $\mathcal{R} \in \mathfrak{S}_\pi$. Nach Korollar 2.47 ist $M^{\mathcal{R}}$ wieder rekurrent mit stationärem Maß $\pi(\cdot \cap \mathcal{R})$, das bis auf skalares Vielfaches eindeutig und außerdem endlich ist. Wiederum aus Satz 2.33(a) folgt somit, dass $M^{\mathcal{R}}$ positiv rekurrent und $\pi(\cdot \cap \mathcal{R})/\pi(\mathcal{R})$ die zugehörige eindeutig stationäre Verteilung bildet. \square

Mit null-rekurrenten und positiv rekurrenten DMK verhält es sich also ähnlich wie mit σ -endlichen und endlichen Maßen: Während eine positiv rekurrente DMK M eine eindeutig bestimmte stationäre Verteilung besitzt, existiert im null-rekurrenten Fall lediglich eine aufsteigende Folge von π -endlichen Teilmengen $\mathcal{S}_n \subset \mathcal{S}$, $\mathcal{S}_n \uparrow \mathcal{S}$, so dass die Restriktion $M^{\mathcal{S}_n}$ von M auf \mathcal{S}_n für jedes n positiv rekurrent ist mit eindeutig bestimmter stationärer Verteilung, die sich durch Einschränkung und anschließende Normierung des stationären Maßes von M auf \mathcal{S}_n ergibt.

2.5.4 Wie viele stationäre Maße hat eine DMK?

Zum Ende dieses Abschnitts wollen wir uns noch kurz mit der generell interessanten Frage auseinandersetzen, wie viele stationäre Maße eine DMK haben kann. Als erstes notieren wir, dass eine reduzible DMK M , deren Zustandsraum in (notwendig abgeschlossene) Rekurrenzklassen \mathcal{R}_α zerfällt, gleich unendlich viele stationäre Maße besitzt, die sich nicht bloß durch ein skalares Vielfaches unterscheiden. M ist nämlich auf jedem \mathcal{R}_α eine irreduzible rekurrente DMK mit einem bis auf skalares Vielfaches eindeutig bestimmten stationären Maß π^α , d.h. $\pi_i^\alpha \in (0, \infty)$ für $i \in \mathcal{R}_\alpha$ und $= 0$ sonst. Dann bildet aber, wie schon in Abschnitt 1.5 bemerkt, auch jede Linearkombination $\sum_\alpha c_\alpha \pi^\alpha$ mit $c_\alpha \geq 0$ und $\sum_\alpha c_\alpha > 0$ wieder ein stationäres Maß. Da im übrigen jede DMK M mit mindestens einem rekurrenten Zustand i immer ein stationäres Maß besitzt, nämlich ${}^{(i)}\pi$, bleibt für weitere Betrachtungen nur noch der transiente Fall, in dem tatsächlich alles möglich ist, insbesondere auch, dass überhaupt kein stationäres Maß existiert. Zur Illustration geben wir drei Beispiele:

Beispiel 2.49. (Irrfahrten auf \mathbb{Z}) Sei $M = (M_n)_{n \geq 0}$ eine Irrfahrt auf \mathbb{Z} mit Parametern $p, q \in (0, 1)$, d.h. $M_n = M_0 + \sum_{k=1}^n X_k$ mit unter jedem \mathbb{P}_i unabhängigen, identisch verteilten X_k ,

$$\mathbb{P}_i(X_1 = 1) = p \quad \text{und} \quad \mathbb{P}_i(X_1 = -1) = q = 1 - p.$$

Wie in Abschnitt 2.3 gezeigt wurde, ist M genau dann rekurrent, wenn $p = q = \frac{1}{2}$. Jedes stationäre Maß π von M ist eine (nichttriviale) Lösung des Gleichungssystems $\pi_j = \sum_{i \in \mathbb{Z}} \pi_i p_{ij}$, $j \in \mathbb{Z}$, d.h. hier

$$\pi_j = p\pi_{j-1} + q\pi_{j+1}, \quad j \in \mathbb{Z}.$$

Schreiben wir dieses in der Form

$$\pi_{j+1} - \pi_j = \frac{p}{q}(\pi_j - \pi_{j-1}), \quad j \in \mathbb{Z},$$

ergeben sich offenkundig

$$\pi^{(1)} = \text{Zählmaß auf } \mathbb{Z} \quad \text{und} \quad \pi^{(2)} = \left(\left(\frac{p}{q} \right)^i \right)_{i \in \mathbb{Z}}$$

als linear unabhängige Lösungen, die nur im symmetrischen Fall $p = q = \frac{1}{2}$ – in dem es nach Satz 2.33(a) ja auch nur ein stationäres Maß bis auf skalares Vielfaches geben kann – zusammenfallen.

Beispiel 2.50. (Geburtsprozesse) Eine DMK $M = (M_n)_{n \geq 0}$ mit Zustandsraum \mathbb{Z} heißt *Geburtsprozess auf \mathbb{Z}* , wenn ihre Übergangswahrscheinlichkeiten p_{ij} die Form

$$p_{ij} = \begin{cases} \alpha_i, & \text{falls } j = i + 1 \\ \beta_i, & \text{falls } j = i \\ 0, & \text{sonst} \end{cases} \quad (\alpha_i > 0, \beta_i \geq 0, \alpha_i + \beta_i = 1),$$

haben. Da M nur Übergänge $i \rightarrow i$ und $i \rightarrow i + 1$ erlaubt und alle α_i positiv sind, strebt M_n unter jedem \mathbb{P}_i f.s. gegen unendlich, wobei die Verweildauer in einem beliebigen Zustand j geometrisch verteilt ist mit Parameter α_j . M hat also nur transiente Zustände. Zur Bestimmung des oder der stationären Maße lösen wir das zugehörige Gleichungssystem, das hier die besonders einfache Form

$$\pi_j = \alpha_{j-1}\pi_{j-1} + \beta_j\pi_j,$$

das heißt

$$\pi_j = \frac{\alpha_{j-1}}{\alpha_j}\pi_{j-1}, \quad j \in \mathbb{Z}$$

besitzt mit der bis auf skalares Vielfaches *eindeutigen* Lösung

$$\pi_0 = 1 \quad \text{und} \quad \pi_j = \frac{\alpha_0}{\alpha_j} \quad \text{für } j \neq 0.$$

Als letztes geben wir ein Beispiel, für das überhaupt kein stationäres Maß existiert.

Beispiel 2.51 (Die "Strähnen"-Kette). Sei $M = (M_n)_{n \geq 0}$ eine DMK mit Zustandsraum \mathbb{N}_0 und Übergangsmatrix

$$P = \begin{pmatrix} q_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & 0 & p_1 & 0 & 0 & \dots \\ q_2 & 0 & 0 & p_2 & 0 & \dots \\ \vdots & & & & \ddots & \end{pmatrix},$$

wobei $p_i \in (0, 1)$ für alle $i \in \mathbb{N}_0$. Zur Rechtfertigung des Namens “Strähnen”-Kette betrachtet man am besten den Fall $p_0 = p_1 = \dots = p$ und interpretiert p als Wahrscheinlichkeit, mit der ein Spieler in einem bestimmten unbegrenzt andauernden Spiel pro Runde gewinnt. $M_n = i$ bedeutet dann offenbar, dass er i aufeinanderfolgende Runden nach einer Niederlage oder nach Spielbeginn für sich entscheidet, was bekanntlich, zumindest für hinreichend große i , als *Lauf* oder *Glückssträhne* bezeichnet wird (engl. “run” oder “success run”). Gegeben eine unabhängige Folge $(X_n)_{n \geq 1}$ Bern(p)-verteilter Zufallsgrößen, wobei $X_n = 1$, falls der Spieler die n -te Runde gewinnt, ergibt sich $(M_n)_{n \geq 0}$ zu

$$M_n = (M_{n-1} + 1) \mathbf{1}_{\{X_n=1\}}.$$

Zurückkehrend zur allgemeinen Situation variabler p_i ist intuitiv klar, dass M nur dann transient ist, d.h. nur endlich oft in den Zustand 0 zurückkehrt (M irreduzibel), wenn die p_i für $i \rightarrow \infty$ hinreichend schnell gegen 1 streben, was bedeutet, dass mit zunehmender Länge einer Glückssträhne eine gegen 1 wachsende Wahrscheinlichkeit besteht, auch die nächste Runde zu gewinnen. Hier ist die formale Begründung: Es gilt offenbar für jedes $n \geq 1$

$$\begin{aligned} f_{00}^{(n)} &= \mathbb{P}_0(M_1 = 1, \dots, M_{n-1} = n-1, M_n = 0) \\ &= p_0 p_1 \cdot \dots \cdot p_{n-2} q_{n-1} \\ &= \left(\prod_{i=0}^{n-2} p_i \right) - \left(\prod_{i=0}^{n-1} p_i \right) \end{aligned} \tag{2.58}$$

und folglich

$$\mathbb{P}_0(\tau(0) \leq n) = \sum_{k=1}^n f_{00}^{(k)} = 1 - \left(\prod_{i=0}^{n-1} p_i \right).$$

0 ist genau dann rekurrent, wenn $f_{00}^* = \mathbb{P}_0(\tau(0) < \infty) = 1$, also, wenn

$$\lim_{n \rightarrow \infty} \prod_{i=0}^{n-1} p_i = 0.$$

Durch Logarithmieren und Benutzung von $\log(1-x) \simeq -x$ für $x \rightarrow 1$ erweist sich dies wiederum als äquivalent zu

$$\sum_{i \geq 1} (1 - p_i) = \sum_{i \geq 1} q_i = \infty,$$

was insbesondere die erwartete Eigenschaft $p_i \rightarrow 1$ für $i \rightarrow \infty$ im transienten Fall bestätigt.

Wenden wir uns schließlich dem Gleichungssystem für stationäre Maße zu, das hier die Form

$$\pi_0 = \sum_{i \geq 0} q_i \pi_i \quad \text{und} \quad \pi_j = p_{j-1} \pi_{j-1} \quad \text{für } j \geq 1$$

annimmt. Ignoriert man zunächst die Invarianzgleichung für π_0 , so erhält man als eindeutige Lösung der übrigen leicht

$$\pi_j = \pi_0 \prod_{i=0}^{j-1} p_i.$$

Wenn ein stationäres Maß π existiert, wobei wir $\pi_0 = 1$ wählen dürfen, so folgt aus der verbliebenen Gleichung unter Hinweis auf (2.58)

$$1 = \sum_{n \geq 0} q_n \prod_{i=0}^{n-1} p_i = \sum_{n \geq 0} f_{00}^{(n+1)} = f_{00}^*,$$

d.h. die Rekurrenz des Zustands 0. Im transienten Fall existiert also kein stationäres Maß für die Kette M .

2.6 Kopplung und gleichmäßige Verteilungskonvergenz (Ergodizität) diskreter Markov-Ketten

Nachdem wir im vorherigen Abschnitt die Frage der Konvergenz im Zeitmittel ausgiebig diskutiert haben, kommen wir nun zu der noch spannenderen Frage, wann eine positiv rekurrente DMK M sogar in Totalvariation konvergiert, d.h., wann

$$\lim_{n \rightarrow \infty} \|\mathbb{P}_\lambda^{M_n} - \pi\| = \lim_{n \rightarrow \infty} \sup_{A \subset \mathcal{S}} |\mathbb{P}_\lambda(M_n \in A) - \pi(A)| = 0 \quad (2.59)$$

für alle $\lambda \in \mathcal{P}(\mathcal{S})$ gilt, wobei π die eindeutig bestimmte stationäre Verteilung von M bezeichnet. Die folgende einfache Überlegung zeigt, dass dies nur im aperiodischen Fall möglich ist. Hat M nämlich die Periode $d \geq 2$, so folgt unter Hinweis auf $\pi_i > 0$ für alle $i \in \mathcal{S}$ (Satz 2.33)

$$\lim_{n \rightarrow \infty} \mathbb{P}_i(M_{nd+r} = i) = \lim_{n \rightarrow \infty} p_{ii}^{(nd+r)} = 0 \neq \pi_i$$

für jedes $0 < r < d$. Wir können uns somit bei der Untersuchung der Gültigkeit von (2.59) auf die *aperiodischen, positiv rekurrenten DMK* beschränken.

2.6.1 Die Kopplungsmethode

Die Methode, mit der wir das Konvergenzresultat beweisen werden, basiert auf einer wunderbaren Idee von WOLFGANG DOEBLIN (ein Sohn des bekannten Schriftstellers Alfred Döblin), publiziert im Jahre 1938 in einer Arbeit mit dem Titel “*Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états*”. Doeblins früher Tod¹ und die schwere Zugänglichkeit der Zeitschrift, in der die genannte Arbeit erschien, waren vermutlich die Ursache, dass seine Idee der *Kopplung* von Markov-Ketten mehr als 30 Jahre unbeachtet blieb und erst in den siebziger Jahren durch Arbeiten u.a. von PITMAN [49], GRIFFEATH [27, 28] und LINDVALL [41] eine Renaissance erfuhr, diesmal allerdings mit nachhaltiger Wirkung bis zum heutigen Tag. Die Kopplungsmethode als ureigenes Instrument der W-Theorie hat sich nämlich mittlerweile weit über die Theorie der Markov-Ketten hinaus als äußerst wirkungsvolles und elegantes Instrument erwiesen, Grenzwertsätze für stochastische Prozesse zu beweisen, die vorher mit anderen, meist analytischen Methoden weitaus schwerer, wenn überhaupt erzielt werden konnten. Die Monographien von LINDVALL [43] und THORISSON [59] geben einen Einblick in diese Entwicklungen. Das Faszinierende an der Methode ist ihr der Anschauung im Nu zugängliches Wesen, das selbst in äußerst komplexen Modellen immer erkennbar bleibt. Im folgenden stellen wir kurz die für unsere Zwecke notwendigen Grundlagen bereit:

Q und Q' seien zwei W-Maße auf einem messbaren Raum (E, \mathfrak{E}) . Ein Paar (X, X') von Zufallsvariablen auf demselben W-Raum $(\Omega, \mathfrak{A}, \mathbb{P})$ mit Werten in (E, \mathfrak{E}) heißt *Kopplung* von (Q, Q') , wenn


$$\mathbb{P}^X = Q \quad \text{und} \quad \mathbb{P}^{X'} = Q'.$$

Die Nützlichkeit der Kopplung zum Vergleich von W-Maßen manifestiert sich in der sogenannten *Kopplungsungleichung*

$$\|Q - Q'\| \leq \mathbb{P}(X \neq X'), \quad (2.60)$$

die sich sofort aus der Abschätzung

$$\begin{aligned} \|Q - Q'\| &= \sup_{A \in \mathfrak{E}} |\mathbb{P}(X \in A) - \mathbb{P}(X' \in A)| \\ &= \sup_{A \in \mathfrak{E}} |\mathbb{P}(X \in A, X = X') + \mathbb{P}(X \in A, X \neq X') \\ &\quad - \mathbb{P}(X' \in A, X = X') - \mathbb{P}(X' \in A, X \neq X')| \\ &= \sup_{A \in \mathfrak{E}} |\mathbb{P}(X \in A, X \neq X') - \mathbb{P}(X' \in A, X \neq X')| \\ &\leq \mathbb{P}(X \neq X') \end{aligned}$$

¹ Er beging am 26. Juni 1940, vier Tage vor der Kapitulation Frankreichs, im Alter von 25 Jahren Selbstmord, nachdem er seine von Deutschen umzingelte frz. Truppeneinheit verlassen hatte und nur noch die Alternativen sah, zu sterben oder sich den Deutschen auszuliefern;  LINDVALL [42] für eine ausführlichere Biographie Wolfgang Doeblins.

ergibt. Der Variationsabstand zwischen Q und Q' lässt sich also durch die wesentlich handlichere Wahrscheinlichkeit $\mathbb{P}(X \neq X')$ abschätzen und ist dementsprechend klein, wenn die gekoppelten Variablen X und X' mit nur kleiner Wahrscheinlichkeit verschieden sind. Das Problem besteht nun natürlich darin, solche möglichst stark gekoppelten Variablen zu konstruieren, was allerdings vom Einzelfall abhängig ist.

Nach dieser sehr allgemeinen Kurzeinführung richten wir unseren Blick auf die folgende, für unsere Zwecke relevante Situation: Gegeben einen messbaren Raum $(\mathcal{S}, \mathfrak{S})$, seien Q und Q' W-Maße auf $(\mathcal{S}^\infty, \mathfrak{S}^\infty)$, üblicherweise Verteilungen irgendwelcher Folgen $Y = (Y_n)_{n \geq 0}$ und $Y' = (Y'_n)_{n \geq 0}$ von Zufallsvariablen mit Wertebereich $(\mathcal{S}, \mathfrak{S})$. Im weiteren Verlauf werden Q und Q' die Verteilungen derselben DMK unter verschiedenen Anfangsverteilungen sein. Die eindimensionalen Randverteilungen (Verteilungen der Y_n bzw. Y'_n) bezeichnen wir mit Q_n bzw. Q'_n , $n \geq 0$. Nehmen wir an, unser Ziel ist der Vergleich von Q_n und Q'_n für $n \rightarrow \infty$, etwa der Nachweis von $\|Q_n - Q'_n\| \rightarrow 0$. In diesem Fall wird man i.A. nicht für jedes (Q_n, Q'_n) eine Kopplung konstruieren, sondern vielmehr für (Q, Q') , und zwar in folgender Weise: Seien $X = (X_n)_{n \geq 0}$ und $X' = (X'_n)_{n \geq 0}$ Folgen von Zufallsvariablen auf demselben W-Raum $(\Omega, \mathfrak{A}, \mathbb{P})$ (dies muss für Y und Y' keineswegs gelten) derart, dass (X, X') eine Kopplung von (Q, Q') bildet. Dann heißt

$$T := \inf\{n \geq 0 : X_k = X'_k \text{ für alle } k \geq n\} \quad (2.61)$$

die zu (X, X') gehörende *Kopplungszeit*, und es gilt unter Verwendung von (2.60) die ebenfalls *Kopplungsungleichung* genannte Abschätzung

$$\|Q_n - Q'_n\| \leq \mathbb{P}(X_n \neq X'_n) \leq \mathbb{P}(T > n). \quad (2.62)$$

Aus $\mathbb{P}(T < \infty) = 1$ folgt offenbar $\|Q_n - Q'_n\| \rightarrow 0$, so dass das Problem nun darin besteht, die Prozesse X und X' so zu konstruieren, dass sie sich f.s. irgendwann treffen und danach übereinstimmen. Es bleibt erneut offen, wie dies bewerkstelligt werden kann, aber wir werden bald sehen, dass positiv rekurrente DMK hierfür ideal geeignete Objekte bilden.

2.6.2 Der Ergodensatz für aperiodische, positiv rekurrente DMK

Der nachfolgende Satz, der in der Literatur oft *Ergodensatz für positiv rekurrente DMK* oder einfach *Ergodensatz für DMK* genannt wird, darf als Perle der Theorie diskreter MK angesehen werden und beinhaltet das eingangs angekündigte Konvergenzresultat für aperiodische, positiv rekurrente DMK:

Satz 2.52. (Ergodensatz für DMK) Sei $M = (M_n)_{n \geq 0}$ eine aperiodische, positiv rekurrente DMK mit Übergangsmatrix $P = (p_{ij})_{i,j \in \mathcal{S}}$ und stationärer Verteilung $\pi = (\mu_i^{-1})_{i \in \mathcal{S}}$. Dann gilt (2.59), also $\lim_{n \rightarrow \infty} \|\mathbb{P}_\lambda^{M_n} - \pi\| = 0$ für

jede Anfangsverteilung λ sowie insbesondere

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j = \frac{1}{\mu_{jj}}$$

für alle $i, j \in \mathcal{S}$ (gleichmäßig in j).

Beweis. Sei $M \otimes M' := (M_n, M'_n)_{n \geq 0}$ eine DMK mit Zustandsraum \mathcal{S}^2 , kanonischer Filtration $(\mathcal{F}_n)_{n \geq 0}$ und Übergangswahrscheinlichkeiten

$$P_{(i_1, i_2), (j_1, j_2)} = P_{i_1 j_1} P_{i_2 j_2},$$

für die wir ein Standardmodell

$$(\Omega, \mathfrak{A}, M \otimes M', (\mathbb{P}_v)_{v \in \mathcal{P}(\mathcal{S}^2)})$$

zugrundelegen. Im Fall $v = \lambda \otimes \mu$ schreiben wir $\mathbb{P}_{\lambda, \mu}$ für $\mathbb{P}_{\lambda \otimes \mu}$, im Fall $v = \delta_{(i, j)} = \delta_i \otimes \delta_j$ entsprechend $\mathbb{P}_{i, j}$ für $\mathbb{P}_{\delta_{(i, j)}}$. Elementare Rechnungen zeigen, dass $M = (M_n)_{n \geq 0}$ und $M' = (M'_n)_{n \geq 0}$ jeweils DMK bezüglich $(\mathcal{F}_n)_{n \geq 0}$ mit Übergangsmatrix P bilden, die unter jedem \mathbb{P}_v mit $v = \lambda \otimes \mu$ stochastisch unabhängig sind. Da unter $\mathbb{P}_{\lambda, \mu}$ Wahrscheinlichkeiten von Ereignissen, die nur die Kette M bzw. M' betreffen, nicht von μ bzw. λ abhängen, schreiben wir in einem solchen Fall zur Kennzeichnung $\mathbb{P}_{\lambda, \bullet}$ bzw. $\mathbb{P}_{\bullet, \mu}$. Schließlich notieren wir noch, dass

$$\mathbb{P}_{\lambda, \bullet}^M = \mathbb{P}_{\lambda, \mu}^M = \mathbb{P}_{\mu, \lambda}^{M'} = \mathbb{P}_{\bullet, \lambda}^{M'} \quad (2.63)$$

für alle $\lambda, \mu \in \mathcal{P}(\mathcal{S})$ gilt.

Da M aperiodisch ist, gilt nach Lemma 2.14 $p_{jj}^{(n)} > 0$ für alle $j \in \mathcal{S}$ und $n \geq n_0(j)$ geeignet. $i \leftrightarrow j$ impliziert ferner $p_{ij}^{(m)} > 0$ für ein $m \geq 1$, so dass auch

$$p_{ij}^{(n)} \geq p_{ij}^{(m)} p_{jj}^{(n-m)} > 0$$

für alle hinreichend großen n . Dies liefert schließlich, dass auch $M \otimes M'$ irreduzibel und aperiodisch ist, denn, gegeben beliebige $i_1, i_2, j_1, j_2 \in \mathcal{S}$, folgt nun

$$p_{(i_1, i_2), (j_1, j_2)}^{(n)} = p_{i_1 j_1}^{(n)} p_{i_2 j_2}^{(n)} > 0$$

für alle hinreichend großen n . $\pi \otimes \pi = (\pi_i \pi_j)_{(i, j) \in \mathcal{S}^2}$ definiert außerdem eine stationäre Verteilung der bivariaten Kette, wie man leicht nachrechnet, so dass aus Satz 2.33(a) und Satz 2.45 die positive Rekurrenz von $M \otimes M'$ sowie die Eindeutigkeit von $\pi \otimes \pi$ folgt.

Nun sind alle Vorbereitungen für die Kopplung getroffen. Sei

$$T = \inf\{n \geq 0 : M_n = M'_n\} = \inf\{n \geq 0 : (M_n, M'_n) \in \{(i, i) : i \in \mathcal{S}\}\}.$$

Aus der Rekurrenz aller (i, i) , $i \in \mathcal{S}$, folgt $\mathbb{P}_v(T < \infty) = 1$ für alle $v \in \mathcal{P}(\mathcal{S}^2)$. Wir definieren den zugehörigen *Kopplungsprozess* $\widehat{M} = (\widehat{M}_n)_{n \geq 0}$ durch

$$\widehat{M}_n = \begin{cases} M'_n, & \text{falls } n \leq T \\ M_n, & \text{falls } n \geq T \end{cases}. \quad (2.64)$$

\widehat{M} folgt demnach dem Pfad der Kette M' , bis diese erstmals gemeinsam mit M

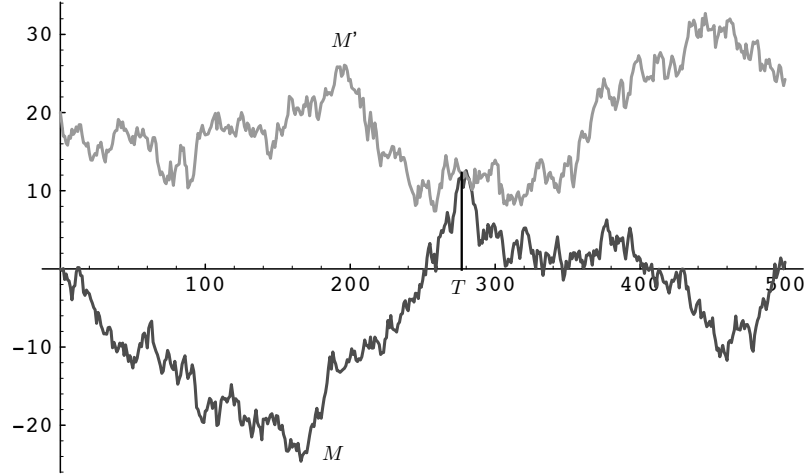


Abb. 2.7 Realisierungen von M und M' mit Kopplungszeit T für (M, M') .

denselben Zustand erreicht, und wechselt dann auf den Pfad von M . Beachte, dass T eine Stopzeit für $M \otimes M'$ bildet. Der Leser sollte sich an dieser Stelle zunächst anschaulich klar machen, dass \widehat{M} und M' unter jedem \mathbb{P}_v dieselbe Verteilung besitzen. Aufgrund der starken Markov-Eigenschaft hängen nämlich die Post- T -Folgen $M^{(T)} = (M_n)_{n \geq T}$ und $M'^{(T)}$ von der Vergangenheit nur über M_T bzw. M'_T ab und stimmen folglich überein. Für die Verteilung der Kette \widehat{M} macht es daher keinen Unterschied, welchem Pfad sie nach T folgt. Hier ist die formale Begründung: Für alle $v \in \mathcal{P}(\mathcal{S}^2)$ und $A_0, A_1, \dots \subset \mathcal{S}$ gilt unter Hinweis auf (2.63)

$$\begin{aligned} \mathbb{P}_v(\widehat{M}_k \in A_k, k \geq 0) &= \sum_{n \geq 0} \mathbb{P}_v(T = n, M'_0 \in A_0, \dots, M'_n \in A_n, M_{n+1} \in A_{n+1}, \dots) \\ &= \sum_{n \geq 0} \sum_{i \in A_n} \int_{\{T=n, M'_0 \in A_0, \dots, M'_{n-1} \in A_{n-1}, M_n = M'_n = i\}} \mathbb{P}(M_{n+1} \in A_{n+1}, \dots | \mathcal{F}_n) d\mathbb{P}_v \\ &= \sum_{n \geq 0} \sum_{i \in A_n} \mathbb{P}_v(T = n, M'_0 \in A_0, \dots, M'_{n-1} \in A_{n-1}, M_n = M'_n = i) \mathbb{P}_{i, \bullet}(M_1 \in A_{n+1}, \dots) \\ &= \sum_{n \geq 0} \sum_{i \in A_n} \mathbb{P}_v(T = n, M'_0 \in A_0, \dots, M'_{n-1} \in A_{n-1}, M'_n = i) \mathbb{P}_{\bullet, i}(M'_1 \in A_{n+1}, \dots) \end{aligned}$$

$$\begin{aligned}
&= \sum_{n \geq 0} \sum_{i \in A_n} \int_{\{T=n, M'_0 \in A_0, \dots, M'_{n-1} \in A_{n-1}, M'_n=i\}} \mathbb{P}(M'_{n+1} \in A_{n+1}, \dots | \mathcal{F}_n) d\mathbb{P}_v \\
&= \sum_{n \geq 0} \mathbb{P}_v(T = n, M'_0 \in A_0, \dots, M'_n \in A_n, M'_{n+1} \in A_{n+1}, \dots) \\
&= \mathbb{P}_v(M'_k \in A_k, k \geq 0).
\end{aligned}$$

RESNICK [51, S. 130]) gibt folgende amüsante Veranschaulichung der Kopplungsidee:

Stellen wir uns vor, zwei Frösche, Sam und Suzie, hüpfen von Stein zu Stein, wobei M die Sprünge von Sam und M' zunächst die Sprünge von Suzie beschreibt. Allerdings gibt es einen Haken. Wenn beide auf demselben Stein landen, so hüpfet Suzie auf Sams Rücken und von diesem Zeitpunkt an (der Kopplungszeit) mit ihm gemeinsam von Stein zu Stein. Da jedoch beide gemäß derselben Übergangsmatrix springen, ändert es nichts an der Verteilung von Suzies Wanderung, ob sie sich nun auf Sams Rücken weiter bewegt oder davon unabhängig gemäß der Kette M' mit gelegentlichen Treffen auf demselben Stein.

Gegeben das Paar (M, \widehat{M}) mit den Eigenschaften

1. $M_n = \widehat{M}_n$ für alle $n \geq T$, T die Kopplungszeit,
2. $\widehat{M} \sim M'$ unter jedem $\mathbb{P}_{\lambda, \mu}$ (sogar unter jedem \mathbb{P}_v),

wählen wir nun $\mu = \pi$. Dann ist \widehat{M} unter jedem $\mathbb{P}_{\lambda, \pi}$ stationär, insbesondere $\mathbb{P}_{\lambda, \pi}^{\widehat{M}_n} = \pi$ für alle $n \geq 0$, und es folgt vermöge der Kopplungsungleichung (2.62)

$$\|\mathbb{P}_{\lambda, \bullet}^{M_n} - \pi\| = \|\mathbb{P}_{\lambda, \pi}^{M_n} - \mathbb{P}_{\lambda, \pi}^{\widehat{M}_n}\| \leq \mathbb{P}_{\lambda, \pi}(T > n), \quad (2.65)$$

wegen $\mathbb{P}_{\lambda, \pi}(T < \infty) = 1$ also (2.59).

Da die Post- n -Prozesse $M^{(n)}$ und $\widehat{M}^{(n)}$ für alle $n \geq T$ übereinstimmen und $\mathbb{P}_{\pi}^{M^{(n)}} = \mathbb{P}_{\pi}^M$ für alle $n \geq 0$ gilt, ergibt sich ohne Zusatzargumente die folgende Verschärfung von (2.59):

Korollar 2.53. *In der Situation von Satz 2.52 gilt ferner*

$$\lim_{n \rightarrow \infty} \|\mathbb{P}_{\lambda}^{M^{(n)}} - \mathbb{P}_{\pi}^M\| = 0$$

für jede Anfangsverteilung λ .

Beweis. Es genügt der Hinweis, dass anstelle von (2.65) auch

$$\|\mathbb{P}_{\lambda, \bullet}^{M^{(n)}} - \mathbb{P}_{\lambda, \pi}^{\widehat{M}}\| = \|\mathbb{P}_{\lambda, \pi}^{M^{(n)}} - \mathbb{P}_{\lambda, \pi}^{\widehat{M}^{(n)}}\| \leq \mathbb{P}_{\lambda, \pi}(T > n) \quad (2.66)$$

für alle $n \geq 0$ gilt, denn $\mathbb{P}_{\pi, \bullet}^M = \mathbb{P}_{\lambda, \pi}^{M'} = \mathbb{P}_{\lambda, \pi}^{\widehat{M}}$.

Als weitere zu (2.59) bzw. (2.66) äquivalente Aussagen hinsichtlich sogenannter *Funktionale* $\mathbb{E}_\lambda f(M_n)$ bzw. $\mathbb{E}_\lambda f(M^{(n)}) = \mathbb{E}_\lambda f(M_n, M_{n+1}, \dots)$ der Kette können wir festhalten (vgl. Anmerkung 2.35):

Korollar 2.54. *In der Situation von Satz 2.52 gelten ferner*

$$\lim_{n \rightarrow \infty} \sup_{f \in b\mathfrak{G}, \|f\|_\infty \leq 1} \left| \mathbb{E}_\lambda f(M_n) - \int_{\mathcal{S}} f(s) \pi(ds) \right| = 0,$$

$$\lim_{n \rightarrow \infty} \sup_{f \in b\mathfrak{G}^\infty, \|f\|_\infty \leq 1} \left| \mathbb{E}_\lambda f(M^{(n)}) - \mathbb{E}_\pi f(M) \right| = 0$$

für jede Anfangsverteilung λ .

Die nachfolgende Erweiterung von Satz 2.52, die zusätzlich eine Klasse transienter Zustände zulässt, von der aus die Kette f.s. irgendwann in einen rekurrenten Zustand eintritt, bildet das Pendant zu Satz 2.44, der entsprechenden Verallgemeinerung von Satz 2.33. Den einfachen Beweis überlassen wir dem Leser.

Satz 2.55. *Sei $M = (M_n)_{n \geq 0}$ eine DMK mit Zustandsraum \mathcal{S} , der in eine Menge (nicht notwendig Klasse) \mathcal{T} transienter und eine Klasse \mathcal{R} aperiodischer, positiv rekurrenter Zustände zerfällt, wobei ferner $\mathbb{P}_i(\tau(\mathcal{R}) < \infty) = 1$ für alle $i \in \mathcal{T}$. Dann gelten weiterhin die Aussagen in Satz 2.52, Korollar 2.53 und Korollar 2.54.*

Zum Abschluss dieses Teilabschnitts noch ein wenig Nomenklatur: Eine aperiodische, positiv rekurrente DMK nennt man aufgrund der soeben gezeigten Resultate auch kurz *ergodisch*, wobei dasselbe Attribut für Zustände verwendet wird. Die gleichmäßige Konvergenz der $\mathbb{P}_\lambda^{M_n}$ gegen die stationäre Verteilung bezeichnet man als *starke Ergodizität*, dieselbe im Zeitmittel dagegen als *schwache Ergodizität*. Damit können wir zusammenfassend festhalten, dass positiv rekurrente DMK stets schwach ergodisch sind (Satz 2.33) und ergodische DMK sogar stark ergodisch (Satz 2.52).

2.6.3 Asymptotisches Verhalten im periodischen Fall

Nach den zuvor erzielten Ergebnissen bedarf es keiner weiteren tiefen Einsichten, um das Langzeitverhalten positiv rekurrenter d -periodischer DMK zu beschreiben. Man kann es sich wie eine Schwingung mit Frequenz d vorstellen, wobei das Verhalten zu Zeitpunkten derselben Phase dem einer ergodischen DMK entspricht. Dies gilt es im Folgenden zu präzisieren, und zwar mittels sogenannter *zyklischer Klassen*.

Satz 2.56. Gegeben eine d -periodische DMK $M = (M_n)_{n \geq 0}$, existiert eine Zerlegung $\mathcal{S} = \sum_{r=0}^{d-1} \mathcal{S}_r$ des Zustandsraums in disjunkte, nichtleere Teilmengen $\mathcal{S}_0, \dots, \mathcal{S}_{d-1}$, eindeutig bis auf zyklische Vertauschung, so dass

$$\mathbb{P}_i(M_1 \in \mathcal{S}_{r+1}) = 1 \quad \text{für alle } i \in \mathcal{S}_r \text{ und } 0 \leq r < d,$$

wobei $\mathcal{S}_d = \mathcal{S}_0$. Die \mathcal{S}_r heißen **zyklische Klassen**, weil sie von M in zyklischer Weise durchlaufen werden ($\dots \rightarrow \mathcal{S}_0 \rightarrow \mathcal{S}_1 \rightarrow \dots \rightarrow \mathcal{S}_{d-1} \rightarrow \mathcal{S}_d = \mathcal{S}_0 \rightarrow \mathcal{S}_1 \rightarrow \dots$).

Beweis. Wähle irgendein $i_0 \in \mathcal{S}$ und setze

$$\mathcal{S}_r := \left\{ j \in \mathcal{S} : p_{i_0 j}^{(nd+r)} > 0 \text{ für ein } n \in \mathbb{N}_0 \right\}$$

sowie $\mathcal{S}_{nd+r} = \mathcal{S}_r$ für alle $0 \leq r < d$ und $n \geq 1$. $\mathcal{S}_r \neq \emptyset$ für alle r ist ebenso wie $\mathcal{S}_0 \cup \dots \cup \mathcal{S}_{d-1} = \mathcal{S}$ offensichtlich, letzteres wegen der Irreduzibilität von M . Sei nun $j \in \mathcal{S}_q \cap \mathcal{S}_r$ angenommen und o.E. $q > r$. Dann existieren $m, n \geq 0$, so dass $p_{i_0 j}^{(md+q)} > 0$ und $p_{i_0 j}^{(nd+r)} > 0$. Wähle ein $k \geq 1$ mit $p_{j i_0}^{(k)} > 0$. Es folgt $p_{i_0 i_0}^{(md+q+k)} \geq p_{i_0 j}^{(md+q)} p_{j i_0}^{(k)} > 0$ und analog $p_{i_0 i_0}^{(nd+r+k)} > 0$. $md+q+k$ und $nd+r+k$ müssen folglich beide Vielfache von d sein, was $q-r \in d\mathbb{N}_0$ und dann $\mathcal{S}_q = \mathcal{S}_r$ impliziert. $\mathcal{S}_0, \dots, \mathcal{S}_{d-1}$ sind also paarweise disjunkt. Zum Nachweis von $\mathbb{P}_i(M_1 \in \mathcal{S}_{r+1}) = 1$ für $i \in \mathcal{S}_r$ notieren wir zuerst, dass aus $i \in \mathcal{S}_r$ offensichtlich $\{j \in \mathcal{S} : p_{ij} > 0\} \subset \mathcal{S}_{r+1}$ folgt. Wir erhalten deshalb wegen

$$1 = \sum_{j: p_{ij} > 0} p_{ij} \leq \sum_{j \in \mathcal{S}_{r+1}} p_{ij} = \mathbb{P}_i(M_1 \in \mathcal{S}_{r+1})$$

das Gewünschte. Wir notieren, dass natürlich analog $\mathbb{P}_i(M_n \in \mathcal{S}_{r+n}) = 1$ für alle $0 \leq r < d$ und $n \geq 2$ folgt.

Gegeben irgendeine weitere zyklische Zerlegung $\mathcal{S}'_0, \dots, \mathcal{S}'_{d-1}$ des Zustandsraums, wobei wiederum $\mathcal{S}'_{nd+r} = \mathcal{S}'_r$ für $n \geq 1$ und $0 \leq r < d$, sei $0 \leq q < d$ derart gewählt, dass $i_0 \in \mathcal{S}_0 \cap \mathcal{S}'_q$. Für jedes $j \in \mathcal{S}_r$, $0 < r < d$ existiert dann ein $n \geq 1$, so dass $p_{i_0 j}^{(nd+r)} > 0$. Damit folgt aber $j \in \mathcal{S}'_{q+r}$, wegen $q+r \not\equiv q \pmod{d}$ also $j \notin \mathcal{S}'_q$, was $\mathcal{S}_r \cap \mathcal{S}'_q = \emptyset$ für alle $0 < r < d$ zeigt, d.h. $\mathcal{S}_0 = \mathcal{S}'_q$. Letzteres ist aber gleichbedeutend mit $\mathcal{S}_r = \mathcal{S}'_{q+r}$ für alle $0 \leq r < d$. \square

Das Verhalten periodischer DMK beschreibt nun der folgende

Satz 2.57. Gegeben eine d -periodische ($d \geq 2$), positiv rekurrente DMK M mit stationärer Verteilung π und zyklischer Zerlegung $\mathcal{S}_0, \dots, \mathcal{S}_{d-1}$ des Zustandsraums gilt: Das **d -Skelett** $(M_{nd})_{n \geq 0}$ bildet auf jeder zyklischen Klasse \mathcal{S}_r , $0 \leq r < d$, eine ergodische DMK mit stationärer Verteilung $\pi^{(r)} :=$

$d\pi(\cdot \cap \mathcal{S}_r)$. Insbesondere folgt

$$\mathbb{P}_\pi(M_n \in \mathcal{S}_r) = \frac{1}{d}$$

für alle $0 \leq r < d$ und $n \geq 0$.

Beweis. Die Behauptungen ergeben sich leicht aus den vorherigen Ergebnissen. Die Details mag der Leser selbst ausführen. \square

Unter der stationären Verteilung besitzt somit jede zyklische Klasse dieselbe Wahrscheinlichkeit $1/d$, und M verhält sich bei Einschränkung der Zeitachse auf Vielfache von d , d.h. bei Betrachtung des d -Skeletts $(M_{nd})_{n \geq 0}$, sowie einer auf nur eine zyklische Klasse konzentrierten Anfangsverteilung wie eine ergodische DMK, auf die die Ergebnisse des vorherigen Teilabschnitts hinsichtlich ihres Langzeitverhaltens angewendet werden können.

2.6.4 Nochmals der null-rekurrenente Fall

Für eine null-rekurrenente DMK M mit (bis auf skalares Vielfaches eindeutigem) stationären Maß π ist der Césaro-Limes $C\text{-}\lim_{n \rightarrow \infty} \mathbb{P}_\lambda(M_n \in A)$ nach Satz 2.33 gleich 0 für jedes $A \in \mathfrak{S}_\pi$ und jede Anfangsverteilung λ . Wir wollen nun zeigen, dass dasselbe auch für den gewöhnlichen Limes gilt.

Satz 2.58. Gegeben eine null-rekurrenente DMK M mit stationärem Maß π , gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}_\lambda(M_n \in A) = 0 \quad (2.67)$$

für alle $A \in \mathfrak{S}_\pi$ und Anfangsverteilungen λ .

Beweis. Angenommen, es gibt ein π -endliches $A \subset \mathcal{S}$, ein $\lambda \in \mathcal{P}(\mathcal{S})$ und eine Teilfolge $(n(k))_{k \geq 1}$, so dass

$$\lim_{k \rightarrow \infty} \mathbb{P}_\lambda(M_{n(k)} \in A) = \lim_{k \rightarrow \infty} \sum_{i \in \mathcal{S}} \sum_{j \in A} \lambda_i p_{ij}^{(n(k))} > 0.$$

Ein einfaches Kompaktheitsargument (analog zu dem im Auswahlssatz von Helly) zeigt die Existenz einer weiteren Teilfolge $(n'(k))_{k \geq 1}$ von $(n(k))_{k \geq 1}$, so dass $p_{ij}^{(n'(k))}$ für alle $i, j \in \mathcal{S}$ konvergiert, kurz

$$\lim_{k \rightarrow \infty} P^{n'(k)} = Q$$

für eine Matrix $Q = (q_{ij})_{i,j \in \mathcal{S}}$. Aufgrund majorisierter Konvergenz folgt dann weiter

$$0 < c = \lim_{k \rightarrow \infty} \sum_{i \in \mathcal{S}} \sum_{j \in A} \lambda_i p_{ij}^{(n'(k))} = \sum_{i \in \mathcal{S}} \sum_{j \in A} \lambda_i q_{ij}$$

und damit $q_{i_0 j_0} > 0$ für mindestes ein Paar $(i_0, j_0) \in \mathcal{S} \times A$, d.h. $Q \neq 0$. Beachte ferner, dass sich unter Verwendung des Fatouschen Lemmas

$$1 = \lim_{k \rightarrow \infty} \sum_{j \in \mathcal{S}} p_{ij}^{(n'(k))} \geq \sum_{j \in \mathcal{S}} q_{ij} \quad (2.68)$$

für alle $i \in \mathcal{S}$ ergibt. Alle Zeilensummen in Q sind demnach ≤ 1 (substochastische Matrix).

Betrachte nun die bivariate Kette $M \otimes M'$ aus dem Beweis von Satz 2.52, die bekanntlich wiederum irreduzibel ist. Aus $p_{i_0 j_0}^{(n'(k))} \rightarrow q_{i_0 j_0} > 0$ folgt

$$\sum_{n \geq 1} p_{(i_0, i_0), (j_0, j_0)}^{(n)} = \sum_{n \geq 1} (p_{i_0 j_0}^{(n)})^2 = \infty,$$

gemäß Satz 2.22 also die Rekurrenz von (j_0, j_0) für $M \otimes M'$ und damit aus Solidarität die Rekurrenz der Kette selbst. M und M' lassen sich folglich in endlicher Zeit T erfolgreich koppeln, und es ergibt sich mittels der Kopplungsungleichung (\widehat{M} der Kopplungsprozess)

$$\begin{aligned} |p_{i_1 j}^{(n)} - p_{i_2 j}^{(n)}| &= |\mathbb{P}_{i_1, i_2}(M_n = j) - \mathbb{P}_{i_1, i_2}(\widehat{M}_n = j)| \\ &\leq \mathbb{P}_{i_1, i_2}(T > n) \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

für alle $i_1, i_2, j \in \mathcal{S}$. $q_{ij} = q_{\bullet j}$ hängt also gar nicht von i ab, was insbesondere bedeutet, dass jede Spalte von Q konstante Komponenten besitzt. Multipliziert man nun Q von links mit der stochastischen Matrix P (alle Zeilensummen = 1), folgt $PQ = Q$ und daraus weiter mit dem Satz von der majorisierten Konvergenz

$$P^{n'(k)+1} = PP^{n'(k)} \xrightarrow{k \rightarrow \infty} PQ = Q.$$

Schließlich liefert das Fatousche Lemma für alle $j \in \mathcal{S}$

$$q_{\bullet j} = \lim_{k \rightarrow \infty} p_{ij}^{(n'(k)+1)} = \lim_{k \rightarrow \infty} \sum_{l \in \mathcal{S}} p_{il}^{(n'(k))} p_{lj} \geq \sum_{l \in \mathcal{S}} q_{\bullet l} p_{lj},$$

so dass aus

$$0 \leq \sum_{j \in \mathcal{S}} \left(q_{\bullet j} - \sum_{l \in \mathcal{S}} q_{\bullet l} p_{lj} \right) = \sum_{j \in \mathcal{S}} q_{\bullet j} - \sum_{l \in \mathcal{S}} q_{\bullet l} \underbrace{\sum_{j \in \mathcal{S}} p_{lj}}_{=1} = 0$$

$q_{\bullet j} = \sum_{l \in \mathcal{S}} q_{\bullet l} p_{lj}$, d.h. die Invarianz des gemäß (2.68) endlichen Maßes $(q_{\bullet j})_{j \in \mathcal{S}}$ für M folgt. Da M andererseits als null-rekurrent vorausgesetzt wurde, besitzt M gemäß Satz 2.33 nur stationäre Maße unendlicher Gesamtmasse, die sich zudem nur durch ein skalares Vielfaches unterscheiden. Wir haben demnach einen Widerspruch produziert. \square

Als unmittelbare Konsequenz des vorherigen Satzes notieren wir (ohne Beweis) die folgende Verschärfung von Korollar 2.39.

Korollar 2.59. *Ein Zustand $i \in \mathcal{S}$ ist genau dann null-rekurrent, wenn*

$$\sum_{n \geq 0} p_{ii}^{(n)} = \infty \quad \text{und} \quad \lim_{n \rightarrow \infty} p_{ii}^{(n)} = 0$$

gilt. In diesem Fall folgt sogar $\lim_{n \rightarrow \infty} p_{ji}^{(n)} = 0$ für alle $j \in \mathcal{S}$.

Eine Kombination der Sätze 2.22, 2.52, 2.57 und 2.58 erlaubt uns nun auch eine Antwort auf die Frage nach dem asymptotischen Verhalten der n -Schritt-Übergangswahrscheinlichkeiten $p_{ij}^{(n)}$ einer beliebigen DMK.

Satz 2.60. *Gegeben eine DMK $M = (M_n)_{n \geq 0}$ mit Zustandsraum \mathcal{S} , gilt*

$$\lim_{n \rightarrow \infty} p_{ij}^{(nd(j)+r)} = \frac{d(j) \mathbb{P}_i(\tau(j) \in d(j)\mathbb{N}_0 + r)}{\mu_{jj}}$$

für alle $i, j \in \mathcal{S}$ und $0 \leq r < d(j)$.

Beweis. Zunächst notieren wir, dass

$$p_{ij}^{(nd(j)+r)} = \sum_{k=0}^{nd(j)+r} f_{ij}^{(k)} p_{jj}^{(nd(j)+r-k)} = \sum_{k=0}^n f_{ij}^{(kd(j)+r)} p_{jj}^{((n-k)d(j)}$$

wegen $p_{jj}^{(k)} = 0$ für $k \notin d(j)\mathbb{N}_0$ gilt. Nun gilt aber nach den oben genannten Resultaten (Satz 2.22, falls j transient, Satz 2.58, falls j null-rekurrent, Satz 2.52, falls j ergodisch, und Satz 2.57, falls j positiv rekurrent und periodisch ist) stets $\lim_{n \rightarrow \infty} p_{jj}^{((n-k)d(j))} = d(j)/\mu_{jj}$ und folglich aufgrund majorisierter Konvergenz

$$\lim_{n \rightarrow \infty} p_{ij}^{(nd(j)+r)} = \sum_{k \geq 0} f_{ij}^{(kd(j)+r)} \lim_{n \rightarrow \infty} p_{jj}^{((n-k)d(j))} = \frac{d(j)}{\mu_{jj}} \sum_{k \geq 0} f_{ij}^{(kd(j)+r)},$$

was wegen $\sum_{k \geq 0} f_{ij}^{(kd(j)+r)} = \mathbb{P}_i(\tau(j) \in d(j)\mathbb{N}_0 + r)$ den Beweis abschließt. \square

2.6.5 Gleichmäßige und exponentielle Ergodizität

Es wäre eine Sünde, nicht wenigstens kurz auf eine weitere Stärke der Kopplungsmethode einzugehen, die in der Möglichkeit besteht, auf elegante Weise schärfere Aussagen über die Konvergenz in (2.59) zu gewinnen, und zwar vermöge einer genaueren Abschätzung der Überlebensfunktion $\mathbb{P}_{\lambda,\pi}(T > n)$ der Kopplungszeit T in der entscheidenden Ungleichung (2.65). Natürlich bedarf es dazu geeigneter Zusatzvoraussetzungen an die betreffende Markov-Kette. Verschärfungen der angedeuteten Art besitzen zwei Stoßrichtungen, die zudem kombinierbar sind:

- (1) Eine Abschätzung von $\mathbb{P}_{\lambda,\pi}(T > n)$ durch eine nicht mehr von der Anfangsverteilung λ abhängigen Schranke mit dem Ziel des Nachweises von

$$\lim_{n \rightarrow \infty} \sup_{\lambda \in \mathcal{P}(\mathcal{S})} \|\mathbb{P}_{\lambda}^{M_n} - \pi\| = 0. \quad (2.69)$$

In diesem Fall heißt M *gleichmäßig ergodisch*.

- (2) Eine Abschätzung von $\mathbb{P}_{\lambda,\pi}(T > n)$ durch $C(\lambda)f(n)$ für alle $\lambda \in \mathcal{P}_0(\mathcal{S}) \subset \mathcal{P}(\mathcal{S})$, $n \geq 0$ und eine geeignete Konstante $C(\lambda) > 0$, wobei $f: \mathbb{N}_0 \rightarrow [0, \infty)$ eine für $n \rightarrow \infty$ gegen 0 konvergente Funktion bezeichnet, typischerweise $f(n) = n^{-\beta}$ oder $f(n) = e^{-\beta n}$ für ein $\beta > 0$. Dies liefert eine Aussage über die Konvergenzrate in (2.59), nämlich

$$\|\mathbb{P}_{\lambda}^{M_n} - \pi\| \leq C(\lambda)f(n) \quad (2.70)$$

für alle $\lambda \in \mathcal{P}_0(\mathcal{S})$ und $n \geq 0$. In aller Regel betrachtet man $\mathcal{P}_0(\mathcal{S}) = \{\delta_i : i \in \mathcal{S}\}$. Im Fall dieser Klasse von Anfangsverteilungen und $f(n) = e^{-\beta n}$ heißt M *exponentiell* oder auch *geometrisch ergodisch*, denn

$$\lim_{n \rightarrow \infty} e^{\gamma n} \|\mathbb{P}_i^{M_n} - \pi\| = 0$$

für alle $i \in \mathcal{S}$ und $\gamma < \beta$.

Eine Kombination von (2.69) mit (2.70) für $f(n) = e^{-\beta n}$ und $\mathcal{P}_0(\mathcal{S}) = \mathcal{P}(\mathcal{S})$ führt zur besonders starken *gleichmäßig exponentiellen Ergodizität*:

$$\sup_{\lambda \in \mathcal{P}(\mathcal{S})} \|\mathbb{P}_{\lambda}^{M_n} - \pi\| \leq C e^{-\beta n} \quad (2.71)$$

für alle $n \geq 0$ und damit

$$\lim_{n \rightarrow \infty} e^{\gamma n} \sup_{\lambda \in \mathcal{P}(\mathcal{S})} \|\mathbb{P}_{\lambda}^{M_n} - \pi\| = 0$$

für alle $\gamma < \beta$. Diese wollen wir im Folgenden unter der sogenannten

Doebelin-Bedingung: $\exists i_0 \in \mathcal{S}, n_0 \geq 1 : \alpha(i_0, n_0) := \inf_{i \in \mathcal{S}} p_{i i_0}^{(n_0)} > 0.$

beweisen, die insbesondere für jede ergodische EMK M erfüllt ist, weil dann offenbar

$$\lim_{n \rightarrow \infty} \min_{i \in \mathcal{S}} p_{ij}^{(n)} = \pi_j > 0$$

für alle $j \in \mathcal{S}$ gilt.

Satz 2.61. *Sei M eine ergodische DMK, die die obige Doeblin-Bedingung für ein $i_0 \in \mathcal{S}$ und $n_0 \geq 1$ erfüllt ($\alpha = \alpha(i_0, n_0)$). Dann ist M gleichmäßig exponentiell ergodisch, und zwar gilt (2.71) mit $C = (1 - \alpha^2)^{-(n_0-1)/n_0}$ und $\beta = -\log(1 - \alpha^2)^{1/n_0}$, d.h.*

$$\sup_{\lambda \in \mathcal{P}(\mathcal{S})} \|\mathbb{P}_\lambda^{M_n} - \pi\| \leq (1 - \alpha^2)^{(n-n_0+1)/n_0} \quad (2.72)$$

für alle $n \geq 0$.

Den Schlüssel zum Beweis des Satz geben wir mit einem Lemma:

Lemma 2.62. *Gegeben eine DMK M , die die obige Doeblin-Bedingung für ein $i_0 \in \mathcal{S}$ und $n_0 \geq 1$ erfüllt, gilt mit $\alpha = \alpha(i_0, n_0)$*

$$\mathbb{P}_\lambda(\tau(i_0) > kn_0) \leq (1 - \alpha)^k \quad (2.73)$$

für alle $k \geq 0$ und $\lambda \in \mathcal{P}(\mathcal{S})$.

Beweis. Es genügt offensichtlich, die Behauptung für jedes $\lambda = \delta_i$, $i \in \mathcal{S}$, zu zeigen. Dann ergibt sich mit Hilfe der Markov-Eigenschaft für alle $k \geq 1$ die rekursive Abschätzung

$$\begin{aligned} \mathbb{P}_i(\tau(i_0) > kn_0) &\leq \mathbb{P}_i(\tau(i_0) > (k-1)n_0, M_{kn_0} \neq i_0) \\ &= \sum_{j \neq i_0} \mathbb{P}_i(\tau(i_0) > (k-1)n_0, M_{(k-1)n_0} = j, M_{kn_0} \neq i_0) \\ &= \sum_{j \neq i_0} \mathbb{P}_i(\tau(i_0) > (k-1)n_0, M_{(k-1)n_0} = j)(1 - p_{ji_0}^{(n_0)}) \\ &\leq (1 - \alpha) \mathbb{P}_i(\tau(i_0) > (k-1)n_0) \end{aligned}$$

und daraus die Behauptung per Induktion über k . \square

Beweis (von Satz 2.61). Wir betrachten wieder das im Beweis von Satz 2.52 eingeführte Kopplungsmodell $(\Omega, \mathfrak{A}, M \otimes M', (\mathbb{P}_v)_{v \in \mathcal{P}(\mathcal{S}^2)})$ und notieren als erstes, dass auch die bivariate Kette $M \otimes M'$ die Doeblin-Bedingung erfüllt, und zwar mit demselben n_0 wie M und (i_0, i_0) anstelle von i_0 . Es gilt nämlich vermöge der Unabhängigkeit von M und M'

$$\inf_{i,j \in \mathcal{S}} P_{(i,j),(i_0,i_0)}^{(n_0)} = \inf_{i,j \in \mathcal{S}} P_{ii_0}^{(n_0)} P_{ji_0}^{(n_0)} \geq \alpha^2. \quad (2.74)$$

Beachten wir nun, dass die Kopplungszeit $T = \inf\{n \geq 0 : M_n = M'_n\}$ durch die Ersteintrittszeit $\tau^{M \otimes M'}(i_0, i_0) = \inf\{n \geq 1 : (M_n, M'_n) = (i_0, i_0)\}$ der bivariaten Kette $M \otimes M'$ in den Zustand (i_0, i_0) beschränkt ist, so folgt aus (2.74) und Lemma 2.62 für alle $\lambda \in \mathcal{P}(\mathcal{S})$ und $n \geq 0$, wobei $n = kn_0 + r$ mit $k \geq 0$ und $0 \leq r < n_0$,

$$\begin{aligned} \mathbb{P}_{\lambda, \pi}(T > n) &\leq \mathbb{P}_{\lambda, \pi}(\tau^{M \otimes M'}(i_0, i_0) > n) \\ &\leq \mathbb{P}_{\lambda, \pi}(\tau^{M \otimes M'}(i_0, i_0) > kn_0) \\ &\leq (1 - \alpha^2)^k = e^{-\beta kn_0} \leq C e^{-\beta n}, \end{aligned}$$

was zusammen mit (2.65) die Behauptung des Satzes beweist. \square

Gilt die Doeblin-Bedingung, bei festem n_0 , für mehr als einen Zustand i_0 , so läßt sich die obere Schranke in (2.72) leicht verbessern. Der Leser beweise als Übung, dass generell gilt:

$$\sup_{\lambda \in \mathcal{P}(\mathcal{S})} \|\mathbb{P}_{\lambda}^{M_n} - \pi\| \leq \left(1 - \sum_{j \in \mathcal{S}} \alpha_j^2\right)^{(n-n_0+1)/n_0} \quad (2.75)$$

für alle $n \geq 0$ und $n_0 \geq 1$, wobei $\alpha_j = \alpha(j, n_0)$. Ist die Doeblin-Bedingung verletzt, d.h. $\alpha(j, n_0) = 0$ für alle j, n_0 , so hat die obere Schranke offenkundig stets den bedeutungslosen Wert 1. Ferner erwähnen wir, dass die erzielten Abschätzungen wiederum auch für die Post- n -Folgen $M^{(n)}$ gültig bleiben, d.h. für $\|\mathbb{P}_{\lambda}^{M^{(n)}} - \mathbb{P}_{\pi}^M\|$ anstelle von $\|\mathbb{P}_{\lambda}^{M_n} - \pi\|$ (vgl. Korollar 2.53). Eine Darstellung weiterer Methoden und Ergebnisse zur Konvergenzgeschwindigkeit einer DMK gegen ihre stationäre Verteilung werden wir in Kapitel 3 geben.

2.7 Absorptionswahrscheinlichkeiten

Die Frage nach dem Zeitpunkt, zu dem eine DMK, ausgehend von einem beliebigen Anfangszustand, einen bestimmten kritischen Zustand erreicht, ist in vielen Anwendungsbeispielen von Interesse. Man denke beispielsweise an den Ausfallszeitpunkt eines technischen Systems, an den Ruinzeitpunkt in einer Spielsituation oder in der Geschäftswelt, oder auch an den Aussterbezeitpunkt einer Population, den es zu untersuchen gilt.

Probleme dieser Art lassen sich oft als *Absorptionsprobleme* formulieren unter Zugrundelegung folgender Situation: Sei $\mathcal{S} = \mathcal{T} + \mathcal{R}$ die nach Satz 2.28 immer existierende eindeutige Zerlegung des Zustandsraums einer DMK $M = (M_n)_{n \geq 0}$ in die Mengen \mathcal{T} und \mathcal{R} ihrer transienten bzw. rekurrenten Zustände, wobei \mathcal{R} möglicherweise weiter in höchstens abzählbar unendlich viele abgeschlossene Rekurrenzklassen \mathcal{R}_{α} zerfällt, d.h. $\mathcal{R} = \sum_{\alpha} \mathcal{R}_{\alpha}$. Sei außerdem

$$\tau = \tau^0(\mathcal{T}^c) = \inf\{n \geq 0 : M_n \notin \mathcal{T}\}$$

der Zeitpunkt, zu dem die DMK M die Menge \mathcal{T} verlässt. Beachte, dass durchaus $\mathbb{P}_i(\tau = \infty) > 0$ gelten kann, etwa, wenn M nur transiente Zustände besitzt. Sobald die Kette die Menge \mathcal{T} verlässt, erreicht sie irgendeinen rekurrenten Zustand und kehrt nie mehr nach \mathcal{T} zurück. Im folgenden setzen wir voraus, dass mindestens ein rekurrenter Zustand existiert ($\mathcal{R} \neq \emptyset$), weil die nachfolgenden Untersuchungen andernfalls keinen Sinn machen.

Bei entsprechender Anordnung der Zustände hat die Übergangsmatrix P die Gestalt

$$P = \begin{pmatrix} Q & R \\ 0 & * \end{pmatrix},$$

wobei Q die Restriktion von P auf die Zustände in \mathcal{T} bezeichnet, d.h.

$$Q := (q_{ij})_{i,j \in \mathcal{T}} = (p_{ij})_{i,j \in \mathcal{T}},$$

und

$$R := (r_{ij})_{i \in \mathcal{T}, j \in \mathcal{R}} = (p_{ij})_{i \in \mathcal{T}, j \in \mathcal{R}}.$$

Für $i \in \mathcal{T}$ und $j \in \mathcal{R}$ definieren wir nun

$$u_{ij} = \mathbb{P}_i(M_\tau = j, \tau < \infty).$$

was der Wahrscheinlichkeit entspricht, dass die Kette, ausgehend von $i \in \mathcal{T}$, im Zustand $j \in \mathcal{R}$ absorbiert wird. Durch Summation der u_{ij} über $j \in \mathcal{C}$ erhält man dann auch die Absorptionswahrscheinlichkeiten zu beliebigen Teilmengen \mathcal{C} von \mathcal{R} , d.h.

$$u_i(\mathcal{C}) = \mathbb{P}_i(M_\tau \in \mathcal{C}, \tau < \infty) = \sum_{j \in \mathcal{C}} u_{ij}.$$

Bezeichnen $q_{ij}^{(n)}$ wie üblich die Komponenten von Q^n , so gilt für alle $n \geq 1$ und $i, j \in \mathcal{T}$

$$q_{ij}^{(n)} = p_{ij}^{(n)}. \quad (2.76)$$

Da Übergänge von \mathcal{R} nach \mathcal{T} unmöglich sind, folgt nämlich

$$\begin{aligned} q_{ij}^{(n)} &= \sum_{j_1, \dots, j_{n-1} \in \mathcal{T}} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j} \\ &= \sum_{j_1, \dots, j_{n-1} \in \mathcal{T}} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j} = p_{ij}^{(n)} \end{aligned}$$

für alle $j \in \mathcal{T}$ und $n \geq 0$. Des Weiteren liefert (2.76) in Verbindung mit Satz 2.22

$$\sum_{n \geq 0} q_{ij}^{(n)} = \mathbb{E}_i \left(\sum_{n \geq 0} \mathbf{1}_{\{M_n = j\}} \right) < \infty. \quad (2.77)$$

$\sum_{n \geq 0} q_{ij}^{(n)}$ gibt also die erwartete Anzahl von Aufenthalten in $j \in \mathcal{T}$ bei Start in i an.

Zur Berechnung der u_{ij} , die wir in der Matrix $U = (u_{ij})_{i \in \mathcal{T}, j \in \mathcal{R}}$ zusammenfassen, verwenden wir eine einfache *Ein-Schritt-Analyse*, deren Ergebnis das folgende Lemma zusammenfasst.

Lemma 2.63. *In den vorherigen Bezeichnungen gilt*

$$u_{ij} = r_{ij} + \sum_{k \in \mathcal{T}} q_{ik} u_{kj}$$

für alle $i \in \mathcal{T}$ und $j \in \mathcal{R}$, d.h. in Matrix-Schreibweise

$$U = R + QU. \quad (2.78)$$

Falls $|\mathcal{T}| < \infty$, ist die Matrix $I - Q$ invertierbar (I die Einheitsmatrix) und

$$U = (I - Q)^{-1}R \quad (2.79)$$

die eindeutige Lösung von (2.78).

Beweis. Die besagte Ein-Schritt-Analyse basiert auf der Zerlegung

$$\{M_\tau = j, \tau < \infty\} = \{M_1 = j\} + \sum_{k \in \mathcal{T}} \{M_1 = k, M_\tau = j, \tau < \infty\}.$$

Mittels der starken Markov-Eigenschaft ergibt sich daraus

$$\begin{aligned} u_{ij} &= \mathbb{P}_i(M_1 = j) + \sum_{k \in \mathcal{T}} \mathbb{P}_i(M_1 = k) \mathbb{P}_k(M_\tau = j, \tau < \infty) \\ &= p_{ij} + \sum_{k \in \mathcal{T}} q_{ik} u_{kj} \end{aligned}$$

für alle $i \in \mathcal{T}$ und $j \in \mathcal{R}$, also (2.78).

Sei nun $|\mathcal{T}| < \infty$. Schreibt man (2.78) um in $(I - Q)U = R$, folgt auch (2.79), sofern wir noch zeigen, dass $I - Q$ invertierbar ist. Gemäß (2.77) bildet $\sum_{n \geq 0} Q^n$ eine endliche Matrix mit den Komponenten $\mathbb{E}_i(\sum_{n \geq 0} \mathbf{1}_{\{M_n = j\}})$, und eine Multiplikation von rechts oder links mit $I - Q$ ergibt I , d.h.

$$(I - Q)^{-1} = \sum_{n \geq 0} Q^n \quad \text{und} \quad ((I - Q)^{-1})_{ij} = \mathbb{E}_i \left(\sum_{n \geq 0} \mathbf{1}_{\{M_n = j\}} \right)$$

für alle $i, j \in \mathcal{T}$. □

Aufgrund ihrer offenkundigen Bedeutung im Zusammenhang mit Absorptionswahrscheinlichkeiten heißt $(I - Q)^{-1}$ *Fundamentalmatrix*. Gemäß (2.77) existiert die Neumannsche Reihe $T := \sum_{n \geq 0} Q^n$ auch dann noch, wenn $|\mathcal{T}| = \infty$, und bildet

weiterhin die formale Inverse von $S := I - Q$, d.h. $ST = TS = I$. Auf dem unendlichdimensionalen Vektorraum $\mathbb{R}^{\mathcal{T}}$ bildet S jedoch i.A. keinen linearen Operator mehr, weil Sx gar nicht zu existieren braucht. Selbst als Endomorphismus auf dem normierten Teilraum $(\ell_\infty, |\cdot|_\infty)$ der beschränkten reellen Folgen ist S nicht notwendig injektiv und folglich das Gleichungssystem (2.78) nicht eindeutig lösbar. Bezeichnen

$$\mathcal{D}(S) = \{x \in \mathbb{R}^{\mathcal{T}} : Sx \in \mathbb{R}^{\mathcal{T}}\} \quad \text{und} \quad \mathcal{B}(S) = \{y \in \mathbb{R}^{\mathcal{T}} : Sx = y \text{ für ein } x \in \mathcal{D}(S)\}$$

den Definitions- bzw. Bildbereich von S , beides lineare Teilräume von $\mathbb{R}^{\mathcal{T}}$, so gilt $STx = x$ nur für diejenigen $x \in \mathcal{D}(T)$ mit $Tx \in \mathcal{D}(S)$. Dies mag als Erläuterung dafür genügen, dass der unendlichdimensionale Fall $(|\mathcal{T}| = \infty)$ größere Sorgfalt verlangt. Wir werden weiter unten eine notwendige und hinreichende Bedingung angeben, unter der das Gleichungssystem (2.78) eine eindeutige Lösung besitzt. Zuvor betrachten wir jedoch das Beispiel einer EMK, für die dies bereits mit Lemma 2.63 klar ist.

Beispiel 2.64 (Gambler's ruin). Sei $(M_n)_{n \geq 0}$ eine Irrfahrt auf $\{0, \dots, N\}$ mit den absorbierenden Barrieren 0 und N und der Übergangsmatrix P aus Beispiel 2.1.4, d.h. $p_{00} = p_{NN} = 1$ und $p_{i,i+1} = 1 - p_{i,i-1} = p \in (0, 1)$ für alle $1 \leq i \leq N-1$. Diese EMK liegt insbesondere folgender einfachen Spielsituation zugrunde: Zwei Spieler A und B spielen irgendein Spiel über mehrere Runden, allerdings nur so lange, bis einer der beiden ruiniert ist. Sei M_n das Kapital von A und $N - M_n$ dasjenige von Spieler B zum Zeitpunkt n . Spieler A gewinnt eine Runde, und damit eine Geldeinheit, unabhängig von allen anderen Runden mit Wahrscheinlichkeit p . Wir fragen nach der Wahrscheinlichkeit, mit der Spieler A bankrott geht.

Das Problem trägt in der Literatur den Namen "Gambler's ruin" und ist offensichtlich genau vom zuvor beschriebenen Typ. Wir fragen nämlich nach $u_{i0} = \mathbb{P}_i(M_\tau = 0)$ für $i \in \mathcal{T} = \{1, \dots, N-1\}$, wobei $\tau = \inf\{n \geq 0 : M_n = 0 \text{ oder } = N\}$. Beachte, dass τ unter jedem \mathbb{P}_i f.s. endlich ist. Setzen wir $u_i = u_{i0}$ für $i \in \mathcal{T}$, $u_0 = 1$, $u_N = 0$ (sofortiger Ruin von Spieler A bzw. B) sowie $q = 1 - p$, so lässt sich (2.78) in der Form

$$\begin{aligned} u_0 &= 1, \quad u_N = 0, \\ u_i &= pu_{i+1} + qu_{i-1}, \quad 1 \leq i \leq N-1 \end{aligned}$$

schreiben, woraus leicht

$$p(u_{i+1} - u_i) = q(u_i - u_{i-1}), \quad 1 \leq i \leq N-1,$$

und dann

$$u_{i+1} - u_i = \left(\frac{q}{p}\right)^i (u_1 - 1), \quad 0 \leq i \leq N-1,$$

folgt. Summation über $i = 0, \dots, N-1$ liefert weiter

$$1 = 1 - u_N = \sum_{i=0}^{N-1} (u_i - u_{i+1}) = (1 - u_1) \sum_{i=0}^{N-1} \left(\frac{q}{p}\right)^i$$

und somit

$$u_1 = \begin{cases} 1 - \frac{1}{N}, & \text{falls } p = q, \\ \frac{\frac{q}{p} - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}, & \text{falls } p \neq q. \end{cases}$$

Kombiniert man dies mit

$$1 - u_j = \sum_{i=0}^{j-1} (u_i - u_{i+1}) = (1 - u_1) \sum_{i=0}^{j-1} \left(\frac{q}{p}\right)^i$$

folgt schließlich allgemein

$$u_j = \begin{cases} 1 - \frac{j}{N}, & \text{falls } p = q, \\ \frac{\left(\frac{q}{p}\right)^j - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}, & \text{falls } p \neq q. \end{cases} \quad (2.80)$$

für alle $0 \leq j \leq N$.

Wenden wir uns nun, wie bereits angekündigt, der Frage nach der Eindeutigkeit der Lösung von (2.78) im Fall “ $|\mathcal{S}| = \infty$ ” zu. Ein weiteres Beispiel soll zunächst zeigen, dass dies tatsächlich nicht immer der Fall zu sein braucht:

Beispiel 2.65 (Die “Strähnen”-Kette). Gegeben die DMK $(M_n)_{n \geq 0}$ aus Beispiel 2.51 mit Zustandsraum $\mathcal{S} = \mathbb{N}_0$ und Übergangsmatrix

$$P = \begin{pmatrix} q_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & 0 & p_1 & 0 & 0 & \dots \\ q_2 & 0 & 0 & p_2 & 0 & \dots \\ \vdots & & & & \ddots & \end{pmatrix},$$

wobei $p_i > 0$ für alle $i \in \mathbb{N}$ und $\sum_{i \geq 0} (1 - p_i) < \infty$ (transienter Fall), interessieren wir uns im Folgenden für die Wahrscheinlichkeit $f_{i0}^* = \mathbb{P}_i(\tau(0) < \infty)$, von $i \in \mathbb{N}$ aus den Zustand 0 irgendwann zu erreichen. Um dies im Rahmen unserer vorhergehenden Untersuchungen tun zu können, machen wir 0 einfach zu einem absorbierenden Zustand und setzen $p_0 = 0$. Dann ist $u_i = f_{i0}^*$ nichts anderes als die Wahrscheinlichkeit dafür, in 0 absorbiert zu werden. Durch Streichen der ersten Zeile und Spalte in P ergibt sich

$$Q = \begin{pmatrix} 0 & p_1 & 0 & 0 & \dots \\ 0 & 0 & p_2 & 0 & \dots \\ \vdots & & & \ddots & \end{pmatrix}.$$

Das Gleichungssystem (2.78) hat damit die Form ($\mathcal{T} = \mathbb{N}$, $u_{i0} = u_i$)

$$u_i = q_i + p_i u_{i+1}, \quad i \geq 1. \quad (2.81)$$

Für $v_i := 1 - u_i$ ergibt dies

$$v_i = p_i v_{i+1}, \quad i \geq 1, \quad (2.82)$$

und man erkennt nun sofort, dass

$$v_i = c \prod_{j \geq i} p_j, \quad i \geq 1,$$

für jedes $0 \leq c \leq 1$ eine Lösung von (2.82) bildet und somit

$$u_i = 1 - c \prod_{j \geq i} p_j, \quad i \geq 1,$$

eine Lösung von (2.81). Die gesuchte Lösung erhalten wir für $c = 1$, denn

$$1 - \prod_{j \geq i} p_j = 1 - \mathbb{P}_i(\tau(0) = \infty) = \mathbb{P}_i(\tau(0) < \infty).$$

Zurückkehrend zur allgemeinen Situation, zeigen wir als nächstes, dass die von uns gesuchte Lösung $u_{ij} = \mathbb{P}_i(M_\tau = j, \tau < \infty)$, $i \in \mathcal{T}$, $j \in \mathcal{R}$, gerade die minimale Lösung von (2.78) bildet:

$$U^{\min} := (I - Q)^{-1}R = \sum_{n \geq 0} Q^n R.$$

Lemma 2.66. *Es gilt $(U^{\min})_{ij} = \mathbb{P}_i(M_\tau = j, \tau < \infty)$ für alle $i \in \mathcal{T}$ und $j \in \mathcal{R}$. Jede weitere Lösung U von (2.78) mit $0 \leq U \leq 1$ (komponentenweise) erfüllt $U \geq U^{\min}$.*

Beweis. Für alle $i \in \mathcal{T}$ und $j \in \mathcal{R}$ gilt

$$\begin{aligned} \left(\sum_{n \geq 0} Q^n R \right)_{ij} &= \sum_{n \geq 0} \sum_{k \in \mathcal{T}} q_{ik}^{(n)} r_{kj} \\ &= \sum_{n \geq 0} \sum_{k \in \mathcal{T}} \mathbb{P}_i(M_1 \in \mathcal{T}, \dots, M_{n-1} \in \mathcal{T}, M_n = k, M_{n+1} = j) \\ &= \sum_{n \geq 0} \mathbb{P}_i(M_1 \in \mathcal{T}, \dots, M_n \in \mathcal{T}, M_{n+1} = j) \\ &= \sum_{n \geq 0} \mathbb{P}_i(\tau = n + 1, M_\tau = j) \end{aligned}$$

$$= \mathbb{P}_i(\tau < \infty, M_\tau = j),$$

was die erste Behauptung sowie insbesondere die Konvergenz der U^{\min} definierenden Reihe beweist. Bildet U eine weitere Lösung von (2.78) mit $0 \leq U \leq 1$, folgt

$$U = R + QU \geq R$$

und damit

$$U \geq R + QR.$$

Wiederholt man dieses Vorgehen, so ergibt sich

$$U \geq Q(R + QR) + R = Q^2R + QR + R \geq \dots \geq \sum_{n=0}^N Q^n R$$

für alle $N \geq 0$ und somit $U \geq U^{\min}$ per Grenzübergang $N \rightarrow \infty$. \square

Als Konsequenz von Lemma 2.66 besitzt (2.78) dann und nur dann die eindeutige Lösung U^{\min} zwischen 0 und 1, wenn für jede weitere Lösung $0 \leq U \leq 1$ schon $U = U^{\min}$ gilt, was durch die Gleichung

$$U - U^{\min} = Q(U - U^{\min})$$

zu der hinreichenden Bedingung

$$Qx = x, \quad 0 \leq x = (x_i)_{i \in \mathcal{I}} \leq 1 \quad \Leftrightarrow \quad x = 0 \quad (2.83)$$

führt. Darüber hinaus können wir nun zeigen:

Satz 2.67. *Genau dann existiert eine eindeutige Lösung $U \in [0, 1]$ (komponentenweise) der Gleichung (2.78), d.h. von $U = QU + R$, wenn nur $x = 0$ die Gleichung (2.83) löst. Letzteres ist wiederum äquivalent zu*

$$\mathbb{P}_i(\tau = \infty) = \mathbb{P}_i(M_n \in \mathcal{I} \text{ für alle } n \in \mathbb{N}_0) = 0 \quad (2.84)$$

für alle $i \in \mathcal{I}$.

Beweis. Wir zeigen zuerst, dass

$$x^{\max} := (\mathbb{P}_i(\tau = \infty))_{i \in \mathcal{I}}$$

die maximale Lösung zwischen 0 und 1 von (2.83) bildet: Für jedes $i \in \mathcal{I}$ gilt

$$\begin{aligned} \mathbb{P}_i(\tau = \infty) &= \sum_{k \in \mathcal{I}} \mathbb{P}_i(M_1 = k, M_n \in \mathcal{I} \text{ für alle } n \geq 2) \\ &= \sum_{k \in \mathcal{I}} \mathbb{P}_i(M_1 = k) \mathbb{P}_k(M_n \in \mathcal{I} \text{ für alle } n \geq 1) \end{aligned}$$

$$= \sum_{k \in \mathcal{T}} q_{ik} \mathbb{P}_k(\tau = \infty).$$

x^{\max} ist demnach eine Lösung von (2.83) und erfüllt außerdem

$$\begin{aligned} x_i^{\max} &= \mathbb{P}_i(\tau = \infty) = \lim_{n \rightarrow \infty} \mathbb{P}_i(\tau > n) \\ &= \lim_{n \rightarrow \infty} \sum_{k \in \mathcal{T}} \mathbb{P}_i(M_n = k, \tau > n) = \lim_{n \rightarrow \infty} \sum_{k \in \mathcal{T}} q_{ik}^{(n)} \end{aligned}$$

für alle $i \in \mathcal{T}$, in Vektorschreibweise

$$x^{\max} = \lim_{n \rightarrow \infty} Q^n \mathbf{1}_{\mathcal{T}}.$$

Dies liefert aber sofort die Maximalität von x^{\max} , denn für jede weitere Lösung x von (2.83) mit $0 \leq x \leq 1$ folgt per Iteration $x = Q^n x$ für alle $n \geq 0$ und dann

$$x = Q^n x \leq Q^n \mathbf{1}_{\mathcal{T}},$$

also bei Grenzübergang $n \rightarrow \infty$

$$x \leq \lim_{n \rightarrow \infty} Q^n \mathbf{1}_{\mathcal{T}} = x^{\max}.$$

(2.83) besitzt also genau dann nur eine Lösung zwischen 0 und 1, wenn $x^{\max} = 0$, d.h.

$$\mathbb{P}_i(\tau = \infty) = \mathbb{P}_i(M_n \in \mathcal{T} \text{ für alle } n \in \mathbb{N}_0) = 0$$

für alle $i \in \mathcal{T}$ gilt, und dies impliziert unter Hinweis auf die Überlegungen vor dem Satz auch die eindeutige Lösbarkeit von (2.78).

Nehmen wir umgekehrt $x^{\max} \neq 0$ an, so definiert neben $U^{\min} = (u_{ij}^{\min})_{i \in \mathcal{T}, j \in \mathcal{R}}$ auch jedes $U(c) := (u_{ij}(c))_{i \in \mathcal{T}, j \in \mathcal{R}}$, $0 < c \leq 1$, gegeben durch

$$u_{ij}(c) := u_{ij}^{\min} + c x_i^{\max} = \mathbb{P}_i(M_\tau = j, \tau < \infty) + c \mathbb{P}_i(\tau = \infty) \in [0, 1]$$

eine Lösung von (2.78), denn

$$\begin{aligned} (R + QU(c))_{ij} &= r_{ij} + (QU^{\min})_{ij} + c(Qx^{\max})_i \\ &= u_{ij}^{\min} + c x_i^{\max} \\ &= u_{ij}(c) \end{aligned}$$

für alle $i \in \mathcal{T}, j \in \mathcal{R}$. Der Beweis des Satzes ist hiermit vollständig. \square

Zum Abschluss bemerken wir, dass im Rahmen der vorherigen Überlegungen auch Funktionale der Form

$$w_i := \mathbb{E}_i \left(\sum_{n=0}^{\tau-1} g(M_n) \right), \quad i \in \mathcal{T},$$

für $g : \mathcal{I} \rightarrow [0, \infty)$ behandelt werden können. $w := (w_i)_{i \in \mathcal{I}}$ erfüllt das Gleichungssystem

$$w_i = g(i) + \sum_{j \in \mathcal{I}} p_{ij} w_j = g(i) + \sum_{j \in \mathcal{I}} q_{ij} w_j, \quad i \in \mathcal{I},$$

bzw. in Vektorschreibweise

$$w = g + Qw,$$

was unter Voraussetzung von (2.84)

$$w = (I - Q)^{-1}g \quad (2.85)$$

liefert. Für Details [51, Seite 107ff]. Wählt man speziell $g \equiv 1$, ergibt sich w als Vektor der mittleren Absorptionszeiten, d.h. $w = (\mathbb{E}_i \tau)_{i \in \mathcal{I}}$, und es gilt gemäß (2.85)

$$(\mathbb{E}_i \tau)_{i \in \mathcal{I}} = (I - Q)^{-1} \mathbf{1}_{\mathcal{I}} \quad (2.86)$$

2.8 Reversibilität: Der Blick zurück

Zeitliche Reversibilität oder kurz *Reversibilität* einer DMK bedeutet anschaulich, dass es für ihre Evolution keinen Unterschied macht, ob man die Zeit vorwärts oder rückwärts liest. Im folgenden wollen wir kurz diskutieren, unter welchen Voraussetzungen diese Eigenschaft vorliegt. Gegeben sei wieder ein Standardmodell $(\Omega, \mathfrak{A}, M = (M_n)_{n \geq 0}, (\mathbb{P}_\lambda)_{\lambda \in \mathcal{P}(\mathcal{I})})$ mit Übergangsmatrix $P = (p_{ij})_{i, j \in \mathcal{I}}$. Den zum Zeitpunkt N zeitlich invertierten Prozess bezeichnen wir mit $\widehat{M}(N) = (\widehat{M}_n(N))_{0 \leq n \leq N}$, d.h.

$$\widehat{M}_n(N) := M_{N-n}.$$

Wir setzen außerdem $\lambda_i^{(n)} = \mathbb{P}_\lambda(M_n = i)$ für $i \in \mathcal{I}$ und $n \geq 0$. Dann gilt:

Satz 2.68.

- (a) Für jedes $N \geq 1$ und $\lambda \in \mathcal{P}(\mathcal{I})$ mit $\lambda_i^{(n)} > 0$ für alle $n \geq 0$ und $i \in \mathcal{I}$ bildet $\widehat{M}(N)$ unter \mathbb{P}_λ eine DMK mit zeitabhängigen Übergangswahrscheinlichkeiten

$$\mathbb{P}_\lambda(\widehat{M}_{n+1}(N) = j | \widehat{M}_n(N) = i) = \widehat{p}_{ij}^{n, n+1}(N, \lambda) := \frac{\lambda_j^{(N-n-1)} p_{ji}}{\lambda_i^{(N-n)}}.$$

- (b) Ist λ eine stationäre Verteilung für M , so ist $\widehat{M}(N)$ für jedes N unter \mathbb{P}_λ zeitlich homogen, d.h. $\widehat{p}_{ij}^{n, n+1}(N, \lambda)$ unabhängig von n und N für alle $i, j \in \mathcal{I}$.

(c) Sind alle $p_{ij} > 0$ ($\Rightarrow M$ irreduzibel und aperiodisch), gilt hiervon auch die Umkehrung: $\widehat{M}(N)$ ist genau dann zeitlich homogen unter \mathbb{P}_λ , wenn M ergodisch und λ die eindeutig bestimmte stationäre Verteilung ist.

Beweis. (a) Die erste Behauptung folgt, weil

$$\begin{aligned} & \mathbb{P}_\lambda(\widehat{M}_{n+1}(N) = j | \widehat{M}_n(N) = i, \widehat{M}_{n-1}(N) = i_{n-1}, \dots, \widehat{M}_0(N) = i_0) \\ &= \frac{\mathbb{P}_\lambda(M_{N-n-1} = j, M_{N-n} = i, M_{N-n+1} = i_{n-1}, \dots, M_N = i_0)}{\mathbb{P}_\lambda(M_{N-n} = i, M_{N-n+1} = i_{n-1}, \dots, M_N = i_0)} \\ &= \frac{\lambda_j^{(N-n-1)} p_{ji} p_{i i_{n-1}} \cdots p_{i_1 i_0}}{\lambda_i^{(N-n)} p_{i i_{n-1}} \cdots p_{i_1 i_0}} \\ &= \frac{\lambda_j^{(N-n-1)} p_{ji}}{\lambda_i^{(N-n)}} \\ &= \mathbb{P}_\lambda(\widehat{M}_{n+1}(N) = j | \widehat{M}_n(N) = i) \end{aligned}$$

für alle $i, i_0, \dots, i_{n-1} \in \mathcal{S}$ mit $\mathbb{P}_\lambda(\widehat{M}_n(N) = i, \widehat{M}_{n-1}(N) = i_{n-1}, \dots, \widehat{M}_0(N) = i_0) > 0$.

(b) Ist λ eine stationäre Verteilung für M , folgt $\lambda_i^{(n)} = \lambda_i$ für alle $i \in \mathcal{S}$, $n \geq 0$ und daraus die Unabhängigkeit der $\widehat{p}_{ij}^{n,n+1}(N, \lambda)$ von n und N .

(c) Hängt umgekehrt $\widehat{p}_{ij}^{n,n+1}(N, \lambda) = \widehat{p}_{ij}$ nicht von n, N ab, betrachte zunächst den Fall $i = j$: Offensichtlich folgt dann aus $\widehat{p}_{ii} = \lambda_i^{(n-1)} p_{ii} / \lambda_i^{(n)} > 0$ für alle n die Existenz eines $\alpha_i > 0$, so dass $\lambda_i^{(n)} = \alpha_i \lambda_i^{(n-1)}$, also $\lambda_i^{(n)} = \alpha_i^n \lambda_i$ für alle $n \geq 0$ und $i \in \mathcal{S}$. Dies liefert weiter

$$\widehat{p}_{ij} = \left(\frac{\alpha_j}{\alpha_i} \right)^n \cdot \frac{p_{ji}}{\alpha_i} \cdot \frac{\lambda_j}{\lambda_i}$$

für alle $n \geq 0$ und somit die Unabhängigkeit der α_i von i , d.h. $\lambda_i^{(n)} = \alpha^n \lambda_i$ für alle n, i . Beachten wir abschließend $1 = \sum_{i \in \mathcal{S}} \lambda_i^{(n)} = \alpha^n \sum_{i \in \mathcal{S}} \lambda_i = \alpha^n$, folgt $\alpha = 1$ und damit $\lambda = \lambda^{(n)}$ für alle $n \geq 0$, was die Stationarität von M unter \mathbb{P}_λ beweist. Als irreduzible und aperiodische DMK ist M folglich ergodisch und λ die eindeutige stationäre Verteilung. \square

Satz 2.68 deckt zwar aufgrund der Voraussetzung $\lambda_i^{(n)} > 0$ sowie $p_{ij} > 0$ in Teil c) nicht alle denkbaren Fälle ab, reicht aber für unsere Zwecke aus. Wir richten unser Augenmerk im Folgenden ohnehin auf den stationären Fall und betrachten eine positiv rekurrente DMK M mit stationärer Verteilung π . Zur besseren Veranschaulichung der Zeitumkehrung führen wir die *doppelt unendliche* Folge $M^* = (M_n^*)_{n \in \mathbb{Z}}$ ein, definiert (auf irgendeinem W-Raum $(\Omega, \mathfrak{A}, \mathbb{P})$) durch

$$\mathbb{P}^{(M_n^*, M_{n+1}^*, \dots)} = \mathbb{P}_\pi^M$$

für alle $n \in \mathbb{Z}$. Die Existenz eines solchen Prozesses ergibt sich aus dem Konsistenzsatz von Kolmogorov ([2, Satz 54.7]): \mathbb{P}^{M^*} ist der projektive Limes der Familie

$$Q_I = \mathbb{P}_\pi^{(M_n)_{n \in I_0}}, \quad I \subset \mathbb{Z} \text{ endlich,}$$

wobei $I_0 = \{0, i_1 - i_0, \dots, i_m - i_0\}$, falls $I = \{i_0, \dots, i_m\}$ mit $i_0 < \dots < i_m$. Anschaulich können wir uns vorstellen, dass M^* aus M entsteht, indem wir den Zeitpunkt 0 nach $-\infty$ verschieben. Mit anderen Worten: M^* repräsentiert die DMK M unter der Voraussetzung, dass diese vor "langer, langer Zeit" im Gleichgewicht gestartet ist.

Nach Satz 2.68(b) folgt, dass für jedes $N \in \mathbb{Z}$ der in N zeitumgekehrte Prozess $(M_{N-n}^*)_{n \in \mathbb{Z}}$ eine zeitlich homogene stationäre DMK mit Übergangswahrscheinlichkeiten

$$\hat{p}_{ij} = \frac{\pi_j p_{ji}}{\pi_i}, \quad i, j \in \mathcal{S}, \quad (2.87)$$

bildet. Es gilt nämlich

$$\mathbb{P}^{(M_{N-n}^*)_{0 \leq n \leq N}} = \mathbb{P}_\pi^{(\hat{M}_n(N))_{0 \leq n \leq N}}$$

für alle $N \geq 0$. Reversibilität bedeutet nun folgendes:

Definition 2.69. Eine DMK $M = (M_n)_{n \geq 0}$ heißt (zeitlich) reversibel, wenn

$$(M_0, M_1, \dots, M_n) \stackrel{d}{=} (M_n, M_{n-1}, \dots, M_0)$$

für alle $n \geq 0$.

Da (13.4) insbesondere $M_0 \stackrel{d}{=} M_n$ für alle $n \geq 0$ impliziert, ist jede reversible DMK notwendig stationär. Für die doppelt unendliche Folge M^* ist (13.4) gleichbedeutend mit

$$(M_{N-n}^*)_{n \in \mathbb{Z}} \stackrel{d}{=} (M_{N+n}^*)_{n \in \mathbb{Z}}$$

für alle $N \in \mathbb{Z}$. Vorwärts- und Rückwärtsprozess bilden also identisch verteilte zeitlich homogene DMK mit folglich gleicher Übergangsmatrix, was $p_{ij} = \pi_j p_{ji} / \pi_i$ oder umgeschrieben

$$\pi_i p_{ij} = \pi_j p_{ji} \quad (2.88)$$

für alle $i, j \in \mathcal{S}$ liefert. Diese Gleichungen, genannt *detaillierte Gleichgewichtsgleichungen*, besagen anschaulich, dass für je zwei Zustände i und j die Flusswahrscheinlichkeit von i nach j mit der für die umgekehrte Richtung, also von j nach i , übereinstimmt. Abb. 2.8 stellt dies schematisch dar.

Summiert man in (2.88) auf beiden Seiten über $i \in \mathcal{S}$, ergibt sich offenbar

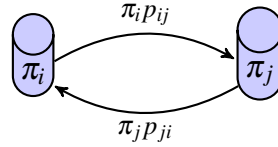


Abb. 2.8 Detailliertes Gleichgewicht veranschaulicht in einem *Flüssigkeitsmodell*: Stellen wir uns die Zustände $i \in \mathcal{S}$ als Wasserbehälter vor. Behälter i enthält im Gleichgewicht π_i Liter Wasser. In jedem Zeitschritt werden $\pi_i p_{ij}$ Liter von Behälter i in Behälter j und $\pi_j p_{ji}$ Liter von j nach i umgefüllt. Liegt detailliertes Gleichgewicht vor, so bleiben die Flüssigkeitsmengen in zwei Behältern bereits gleich, wenn nur zwischen diesen beiden der Flüssigkeitsaustausch vorgenommen worden ist. Im nicht-reversiblen Fall muss dagegen erst zwischen allen Behältern der Austausch vorgenommen worden sein, bevor wieder alle Behälter dieselbe Menge wie vor dem Austausch enthalten.

$$\sum_{i \in \mathcal{S}} \pi_i p_{ij} = \pi_j \sum_{i \in \mathcal{S}} p_{ji} = \pi_j$$

für alle $j \in \mathcal{S}$. Jede normierte (σ -endliche) Lösung $\pi \neq 0$ der detaillierten Gleichgewichtsgleichungen ist also notwendig eine stationäre Verteilung (ein stationäres Maß) der betrachteten DMK. Wir halten fest:

Satz 2.70. Eine irreduzible DMK M mit Übergangsmatrix $P = (p_{ij})_{i,j \in \mathcal{S}}$ ist genau dann reversibel, wenn sie stationär ist und die stationäre Verteilung, π , den detaillierten Gleichgewichtsgleichungen (2.88) genügt. M ist dann also positiv rekurrent.

Beweis. “ \Rightarrow ” Eine reversible DMK M ist notwendig stationär, wie bereits oben bemerkt, und besitzt somit eine stationäre Verteilung π , die aufgrund der Irreduzibilität außerdem eindeutig bestimmt und positiv ist (Satz 2.33(a) und Satz 2.45). Dass die detaillierten Gleichgewichtsgleichungen gelten, folgt aus den Überlegungen unmittelbar vor diesem Satz.

“ \Leftarrow ” Für diese Richtung ist wegen Satz 2.68 nichts mehr zu zeigen. □

Wir kommen zu einigen Folgerungen aus den detaillierten Gleichgewichtsgleichungen. Sie besagen offenkundig nichts anderes als die Symmetrie der Funktion $f(i, j) = \pi_i p_{ij}$ oder $g(i, j) = p_{ij} / \pi_j$ oder auch $h(i, j) = \pi_i^{1/2} p_{ij} / \pi_j^{1/2}$.

Satz 2.71. Eine positiv rekurrente DMK M ist genau dann im Gleichgewicht reversibel, wenn ihre Übergangsmatrix $P = (p_{ij})_{i,j \in \mathcal{S}}$ die Darstellung

$$P = D^{-1}AD \tag{2.89}$$

für eine symmetrische Matrix A und eine Diagonalmatrix $D = \text{diag}(d_i, i \in \mathcal{S})$ mit $\text{Spur } D^2 = \sum_{i \in \mathcal{S}} d_i^2 < \infty$ besitzt.

Beweis. “ \Rightarrow ” Ist M positiv rekurrent mit stationärer Verteilung π und reversibel, gelten die detaillierten Gleichgewichtsgleichungen, und daher (2.89) mit $A = (h(i, j))_{i, j \in \mathcal{S}}$ und $D = \text{diag}(\pi_i^{1/2}, i \in \mathcal{S})$.

“ \Leftarrow ” Bei Gültigkeit von (2.89) dürfen wir o.B.d.A. $D \geq 0$ und $d^2 := \sum_{i \in \mathcal{S}} d_i^2 = 1$ voraussetzen; andernfalls gehen wir zu der Darstellung $D'^{-1}A'D'$ mit $D' = \text{diag}(|d_i|/d, i \in \mathcal{S})$ und $A' = (\text{sign}(d_i)a_{ij}\text{sign}(d_j))_{i, j \in \mathcal{S}}$ über. Aufgrund der Symmetrie von A erhalten wir

$$p_{ij} = d_i^{-1}a_{ij}d_j = d_j a_{ji} d_i^{-1} = d_j^2 p_{ji} d_i^{-2},$$

also die Gültigkeit der detaillierten Gleichgewichtsgleichungen mit $\pi = (d_i^2)_{i \in \mathcal{S}}$. π bildet folglich die eindeutig bestimmte stationäre Verteilung (M positiv rekurrent), und M ist unter \mathbb{P}_π (im Gleichgewicht) reversibel (Satz 2.70). \square

Nach dieser algebraischen Charakterisierung von Reversibilität geben wir noch ein anderes, von KOLOMOGOROV stammendes Kriterium probabilistischer Natur:

Satz 2.72. (Kolmogorov-Kriterium für Reversibilität) *Eine positiv rekurrente DMK M mit Übergangsmatrix $P = (p_{ij})_{i, j \in \mathcal{S}}$ ist genau dann im Gleichgewicht reversibel, wenn*

$$\prod_{k=1}^n p_{i_{k-1}i_k} = \prod_{k=1}^n p_{i_k i_{k-1}} \quad (2.90)$$

für alle $i_0, \dots, i_n \in \mathcal{S}$, $i_0 = i_n$, und $n \geq 1$.

Beweis. “ \Rightarrow ” Unter \mathbb{P}_π gilt nach Definition von Reversibilität

$$\begin{aligned} \pi_{i_0} \prod_{k=1}^n p_{i_{k-1}i_k} &= \mathbb{P}_\pi(M_0 = i_0, \dots, M_{n-1} = i_{n-1}, M_n = i_0) \\ &= \mathbb{P}_\pi(M_0 = i_0, M_1 = i_{n-1}, \dots, M_n = i_0) = \pi_{i_0} \prod_{k=1}^n p_{i_k i_{k-1}} \end{aligned}$$

für alle $i_0, \dots, i_n \in \mathcal{S}$, $i_0 = i_n$, und $n \geq 1$, woraus (2.90) nach Division durch $\pi_{i_0} > 0$ folgt.

“ \Leftarrow ” Um zu zeigen, dass (2.90) auch hinreichend ist für Reversibilität, wähle dort $n \geq 2$, $i_0 = i$ und $i_{n-1} = j$ beliebig und summiere im Fall $n \geq 3$ außerdem über alle $(i_1, \dots, i_{n-2}) \in \mathcal{S}^{n-2}$. Dann ergibt sich

$$p_{ij}^{(n-1)} p_{ji} = p_{ij} p_{ji}^{(n-1)}$$

für alle $n \geq 2$ und bei Summation über $n = 2, \dots, N+1$

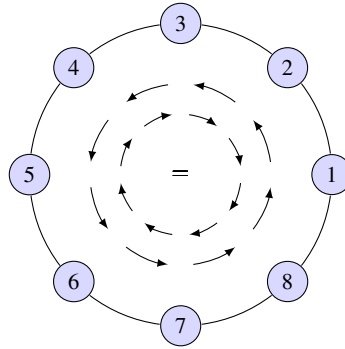


Abb. 2.9 Das Kolmogorov-Kriterium für Reversibilität: Zyklische Pfade haben für jede Laufrichtung die gleiche Wahrscheinlichkeit.

$$\left(\frac{1}{N} \sum_{n=2}^{N+1} p_{ij}^{(n-1)} \right) p_{ji} = p_{ij} \left(\frac{1}{N} \sum_{n=2}^{N+1} p_{ij}^{(n-1)} \right).$$

Lässt man nun noch N gegen ∞ streben und beachtet $C\text{-}\lim_{N \rightarrow \infty} p_{ij}^{(N)} = \pi_j$ gemäß (2.51) (M positiv rekurrent), so folgen die detaillierten Gleichgewichtsgleichungen und damit die Reversibilität von M unter \mathbb{P}_π . \square

Anschaulich besagt das Kolmogorov-Kriterium, dass eine reversible DMK jeden zyklischen Pfad $i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_{n-1} \rightarrow i_0$ mit gleicher Wahrscheinlichkeit vorwärts wie rückwärts durchläuft [Abb. 2.9].

Als einfache, aber interessante Konsequenz des Kolmogorov-Kriteriums notieren wir noch:

Korollar 2.73. Die Periode einer positiv rekurrenten, im Gleichgewicht reversiblen DMK beträgt höchstens 2.

Beweis. Den Trivialfall $|\mathcal{S}| = 1$ ausgeschlossen, wähle in (2.90) irgendeinen zyklischen Pfad positiver Wahrscheinlichkeit mit $i_0 = i$ und $i_{n-1} = j \neq i$. Dann folgt insbesondere $p_{ij} > 0$ und $p_{ji} > 0$ und somit $p_{ii}^{(2)} \geq p_{ij} p_{ji} > 0$, d.h. i und folglich alle Zustände sind höchstens 2-periodisch. \square

Reversibilität vereinfacht häufig ganz erheblich die Analyse stochastischer Systeme im Gleichgewicht und spielt eine wichtige Rolle z.B. bei der Untersuchung stochastischer Netzwerke (Warteschlangentheorie), von Genfrequenzen in Populationen oder auch in der statistischen Physik. Eine exzellente Monographie zu diesem Thema bildet [38]. Betrachten wir einige Beispiele:

Beispiel 2.74 (Doppelt stochastische Matrizen). Sei $M = (M_n)_{n \geq 0}$ eine EMK in einem Standardmodell mit Zustandsraum \mathcal{S} und Übergangsmatrix $P = (p_{ij})_{i,j \in \mathcal{S}}$.

Wir nehmen an, dass die Gleichverteilung auf \mathcal{S} die stationäre Verteilung der Kette bildet, also $\pi_i = 1/|\mathcal{S}|$ für alle $i \in \mathcal{S}$. Aus den Invarianzgleichungen ergibt sich dann

$$\frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} p_{ij} = \frac{1}{|\mathcal{S}|}, \quad \text{also} \quad \sum_{i \in \mathcal{S}} p_{ij} = 1$$

für alle $i \in \mathcal{S}$ und somit für die Übergangsmatrix P , dass nicht nur ihre Zeilensummen, sondern auch ihre Spaltensummen stets 1 betragen. Man bezeichnet P in diesem Fall als *doppelt stochastisch*. Beachte, dass mit P auch jedes P^n doppelt stochastisch ist, denn P^n erfüllt ebenfalls die Invarianzgleichungen. Ein Blick auf die detaillierten Gleichgewichtsgleichungen (2.88) zeigt, dass diese unter den getroffenen Voraussetzungen nur dann gelten, wenn

$$p_{ij} = p_{ji}$$

für alle $i, j \in \mathcal{S}$, wenn also P symmetrisch ist ($P = P^\top$). Allgemein folgt in der gegebenen Situation nach (2.87), dass $P^\top = (p_{ji})_{i, j \in \mathcal{S}}$ gerade die Übergangsmatrix der doppelt unendlichen stationären Version von M bei rückwärts laufender Zeit, d.h. von $(M_{-n}^*)_{n \in \mathbb{Z}}$ bildet.

Als spezielle Beispiele endlicher MK mit doppelt stochastischer und symmetrischer Übergangsmatrix erwähnen wir symmetrische Irrfahrten auf endlichen zyklischen Gruppen $G = \{a^n : 1 \leq n \leq |G|\}$ mit den Übergangswahrscheinlichkeiten

$$p_{a^n a^{n+1}} = p_{a^n a^{n-1}} = \frac{1}{2}.$$

Beispiel 2.75 (Einfache Irrfahrten auf einem Graphen). [☞ 2.1.5] Eine einfache Irrfahrt $M = (M_n)_{n \geq 0}$ auf einem endlichen, einfachen, ungerichteten und zusammenhängenden Graphen $G = (V, E)$ springt von einem beliebigen Knoten v in einen Nachbarknoten w mit Wahrscheinlichkeit $\frac{1}{d(v)}$, wobei $d(v)$ die Zahl der Nachbarn von v bezeichnet. Dabei gelten zwei Knoten als benachbart ($v \sim w$), wenn es eine Kante zwischen ihnen gibt. Da G zusammenhängend ist, folgt $d(v) \geq 1$ für alle $v \in V$ sowie die Irreduzibilität von M .

Es leuchtet intuitiv ein, dass derartige Irrfahrten unter der stationären Verteilung stets reversibel sind. Um dies auch formal zu verifizieren, betrachten wir wieder die detaillierten Gleichgewichtsgleichungen (2.88), die hier die Form $\frac{\pi_v}{d(v)} = \frac{\pi_w}{d(w)}$ für alle v, w mit $v \sim w$ haben. Wie man sofort erkennt, ergibt sich eine normierte nichtnegative Lösung π durch

$$\pi_v = \frac{d(v)}{\sum_{w \in V} d(w)} = \frac{d(v)}{2|E|}$$

für alle $v \in V$ und somit die erwartete Reversibilität von M unter \mathbb{P}_π . Beachte hierbei, dass die Bestimmung der stationären Verteilung nicht vorab, sondern direkt mit dem Lösen der detaillierten Gleichgewichtsgleichungen erfolgt ist.

Beispiel 2.76 (Markov-Ketten auf einem Baum). Ein Graph der zuvor betrachteten Form, in dem ferner je zwei Knoten durch genau einen Pfad minimaler Länge (= Anzahl durchlaufener Kanten) verbunden sind, heißt (endlicher) Baum. Dessen Knoten lassen sich wie folgt in lexikographischer Weise durch endliche "Wörter" mit Alphabet \mathbb{N} benennen (Ulam-Harris-Markierung): Wähle irgendeinen Knoten als Wurzel und bezeichne ihn mit \emptyset (leeres Wort). Betrachte als nächstes dessen $d(\emptyset)$ Nachbarn und markiere diese mit $1, \dots, d(\emptyset)$ (Wörter der Länge 1), wobei die Reihenfolge keine Rolle spielt. Für jeden der Knoten $v \in \{1, \dots, d(\emptyset)\}$ bezeichne dessen $n(v) := d(v) - 1$ (sofern $\neq 0$) noch nicht numerierten Nachbarn mit $(v, 1), \dots, (v, n(v))$ (Wörtern der Länge 2), wobei diese zumeist in der verkürzten Form $v1, \dots, 1n(v)$ geschrieben werden. So fortfahrend bis alle Knoten markiert sind, erhalten wir einen Baum, in dem jeder Knoten durch ein endliches Wort der Form $v = i_1 \dots i_m \in \mathbb{N}^m$ markiert ist und $v \rightarrow i_1 \dots i_{m-1} \rightarrow \dots \rightarrow i_1 \rightarrow \emptyset$ den minimalen Pfad von v zur Wurzel angibt. Abb. 2.10 illustriert dies an einem einfachen Beispiel.

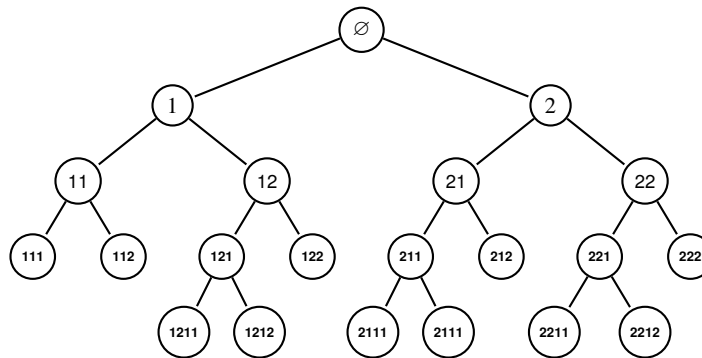


Abb. 2.10 Ein endlicher binärer Baum mit Ulam-Harris-Markierung.

Sei $M = (M_n)_{n \geq 0}$ eine irreduzible MK auf einem derartigen Baum mit Übergangswahrscheinlichkeiten p_{vw} , $v \neq w$, die genau dann positiv sind, wenn $v \sim w$ gilt, ansonsten aber keiner Einschränkung unterliegen. Falls $v = i_1 \dots i_m$, $m \geq 2$, bedeutet dies $w = i_1 \dots i_{m-1}$ oder $w \in \{i_1, \dots, i_m j : 1 \leq j < d_{i_1 \dots i_m}\}$. Wir machen keine Annahmen über die Verharrungswahrscheinlichkeiten p_{vv} . Dass jede solche EMK M reversibel ist, springt vielleicht nicht sofort ins Auge, lässt sich aber leicht mittels des Kolmogorov-Kriteriums beweisen: Gegeben einen zyklischen Pfad $v \rightarrow v_1 \rightarrow \dots \rightarrow v_{n-1} \rightarrow v$, folgt aus der Baumeigenschaft, dass jede Kante dieses Pfades genauso oft vorwärts wie rückwärts durchlaufen wird (Abb. 2.11) und daher stets

$$p_{vv_1} \cdot \dots \cdot p_{v_{n-1}v} = p_{vv_{n-1}} \cdot \dots \cdot p_{v_1v}$$

gilt. Auf der Basis der vorgenommenen Knotennumerierung und mit Hilfe der detaillierten Gleichgewichtsgleichungen können wir auch die stationäre Verteilung bestimmen: Aus

Beispiel 2.77 (Geburts- und Todesprozesse). Sei also M eine DMK auf $\mathcal{S} = \mathbb{N}_0$ oder $\mathcal{S} = \{0, \dots, N\}$ mit Übergangsmatrix P der Form

$$P = \begin{pmatrix} p_{00} & p_{01} & 0 & 0 & 0 & \dots \\ p_{10} & p_{11} & p_{12} & 0 & 0 & \dots \\ 0 & p_{21} & p_{22} & p_{23} & 0 & \dots \\ \vdots & & & \ddots & & \end{pmatrix},$$

wobei $p_{ij} > 0$ für alle $i, j \in \mathcal{S}$ mit $|i - j| = 1$ gelte. Dies garantiert die Irreduzibilität von M .

Unter Hinweis auf (2.91) und die Bemerkung vor Satz 2.70 bildet

$$\pi_0 = 1 \quad \text{und} \quad \pi_i = \frac{p_{01}p_{12} \cdot \dots \cdot p_{i-1,i}}{p_{10}p_{21} \cdot \dots \cdot p_{i,i-1}} \quad \text{für } i \geq 1 \quad (2.93)$$

ein stationäres Maß. Hat π endliche Gesamtmasse, d.h.

$$\pi(\mathcal{S}) = 1 + \sum_{0 \neq i \in \mathcal{S}} \frac{p_{01}p_{12} \cdot \dots \cdot p_{i-1,i}}{p_{10}p_{21} \cdot \dots \cdot p_{i,i-1}} < \infty, \quad (2.94)$$

so ist M nach Satz 2.45 positiv rekurrent und folglich im Gleichgewicht reversibel.

Wie der Leser nun leicht einsieht, gelten entsprechende Resultate in der Tat für beliebige irreduzible, sprungfreie MK auf $\mathcal{S} = \mathbb{Z}$ und $\mathcal{S} = \{m, \dots, n\}$, $m, n \in \mathbb{Z}$. Im ersten Fall fasse man \mathbb{Z} als Baum mit Wurzel 0 auf, im zweiten genügt gar eine einfache Verschiebung (Umbenennung) der Zustände (Übergang zu $\mathcal{S} = \{0, \dots, n - m\}$). Als spezielle Beispiele, die uns zuvor bereits begegnet sind, erwähnen wir die *Irrfahrten mit reflektierenden Barrieren* (Beispiel 2.1.3) sowie das *Ehrenfest-Modell* für Wärmeaustausch (Beispiel 2.1.6), auf das wir im nächsten Abschnitt nochmals zurückkommen (☞ 2.9.2).

2.9 Und nochmals Beispiele – alte und neue

Es folgt ein weiterer Abschnitt mit einer Reihe von Anwendungsbeispielen, wobei wir vor allem die Beispiele aus Abschnitt 2.1 nochmals aufgreifen, aber am Ende auch einige neue hinzugefügt haben.

2.9.1 Markov-Ketten mit zwei Zuständen

[☞ 2.1.1] Sei M eine EMK mit Zustandsraum $\mathcal{S} = \{0, 1\}$ und Übergangsmatrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

wobei $0 < p, q < 1$ gelte. Die ausgeschlossenen Fälle kann der Leser leicht selbst untersuchen. Sie sind aber von nur geringem Interesse. P hat die Eigenwerte 1 und $1 - p - q$ und läßt sich diagonalisieren:

$$P = SDS^{-1}, \quad \text{wobei}$$

$$S = \begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix}, \quad S^{-1} = \frac{1}{p+q} \begin{pmatrix} q & p \\ -1 & 1 \end{pmatrix} \quad \text{und} \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix}.$$

Die Spalten von S bilden rechte, die Zeilen von S^{-1} linke Eigenvektoren von P . Sie sind bis auf skalares Vielfaches eindeutig bestimmt, wobei wir den Skalar so gewählt haben, dass der linke Eigenvektor zum Eigenwert 1 eine Verteilung, nämlich die stationäre Verteilung π definiert, d.h.

$$\pi_0 = \frac{q}{p+q} \quad \text{und} \quad \pi_1 = \frac{p}{p+q}.$$

Mittels der Diagonalisierung kann man nun leicht alle Potenzen P^n von P berechnen und erhält

$$\begin{aligned} P^n &= SD^nS^{-1} = S \begin{pmatrix} 1 & 0 \\ 0 & (1-p-q)^n \end{pmatrix} S^{-1} \\ &= \begin{pmatrix} \pi_0 & \pi_1 \\ \pi_0 & \pi_1 \end{pmatrix} + (1-p-q)^n \begin{pmatrix} \pi_1 & -\pi_1 \\ -\pi_0 & \pi_0 \end{pmatrix}. \end{aligned} \quad (2.95)$$

M ist ergodisch und im Gleichgewicht reversibel. Aus (2.95) ergibt sich außerdem die exakte Konvergenzgeschwindigkeit im Ergodensatz:

$$|p_{0i}^{(n)} - \pi_i| = \pi_1(1-p-q)^n \quad \text{und} \quad |p_{1i}^{(n)} - \pi_i| = \pi_0(1-p-q)^n$$

für $i \in \{0, 1\}$. Es ist also gerade der zweite Eigenwert von P , der die geometrische Rate im Ergodensatz determiniert. Kein Zufall, wie sich im nächsten Kapitel herausstellt (☞ Satz 3.10 und Korollar 3.12)! Der zweitgrößte Eigenwert einer ergodischen Übergangsmatrix ist auch allgemein der entscheidende Wert für die Konvergenzrate gegen die stationäre Verteilung.

MK mit zwei Zuständen gehören zu den wenigen Beispielen, in denen sich alle Potenzen der Übergangsmatrix explizit berechnen lassen. Entscheidend ist die niedrige Dimension, welche den Aufwand für die Diagonalisierung einschließlich der Berechnung der Eigenvektormatrizen in Grenzen hält.

2.9.2 Die Diffusionsmodelle von Ehrenfest und Bernoulli-Laplace

[☞ 2.1.6] Urnenmodelle sind in vielen Anwendungsbereichen ein beliebtes und mitunter faszinierendes Mittel zur Beschreibung mehr oder minder komplexer zufallsabhängiger Sachverhalte. Ihre Faszination beziehen sie vor allem aus der Ein-

fachheit des durchzuführenden Experiments, das schlicht im Ziehen von Kugeln besteht. Doch lässt sich dies in derartig mannigfaltiger Weise bewerkstelligen (eine oder mehrere Ziehungen von einer oder mehreren Kugeln mit oder ohne Zurücklegen aus einer oder mehreren Urnen), dass in der Tat selbst äußerst komplexe stochastische Phänomene durch sie modelliert werden können.

Die folgenden beiden Modelle, von denen wir das erste bereits aus 2.1.6 kennen, dienen der Beschreibung der Diffusion von Teilchen:

Ehrenfest-Modell

Aus einer Urne mit $2N$ Kugeln der Farbe rot oder schwarz wird sukzessiv immer eine Kugel gezogen und durch eine Kugel der jeweils anderen Farbe ausgetauscht. Bezeichnet M_n die Anzahl roter Kugeln nach n Ziehungen, so bildet $M = (M_n)_{n \geq 0}$ eine EMK mit Zustandsraum $\mathcal{S} = \{0, \dots, 2N\}$ und Übergangswahrscheinlichkeiten

$$p_{ij} = \begin{cases} 1 - \frac{i}{2N}, & \text{falls } j = i + 1, \\ \frac{i}{2N}, & \text{falls } j = i - 1, \\ 0, & \text{sonst.} \end{cases}$$

Auf den ersten Blick scheint dies ein anderes Modell als das in 2.1.6 zu sein. Zum einen war dort der Zustandsraum $\{-N, \dots, 0, \dots, N\}$, zum anderen hatten wir es mit zwei Urnen A und B statt mit nur einer zu tun. Eine Umbenennung der Zustände ($i \rightarrow i + N$ für $i = -N, \dots, N$) sowie die Identifikation der roten bzw. schwarzen Kugeln hier mit denen in Urne A bzw. B dort sollte den Leser aber von der tatsächlichen Äquivalenz der beiden Modelle überzeugen.

Es handelt sich erneut um einen Geburts- und Todesprozess mit endlichem Zustandsraum, der außerdem offenkundig irreduzibel und damit positiv rekurrent ist. Es folgt die Reversibilität im Gleichgewicht, und die stationäre Verteilung ergibt sich vermöge (2.93) und (2.94) leicht zu

$$\pi_i = \binom{2N}{i} \frac{1}{2^{2N}}, \quad i = 0, \dots, 2N,$$

d.h. $\pi = \text{Bin}(2N, \frac{1}{2})$.

Die Physiker P. und T. EHRENFEST im Jahr 1907 und später SMOLUCHOWSKI im Jahr 1916 benutzten das Modell, um ein scheinbares Paradoxon, aufgezeigt von Zermelo, zu erklären, das um die Jahrhundertwende BOLTZMANNs *kinetische Theorie der Materie* zu zerstören drohte. In der kinetischen Theorie bildet Wärmeaustausch einen zufälligen Prozess, in der *Thermodynamik* dagegen einen geordneten, irreversiblen Vorgang gegen einen Gleichgewichtszustand. Im vorliegenden Modell wird dieses Gleichgewicht erreicht, wenn die Temperaturen beider Körper (zumindest approximativ oder makroskopisch) übereinstimmen, d.h. die Anzahl roter und schwarzer Kugeln übereinstimmen. Andererseits gilt in einem kinetischen Modell wie diesem wegen der *Rekurrenzeigenschaft*, dass das System vom thermodynamischen Gleichgewicht $i = N$ irgendwann in einen Zustand extremen

Ungleichgewichts ($i = 0$ oder $i = 2N$) zurückkehrt. Es war POINCARÉ, der als erster nachwies, dass jedes statistisch-mechanische System diese Eigenschaft besitzt, und man glaubte dies im Widerspruch zur Irreversibilität in der Thermodynamik. Kurioserweise bildete eines der Hauptanliegen der kinetischen Theorie, ausgehend von der Molekulartheorie der Materie, das bessere Verständnis der Thermodynamik, die eine weitgehend phänomenologische Theorie auf makroskopischer Skala darstellt.

Ein Blick auf die stationäre Verteilung zeigt, dass unter dieser eine in etwa gleichgroße Anzahl roter und schwarzer Kugeln wesentlich wahrscheinlicher ist als ein starkes Übergewicht der einen oder anderen Farbe. Insbesondere gilt im Mittel gerade

$$\mathbb{E}_\pi(M_n - N) = 0,$$

und man kann ferner leicht zeigen, dass unter beliebigem \mathbb{P}_i für die Abweichung $M_n - N$

$$\mathbb{E}_i(M_n - N) = \left(1 - \frac{1}{N}\right)^n (i - N), \quad (2.96)$$

also für $n \rightarrow \infty$ exponentielle Konvergenz gegen 0 vorliegt. Bezeichnet τ im physikalischen Modell die Austauschfrequenz pro Sekunde, so hat man $n = \tau t$ Übergänge (Ziehungen) im Zeitraum t . Mit $v = \log\left(1 - \frac{1}{N}\right)^\tau$ erhält man so in (2.96) das *Newtonsche Abkühlungsgesetz*:

$$\mathbb{E}_i(M_n - N) = (i - N)e^{vt}. \quad (2.97)$$

Der Grund für dieses Mittelwertverhalten besteht in einer starken Asymmetrie der Übergangswahrscheinlichkeiten $p_{i,i+1}$ und $p_{i,i-1}$, die aber nicht im Widerspruch zur Reversibilität des ablaufenden Prozesses steht. Letztere erfordert nämlich eine Symmetrie der *gemeinsamen Verteilungen*,

$$\mathbb{P}_\pi^{(M_n, \dots, M_{n+m})} = \mathbb{P}_\pi^{(M_{n+m}, \dots, M_n)}$$

für alle n, m , und besagt anschaulich, dass es keinen Unterschied macht, ob man den Prozess rückwärts oder vorwärts ablaufen lässt. Zustände starken Ungleichgewichts wird man in jedem Fall äußerst selten sehen, und ausgehend von solchen, besteht zeitlich *in beiden Richtungen* ein exponentieller Trend zum Gleichgewicht, manifestiert durch (2.96) oder (2.97). Für eine weitergehende Diskussion des Themas \mathbb{E}^{\otimes} [60].

Bernoulli-Laplace-Modell

Eine Variante zum vorherigen Modell sieht wie folgt aus: Es befinden sich $2N$ Kugeln, davon N rot und N schwarz, je zur Hälfte in zwei Urnen A und B . Eine Ziehung besteht darin, aus jeder Urne eine Kugel auszuwählen und sie in die jeweils andere Urne zu legen. Sei M_n die Anzahl roter Kugeln in Urne A nach n Ziehungen. Dann bildet $M = (M_n)_{n \geq 0}$ wiederum einen Geburts- und Todesprozess mit Zustandsraum $\mathcal{S} = \{0, \dots, N\}$. Seine Übergangswahrscheinlichkeiten folgen mittels elementarer Kombinatorik:

$$p_{ij} = \begin{cases} (N-i)^2/N^2, & \text{falls } j = i+1, \\ 2i(N-i)/N^2, & \text{falls } j = i, \\ i^2/N^2, & \text{falls } j = i-1, \\ 0, & \text{sonst.} \end{cases}$$

M ist damit offenkundig irreduzibel, aperiodisch und positiv rekurrent, also ergodisch und auch reversibel. Als stationäre Verteilung erhält man vermöge (2.93) und (2.94)

$$\pi_i = \frac{\binom{N}{i} \binom{N}{i}}{\binom{2N}{N}}, \quad (2.98)$$

d.h. $\pi = H(2N, N, N)$, die hypergeometrische Verteilung, die sich auch beim Ziehen von N Kugeln ohne Zurücklegen aus einer Urne mit $2N$ Kugeln, je zur Hälfte rot und schwarz, ergibt.

Wir hatten bereits in Unterabschnitt 2.1.7 recht ausgiebig eine andere Klasse von Urnenmodellen diskutiert, die in der Populationsgenetik von Bedeutung sind. Wir kehren nochmals kurz zu diesen zurück:

2.9.3 Markov-Ketten in der Genetik

[☞ 2.1.7] Die Modelle von Wright-Fisher und Moran dienen der Beschreibung der Evolution einer endlichen Population von $2N$ Genen, die in zwei Typen a und A , genannt Allele, auftreten können. Sofern keine Mutations- und Selektionseffekte auftreten, lauten die Übergangswahrscheinlichkeiten der jeweiligen EMK

$$p_{ij} = \binom{2N}{j} \alpha_i^j \beta_i^{2N-j} \quad \text{bzw.} \quad p_{ij} = \begin{cases} \alpha_i \beta_i, & \text{falls } j = i \pm 1, \\ \alpha_i^2 + \beta_i^2 & \text{falls } i = j, \\ 0, & \text{sonst.} \end{cases}$$

wobei $\alpha_i = 1 - \beta_i = \frac{i}{2N}$ für $i = 0, \dots, 2N$. Die Zustände 0 und $2N$ sind absorbierend, insbesondere rekurrent, die übrigen transient, wie man sofort einsieht. Damit tritt, unabhängig von der Anfangsverteilung, Fixierung ein, und es stellt sich nur noch Frage, mit welcher Wahrscheinlichkeit sich das eine oder das andere Allel durchsetzt. Zur Berechnung der Absorptionswahrscheinlichkeiten $u_{i,0}$ und $u_{i,2N} = 1 - u_{i,0}$ unter \mathbb{P}_i könnte man natürlich auf die Ergebnisse in Abschnitt 2.7 zurückgreifen, was allerdings einigen Rechenaufwand erforderte. Stattdessen verwenden wir eine andere, wesentlich einfachere Methode, die darauf beruht, dass in beiden Modellen $\mathbb{E}(M_n | M_{n-1} = i) = i$, also insbesondere $\mathbb{E}_i M_n = i$ für alle $i = 0, \dots, 2N$ und $n \geq 1$ gilt (dies bedeutet, dass $(M_n)_{n \geq 0}$ ein sogenanntes Martingal bildet, ein Prozesstyp, den wir in Kapitel 4 eingehend studieren werden). Aufgrund der f.s. Absorption folgt außerdem $M_n \rightarrow M_\infty$ \mathbb{P}_i -f.s. für alle $i = 0, \dots, 2N$, wobei M_∞ nur die Werte 0 und $2N$ mit Wahrscheinlichkeit $u_{i,0}$ bzw. $u_{i,2N}$ annimmt. Kombiniert man dies, ergibt sich mit dem Satz von der majorisierten Konvergenz (\mathcal{S} endlich)

$$i = \lim_{n \rightarrow \infty} \mathbb{E}_i M_n = \mathbb{E}_i M_\infty = (1 - u_{i,0})2N,$$

und somit

$$u_{i,0} = 1 - \frac{i}{2N} \quad \text{und} \quad u_{i,2N} = \frac{i}{2N}. \quad (2.99)$$

Es liegt der Versuch nahe, dieselbe Methode auch im Wright-Fisher-Modell mit Selektionsdruck zu verwenden (α_i, β_i gemäß (2.6)), was aber daran scheitert, dass man $\mathbb{E}_i M_n$ nicht in verwertbarer Form berechnen kann.

2.9.4 Ein Bedienungssystem mit konstanten Bedienungszeiten

[☞ 2.1.11] Wir fassen noch einmal die wesentlichen Fakten des in 2.1.11 vorgestellten Bedienungssystems zusammen: Kunden, die das System betreten, reihen sich in die Warteschlange ein, wenn der Bedienungsschalter besetzt ist. Der Server arbeitet in einem festen Zeittakt; das soll heißen, die Zeitachse ist in Intervalle (Bedienungsperioden) der Länge 1 eingeteilt, wobei am Anfang jeder Periode der erste Kunde in der Schlange (sofern das System nicht leer ist) an den Schalter tritt und während dieser Periode bedient wird. Die Bedienungszeit pro Kunde beträgt also stets eine Zeiteinheit, und der Arbeitsmodus ist “first in first out”. Für $n \geq 1$ bezeichne M_n die Anzahl wartender Kunden am Ende und X_n die Anzahl ankommender Kunden während der n -ten Bedienungsperiode. Wir nehmen an, dass $(X_n)_{n \geq 1}$ eine unabhängige Folge identisch gemäß $\xi = (p_k)_{k \geq 0}$ verteilter Zufallsgrößen bildet. $(X_n)_{n \geq 1}$ sei außerdem unabhängig von M_0 , dem Anfangszustand des Systems. Wir hatten bereits festgestellt, dass $(M_n)_{n \geq 0}$ eine DMK bildet mit Zustandsraum \mathbb{N}_0 und Übergangswahrscheinlichkeiten

$$p_{ij} = \mathbb{P}(M_{n+1} = j | M_n = i) = \mathbb{P}(X_{n+1} = j - (i - 1)^+) = p_{j - (i - 1)^+} \quad (2.100)$$

für alle $j + 1 \geq i \geq 0$. Wir wollen im Folgenden die intuitiv naheliegende Vermutung beweisen, dass $(M_n)_{n \geq 0}$ eine stationäre Verteilung besitzt, also asymptotisch stabil ist, wenn die mittlere Anzahl der pro Bedienungsperiode eintreffenden Kunden $\mu := \sum_{k \geq 1} k p_k$ kleiner als 1 ist. Um Trivialfälle auszuschließen, setzen wir $0 < p_0 \leq p_0 + p_1 < 1$ voraus.

Wie der Leser leicht nachprüft, ist die Kette irreduzibel und aperiodisch ($p_{00} = p_0 > 0$). Ergodizität liegt also genau dann vor, wenn es ein endliches stationäres Maß π gibt. Angesichts der wenigen Voraussetzungen an die Verteilung $(p_j)_{j \geq 0}$ wird es uns zwar unmöglich sein, π vollständig zu bestimmen, dennoch lassen sich mittels erzeugender Funktionen eine Reihe wichtiger Informationen gewinnen. Nehmen wir an, dass eine Lösung π der Invarianzgleichungen mit $\pi_0 > 0$ existiert und setzen

$$\Pi(s) = \sum_{i \geq 0} \pi_i s^i \quad \text{und} \quad F(s) = \sum_{i \geq 0} p_i s^i, \quad s \in [0, R \wedge 1),$$

wobei wir ferner unterstellen, dass der Konvergenzradius R von Π positiv ist. Vermöge der Invarianzgleichungen sowie (2.100) ergibt sich zwischen diesen Funktionen die Beziehung

$$\begin{aligned}
\Pi(s) &= \sum_{j \geq 0} \left(\sum_{i=0}^{j+1} \pi_i p_{ij} \right) s^j \\
&= \sum_{j \geq 0} \sum_{i=1}^{j+1} \pi_i p_{j-i+1} s^j + \pi_0 \sum_{j \geq 0} p_j s^j \\
&= \sum_{i \geq 1} \pi_i s^{i-1} \sum_{j \geq i-1} p_{j-i+1} s^{j-i+1} + \pi_0 F(s) \\
&= s^{-1} (\Pi(s) - \pi_0) F(s) + \pi_0 F(s)
\end{aligned}$$

und daraus für $s \in [0, R \wedge 1)$

$$\Pi(s) = \frac{\pi_0(1-s)F(s)}{F(s)-s} = \frac{\pi_0 F(s)}{1 - \frac{1-F(s)}{1-s}} = \pi_0 F(s) \sum_{n \geq 0} \left(\frac{1-F(s)}{1-s} \right)^n.$$

Da $\frac{1-F(s)}{1-s} = \sum_{n \geq 0} (\sum_{j > n} p_j) s^n$, wie man leicht nachrechnet, und $\lim_{s \uparrow 1} \frac{1-F(s)}{1-s} = F'(1-) = \mu$, folgen $R = \sup\{s : \frac{1-F(s)}{1-s} < 1\} \geq 1$, falls $\mu \leq 1$, sowie $\pi \geq 0$,

$$\Pi(1-) = \frac{\pi_0}{1-\mu}$$

und schließlich die positive Rekurrenz genau dann, wenn $\mu < 1$ gilt. Für die stationäre Verteilung π^* mit erzeugender Funktion $\Pi^*(s)$ ergibt sich mittels Normierung ($\Pi^*(1-) = 1$)

$$\pi_0^* = 1 - \mu \quad \text{und} \quad \Pi^*(s) = \frac{(1-\mu)(1-s)F(s)}{F(s)-s}, \quad s \in [0, 1).$$

π_0^* gibt zum einen die Wahrscheinlichkeit im Gleichgewicht an, mit der der Bediener in einer Periode unbeschäftigt ist (Leerperiode), zum anderen wegen

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \mathbf{1}_{\{M_k=0\}} = \pi_0^* \quad \mathbb{P}_\lambda\text{-f.s.}$$

für alle $\lambda \in \mathcal{P}(\mathbb{N}_0)$ (Satz (2.33)) auch den langfristigen Anteil von Leerperioden.

Eine weitere wichtige Größe ist die mittlere Anzahl wartender Kunden im Gleichgewicht, gegeben durch den Erwartungswert $\sum_{k \geq 0} k \pi_k^* = \Pi^{*'}(1-)$ von π^* . Statt jedoch die Ableitung von Π^* zu betrachten, greifen wir auf eine andere Charakterisierung der stationären Verteilung zurück, die sich aus der Beziehung (2.8), nämlich

$$M_{n+1} = (M_n - 1)^+ + X_{n+1}$$

ergibt. Da M_n und X_{n+1} unter jedem \mathbb{P}_λ unabhängig sind, gilt offenbar unter \mathbb{P}_{π^*}

$$M \stackrel{d}{=} (M - 1)^+ + X, \quad (2.101)$$

wobei $(M, X) \stackrel{d}{=} \pi^* \otimes \xi$. Eine solche Beziehung nennt man *stochastische Fixpunktgleichung*. Bilden wir auf beiden Seiten den Erwartungswert, erhalten wir

$$\mathbb{E}M = \mathbb{E}(M-1)^+ + \mu = \mathbb{E}(M-1)\mathbf{1}_{\{M>0\}} + \mu = \mathbb{E}M - \mathbb{P}(M > 0) + \mu$$

und somit die schon bekannte Beziehung

$$1 - \pi_0^* = \mathbb{P}(M > 0) = \mu.$$

Um an den Wert von $\mathbb{E}M$ heranzukommen, betrachten wir als nächstes das zweite Moment unter Voraussetzung von $\mu_2 := \mathbb{E}X^2 = \sum_{k \geq 1} k^2 p_k < \infty$. Es ergibt sich vermöge (2.101)

$$\begin{aligned} \mathbb{E}M^2 &= \mathbb{E}M^2 + \mathbb{P}(M > 0) - 2\mathbb{E}M + \mathbb{E}X^2 + 2\mathbb{E}X(\mathbb{E}M - \mathbb{E}X) \\ &= \mathbb{E}M^2 - 2(1 - \mu)\mathbb{E}M + \mu_2 + \mu(1 - 2\mu) \end{aligned}$$

und somit

$$\sum_{k \geq 1} k \pi_k^* = \mathbb{E}M = \frac{\mu(1 - 2\mu) + \mu_2}{2(1 - \mu)} = \frac{\mu}{2} + \frac{\sigma^2}{2(1 - \mu)},$$

wobei $\sigma^2 := \mu_2 - \mu^2 = \text{Var}X$. Man kann ferner zeigen, dass im Fall $\mu_2 = \infty$ auch $\mathbb{E}M = \infty$ gilt. Beachten wir, dass ein Kunde, der zu Beginn der n -ten Bedienungsperiode das System betritt, genau M_{n-1} Kunden vor sich in der Schlange antrifft, so folgt, dass $\mathbb{E}M + 1 = \sum_{k \geq 0} (k+1)\pi_k^*$ gerade seine Wartezeit (inklusive der eigenen Bedienungszeit) im Gleichgewicht angibt.

Abschließend wollen wir uns der Frage nach dem Systemverhalten in den Fällen “ $\mu = 1$ ” und “ $\mu > 1$ ” widmen, wobei wir erneut auf die rekursive Gleichung (2.8) zurückgreifen. Wegen der Irreduzibilität von $(M_n)_{n \geq 0}$ genügt es, den Zustand 0 zu untersuchen. Bezeichnet $\tau(0)$ wie üblich den Zeitpunkt des ersten Eintritts (≥ 1) in diesen Zustand, und ist $M_0 = 0$, so folgt wegen $M_n \geq 1$ für $1 \leq n < \tau(0)$

$$M_1 = X_1 \text{ auf } \{\tau(0) \geq 1\}, \quad M_2 = (M_1 - 1) + X_2 = X_1 + X_2 - 1 \text{ auf } \{\tau(0) \geq 2\}$$

sowie allgemein

$$M_n = \left(\sum_{k=1}^n X_k \right) - (n-1) = 1 + \sum_{k=1}^n (X_k - 1) \text{ auf } \{\tau(0) \geq n\}.$$

Damit erhalten wir für alle $n \geq 1$ (immer $M_0 = 0$ vorausgesetzt)

$$\{\tau(0) > n\} = \{M_1 \geq 1, \dots, M_n \geq 1\} = \left\{ \sum_{k=1}^n (X_k - 1) \geq 0, 1 \leq k \leq n \right\} = \{\sigma^< > n\},$$

wobei $\sigma^< := \inf\{n \geq 1 : \sum_{k=1}^n (X_k - 1) < 0\}$ den sogenannten *ersten streng absteigenden Leiterindex* des in 0 startenden diskreten Random Walks (☞ 2.1.10)

$(\sum_{k=1}^n (X_k - 1))_{n \geq 0}$ bezeichnet. Insbesondere gilt somit

$$f_{00}^* = \mathbb{P}_0(\tau(0) < \infty) = \mathbb{P}_0(\sigma^< < \infty) \quad \text{und} \quad \mu_{00} = \mathbb{E}_0 \tau(0) = \mathbb{E} \sigma^<.$$

Wir werden in [4, Satz 1.13] zeigen, dass

$$\mu - 1 = \mathbb{E}(X_1 - 1) \begin{cases} > 0 \\ = 0 \\ < 0 \end{cases} \Leftrightarrow \begin{cases} \mathbb{P}(\sigma^< < \infty) < 1, \\ \mathbb{P}(\sigma^< < \infty) = 1, \mathbb{E} \sigma^< = \infty \\ \mathbb{P}(\sigma^< < \infty) = 1, \mathbb{E} \sigma^< < \infty \end{cases} \quad (2.102)$$

Also ist 0 und damit auch $(M_n)_{n \geq 0}$ genau dann positiv rekurrent, null-rekurrent bzw. transient, wenn $\mu < 1$, $\mu = 1$ bzw. $\mu > 1$ gilt.

2.9.5 Das DM/DM/1-Bedienungssystem

Auch das folgende Modell stammt aus der Warteschlangentheorie. Stellen wir uns vor, Kunden betreten ein Bedienungssystem nur zu diskreten Zeitpunkten $0, 1, 2, \dots$. Eine solche Situation liegt beispielsweise in digitalen Telekommunikationsnetzen vor, in denen sogenannte Datenpäckchen (“packets”) die Kunden bilden und in festen Zeitschlitzen (“time slots”) übertragen werden. Es gebe einen Bedienungsschalter, vor dem sich eintreffende Kunden in die Schlange einreihen, wenn dieser besetzt ist. Sie werden der Reihe nach bedient (FIFO).

Seien $T_0 = 0$ der Beobachtungsstartpunkt, T_1, T_2, \dots die Ankunftszeiten der sukzessiv eintreffenden Kunden $1, 2, \dots$ mit zugehörigen Zwischenankunftszeiten $A_1 = T_1, A_2 = T_2 - T_1, \dots$ sowie B_1, B_2, \dots die zugehörigen Bedienzeiten. Ferner bezeichne M_0 die Anzahl der wartenden Kunden zum Zeitpunkt 0 (inklusive des gerade am Schalter stehenden), B_{-M_0+1}, \dots, B_0 deren Bedienzeiten und $V_0 = \sum_{k=1}^{M_0} B_{-M_0+k}$ die zugehörige *anstehende Arbeit*, gegeben durch die für sie insgesamt benötigte Bedienzeit. Es werden folgende Annahmen gemacht:

- (A1) $A_1 - 1, A_2 - 1, \dots$ sind stochastisch unabhängig und jeweils geometrisch verteilt mit Parameter $\alpha \in (0, 1)$, d.h. $\mathbb{P}(A_1 = n) = \alpha(1 - \alpha)^{n-1}$ für $n \in \mathbb{N}$.
- (A2) $B_n - 1, n \in \mathbb{Z}$, sind stochastisch unabhängig und jeweils geometrisch verteilt mit Parameter $\beta \in (0, 1)$, d.h. $\mathbb{P}(B_1 = n) = \beta(1 - \beta)^{n-1}$ für $n \in \mathbb{N}$.
- (A3) $(A_n)_{n \geq 1}, (B_n)_{n \geq 1}$ und $(M_0, V_0, (B_n)_{n \leq 0})$ sind stochastisch unabhängig.

Beachte, dass $\mathbb{E}A_1 = 1/\alpha$ und $\mathbb{E}B_1 = 1/\beta$. Als weitere Bezeichnungen führen wir ein ($n \in \mathbb{N}_0$):

- M_n Schlängellänge zum Zeitpunkt $n =$ Anzahl der wartenden Kunden zum Zeitpunkt n .
- W_n Wartezeit des n -ten Kunden, bis er zum Schalter vorgelassen wird.
- V_n Anstehende Arbeit zum Zeitpunkt n , auch *virtuelle Wartezeit* genannt, da sie im Fall eines zum Zeitpunkt n eintreffenden Kunden dessen Wartezeit angibt).

ρ Verkehrsintensität, gegeben durch $\mathbb{E}B_1/\mathbb{E}A_1 = \alpha/\beta$.

Betrachten wir die Folge $(M_n)_{n \geq 0}$ und argumentieren zunächst, dass diese eine zeitlich homogene DMK bildet. Befinden sich $i \geq 1$ Kunden zum Zeitpunkt n in der Schlange, so ist klar, dass M_{n+1} nur die Werte $i-1$, i und $i+1$ annehmen kann, und zwar gilt:

- $i \rightarrow i-1 \Leftrightarrow$ zum Zeitpunkt $n+1$ trifft kein neuer Kunde ein, und der am Schalter stehende Kunde verlässt das System.
- $i \rightarrow i+1 \Leftrightarrow$ zum Zeitpunkt $n+1$ trifft ein neuer Kunde ein, und der am Schalter stehende Kunde wird weiter bedient.
- $i \rightarrow i \Leftrightarrow$ zum Zeitpunkt $n+1$ trifft kein neuer Kunde ein, und der am Schalter stehende Kunde wird weiter bedient; oder es trifft ein neuer Kunde ein, und der am Schalter stehende Kunde verlässt das System.

Im Fall “ $i=0$ ” (System leer) entfällt offenkundig die Möglichkeit “ $i \rightarrow i-1$ ”.

Mit welchen bedingten Wahrscheinlichkeiten die drei Varianten eintreten, hängt ab von den bedingten Verteilungen der verbleibenden Wartezeit bis zur Ankunft des nächsten Kunden, $T(n)-n$, und der verbleibenden Bedienungszeit des zum Zeitpunkt n am Schalter stehenden Kunden, $B(n)-n$ ($:=0$, falls System leer), d.h. $B(n)$ gibt den Zeitpunkt an, zu dem dieser Kunde das System verlässt. Es gilt nämlich offenbar die rekursive Beziehung

$$M_{n+1} = M_n + \mathbf{1}_{\{T(n)-n=1\}} - \mathbf{1}_{\{B(n)-n=1\}}. \quad (2.103)$$

Die σ -Algebra \mathcal{F}_n der bis zum Zeitpunkt n beobachtbaren Ereignisse (verfügbare Information) wird erzeugt von (M_0, V_0) , den Ankunftszeiten aller Kunden im Intervall $(0, n]$, den Bedienungszeiten aller bis zum Zeitpunkt n vollständig bedienten Kunden sowie von der bereits verstrichenen Bedienungszeit des gerade am Schalter stehenden Kunden, vorausgesetzt $M_n \geq 1$. Leider lässt sich dies formal nicht so einfach präzisieren, wie der Leser bei einem Versuch sehr schnell merken wird. Nichtsdestotrotz ist aufgrund der Unabhängigkeitsannahmen intuitiv klar, dass die Variablen $T(n)-n$ und $B(n)-n$ bedingt unter \mathcal{F}_n wieder unabhängig sind mit derselben Verteilung wie A_1 bzw. B_1 (falls das System nicht leer ist), letzteres wegen der *Gedächtnislosigkeit* der geometrischen Verteilung (\mathbb{E} [2, Satz 38.4]).

Der formale Nachweis dieser intuitiv offenkundigen Tatsache bereitet Studenten erfahrungsgemäß gewisse Schwierigkeiten. Wir fügen ihn hier deshalb für die $T(n)-n$ ein und notieren als erstes, dass

$$T(n)-n = T_{\tau(n)} - n = \sum_{m \geq 1} \mathbf{1}_{\{\tau(n)=m\}} (T_m - n),$$

wobei $\tau(n) := \inf\{m \geq 1 : T_m > n\}$. Da Ankunftsprozess und Bedienungszeiten nach Voraussetzung unabhängig sind und da $T(n)-n = A_{\tau(n)} - (n - T_{\tau(n)-1})$, folgt nun unter Beachtung der \mathcal{F}_n -Messbarkeit von $n - T_{\tau(n)-1}$

$$\mathbb{P}(T(n)-n > k | \tau(n), \mathcal{F}_n) = \mathbb{P}(A_{\tau(n)} > k + (n - T_{\tau(n)-1}) | \tau(n), n - T_{\tau(n)-1})$$

und dann weiter

$$\begin{aligned}
\mathbb{P}(A_{\tau(n)} \geq k + (n - T_{\tau(n)-1}) | \tau(n) = m, n - T_{\tau(n)-1} = l) \\
&= \mathbb{P}(A_m \geq k + l | A_m > l, n - T_{m-1} = l) \\
&= \mathbb{P}(A_m \geq k + l | A_m > l) \\
&= \mathbb{P}(A_m \geq k) = \mathbb{P}(A_1 \geq k)
\end{aligned}$$

für alle $m \geq 1, l = 0, \dots, n$ und $k \geq 0$, wobei die letzte Zeile aufgrund der besagten Gedächtnislosigkeit der geometrischen Verteilung gilt. $T(n) - n$ ist also unabhängig sowohl von \mathcal{F}_n als auch $\tau(n)$ und besitzt die behauptete Verteilung.

Mittels (2.103) folgt nun leicht, dass $(M_n)_{n \geq 0}$ bezüglich $(\mathcal{F}_n)_{n \geq 0}$ einen Geburts- und Todesprozess bildet mit der Übergangsmatrix

$$P = \begin{pmatrix} 1 - \alpha & \alpha & 0 & 0 & \dots \\ \beta(1 - \alpha) & \alpha\beta + (1 - \alpha)(1 - \beta) & \alpha(1 - \beta) & 0 & \dots \\ 0 & \beta(1 - \alpha) & \alpha\beta + (1 - \alpha)(1 - \beta) & \alpha(1 - \beta) & \dots \\ \vdots & \vdots & \ddots & \vdots & \dots \end{pmatrix}.$$

Eine andere Herleitung basiert auf der folgenden Beschreibung der Gedächtnislosigkeit: Die Annahme geometrisch verteilter (mit Parameter α) Zwischenankunftszeiten lässt sich auch dadurch charakterisieren, dass in jedem Zeitschlitz unabhängig vom vorherigen Geschehen mit Wahrscheinlichkeit α ein Kunde das System betritt. Entsprechend implizieren geometrisch verteilte (mit Parameter β) Bedienungszeiten in Kombination mit (A3), dass in jedem Zeitschlitz, wiederum unabhängig vom bisherigen Geschehen sowie davon, ob ein neuer Kunde eintrifft oder nicht, die Bedienung eines Kunden mit Wahrscheinlichkeit β abgeschlossen wird, vorausgesetzt natürlich, das System ist nicht leer. Wir gelangen so erneut zur Identität (2.104) mit bedingt unter \mathcal{F}_n unabhängigen Bernoulli-Variablen $Y_{n+1} := \mathbf{1}_{\{T(n)-n=1\}}$ und $Z_{n+1} := \mathbf{1}_{\{B(n)-n=1\}}$ für $n \geq 0$, wobei

$$\mathbb{P}^{Y_{n+1} | \mathcal{F}_n} = \text{Bin}(1, \alpha) \quad \text{und} \quad \mathbb{P}^{Z_{n+1} | \mathcal{F}_n} = \mathbf{1}_{\{M_n=0\}} \delta_0 + \mathbf{1}_{\{M_n \geq 1\}} \text{Bin}(1, \beta).$$

Wegen $0 < \alpha, \beta < 1$ ist $(M_n)_{n \geq 0}$ offensichtlich irreduzibel und aperiodisch und nach unseren Überlegungen in Beispiel 2.77 genau dann ergodisch, wenn

$$\rho > \sum_{i \geq 1} \frac{p_{01} p_{12} \cdots p_{i-1,i}}{p_{10} p_{21} \cdots p_{i,i-1}} = \frac{1}{1 - \beta} \sum_{i \geq 1} \left(\frac{\alpha(1 - \beta)}{\beta(1 - \alpha)} \right)^i,$$

also $\rho < 1$ gilt. Wie der Leser mittels (2.94) leicht selbst nachrechnen kann, bildet

$$\pi_0 = 1 - \rho \quad \text{und} \quad \pi_i = \frac{1 - \rho}{1 - \beta} \left(\frac{1 - \beta}{1 - \rho\beta} \right)^i \rho^i, \quad i \geq 1, \quad (2.104)$$

in diesem Fall die eindeutig bestimmte stationäre Verteilung, und $(M_n)_{n \geq 0}$ ist unter \mathbb{P}_π reversibel.

Betrachten wir als nächstes die Fälle “ $\rho = 1$ ” und “ $\rho > 1$ ”, wobei wir auf die Analogie der folgenden Argumentation zu der im vorherigen Beispiel hinweisen. Aufgrund der Irreduzibilität liegt Transienz genau dann vor, wenn

$$1 - f_{10}^* = \mathbb{P}_1(M_n \geq 1 \text{ für alle } n \geq 0) > 0.$$

Wir setzen $U_n = B_0 + \dots + B_{n-1}$ für $n \geq 1$. Dann gilt

$$1 - f_{10}^* = \mathbb{P}_1(T_1 \leq B_0, T_2 \leq B_0 + B_1, \dots) = \mathbb{P}_1(U_n - T_n \geq 0 \text{ für alle } n \geq 1).$$

$(U_n - T_n)_{n \geq 1}$ bildet einen diskreten Random Walk mit den unabhängigen, identisch verteilten Zuwächsen $B_{n-1} - A_n$, die im Fall “ $\rho > 1$ ” positiven Erwartungswert besitzen, im Fall “ $\rho = 1$ ” dagegen symmetrisch sind mit Erwartungswert 0 und endlicher positiver Varianz. Nach [2, Satz 38.1] folgt deshalb in letzterem Fall

$$\mathbb{P}_1 \left(\liminf_{n \rightarrow \infty} \frac{U_n - T_n}{n^{1/2}} = -\infty \right) = 1$$

und damit weiter

$$1 - f_{10}^* = \mathbb{P}_1(U_n - T_n \geq 0 \text{ für alle } n \geq 1) \leq \mathbb{P}_1 \left(\liminf_{n \rightarrow \infty} \frac{U_n - T_n}{n^{1/2}} \geq 0 \right) = 0,$$

d.h. die Null-Rekurrenz von $(M_n)_{n \geq 0}$. Falls $\rho > 1$, ergibt sich dagegen in der Tat Transienz, denn mit $\sigma^< := \inf\{n \geq 1 : U_n - T_n < 0\}$ folgt unter Hinweis auf (2.102) [ersetze dort $\mathbb{E}X_1 - 1$ durch $\mathbb{E}(B_0 - A_1)$]

$$1 - f_{10}^* = \mathbb{P}(\sigma^< = \infty) > 0.$$

Auch die Folgen $(W_n)_{n \geq 0}$ und $(V_n)_{n \geq 0}$ bilden DMK bezüglich der eingangs angegebenen Filtration $(\mathcal{F}_n)_{n \geq 0}$, denn

$$W_{n+1} = (W_n + B_n - A_{n+1})^+ \quad \text{und} \quad V_{n+1} = (V_n - 1)^+ + Y_{n+1} B_{\tau(n)} \quad (2.105)$$

für alle $n \geq 0$, wie man leicht einsieht, und sowohl $B_n - A_{n+1}$ als auch $Y_{n+1} B_{\tau(n)}$ sind unabhängig von \mathcal{F}_n mit einer von n unabhängigen Verteilung. Betreffend $Y_{n+1} B_{\tau(n)} = (T_{\tau(n)} - n) B_{\tau(n)}$ sollte der Leser allerdings selbst nachprüfen, dass in der Tat

$$(Y_{n+1}, B_{\tau(n)}) \stackrel{d}{=} \text{Bin}(1, \alpha) \otimes \mathbb{P}^{B_1}$$

gilt. $(W_n)_{n \geq 0}$ und $(V_n)_{n \geq 0}$ sind ebenso wie $(M_n)_{n \geq 0}$ irreduzibel und aperiodisch.

Aufgrund der konkreten Bedeutung der Variablen W_n und V_n ist klar, dass diese nur dann gegen eine Gleichgewichtsverteilung konvergieren können, wenn dies für M_n gilt, wenn also $\rho < 1$. Wir betrachten im Folgenden nur die Kette $(W_n)_{n \geq 0}$ und setzen $X_n = B_{n-1} - A_n$ sowie $S_n = X_1 + \dots + X_n$ für $n \geq 1$. Sei $S_0 = 0$. Gemäß (2.105) gilt dann

$$W_{n+1} = (W_n + X_{n+1})^+ = \max\{W_n + X_{n+1}, 0\} \quad (2.106)$$

für alle $n \geq 0$. Vorausgesetzt $W_0 = 0$, liefert eine einfache Induktion über n

$$W_n = \max\{S_n - S_j : 0 \leq j \leq n\}$$

und folglich

$$W_n \stackrel{d}{=} \max\{S_0, \dots, S_n\} \quad (2.107)$$

für alle $n \geq 0$, weil die X_n unabhängig und identisch verteilt sind. Allgemein nennt man eine Folge $(W_n)_{n \geq 0}$, die zu gegebenem Random Walk $(S_n)_{n \geq 0}$ die Beziehung (2.106) erfüllt, den von $(S_n)_{n \geq 0}$ erzeugten *Lindley-Prozess*.

Gilt nun $\rho < 1$, d.h. $\mathbb{E}X_1 < 0$, folgt $S_n \rightarrow -\infty$ f.s. und somit

$$\max\{S_0, \dots, S_n\} \uparrow \max_{k \geq 0} S_k < \infty \quad \mathbb{P}\text{-f.s.}$$

Unter Verwendung von (2.107) zeigt dies

$$W_n \xrightarrow{d} \max_{k \geq 0} S_k, \quad (2.108)$$

also die erwartete Ergodizität der Kette mit stationärer Verteilung $\mathbb{P}(\max_{k \geq 0} S_k \in \cdot)$. Bezeichnet W irgendeine Zufallsgröße mit dieser Verteilung und X eine von W unabhängige Kopie von X_1 , so folgt aus (2.106) unter Hinweis auf die Unabhängigkeit von W_n und X_{n+1} sowie Benutzung [2, Satz 36.11] die sogenannte *Lindley-Gleichung*

$$W \stackrel{d}{=} (W + X)^+.$$

Falls $\rho = 1$ bzw. > 1 , ergibt sich, ebenfalls in Analogie zu $(M_n)_{n \geq 0}$, die Null-Rekurrenz bzw. Transienz von $(W_n)_{n \geq 0}$. Wir wollen darauf jedoch nicht weiter eingehen, sondern unsere Untersuchung an dieser Stelle beenden.

Ein Wort noch zur Bezeichnung “*DM/DM/1*”, die auf der sogenannten *Kendall-schen Notation* zur Kurzcharakterisierung von Bedienungssystemen basiert. Das erste Symbol beschreibt den Typ des Ankunftsprozesses, das zweite den der Bedienzeiten. Die Zahl an der dritten Position gibt die Anzahl der Bedienungsschalter an. “*DM*” steht für “discrete Markovian”. Bezeichnet nämlich N_n die Anzahl der erschienenen Kunden bis zum Zeitpunkt n , so bildet $(N_n)_{n \geq 0}$ genau dann eine Markov-Kette, wenn die Zwischenankunftszeiten geometrisch verteilt sind (bei Verschiebung um -1). Der Grund: Die unter \mathcal{F}_n bedingte Wahrscheinlichkeit dafür, ob zum Zeitpunkt $n+1$ ein Kunde das System betritt ($N_{n+1} = N_n + 1$) oder nicht ($N_{n+1} = N_n$), hängt i.A. von der bereits verstrichenen Zeit seit der letzten Ankunft eines Kunden ab, die offenbar eine Funktion von N_0, \dots, N_n bildet, nicht aber von N_n allein. Nur wenn diese verstrichene Wartezeit irrelevant ist, d.h., im Fall gedächtnisloser Zwischenankunftszeiten, führt das Bedingen unter N_n zur selben Wahrscheinlichkeit wie das Bedingen unter \mathcal{F}_n oder N_0, \dots, N_n . “*DM*” an der ersten Stelle steht also für geometrisch verteilte Zwischenankunftszeiten, entsprechend an der zweiten Stelle für geometrisch verteilte Bedienzeiten. Wir werden später als Pendant in stetiger Zeit das *M/M/1*-Bedienungssystem mit exponentialverteilt-

ten Zwischenankunfts- und Bedienungszeiten kennenlernen. Dabei bedeutet “ M ” gerade “Markovian”.

Der in den beiden letzten Beispielen zu Tage getretene Zusammenhang zwischen der Stabilität der betrachteten Bedienungssysteme und dem Verhalten bestimmter eingebetteter Random Walks mit unabhängigen identisch verteilten Zuwächsen ist kein Zufall, sondern in vielen Anwendungsbeispielen von entscheidender Bedeutung, so auch im nächsten. Ein Blick auf die noch unbewiesene Aussage (2.102) zeigt andererseits, dass es sich hier um Eigenschaften handelt, die nicht aus den schon bekannten, klassischen Grenzwertsätzen folgen. Kapitel 6 befasst sich deshalb eingehender mit derartigen Fragestellungen, die gewisse zyklische Verhaltensmuster von Random Walks betreffen und im weiteren Sinne in den Bereich der *Erneuerungstheorie* fallen.

2.9.6 Ein Lagerhaltungsmodell

[☞ 2.1.12] Wir betrachten ein Lager für irgendein laufend nachgefragtes Gut mit Kapazität $S \in \mathbb{N}$ und Anfangsbestand M_0 . Die pro Periode eingehenden Bestellungen X_1, X_2, \dots seien unabhängig und identisch gemäß $(p_k)_{k \geq 0}$ verteilt ($p_0 < 1$) sowie unabhängig von M_0 . Fällt der Bestand am Ende einer Periode auf ein Niveau $< s < S$, wird das Lager sofort wieder aufgefüllt, d.h., auf den Bestand S aufgestockt. Wie bereits in 2.1.12 festgestellt wurde, ergibt sich der Lagerbestand M_n am Ende der n -ten Periode dann zu

$$M_n = \begin{cases} M_{n-1} - X_n, & \text{falls } s \leq M_{n-1} \leq S \\ S - X_n, & \text{falls } M_{n-1} < s \end{cases}$$

für alle $n \geq 1$ und bildet eine DMK mit Zustandsraum $\{S, S-1, \dots, 1, 0, -1, \dots\}$ und Übergangswahrscheinlichkeiten

$$p_{ij} = \begin{cases} \mathbb{P}(X_{n+1} = i - j) = p_{i-j}, & \text{falls } s \leq M_n \leq S \\ \mathbb{P}(X_{n+1} = S - j) = p_{S-j}, & \text{falls } M_n < s \end{cases}.$$

Um die nachfolgende Diskussion zu vereinfachen, setzen wir voraus, dass sämtliche p_k positiv sind, was offenkundig die Irreduzibilität und Aperiodizität von $(M_n)_{n \geq 0}$ nach sich zieht.

Liegt der Lagerbestand in aufeinanderfolgenden Perioden $n, \dots, n+k$ stets zwischen s und S , verhält sich M_n wie ein Random Walk, denn es gilt dann offenkundig

$$\begin{aligned}
M_{n+1} &= M_n - X_{n+1}, \\
M_{n+2} &= M_n - X_{n+1} - X_{n+2}, \\
&\vdots \\
M_{n+k} &= M_n - \sum_{j=1}^k X_{n+j}.
\end{aligned}$$

Wegen $\mathbb{E}X_1 > 0$ fällt der Bestand aber stets nach endlicher Zeit unter die kritische Marke und führt zu einer Aufstockung (Beginn eines neuen Lagerzyklus'). Damit sollte intuitiv bereits klar sein, dass $(M_n)_{n \geq 0}$ in jedem Fall rekurrent ist. Zur Untermauerung definieren wir $T_0 = 0$ und

$$T_n := \inf\{k > T_{n-1} : M_{k-1} < s\}$$

für $n \geq 1$. Jedes T_n markiert also den Beginn eines neuen Lagerzyklus', und es gilt außerdem

$$M_{T_n} = S - X_{T_n} \quad (2.109)$$

für alle $n \geq 1$. Aus [4, Satz 1.9], den wir an dieser Stelle ohne Beweis voraussetzen, folgt weiter:

- (i) Für jedes $n \geq 1$ bildet $X_{T_n}, X_{T_n+1}, \dots$ eine Folge unabhängiger und identisch wie X_1 verteilter Zufallsgrößen.
- (ii) Die Segmente $(X_{T_n}, \dots, X_{T_{n+1}-1})$, $n \geq 0$, sind unabhängig und für $n \geq 1$ identisch verteilt, woraus insbesondere folgt:
- (iii) $T_{n+1} - T_n$, $n \geq 0$, sind unabhängig und für $n \geq 1$ identisch verteilt.

Wegen (ii) und (2.109) sind natürlich auch die M_{T_n} , $n \geq 1$, unabhängig und identisch verteilt, wobei

$$\mathbb{P}(M_{T_n} = S - k) = p_k \quad (2.110)$$

für alle $n \geq 1$ und $k \geq 0$. Unter Hinweis auf $p_k > 0$ liefert das Lemma von Borel-Cantelli (Lemma 34.4 in [2])

$$\mathbb{P}(M_{T_n} = S - k \text{ u.o.}) = 1$$

für alle $k \geq 0$ und folglich die Rekurrenz der Kette $(M_n)_{n \geq 0}$. Dass sie sogar ergodisch ist, sieht man wie folgt:

Für $k = 0, \dots, T_{n+1} - T_n - 1$, $n \geq 1$, gilt gemäß (2.109) und (2.110)

$$M_{T_n+k} = S - \sum_{j=0}^k X_{T_n+j},$$

so dass

$$T_{n+1} - T_n - 1 = \inf \left\{ k \geq 0 : S - \sum_{j=0}^k X_{T_n+j} < \right\}$$

$$= \inf \left\{ k \geq 0 : \sum_{j=0}^k X_{T_n+j} > S - s \right\}$$

für alle $n \geq 1$. Aufgrund dieser Darstellung der $T_{n+1} - T_n - 1$ als Erstaustrittszeiten für die Partialsummen der unabhängigen und identisch wie X_1 verteilten $X_{T_n}, X_{T_n+1}, \dots$ (☞ (1)) erhalten wir gemäß Lemma 2.32 $\mu := \mathbb{E}(T_{n+1} - T_n) < \infty$ ($n \geq 1$). Damit liegt nun folgende Situation vor: $(M_n)_{n \geq 0}$ zerfällt in die unabhängigen Zyklen

$$Z_n = (M_{T_n}, \dots, M_{T_{n+1}-1}), \quad n \geq 0,$$

die für $n \geq 1$ ferner identisch verteilt sind mit mittlerer Länge μ . Im Unterschied zu denen in Unterabschnitt 2.2.3, die sich vermöge Rückkehr in einen rekurrenten Zustand ergaben, hat hier M_{T_n} , die Variable am Zyklusanfang, keine Einpunktverteilung, sondern die Verteilung $\zeta := (p_{S-k})_{k \leq S}$ auf $\{S, S-1, \dots\}$ (☞ (2.110)). Wir überlassen es dem Leser zu zeigen, dass nichtsdestotrotz durch

$$\pi_j = \frac{1}{\mu} \mathbb{E}_\zeta \left(\sum_{n=0}^{T_1-1} \mathbf{1}_{\{M_n=j\}} \right), \quad j \in \{S, \dots, 1, 0, -1, \dots\},$$

die eindeutig bestimmte stationäre Verteilung definiert wird. Er kann hierzu etwa den Invarianzbeweis in Satz 2.30 durchgehen und überprüfen, dass dieser gültig bleibt.

Ein Nachfrageüberhang liegt vor, wenn am Ende der Periode der Lagerbestand kleiner als 0 ist. Folglich bezeichnet

$$\lim_{N \rightarrow \infty} \frac{1}{N+1} \mathbb{E}_i \left(\sum_{n=0}^N \mathbf{1}_{\{M_n < 0\}} \right) = C\text{-}\lim_{N \rightarrow \infty} \sum_{j < 0} p_{ij}^{(N)}$$

die langfristig mittlere Anzahl von Perioden mit Nachfrageüberhang unter \mathbb{P}_i . Da $(M_n)_{n \geq 0}$ ergodisch ist, folgt

$$C\text{-}\lim_{N \rightarrow \infty} \sum_{j < 0} p_{ij}^{(N)} = \lim_{N \rightarrow \infty} \sum_{j < 0} p_{ij}^{(N)} = \sum_{j < 0} \pi_j$$

für alle $i \in \{S, \dots, 1, 0, -1, \dots\}$. Entsprechend erhalten wir, dass

$$\lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N \mathbb{E}_i M_n = C\text{-}\lim_{N \rightarrow \infty} \mathbb{E}_i M_N = C\text{-}\lim_{N \rightarrow \infty} \sum_{j \leq S} j p_{ij}^{(N)}$$

den mittleren Lagerbestand unter \mathbb{P}_i im Zeitablauf angibt. Um jedoch aus $M_n \xrightarrow{d} \pi$ die naheliegende Aussage

$$C\text{-}\lim_{N \rightarrow \infty} \mathbb{E}_i M_N = \lim_{N \rightarrow \infty} \mathbb{E}_i M_N = \sum_{j \leq S} j \pi_j$$

zu schließen, bedarf es bekanntlich ([2, Satz 50.5]) des Nachweises der gleichgradigen Integrierbarkeit der M_n unter \mathbb{P}_i , wozu uns an dieser Stelle noch die Mittel fehlen. Wir verweisen auf [4, Unterabschnitt 1.10.4], wo auf der Grundlage von Erneuerungstheorie übrigens auch eine Aussage über die Gestalt der stationären Verteilung gewonnen wird.

2.9.7 Der Galton-Watson-Verzweigungsprozess

[2.1.13] Wir betrachten eine Population, in der Individuen derselben Generation unabhängig voneinander und von denen der vorangegangenen Generationen Nachkommen produzieren, und zwar stets gemäß derselben, sogenannten *Reproduktionsverteilung* $(p_k)_{k \geq 0}$ mit Erwartungswert μ , dem *Reproduktionsmittel*. Sei M_n die Anzahl der Mitglieder der n -ten Generation ($n \geq 0$). Dann gilt

$$M_{n+1} = \sum_{k=1}^{M_n} X_{n,k}, \quad n \geq 0, \quad (2.111)$$

wobei $X_{n,k}$ die Zahl der Kinder des k -ten Mitglieds der n -ten Generation bezeichnet. Der obigen Beschreibung zufolge bildet $(X_{n,k})_{n \geq 0, k \geq 1}$ eine Folge unabhängiger und identisch gemäß $(p_k)_{k \geq 0}$ verteilter Zufallsgrößen und $M = (M_n)_{n \geq 0}$ eine DMK, genannt *Galton-Watson-Verzweigungsprozess (GWP)*, mit Zustandsraum \mathbb{N}_0 und Übergangswahrscheinlichkeiten

$$p_{ij} = p_j^{*(i)},$$

$(p_k^{*(i)})_{k \geq 0}$ die i -fache Faltung von $(p_k)_{k \geq 0}$. Der Zustand 0 ist offenkundig absorbierend und gleichbedeutend mit dem Aussterben der Population.

Zwei Fragen, die sich ganz natürlich als erstes stellen, wollen wir im Folgenden nachgehen:

1. Mit welcher Wahrscheinlichkeit stirbt die Population aus?
2. Wie verhält sich die Population im Limes, wenn sie überlebt?

Dabei setzen wir, um Trivialfälle auszuschließen,

$$0 < p_0 \leq p_0 + p_1 < 1$$

voraus. Da sämtliche Individuen einer Generation, insbesondere also die Urahnen, unabhängig voneinander reproduzieren, folgt

$$\mathbb{P}_i^M = (\mathbb{P}_1^M)^{*(i)}$$

für alle $i \in \mathbb{N}_0$, d.h.:

$$M_n = \sum_{k=1}^{M_0} M_n^{(k)}, \quad n \geq 0,$$

wobei $M^{(1)} = (M_n^{(1)})_{n \geq 0}, M^{(2)}, \dots$ unter jedem \mathbb{P}_i stochastisch unabhängige GWP mit einem Urahen und Reproduktionsverteilung $(p_k)_{k \geq 0}$ bilden, die ferner unabhängig von M_0 sind. Bezeichnet nun q_i die Aussterbewahrscheinlichkeit von M unter \mathbb{P}_i und $q = q_1$, so folgt unter Beachtung der Antitonie der Ereignisse $\{M_n = 0\}$

$$\begin{aligned} q_i &= \mathbb{P}_i \left(\bigcup_{n \geq 0} \{M_n = 0\} \right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}_i(M_n = 0) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}_i(M_n^{(k)} = 0, 1 \leq k \leq i) \\ &= \lim_{n \rightarrow \infty} \prod_{k=1}^i \mathbb{P}_i(M_n^{(k)} = 0) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}_1(M_n = 0)^i = q^i. \end{aligned}$$

Es genügt also, hiernach die Aussterbewahrscheinlichkeit q bei einem Urahen zu untersuchen. Da wir $p_0 > 0$ voraussetzen, ist q in jedem Fall positiv, denn

$$q \geq \mathbb{P}_1(M_1 = 0) = \mathbb{P}_1(X_{0,1} = 0) = p_0.$$

Wir führen als nächstes erzeugende Funktionen ein und definieren

$$f(s) = \mathbb{E}_1 s^{M_1} = \sum_{n \geq 0} p_n s^n \quad \text{und} \quad f_n(s) = \mathbb{E}_1 s^{M_n}$$

für $s \in [0, 1]$ und $n \geq 0$, d.h. $f_0(s) = s$. Unter Benutzung von (2.111) erhalten wir

$$\begin{aligned} f_{n+1}(s) &= \sum_{j \geq 0} \mathbb{E}_1 \left(\prod_{k=1}^{M_n} s^{X_{n,k}} \mid M_n = j \right) \mathbb{P}_1(M_n = j) \\ &= \sum_{j \geq 0} \mathbb{E}_1 \left(\prod_{k=1}^j s^{X_{n,k}} \right) \mathbb{P}_1(M_n = j) \\ &= \sum_{j \geq 0} f(s)^j \mathbb{P}_1(M_n = j) \\ &= \mathbb{E}_1 f(s)^{M_n} = f_n(f(s)) \end{aligned}$$

und somit induktiv

$$f_n(s) = f \circ \dots \circ f(s) =: f^{\circ(n)}(s)$$

für alle $s \in [0, 1]$ und $n \geq 0$. Bezüglich der Aussterbewahrscheinlichkeit q ergibt sich nun unter Benutzung der Stetigkeit von f und von $\mathbb{P}_1(M_n = 0) = f_n(0)$

$$q = \lim_{n \rightarrow \infty} f_n(0) = \lim_{n \rightarrow \infty} f(f_{n-1}(0)) = f(\lim_{n \rightarrow \infty} f_{n-1}(0)) = f(q),$$

d.h. q ist ein Fixpunkt der erzeugenden Funktion f von $(p_k)_{k \geq 0}$. Wir präzisieren diese Aussage dahingehend, dass q sogar der *kleinste* Fixpunkt von f in $[0,1]$ ist. Aus $a = f(a)$ folgt nämlich $a = f_n(a)$ für alle $n \geq 0$ und dann

$$a = \lim_{n \rightarrow \infty} f_n(a) \geq \lim_{n \rightarrow \infty} f_n(0) = q \geq a.$$

Im Hinblick auf die erste obige Frage wollen wir nunmehr untersuchen, unter welchen Voraussetzungen $q = 1$ bzw. $q \in (0,1)$ gilt. Werfen wir dazu als erstes einen Blick auf das mittlere Wachstum der Population: Eine ähnliche Rechnung wie weiter oben für $f_{n+1}(s)$ liefert

$$\begin{aligned} \mathbb{E}_1 M_{n+1} &= \sum_{j \geq 0} \mathbb{E}_1 \left(\sum_{k=1}^j X_{n,k} \right) \mathbb{P}_1(M_n = j) \\ &= \sum_{j \geq 1} j \mu \mathbb{P}_1(M_n = j) \\ &= \mu \mathbb{E}_1 M_n \end{aligned}$$

und folglich

$$\mathbb{E}_1 M_n = \mu^n$$

für alle $n \geq 0$. Im Fall $\mu < 1$ strebt demnach die mittlere Populationsgröße exponentiell schnell gegen 0, und wir erhalten mittels der Markov-Ungleichung

$$\mathbb{P}_1(M_n \geq 1) \leq \mu^n \rightarrow 0 \quad (n \rightarrow \infty),$$

also $q = 1$. Dagegen lassen die Fälle $\mu > 1$ und $\mu = 1$, in denen die mittlere Populationsgröße exponentiell explodiert bzw. bei 1 stagniert, keinen entsprechenden Rückschluss auf den Wert von q zu. Stattdessen bedarf es eines genaueren Blicks auf die erzeugende Funktion f .

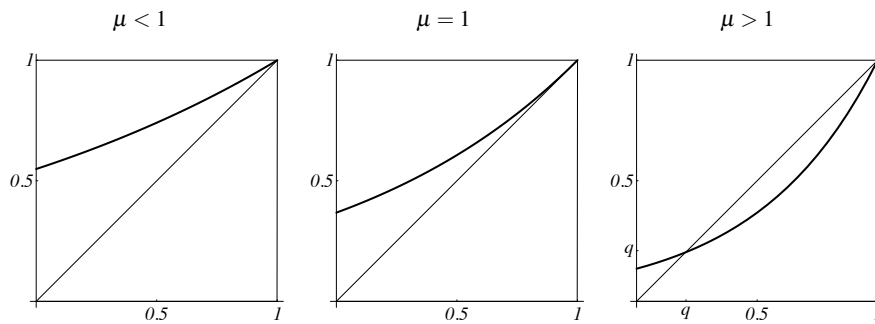


Abb. 2.12 Erzeugende Funktion im subkritischen, kritischen und superkritischen Fall.

Wir weisen zunächst darauf hin, dass f und sämtliche Ableitungen f', f'', \dots auf $[0, 1)$ nichtnegativ, monoton wachsend und strikt konvex sind. Ferner gilt $f'(1-) = \lim_{s \uparrow 1} f'(s) = \mu$. Im Fall “ $\mu = 1$ ” folgt daraus

$$f(s) = f(0) + \int_0^s f'(r) dr > \int_0^s f'(r) dr > s$$

für alle $s \in [0, 1)$, d.h. f besitzt nur den Fixpunkt $q = 1$ auf $[0, 1]$. Dagegen besitzt f im Fall $\mu > 1$ (und $p_0 > 0$) genau einen Fixpunkt in $(0, 1)$, d.h. $q \in (0, 1)$. Die Details kann sich der Leser leicht selbst überlegen, wobei zur Veranschaulichung auf Abb. 2.12 verwiesen sei.

Wir fassen zusammen: Falls $0 < p_0 \leq p_0 + p_1 < 1$, so gilt

$$\begin{array}{lll} \mu < 1 & \text{(subkritischer Fall)} & q = 1 \\ \mu = 1 & \text{(kritischer Fall)} & \Rightarrow q = 1 \\ \mu > 1 & \text{(superkritischer Fall)} & 0 < q < 1 \end{array} .$$

Im kritischen Fall stirbt die Population somit f.s. aus, obwohl die mittlere Populationsgröße stets 1 beträgt. Dies hat zur Konsequenz, dass $\mathbb{E}(M_n | M_n > 0)$ für $n \rightarrow \infty$ gegen ∞ strebt. Eine lange Zeit überlebende kritische Population wird demnach im Mittel sehr groß sein.

Ein Wort zur Berechnung von q im superkritischen Fall. Es gibt leider nur wenige bekannte Reproduktionsverteilungen, für die sich der kleinste Fixpunkt der erzeugenden Funktion explizit durch Lösen der Gleichung $f(s) = s$ berechnen lässt. Man scheitert bereits bei der Poisson-Verteilung ($f(s) = e^{\mu(s-1)}$). Die gute Nachricht aber lautet: Iteriert man f mit Startwert 0, berechnet also $f(0), f(f(0)), \dots$, so konvergiert diese Folge monoton und exponentiell gegen q . Dieses Verfahren lässt sich für jedes in geschlossener Form gegebene f auf jedem PC programmieren. Eine Ausnahme, für die q direkt berechnet werden kann, bildet der “Zellteilungsfall”: $p_n = 0$ für alle $n \geq 3$. Dann ist $f(s) = p_0 + p_1 s + p_2 s^2$ ein Polynom zweiten Grades, d.h. die Berechnung von q erfordert das Lösen einer quadratischen Gleichung. Der Leser prüfe selbst nach, dass in diesem Fall $q = p_0/p_2$, sofern $\mu > 1$.

Wenden wir uns der zweiten Frage zu, die nunmehr nur noch im superkritischen Fall Relevanz besitzt. Da für alle $k \geq 1$

$$\mathbb{P}(M_{n+i} \neq k \text{ für alle } i \geq 1 | M_n = k) \geq \mathbb{P}(M_{n+1} = 0 | M_n = k) = p_0^k > 0,$$

ist jeder Zustand $k \geq 1$ transient, und es folgt

$$\mathbb{P}_1(M_n \neq 0 \text{ für alle } n \geq 1) = \mathbb{P}_1(M_n \rightarrow \infty), \quad (2.112)$$

d.h. im Überlebensfall strebt die Populationsgröße gegen ∞ . *Ein GWP hat also nur die Optionen auszusterben oder zu explodieren.* Es stellt sich natürlich als nächstes die Frage nach einer präziseren Beschreibung des Wachstums von M_n im Überlebensfall. Wegen $\mathbb{E}_1 M_n = \mu^n$ liegt eine Untersuchung von $\mu^{-n} M_n$ nahe, die sich er-

folgreich im Rahmen von Martingalthetheorie bewerkstelligen lässt. Es sei auf Beispiel 4.21 und Abschnitt 5.2 im zweiten Teil über Martingale verwiesen.

Wir beenden an dieser Stelle unsere Diskussion von GWP, versäumen es aber nicht, darauf hinzuweisen, dass wir nur an der Spitze des Eisbergs gekratzt haben. Wer Lust hat auf mehr, sollte die schon früher erwähnte Monographie von JAGERS [36] oder die von ATHREYA & NEY [7] konsultieren.

2.9.8 Der Metropolis-Hastings-Algorithmus

Markov-Ketten Monte Carlo (engl. “*Markov chain Monte Carlo*” genannt und zu meist mit “*MCMC*” abgekürzt) dient als Sammelbegriff für Simulationsverfahren, die auf der Konstruktion von ergodischen MK mit einer vorgegebenen stationären Verteilung beruhen und das Ziel haben, eine Stichprobe dieser Verteilung, der sogenannten *Zielverteilung* (engl. *target distribution*), zu erzeugen. Die Idee dahinter ist einfach: Nach dem Ergodensatz 2.52 konvergiert eine ergodische MK $(M_n)_{n \geq 0}$ gegen ihre stationäre Verteilung π . Startet man die Kette im Simulationslauf in irgendeinem Anfangszustand und wartet eine gewisse Relaxationszeit t , die als “*Burn-in-Phase*” bezeichnet wird und dafür sorgen soll, dass

$$M_t \stackrel{d}{\approx} \pi$$

gilt, so liefern die Realisierungen von M_{t+1}, \dots, M_{t+n} eine Stichprobe von approximativ nach π verteilten Werten, die allerdings, anders als bei der gewöhnlichen Monte Carlo Methode, i.A. nicht unabhängig sind. Auf diese Problematik gehen wir hier jedoch ebenso wenig weiter ein wie auf die gewichtige Frage wie groß t zu wählen ist, was offenbar von der Konvergenzgeschwindigkeit der MK gegen ihre stationäre Verteilung abhängt (☞ hierzu das nächste Kapitel, speziell 3.4.2). Wem dies nicht reicht, dem seien die Übersichtsarbeiten von DIACONIS [16] und RICHEY [52], das – gerade für Neulinge – gut lesbare Buch von HÄGGSTRÖM [30] oder auch die Monographie von ASMUSSEN & GLYNN [6, Ch. 13] über Simulationstechniken empfohlen.

Im Folgenden stellen wir ein spezielles, zu den bekanntesten zählenden MCMC-Verfahren vor, das heutzutage als *Metropolis-Hastings-Algorithmus* oder auch kürzer *Metropolis-Algorithmus* bezeichnet wird (zur Historie ☞ am Ende dieses Unterabschnitts). Wie gerade erklärt, besteht das Ziel darin, Stichproben einer Zielverteilung $\pi = (\pi_i)_{i \in \mathcal{S}}$ zu erzeugen, wobei wir zusätzlich annehmen, dass uns diese Verteilung nur bis auf Normierung bekannt ist. Mit anderen Worten, es sei ein $\hat{\pi} = (\hat{\pi}_i)_{i \in \mathcal{S}}$ gegeben derart, dass $\pi = c^{-1} \hat{\pi}$ gilt, der Normierungsfaktor $c = \sum_{i \in \mathcal{S}} \hat{\pi}_i$ allerdings unbekannt, weil schwer zu berechnen ist. Für den Metropolis-Algorithmus reicht dies aus, weil er nur der Kenntnis von

$$\frac{\pi_i}{\pi_j} = \frac{\hat{\pi}_i}{\hat{\pi}_j} \quad \text{für alle } i, j \in \mathcal{S}$$

bedarf. O.B.d.A. können wir $\pi_i > 0$ für alle $i \in \mathcal{S}$ voraussetzen (andernfalls gehe man einfach von \mathcal{S} zu $\mathcal{S}^+ = \{i : \pi_i > 0\}$ über).

Sei $Q = (q_{ij})_{i,j \in \mathcal{S}}$ eine beliebige irreduzible und symmetrische Übergangsmatrix. Der Metropolis-Algorithmus ist ein zweistufiges Verfahren zur Erzeugung einer reversiblen ergodischen Markov-Kette $(M_n)_{n \geq 0}$ mit stationärer Verteilung π , genannt *Metropolis-Kette*, das in jedem Zeitschritt n zunächst einen neuen Zustand X_{n+1} gemäß $(q_{M_n, j})_{j \in \mathcal{S}}$ vorschlägt (Stufe 1: “proposal step”) und dann mit Wahrscheinlichkeit $a(M_n, X_{n+1})$ bzw. $1 - a(M_n, X_{n+1})$ akzeptiert oder verwirft (Stufe 2: “acceptance/rejection step”). Im ersten Fall ist $M_{n+1} = X_{n+1}$, im zweiten Fall wird der aktuelle Zustand beibehalten, d.h. $M_{n+1} = M_n$. Als Akzeptanzwahrscheinlichkeit wählt man

$$a(i, j) := \min(1, \hat{\pi}_j / \hat{\pi}_i) = \min(1, \pi_j / \pi_i)$$

für $i, j \in \mathcal{S}$, die somit in der Tat von π_i, π_j nur über den Quotienten π_j / π_i abhängt. Ein vorgeschlagener Zustand j wird demnach in jedem Fall akzeptiert, wenn er unter π wahrscheinlicher ist als der aktuelle Zustand i , andernfalls aber nur mit Wahrscheinlichkeit $\pi_j / \pi_i < 1$. Die Metropolis-Kette hat folglich die Übergangsmatrix $P = (p_{ij})_{i,j \in \mathcal{S}}$ mit

$$p_{ij} = \begin{cases} q_{ij} \min(1, \pi_j / \pi_i), & \text{falls } i \neq j, \\ 1 - \sum_{j \neq i} q_{ij} \min(1, \pi_j / \pi_i), & \text{falls } i = j. \end{cases}$$

Wegen

$$\pi_i p_{ij} = q_{ij} \min(\pi_i, \pi_j) = q_{ji} \min(\pi_i, \pi_j) = \pi_j p_{ji}$$

für alle $i, j \in \mathcal{S}$ gelten außerdem die detaillierten Gleichgewichtsgleichungen, so daß $(M_n)_{n \geq 0}$ wie behauptet eine reversible DMK mit stationärer Verteilung π definiert.

Der Metropolis-Algorithmus wurde erstmals 1953 in der Arbeit [46] von METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H. und TELLER, E. für die Simulation von Gibbs-Maßen (in der Chemie und Physik auch “Boltzmann-Verteilungen” genannt) vorgeschlagen und später von HASTINGS [32] auf allgemeinere Situationen erweitert. Hinsichtlich der Frage, wem der Algorithmus zuzuschreiben ist, gibt es Kontroversen. Während E. TELLER [58, S. 328] in seinen Memoiren schreibt, alle fünf Autoren hätten an dem Problem Tage (und Nächte) gearbeitet, bemerkt ROSENBLUTH [53] hierzu in einem schriftlich protokollierten Interview kurz vor seinem Tod, das Problem wäre von E. TELLER vorgeschlagen, von ihm gelöst und von seiner Frau A.W. ROSENBLUTH programmiert worden. A.H. TELLER und METROPOLIS (der Leiter des Computerlabors) wären hingegen in keiner Weise beteiligt gewesen. Diese Darstellung wird zwar durch die Ausführungen in [29] erhärtet, aber eine vollständige Aufklärung der Angelegenheit erscheint nicht mehr möglich zu sein.

2.9.9 Perfekte Simulation: Der Propp-Wilson-Algorithmus

Das oben nur kurz angesprochene Problem der hinreichend schnellen Konvergenz einer MCMC-Kette gegen ihre stationäre Verteilung ließ Fachleute natürlich

nicht ruhen, und es zeichneten sich bereits Anfang der 90er Jahre Ideen für einen grundlegend anderen Ansatz ab. Der Durchbruch gelang 1996 PROPP & WILSON [50] mit der Vorstellung eines neuen Algorithmus', der sich von den bis dahin bekannten MCMC-Verfahren dadurch unterscheidet, als Ausgabe eine Stichprobe zu liefern, deren Werte exakt nach der Zielverteilung π verteilt sind, was heutzutage *perfekte Simulation* oder *perfektes Sampling* genannt wird. Ausgangspunkt bildet wiederum eine EMK mit stationärer Verteilung π , von der jedoch parallel gleich mehrere Realisierungen mit verschiedenen Anfangszuständen erzeugt werden und dies bei rückwärts laufender Zeit, was noch genauer zu erläutern sein wird. Die geniale Einsicht der o.g. Autoren, von ihnen als "*Coupling from the past*" (CFTP) bezeichnet, bestand darin, dass, wenn es gelingt, all diese Realisierungen bis zum Ende der Laufzeit zu koppeln, der gemeinsame terminale Zustand notwendig nach π verteilt ist. Im Folgenden wollen wir dies in seinen Grundzügen kurz darstellen.

Alle Simulationsverfahren basieren auf der Verwendung von auf $(0, 1)$ gleichverteilten Pseudo-Zufallszahlen, die in hinreichender Weise statistischen Tests standhalten, um von diesen als exakt $Unif(0, 1)$ -verteilten und stochastisch unabhängigen Realisierungen zu sprechen. Genaugenommen ist dies natürlich falsch (daher auch der Zusatz "Pseudo"), denn ein Rechner kann nur Zahlen mit endlich vielen Nachkommastellen, also rationale Zahlen erzeugen, wohingegen eine $Unif(0, 1)$ -Zufallsgröße f.s. irrationale Zahlen als Werte annimmt. Spricht man von perfekter Simulation, bedeutet dies folglich, das soeben Gesagte zu vernachlässigen und Pseudo-Zufallszahlen als exakte Stichprobe von unabhängigen, identisch $Unif(0, 1)$ -verteilten Zufallsgrößen aufzufassen.

Sei o.B.d.A. $\mathcal{S} = \{1, \dots, m\}$ für ein $m \geq 2$ der Träger der Zielverteilung π und somit der Zustandsraum der zu simulierenden ergodischen EMK $M = (M_n)_{n \geq 0}$ mit Übergangsmatrix $P = (p_{ij})_{1 \leq i, j \leq m}$ und stationärer Verteilung π . Wir beschreiben zunächst, wie man M bei vorgegebener Anfangsverteilung λ mit Hilfe von unabhängigen, identisch $Unif(0, 1)$ -verteilten Zufallsgrößen U_0, U_1, \dots definieren kann: Sei dazu $M_0 := \Psi^\lambda(U_0)$ und $M_n := \Psi(M_{n-1}, U_n)$ für $n \geq 1$, wobei

$$\Psi^\lambda(u) := \begin{cases} 1, & \text{falls } 0 < u \leq \lambda_1, \\ 2, & \text{falls } \lambda_1 < u \leq \lambda_1 + \lambda_2, \\ \vdots & \\ m, & \text{falls } \sum_{j=1}^{m-1} \lambda_j < u \leq \sum_{j=1}^m \lambda_j = 1 \end{cases}$$

sowie

$$\Psi(i, u) := \begin{cases} 1, & \text{falls } 0 < u \leq p_{i1}, \\ 2, & \text{falls } p_{i1} < u \leq p_{i1} + p_{i2}, \\ \vdots & \\ m, & \text{falls } \sum_{j=1}^{m-1} p_{ij} < u \leq \sum_{j=1}^m p_{ij} = 1. \end{cases}$$

Wie der Leser sich leicht selbst überzeugt, entspricht die so rekursiv definierte Folge

$$M_n = \Psi(\Psi(\dots\Psi(\Psi^\lambda(U_0)), U_1), \dots, U_{n-1}), U_n), \quad n \geq 0$$

den gestellten Anforderungen. Eine etwas schönere und zugleich kompaktere Darstellung dieser Beziehung ergibt sich, wenn man $\Phi_n := \Psi_{U_n}$ setzt und so eine *Zufallsfunktion* $\Phi_n : \mathcal{S} \rightarrow \mathcal{S}$ erhält. Dann gilt nämlich

$$M_n = \Phi_n \circ \dots \circ \Phi_1(M_0) = \Phi_n \circ \dots \circ \Phi_1 \circ \Psi^\lambda(U_0), \quad n \geq 0$$

mit unabhängig, identisch verteilten Φ_1, Φ_2, \dots , die zudem auch von U_0 unabhängig sind. Ganz allgemein wird eine stochastische Folge dieser Gestalt auch *iteriertes Funktionensystem (IFS)* genannt, wobei diese Perspektive zumeist dann eingenommen wird, wenn die Zufallsfunktionen auf metrischen Räumen definiert sind und gewisse Kontraktionseigenschaften besitzen. Eine gut lesbare Einführung mit Bezügen zu vielen interessanten Anwendungen einschließlich des hier im Fokus stehenden Propp-Wilson-Algorithmus' geben DIACONIS & FREEDMAN [18]. Hier ist zunächst lediglich die Tatsache, dass jede EMK als IFS realisierbar ist, von Bedeutung. Allerdings überlegt man sich leicht, dass es viele Zufallsfunktionen geben kann, die zur selben EMK führen (hinsichtlich ihrer Verteilung bzw. Übergangsmatrix). Wir nennen deshalb eine Zufallsfunktion Φ von \mathcal{S} nach \mathcal{S} bzw. eine Folge $(\Phi_n)_n$ von unabhängigen Kopien von Φ *kompatibel mit* $P = (p_{ij})_{i,j \in \mathcal{S}}$, wenn

$$\mathbb{P}(\Phi(i) = j) = p_{ij}$$

für alle $i, j \in \mathcal{S}$ gilt. Unter dieser Voraussetzung ist offenbar das IFS

$$M_n = \Phi_n \circ \dots \circ \Phi_1(M_0), \quad n \geq 0$$

in der Tat eine EMK mit Übergangsmatrix P , sofern wir noch annehmen, dass M_0 unabhängig von den Φ_n ist.

Wir kommen nun zur Beschreibung des CFTP-Gedankens, den wir zuerst verbal umreißen wollen. Sei $1 \leq N_1 < N_2 < \dots$ irgendeine aufsteigende Folge natürlicher Zahlen (eine oft getroffene Wahl ist $N_k = 2^{k-1}$) und $(\Phi_n)_{n \in \mathbb{Z}}$ eine mit P kompatible doppelt unendliche Folge von Zufallsfunktionen. Für die Zeitfenster $\{-N_k, \dots, 0\}$, $k = 1, 2, \dots$, die also anders als üblich in der Vergangenheit beginnen und stets im Zeitpunkt 0 enden, werden parallel für jedes $i \in \mathcal{S}$ Realisierungen der EMK mit Anfangszustand i erzeugt, die dadurch gekoppelt sind, dass sie alle *dieselben* Zufallsfunktionen $\Phi_{-N_k+1}, \dots, \Phi_0$ verwenden (zur Bedeutung dieses Umstands \boxtimes (2.114) weiter unten). Das Verfahren bricht ab, sobald für ein k alle Realisierungen zum Zeitpunkt 0 denselben Wert haben, d.h. gekoppelt sind (Koaleszenz). Dieser Wert ist dann die Ausgabe und, wie noch gezeigt wird, *exakt* nach π verteilt.

Zu den Formalitäten: Wir setzen $\Phi_{k,l} = \Phi_k \circ \dots \circ \Phi_l$ für $k, l \in \mathbb{Z}$. Bezeichnet dann

$$(M_{k,-N_k}^{(i)}, M_{k,-N_k+1}^{(i)}, \dots, M_{k,-1}^{(k,i)}, M_{k,0}^{(i)}) \quad k \geq 0, i \in \mathcal{S}$$

für jedes $k \geq 1$ und $i \in \mathcal{S}$ den Zufallsvektor, der den Simulationslauf zum Zeitfenster $\{-N_k, \dots, 0\}$ bei Anfangszustand i repräsentiert, so können wir diesen durch

$$M_{k,-N_k}^{(i)} := i \quad \text{und} \quad M_{k,-n}^{(i)} := \Phi_{-n:-N_k+1}(i), \quad 0 \leq n < N_k,$$

definieren, was speziell im Fall $N_k = k$ und für $k = 0, 1, 2$

$$\begin{array}{lll} M_{1,-1}^{(i)} = i, & M_{2,-2}^{(i)} = i, & M_{3,-4}^{(i)} = i \\ M_{1,0}^{(i)} = \Phi_0(i), & M_{2,-1}^{(i)} = \Phi_{-1}(i), & M_{3,-3}^{(i)} = \Phi_{-3}(i) \\ & M_{2,0}^{(i)} = \Phi_{0:-1}(i), & M_{3,-2}^{(i)} = \Phi_{-2:-3}(i) \\ & & M_{3,-1}^{(i)} = \Phi_{-1:-3}(i) \\ & & M_{3,0}^{(i)} = \Phi_{0:-3}(i) \end{array}$$

ergibt [138] auch Abb. 2.13].

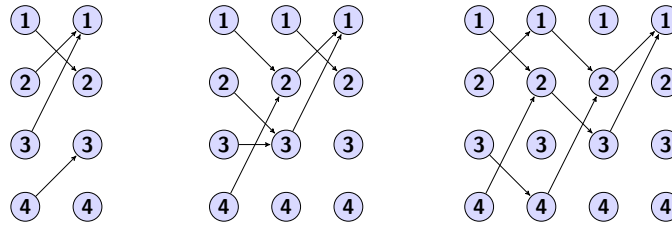


Abb. 2.13 CFTP am Beispiel einer EMK mit vier Zuständen und für $N_k = k$ für $k \geq 1$. Hier gilt $\tau = 3$ und $M_{\tau,0}^{(*)} = 1$. Pfade ohne Bedeutung wie etwa der von 4 nach 3 für $k \geq 2$ werden fortgelassen.

Sei nun

$$\tau := \inf \left\{ k \geq 1 : M_{k,0}^{(1)} = \dots = M_{k,0}^{(m)} \right\} = \inf \{ k \geq 1 : \Phi_{0:-N_k+1} \text{ ist konstant} \}$$

die CFTP-Kopplungszeit, wobei wie üblich $\inf \emptyset := \infty$ gelte. Die Tatsache, dass für jedes k dieselben Φ_{-n} benutzt werden, sichert

$$M_{k,0}^{(1)} = \dots = M_{k,0}^{(m)} =: M_{k,0}^{(*)} \quad \text{für alle } k \geq \tau, \quad (2.113)$$

denn sie impliziert

$$M_{k,0}^{(i)} = \Phi_{0:-N_k+1}(i) = \underbrace{\Phi_{0:-N_{\tau}+1}}_{\text{konstante Funktion}}(\Phi_{-N_{\tau}:-N_k+1}(i)) \quad (2.114)$$

für alle $k > \tau$ und $i \in \mathcal{S}$.

Satz 2.78. In den vorherigen Bezeichnungen gilt

$$M_{\tau,0}^{(*)} \stackrel{d}{=} \pi, \quad (2.115)$$

sofern die CFTP-Kopplungszeit τ f.s. endlich ist.

Beweis. Sofern τ f.s. endlich ist, gilt unter Hinweis auf (2.113) für jedes $j \in \mathcal{S}$

$$\begin{aligned} \mathbb{P}(M_{\tau,0}^{(*)} = j) &= \sum_{i=1}^m \pi_i \mathbb{P}(M_{\tau,0}^{(i)} = j) \\ &= \lim_{k \rightarrow \infty} \sum_{i=1}^m \pi_i \mathbb{P}(M_{k,0}^{(i)} = j) = \lim_{k \rightarrow \infty} \sum_{i=1}^m \pi_i p_{ij}^{(k)} = \pi_j \end{aligned}$$

und somit (2.115) □

Angesichts der Formulierung des Satzes sollte es niemanden überraschen, dass die Kopplungszeit τ nicht immer f.s. endlich ist. Wir überlassen es dem Leser als Übung, ein Gegenbeispiel zu konstruieren, wobei als Hinweis gegeben sei, dass dies schon für den einfachsten Fall $\mathcal{S} = \{1, 2\}$ leicht möglich ist.

Es liegt auf der Hand, dass der Propp-Wilson-Algorithmus für sehr große Zustandsräume sehr aufwendig sein kann, weil man für *jeden* Anfangszustand einen Pfad der Kette simulieren muss. Andererseits gibt es Fälle, in denen sich dieser Aufwand drastisch reduziert, wie wir im Folgenden sehen werden.

Der monotone Fall: \mathcal{S} partiell geordnet und Φ f.s. monoton

Nehmen wir an, dass auf \mathcal{S} eine Halbordnung \preceq existiert mit einem maximalen und einem minimalen Element s_* bzw. s^* . Dies lässt sich durch die folgenden vier Eigenschaften charakterisieren:

- (A1) $s \preceq s$ für jedes $s \in \mathcal{S}$,
- (A2) $r \preceq s$ und $s \preceq r$ impliziert $r = s$ für alle $r, s \in \mathcal{S}$,
- (A3) $r \preceq s$ und $s \preceq t$ impliziert $r \preceq t$ für alle $r, s, t \in \mathcal{S}$,
- (A4) $s_* \preceq s \preceq s^*$ für alle $s \in \mathcal{S}$.

Nehmen wir ferner an, dass zu P eine kompatible Zufallsfunktion Φ existiert, die bezüglich der Halbordnung \preceq monoton wachsend ist, d.h.

$$r \preceq s \quad \Rightarrow \quad \Phi(r) \preceq \Phi(s) \quad \text{f.s.} \quad (2.116)$$

In diesem Fall reicht es, nur Realisierungen der EMK $(M_n)_{n \geq 0}$ mit den Anfangszuständen s_*, s^* zu simulieren, wie der nachfolgende Satz lehrt.

Satz 2.79. *In denselben Bezeichnungen wie in Satz 2.78 gilt für die CFTP-Kopplungszeit*

$$\begin{aligned} \tau &= \inf\{k \geq 1 : M_{k,0}^{(s_*)} = M_{k,0}^{(s^*)}\} \\ &= \inf\{k \geq 1 : \Phi_{0:-N_k+1}(s_*) = \Phi_{0:-N_k+1}(s^*)\}, \end{aligned} \quad (2.117)$$

und diese ist f.s. endlich.

Beweis. Die Monotonieeigenschaft von Φ vererbt sich offenkundig auf die Iterationen $\Phi_{0:-n}$ für alle $n \geq 0$. Folglich sichert $\Phi_{0:-N_k+1}(s_*) = \Phi_{0:-N_k+1}(s^*)$ bereits die f.s. Konstanz der Funktion $\Phi_{0:-N_k+1}$ auf \mathcal{S} , was (??) beweist. Zum Nachweis von $\tau < \infty$ f.s. notieren wir als erstes, dass aus $\Phi_{0:-n}(s^*) = s_*$, wiederum aufgrund der Monotonieannahme sowie der Minimalität von s_* , auch $\Phi_{0:-n}(s) = s_*$ für alle $s \in \mathcal{S}$ folgt und somit

$$\{\tau > k\} \subset \{\Phi_{0:-n}(s^*) \neq s_* \text{ für } 1 \leq n \leq N_k\}$$

für alle $k \geq 1$. Nun gilt aber

$$\begin{aligned} \mathbb{P}(\tau > k) &\leq \mathbb{P}(\Phi_{0:-n}(s^*) \neq s_* \text{ für } 1 \leq n \leq N_k) \\ &= \mathbb{P}_{s^*}(M_n \neq s_* \text{ für } 0 \leq n < N_k) \end{aligned}$$

und diese Wahrscheinlichkeit konvergiert gegen 0, weil $(M_n)_{n \geq 0}$ rekurrent ist. \square

Für den Fall, dass Φ bezüglich \preceq nicht f.s. monoton wachsend, sondern f.s. monoton fallend ist, also

$$r \preceq s \quad \Rightarrow \quad \Phi(s) \preceq \Phi(r) \quad \text{f.s.}$$

gilt, hilft folgende Überlegung. Bezeichnet Φ' eine unabhängige Kopie von Φ , so ist $\Phi \circ \Phi'$ wiederum f.s. monoton wachsend, denn aus $r \preceq s$ folgt $\Phi'(s) \preceq \Phi'(r)$ und daher $\Phi(\Phi'(r)) \preceq \Phi(\Phi'(s))$. Offenkundig bildet $\Phi \circ \Phi'$ eine zu P^2 kompatible Zufallsfunktion, und da P^2 dieselbe stationäre Verteilung besitzt

Ein Beispiel: Das Ising-Modell auf endlichen Gittern

Ein prominentes Beispiel aus der statistischen Physik, auf das die soeben angestellten Überlegungen zutreffen, ist das *Ising-Modell für Ferromagnetismus* auf einem endlichen Gitter $[m] \times [m]$, $[m] := \{1, \dots, m\}$, benannt nach dem Physiker ERNST ISING. Wir folgen der Darstellung in [18, Abschnitt 3] und verweisen wie dort auf KINDERMANN & SNELL [40] für weitergehende Informationen.

Der Zustandsraum ist $\mathcal{S} = \{-1, +1\}^{[m] \times [m]}$, besteht also aus den Funktionen von $[m] \times [m]$ nach $\{-1, +1\}$. Diese beschreiben die möglichen *Spin-Konfigurationen*. Auf \mathcal{S} hat man die natürliche Halbordnung

$$r = (r_{ij})_{1 \leq i, j \leq m} \preceq s = (s_{ij})_{1 \leq i, j \leq m} \quad :\Leftrightarrow \quad r_{ij} \leq s_{ij} \quad \text{für alle } i, j \in [m]$$

mit den extremalen Elementen $s_* \equiv -1$ und $s^* \equiv 1$. Simuliert werden soll das *Gibbs-Maß*

$$\pi_s = \frac{1}{Z_{m, \beta}} \exp(\beta H(s)), \quad s \in \mathcal{S},$$

zum Parameter $\beta \geq 0$, in der Physik als *inverse Temperatur* bezeichnet, und mit der sogenannten *Hamilton-Funktion* (engl. *Hamiltonian*)

$$H(s) := \sum_{(i,j) \sim (k,l)} s_{ij} s_{kl},$$

wobei $(i, j) \sim (k, l)$ genau dann, wenn $|i - k| + |j - l| = 1$. Die Normierungskonstante $Z_{m,\beta}$, auch *Partitionsfunktion* genannt, ist i.A. (für große Gitter) unbekannt. Aus diesem Grund benötigen wir zur Simulation eine MK mit stationärer Verteilung π , die der Kenntnis von $Z_{m,\beta}$ nicht bedarf, was uns natürlich sogleich an den in 2.9.8 vorgestellten Metropolis-Algorithmus erinnert. Bevor wir hierauf weiter eingehen, seien noch folgende Bemerkungen hinsichtlich der Eigenschaften von π bei variierendem β gemacht:

- Falls $\beta = 0$ (unendliche Temperatur), ist π die Gleichverteilung auf \mathcal{S} und damit alle Spins in einer Konfiguration voneinander unabhängig.
- Falls β groß ist, werden Konfigurationen bevorzugt, die nur wenige benachbarte Spins mit unterschiedlicher Ausrichtung aufweisen.
- Im Limes $\beta \rightarrow \infty$ konvergiert π gegen eine Gleichverteilung auf den extremalen Konfigurationen s_* und s^* , d.h. gegen $(\delta_{s_*} + \delta_{s^*})/2$, wie der Leser sich leicht überlegt (beachte dabei, dass $H(s) < H(s_*) = H(s^*)$ für alle $s \in \mathcal{S} \setminus \{s_*, s^*\}$).

Abb. 2.14 zeigt typische Konfigurationen für verschiedene Werte von β .

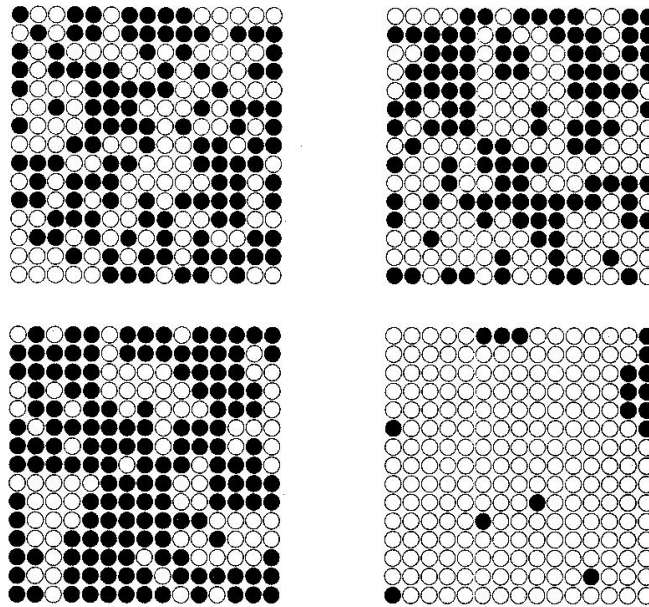


Abb. 2.14 Das Ising-Modell: Typische Konfigurationen für $\beta = 0$ (links oben), $\beta = 0.15$ (rechts oben), $\beta = 0.3$ (links unten) und $\beta = 0.5$ (rechts unten) [dem Buch [30] von HÄGGSTRÖM entnommen].

Wenden wir uns nun der Angabe einer MK mit stationärer Verteilung π zu und nutzen dabei die Gelegenheit zur Vorstellung eines weiteren bekannten MCMC-Verfahrens, des sogenannten *Gibbs-Samplers*. Für eine gegebene Konfiguration s und ein $v = (i, j) \in [m] \times [m]$ seien s^{v+} und s^{v-} die beiden Konfigurationen, die mit s in allen Gitterpunkten $w \neq v$ übereinstimmen und an der Stelle v den Wert $+1$ bzw. -1 aufweisen. Der Gibbs-Sampler wählt in jedem Schritt eine Position v zufällig aus und ersetzt s durch s^{v+} mit Wahrscheinlichkeit

$$\begin{aligned} a(s, s^{v+}) &:= \frac{\pi_{s^{v+}}}{\pi_{s^{v+}} + \pi_{s^{v-}}} \\ &= \frac{\exp(\beta H(s^{v+}))}{\exp(\beta H(s^{v+})) + \exp(\beta H(s^{v-}))} \\ &= \frac{\exp(\beta \sum_{w:w \sim v} s_w)}{\exp(\beta \sum_{w:w \sim v} s_w) + \exp(-\beta \sum_{w:w \sim v} s_w)} \end{aligned} \quad (2.118)$$

und durch s^{v-} mit Wahrscheinlichkeit $a(s, s^{v-}) := 1 - a(s, s^{v+})$, wobei beachtet werde, dass die $a(s, s^{v\pm})$

- ohne Kenntnis von $Z_{m,\beta}$ berechenbar sind,
- de facto nur von den Werten $s_v \sum_{w:w \sim v} s_w$ abhängen, was (2.118) erklärt.

Da natürlich immer einer der beiden Zustände $s^{v\pm}$ mit s übereinstimmt (den anderen nennen wir hiernach $s'(v)$), macht die so definierte EMK pro Zeitschritt nichts anderes als für einen zufällig gewählten Gitterpunkt den Spin zu wechseln oder beizubehalten. Man kann dieses Vorgehen auch als Metropolis-Verfahren mit modifizierten Akzeptanzwahrscheinlichkeiten ansehen: Es schlägt pro Zeitschritt einen Spinwechsel $s \rightarrow s'(v)$ in einem zufällig gewählten Gitterpunkt v vor, d.h. mit Wahrscheinlichkeit $1/m^2$, führt diesen dann aber nur mit Wahrscheinlichkeit $a(s, s'(v))$ aus und verharrt andernfalls in s . Die resultierende, offensichtlich irreduzible EMK hat demnach die Übergangswahrscheinlichkeiten

$$p_{s,s'(v)} = 1 - p_{s,s} = \frac{1}{m^2} a(s, s'(v)) = \frac{1}{m^2} \frac{\pi_{s'(v)}}{\pi_s + \pi_{s'(v)}}$$

für alle $s \in \mathcal{S}$. Darüber hinaus gelten die detaillierten Gleichgewichtsgleichungen

$$\pi_s p_{s,s'(v)} = \frac{1}{m^2} \frac{\pi_s \pi_{s'(v)}}{\pi_s + \pi_{s'(v)}} = p(s, s'(v)) \pi_{s'(v)}, \quad s \in \mathcal{S}, v \in [m] \times [m].$$

Die Kette hat also auch die geforderte stationäre Verteilung π und ist unter dieser reversibel. Ihre Übergangsmatrix bezeichnen wir wie üblich mit P .

Es bleibt schließlich die Aufgabe, eine Repräsentation dieser Kette als IFS mit bezüglich der gegebenen Halbordnung f.s. monotonen Zufallsfunktionen anzugeben. Seien hierzu U, V gleichverteilte Zufallsvariablen auf $(0, 1)$ bzw. $[m] \times [m]$ und $\Psi : \mathcal{S} \times ([m] \times [m]) \times (0, 1) \rightarrow \mathcal{S}$ durch

$$\Psi(s, u, v) := \begin{cases} s^{v+}, & \text{falls } 0 < u \leq a(s, s^{v+}), \\ s^{v-}, & \text{falls } a(s, s^{v+}) < u < 1 \end{cases}$$

definiert. Dann ist $\Phi := \Psi(\cdot, U, V)$ mit P kompatibel, wie man leicht nachprüft, und besitzt auch die geforderte Monotonieeigenschaft (2.117), wie die folgenden Überlegungen zeigen: Gegeben $r \preceq s$, existieren Konfigurationen r^1, \dots, r^k , derart dass $r^{i-1} \preceq r^i$ und r^{i-1}, r^i sich nur in einem Gitterpunkt unterscheiden für $i = 1, \dots, k+1$, wobei $r^0 := r$ und $r^{k+1} := s$. Aus diesem Grund können wir o.B.d.A. annehmen, dass schon r und s letztere Eigenschaft besitzen. Sei also $r_x = -1, s_x = +1$ und $r_y = s_y$ für alle $y \neq x$. Für alle $v \neq x$ mit $v \not\sim x$ gilt dann offenkundig $a(r, r^i(v)) = a(s, s^i(v))$ und somit $\Psi(r, u, v) = \Psi(s, u, v)$ für alle $u \in (0, 1)$. Im Fall $v \sim x$ gilt $a(r, r^{v+}) < a(s, s^{v+})$ und somit $\Psi(r, u, v) \leq \Psi(s, u, v)$, wenn man

$$\sum_{y: y \sim v} r_y = \sum_{y: x \neq y \sim v} r_y - 1 = \sum_{y: x \neq y \sim v} s_y - 1 = \sum_{y: y \sim v} s_y - 2$$

beachtet und dies in (2.118) für $a(r, r^{v+})$ verwendet. Bleibt schließlich noch der Fall $v = x$. Unter nochmaligem Hinweis auf (2.118) gilt aber $a(r, r^{x+}) = a(s, s^{x-}) < a(s, s^{x+}) = a(s, s)$ und daher wie behauptet $\Psi(r, u, x) \leq \Psi(s, u, x)$.

Damit ist alles gezeigt, um den Propp-Wilson-Algorithmus in der für den monotonen Fall beschriebenen Weise anzuwenden.