Kapitel 5

Das Wichtigste über Markov-Ketten

Markov-Ketten sind einfache und anschauliche Beispiele sogenannter *stochastischer Prozesse*, welche die Entwicklung eines stochastischen Modells im Laufe der Zeit beschreiben. Markov-Ketten kann man mit Methoden der linearen Algebra behandeln, weil man es immer mit Eigenschaften spezieller Matrizen, sogenannter *Markov-Matrizen*, zu tun hat. Deshalb wollen wir in unserer knappen Einführung zum Thema auch keine stochastische Definition dieser Prozesse angeben, sondern lediglich mit den zugrundeliegenden Markov-Matrizen arbeiten. Für eine umfangreichere Darstellung sei auf die entsprechenden Lehrbücher verwiesen.

5.1 Grundbegriffe

Zunächst wollen wir eine alternative Beschreibung des (fairen) Münzwurfs besprechen. Offenbar reicht es anzugeben, mit welcher Wahrscheinlichkeit die Kombinationen Kopf-Kopf, Kopf-Zahl, Zahl-Kopf und Zahl-Zahl aufeinander folgen. Dies kann man anhand eines sogenannten Übergangsgraphen veranschaulichen.

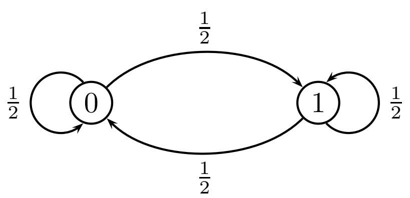


Abbildung 5.1: Markov-Kette für den fairen Münzwurf

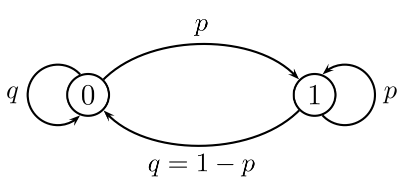
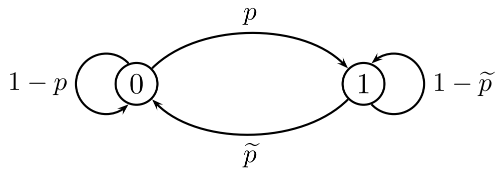


Abbildung 5.2: Markov-Kette für den gezinkten Münzwurf

Nun könnte man in einem solchen Graphen noch allgemeinere Übergangswahrscheinlichkeiten zulassen, etwa wie im nächsten Bild.



Überlegen Sie sich ein einfaches Münzwurf-Modell, welches mit einem solchen Übergangsgraphen beschrieben wird! Die auftretenden Wahrscheinlichkeiten kann man bequem in einer sogenannten *Übergangsmatrix* zusammenfassen. Für die Parameter $0 \leq p, \widetilde{p}, \leq 1$ haben wir $M = \left(\begin{smallmatrix} 1-p & p \\ \widetilde{p} & 1-\widetilde{p} \end{smallmatrix}\right)$, und der Eintrag m_{ij} ist gerade die Übergangswahrscheinlichkeit vom Zustand i in den Zustand j, also deuten wir das als $m_{ij} = \mathbb{P}(i \to j)$. Dies motiviert die folgende Definition.

Definition 5.1 (Markov-Matrix)**.** Eine reelle $n \times n$-Matrix $M = (m_{ij})_{1 \leq i,j \leq n}$ heißt **Markov-Matrix**, wenn gilt:

(1) $m_{ij} \geq 0$ für alle $1 \leq i, j \leq n$,

(2) $\sum_{j=1}^{n} m_{ij} = 1$ für alle $1 \leq i \leq n$.

Offenbar ist die Matrix M vom letzten Beispiel eine allgemeine Markov-Matrix für $n = 2$. Sei M nun eine allgemeine Markov-Matrix, und sei der Spaltenvektor $\mathbf{1} = (1, \ldots, 1)^t$ gegeben. Dann gilt

$$M \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_j M_{1j} \\ \vdots \\ \sum_j M_{nj} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

also ist der Vektor **1** ein Rechts-Eigenvektor (REV) zum Eigenwert[1] (EW) 1.

Da also 1 ein EW von M ist, existiert auch ein Links-Eigenvektor (LEV) für M zum Eigenwert 1, also

$$(p_1, \ldots, p_n) \cdot M = (p_1, \ldots, p_n),$$

wobei dies nicht der Nullvektor ist. Um uns die Bedeutung dieses Gleichungssystems zu erschließen, starten wir von einem beliebigen Wahrscheinlichkeitsvektor $\boldsymbol{q} = (q_1, \ldots, q_n)$, es gelte also $q_i \geq 0$ und $\sum_i q_i = 1$. Den Eintrag q_i solch eines Vektors interpretieren wir als die Wahrscheinlichkeit, die Markov-Kette zu einem

[1]Dies ist der Moment, wo Sie das Kapitel aus der linearen Algebra zu Eigenwerten und Eigenvektoren wiederholen sollten.

gegebenen Zeitpunkt im Zustand i zu finden. (Denken Sie genau über die Bedeutung dieser Aussage nach!) Sei nun $\boldsymbol{r} = \boldsymbol{q}M$, also

$$r_i = \sum_j q_j M_{ji} \geq 0.$$

Dann gilt für die Summe der Einträge

$$\sum_i r_i = \sum_{i,j} q_j M_{ji} = \sum_j q_j \underbrace{\sum_i M_{ji}}_{=1} = \sum_j q_j = 1.$$

Also werden Wahrscheinlichkeitsvektoren auf ebensolche abgebildet! Ein nichtnegativer und normierter LEV zum EW 1 ist somit eine *stationäre* Verteilung von Wahrscheinlichkeiten. Er heißt auch ein **Gleichgewichtsvektor** von M. In den obigen Beispielen treten folgende Gleichgewichtsvektoren auf.

Beispiel. Der faire Münzwurf hat die Übergangsmatrix

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

Diese Matrix ist symmetrisch, also erhalten wir sofort den LEV $\left(\frac{1}{2}, \frac{1}{2}\right)$, welcher offenbar die Häufigkeiten von Kopf bzw. Zahl angibt! Im Fall der obigen allgemeinen 2×2-Markov-Matrix mit Übergangsmatrix $M = \left(\begin{smallmatrix} 1-p & p \\ \widetilde{p} & 1-\widetilde{p} \end{smallmatrix}\right)$ haben wir allgemein den LEV

$$\left(\frac{\widetilde{p}}{p+\widetilde{p}}, \frac{p}{p+\widetilde{p}}\right),$$

wie man leicht nachrechnet. Dies ergibt für den Fall des gezinkten Münzwurfs $\widetilde{p} = 1-p$ den entsprechenden Häufigkeitsvektor $(1-p, p)$. Diese Beobachtung ist kein Zufall, sondern gilt in allgemeiner Form, wie wir weiter unten sehen werden.

Welche "Verhaltensweisen" können Markov-Ketten zeigen? Wir wollen drei charakteristische Effekte betrachten, anhand der folgenden Bilder.

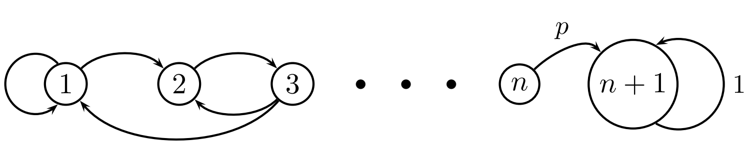


Abbildung 5.3: Absorbierende Markov-Kette: Vom Zustand $n+1$ kommt man nicht mehr fort.

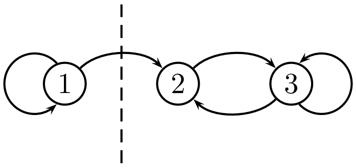


Abbildung 5.4: Reduzible Markov-Kette: Hat man 1 einmal verlassen, spielen nur noch 2 und 3 eine Rolle.

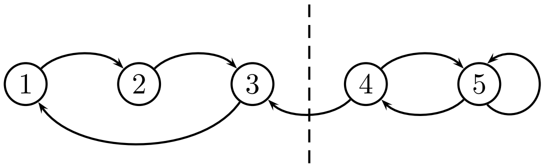


Abbildung 5.5: Markov-Kette mit Zyklen, einer davon absorbierend: Gelangt man einmal zur 3, gerät man in das "Hamsterrad" des Zykels $1 \to 2 \to 3 \to 1 \to \cdots$. Damit ist dieses Beispiel außerdem reduzibel.

5.2 Absorbierende Markov-Ketten

Hier wollen wir ein Beispiel einer absorbierende Markov-Kette im Detail besprechen. Die zu Abbildung 5.6 gehörende Markov-Matrix ist

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5.1}$$

wobei $0 < p < 1$ und $q = 1 - p$ sei. Die Zustände 1 und 5 sind absorbierend.

Definition 5.2 (Absorptionswahrscheinlichkeit)**.** Es sei u ein absorbierender Zustand. Die Wahrscheinlichkeit w_j, von j ausgehend irgendwann in Zustand u zu landen, heißt **Absorptionswahrscheinlichkeit** von j nach u.

Für unser Ziel $u = 5$ in obigem Beispiel wollen wir die entsprechenden Wahrscheinlichkeiten w_j berechnen. Klar sind

$$w_1 = 0 \quad \text{und} \quad w_5 = 1,$$

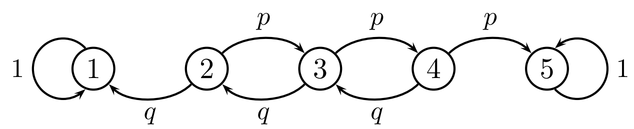


Abbildung 5.6: Absorbierende Markov-Kette zur Gl. (5.1).

denn 1 kommt nie nach 5, und 5 bleibt immer bei 5. Für $j \in \{2,3,4\}$ haben wir offenbar (!) den Zusammenhang

$$\begin{aligned} w_j &= m_{j,j+1} \cdot w_{j+1} + m_{j,j-1} \cdot w_{j-1} \\ &= p \cdot w_{j+1} + q \cdot w_{j-1}. \end{aligned}$$

Machen Sie sich die Bedeutung dieser Konsistenzgleichungen (Stichwort: Zerlegung nach dem ersten Schritt) genau klar! Mit ihrer Hilfe erhalten wir für die w_i folgendes Gleichungssystem.

$$\begin{aligned} w_1 &= 0, \\ w_2 &= p \cdot w_3 + q \cdot w_1 = p \cdot w_3, \\ w_3 &= p \cdot w_4 + q \cdot w_2 = p \cdot w_4 + pq \cdot w_3, \\ w_4 &= p \cdot w_5 + q \cdot w_3 = p + pq \cdot w_4 + pq^2 \cdot w_3, \\ w_5 &= 1. \end{aligned}$$

Also ergibt sich $w_3 = \frac{p}{1-pq}\, w_4$ aus der dritten Gleichung, und wir erhalten für w_4 die Bedingung $(1-pq)w_4 = p + \frac{p^2q^2}{(1-pq)}\, w_4$ aus der vierten Gleichung. Damit erhalten wir die Lösung

$$w_2 = \frac{p^3}{1-2pq}, \qquad w_3 = \frac{p^2}{1-2pq}, \qquad w_4 = \frac{p-p^2+p^3}{1-2pq}.$$

Beachten Sie auch die Gültigkeit der Beziehung $(w_{j+1} - w_j) = \frac{q}{p} \cdot (w_j - w_{j-1})$.

Was haben wir im obigen Beispiel berechnet? Bestimmt haben wir

$$M \begin{pmatrix} w_1 \\ \vdots \\ w_5 \end{pmatrix} = \begin{pmatrix} w_1 \\ \vdots \\ w_5 \end{pmatrix} =: \boldsymbol{w},$$

also einen REV zum Eigenwert 1. Wegen der Absorptionseigenschaft wissen wir, dass zwei linear unabhängige LEVen zum Eigenwert 1 existieren, nämlich

$$\boldsymbol{v}_1 = (1,0,0,0,0) \quad \text{und} \quad \boldsymbol{v}_2 = (0,0,0,0,1).$$

Dies kann man einfach nachrechnen. Jeder Gleichgewichtsvektor ist also (!) von der Form $\alpha \boldsymbol{v_1} + (1-\alpha)\boldsymbol{v}_2$, mit $0 \le \alpha \le 1$. Daraus folgt aber auch, dass es zwei linear unabhängige REVen zum Eigenwert 1 gibt. Dies sind $\boldsymbol{w}$, wie oben bestimmt, und $\mathbf{1} - \boldsymbol{w}$, wie man sofort nachrechnet: $M(\mathbf{1} - \boldsymbol{w}) = M\mathbf{1} - M\boldsymbol{w} = \mathbf{1} - \boldsymbol{w}$.

Aus dieser Überlegung leiten wir folgendes Rezept ab:

> Sei M eine $n \times n$-Markov-Matrix mit zusammenhängendem Übergangsgraphen und den absorbierenden Zuständen
>
> $$a_1, \ldots, a_\ell, \quad \text{mit } \ell \overset{!}{<} n.$$
>
> Zur Berechnung der Absorptionswahrscheinlichkeiten in den Zustand a_k suchen wir den REV $(w_1, \ldots, w_n)^t$ zum EW 1 mit den Nebenbedingungen
>
> $$w_{a_k} = 1 \quad \text{und} \quad w_{a_j} = 0 \quad \text{für alle } 1 \le j \le \ell \text{ mit } j \ne k.$$
>
> Der Eintrag dieses Vektors an der Stelle m ist dann die Wahrscheinlichkeit, vom Zustand m in den absorbierenden Zustand a_k zu gelangen.

Beispiel. Die in Figur 5.7 dargestellte Markov-Kette besitzt die Übergangsmatrix

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}.$$

Für den Vektor der Absorptionswahrscheinlichkeiten in Zustand 1 erhalten wir nach kurzer Rechnung $\boldsymbol{w} = \left(1, 0, 0, 0, \frac{1}{4}\right)^t$. Also haben wir die Absorptionswahrscheinlichkeit $\frac{1}{4}$, vom Zustand 5 in den Zustand 1 zu gelangen. Für die verbleibenden absorbierenden Zustände erhalten wir wegen der Symmetrie des Modells dieselbe Absorptionswahrscheinlichkeit. (Überlegen Sie sich ein Argument, wie man bei diesem Beispiel auch ohne Rechnung zum Ergebnis kommt!)

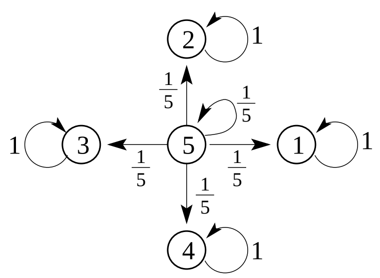


Abbildung 5.7: Markov-Kette mit vier absorbierenden Zuständen

5.3 Markov-Ketten mit stationären Verteilungen

Für eine große und wichtige Klasse von Markov-Ketten können wir ihr asymptotisches Verhalten systematisch beschreiben. Dazu führen wir einige Begriffe ein.

Definition 5.3. Sei M eine Markov-Matrix (oder allgemeiner eine Matrix mit Einträgen $m_{ij} \geq 0$). Dann heißt M

(i) **irreduzibel**, wenn gilt: $\forall i, j : \exists k : (M^k)_{ij} > 0$.

(ii) **primitiv**, wenn gilt: $\exists k : \forall i, j : (M^k)_{ij} > 0$.

(iii) **aperiodisch**, wenn gilt: $\forall i : \mathrm{ggT}\big\{k \geq 1 : (M^k)_{ii} > 0\big\} = 1$.

Bemerkung. Primitivität impliziert Irreduzibilität. Man kann zeigen, dass eine Markov-Matrix M genau dann primitiv ist, wenn M irreduzibel und aperiodisch ist. Alle drei Begriffe haben ihre Entsprechung in Eigenschaften der zugehörigen Übergangsgraphen. Irreduzibilität bedeutet, dass jeder Zustand von jedem anderen auf einem Weg entlang der Pfeile erreicht werden kann. Primitivität bedeutet, dass dies immer in k Schritten geschehen kann, für ein festes k. Aperiodizität heißt, dass für jeden Vertex des Graphen die Längen seiner Zykel von diesem Vertex aus teilerfremd sind.

Beispiel. Als einfaches Beispiel betrachten wir die Markov-Matrix $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Für deren Potenzen gilt

$$M^{2\ell} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad M^{2\ell+1} = M, \qquad \ell \in \mathbb{N}_0.$$

Also ist M irreduzibel, aber nicht primitiv. M ist also (!) auch nicht aperiodisch, wie man sofort an der Beziehung $\mathrm{ggT}(2, 4, 6, \ldots) = 2$ sieht.

Die Asymptotik für Markov-Systeme mit primitiver Markov-Matrix ist durch den folgenden wichtigen Satz festgelegt.

Satz 5.4 (Perron)**.** Sei M eine primitive Markov-Matrix. Dann existiert *genau ein* Wahrscheinlichkeitsvektor $\boldsymbol{p} = (p_1, \ldots, p_n)$, der Linkseigenvektor zum Eigenwert 1 ist. Dabei sind alle p_i positiv, $p_i > 0$, und alle übrigen Eigenwerte λ von M erfüllen $|\lambda| < 1$.

Bemerkung. Dieses Resultat hat viele Facetten:

(i) Zunächst hat der Satz von Perron weitreichende Konsequenzen. Für eine primitive Markov-Matrix M gilt:

- Falls q ein beliebiger Wahrscheinlichkeitsvektor ist, so konvergiert die Folge $\{\boldsymbol{q}M^k\}_{k\in\mathbb{N}}$ gegen die stationäre Verteilung $\boldsymbol{p}$, die Konvergenz ist also unabhängig vom Startvektor! Beispielsweise gilt für

$$M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \text{und} \quad \boldsymbol{r} = (p, q)$$

  mit $p + q = 1$ unter der Wirkung von M die Beziehung

$$(p, q) \mapsto \left(\frac{p+q}{2}, \frac{p+q}{2}\right) = \left(\frac{1}{2}, \frac{1}{2}\right) \mapsto \left(\frac{1}{2}, \frac{1}{2}\right) \mapsto \dots$$

  Der Grund für das Auftreten der Gleichgewichtsverteilung schon im ersten Schritt liegt darin, dass der zweite Eigenwert von M die Null ist.

- M ist "vergesslich": Es gilt

$$M^k \xrightarrow{k\to\infty} \begin{pmatrix} p_1 & p_2 & \dots & p_n \\ \vdots & & & \vdots \\ p_1 & p_2 & \dots & p_n \end{pmatrix},$$

  wobei $\boldsymbol{p} = (p_1, \dots, p_n)$ der eindeutige Gleichgewichtsvektor von M ist.

(ii) Der Beweis für den Satz von Perron beruht genau darauf, für M die Konvergenz gegen den Gleichgewichtsvektor $\boldsymbol{p}$ zu zeigen. Dies kann mit einem Kontraktionsargument durchgeführt werden, etwa unter Einsatz des Banach'schen Fixpunktsatzes aus der Analysis.

Im Fall nicht primitiver, aber noch irreduzibler Markov-Matrizen gilt eine ähnliche, aber etwas schwächere Aussage.

Satz 5.5 (Perron-Frobenius)**.** Sei M eine irreduzible Markov-Matrix. Dann existiert wieder **genau ein** Wahrscheinlichkeitsvektor $\boldsymbol{p} = (p_1, \dots, p_n)$, der LEV zum EW 1 ist. Dabei sind alle Einträge positiv, $p_i > 0$, und alle übrigen EWe λ von M erfüllen $|\lambda| \leq 1$. Im Fall $|\lambda| = 1$ ist λ eine Einheitswurzel, und die EWe mit $|\lambda| = 1$ bilden eine zyklische Gruppe. Kein anderer Eigenvektor ist ein Wahrscheinlichkeitsvektor, und alle EWe λ mit $|\lambda| = 1$ sind einfach.

Beispiel. Als einfaches Beispiel betrachten wir wieder die symmetrische Matrix $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Sie besitzt die Eigenwerte $\lambda = 1$ und $\lambda = -1$, mit zugehörigen (Links- oder Rechts-) Eigenvektoren $v = (1, 1)$ und $w = (1, -1)$.

Bemerkung. Zwischen dem primitiven und dem irreduziblen Fall besteht der folgende wesentliche Unterschied.

- Ist M primitiv, so konvergiert jeder Wahrscheinlichkeitsvektor, den man als Startvektor verwendet, nach der obigen Bemerkung gegen den Gleichgewichtsvektor.

- Ist M irreduzibel, aber mit Periodizität, kann man in Zyklen laufen.

Was mit der letzten Aussage gemeint ist, verdeutlichen wir an obigem Beispiel.

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \begin{tikzcd} 1 \arrow[r, bend left, "1"] & 2 \arrow[l, bend left, "1"] \end{tikzcd}$$

Dann gilt für den Wahrscheinlichkeitsvektor $(1/4, 3/4)$ unter der Wirkung von M:

$$\left(\frac{1}{4}, \frac{3}{4}\right) \mapsto \left(\frac{3}{4}, \frac{1}{4}\right) \mapsto \left(\frac{1}{4}, \frac{3}{4}\right) \mapsto \left(\frac{3}{4}, \frac{1}{4}\right) \mapsto \cdots$$

Dies ist offenbar ein 2-Zyklus, während $\left(\frac{1}{2}, \frac{1}{2}\right)$ der eindeutige Gleichgewichtsvektor von M ist.