

Ausbreitung von Gerüchten – mit Markov-Ketten modellieren

NICOLE BÄUERLE, LEONI GROLL, DIETMAR GRUBER, SAMUEL NEUKIRCH, ANKE RICHERT, KARLSRUHER INSTITUT FÜR TECHNOLOGIE

Zusammenfassung: Der Artikel führt kurz in die Theorie Markovscher Ketten ein und erläutert anschließend, wie diese verwendet werden kann, um die Ausbreitung von Gerüchten in sozialen Netzwerken zu modellieren und zu analysieren.

1 Markov-Ketten

Aus der Schule ist die Bernoulli-Kette gut bekannt. Wesentlich ist hier, dass die einzelnen Versuche *stochastisch unabhängig* durchgeführt werden. Bei einer Markov-Kette ist dies gerade nicht der Fall. Dazu schauen wir uns zunächst bedingte Wahrscheinlichkeiten für Ereignisse an. Seien $A, B \subset \Omega$ Ereignisse über dem Grundraum Ω . Wenn wir bereits wissen, dass A eingetreten ist, wie groß ist dann die Wahrscheinlichkeit, dass B eintritt? Überlegungen an einzelnen Beispielen führen hier zur Definition

$$\mathbb{P}(B|A) := \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

wobei $\mathbb{P}(A) > 0$ sein sollte. Formal ist dann eine Markov-Kette eine Folge von Zufallsvariablen X_0, X_1, X_2, \dots , die Werte in einer endlichen Menge S annehmen können, so dass

$$\begin{aligned} \mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) \\ = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) \end{aligned}$$

gilt, für alle $n = 0, 1, 2, \dots$ und alle möglichen Zustände $i_k \in S, k = 0, 1, 2, \dots$. Die Wahrscheinlichkeitsverteilung des neuen Zustandes (hier zur Zeit $n + 1$) hängt also nur vom aktuellen Zustand ab (hier Zeitpunkt n) und nicht davon, wie wir dort hingekommen sind. Markov-Ketten wurden nach dem russischen Mathematiker Andrey Markov (1856-1922) benannt, der ein stochastisches Modell für die Abfolge von Konsonanten und Vokalen in Pushkins Versroman „Eugen Onegin“ entwarf.

Wir betrachten ein einfaches Beispiel: Angenommen das Wetter kann man jeden Tag durch einen der drei Zustände „sonnig“ (s), „bewölkt“ (w) und „regnerisch“ (r) beschreiben. Wir möchten wissen, mit welcher Wahrscheinlichkeit morgen die Sonne scheint. Modellieren wir das Wetter als Markov-Kette, so nehmen wir an, dass dafür nur das Wetter heute eine Rolle spielt und nicht der ganze Wetterverlauf der

letzten Woche. Außerdem sollen zur Vereinfachung die Übergangswahrscheinlichkeiten unabhängig vom Zeitpunkt des Übergangs sein (also egal ob Mai oder Juni ist). Insgesamt kann man dann die Markov-Kette durch ihre Übergangswahrscheinlichkeiten beschreiben. Z.B. können wir annehmen (oder aus Daten schätzen), dass wenn heute die Sonne scheint, dies mit einer Wahrscheinlichkeit von 70% auch morgen der Fall sein wird, es mit einer Wahrscheinlichkeit von 25% bewölkt sein wird und mit einer Wahrscheinlichkeit von 5% regnet. Diese Wahrscheinlichkeiten können wir dann in einer Matrix P zusammenfassen, also z.B.

$$P = \begin{matrix} & \begin{matrix} s & w & r \end{matrix} \\ \begin{matrix} s \\ w \\ r \end{matrix} & \begin{pmatrix} 0.7 & 0.25 & 0.05 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} \end{matrix}$$

wobei $p_{ss} = 0.7, p_{sw} = 0.25, p_{sr} = 0.05, p_{ws} = 0.3, p_{ww} = 0.4, p_{wr} = 0.3, p_{rs} = 0.1, p_{rw} = 0.4, p_{rr} = 0.5$ die jeweiligen Übergangswahrscheinlichkeiten der drei Zustände s, w, r sind. Wir lesen immer von „Zeile“ nach „Spalte“. Die Summe der Wahrscheinlichkeiten in einer Zeile muss daher immer 1 sein.

Alternativ kann man die Situation in einem Prozessdiagramm festhalten (siehe Abb. 1).

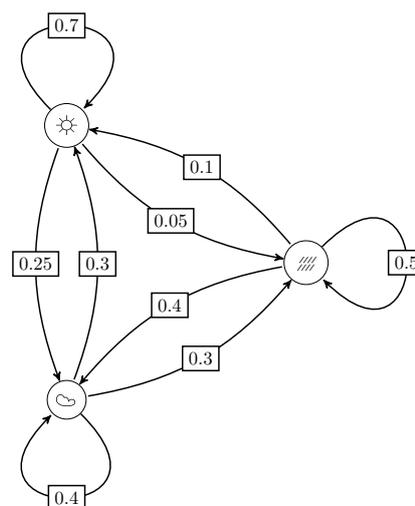


Abb. 1: Prozessdiagramm des Wetterbeispiels.

Die Knoten entsprechen hier den möglichen Zuständen und die Markierungen an den gerichte-

ten Kanten zwischen den Zuständen geben die (positiven) Übergangswahrscheinlichkeiten an. Zu jeder Übergangsmatrix gibt es ein Prozessdiagramm und umgekehrt.

Der Vorteil der Matrixschreibweise ist, dass man nun auch relativ einfach ausrechnen kann, mit welcher Wahrscheinlichkeit in fünf Tagen die Sonne scheint, wenn sie es heute tut. Dazu müssen wir nur die Matrix entsprechend oft mit sich selbst multiplizieren. Mit dem üblichen Matrixprodukt zweier Matrizen $A = (a_{ij})_{1 \leq i, j \leq n}$ und $B = (b_{ij})_{1 \leq i, j \leq n}$ gilt

$$AB = C = (c_{ij})_{1 \leq i, j \leq n}, \quad c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}.$$

Bei unserem Beispiel enthält dann die Matrix $PP = P^2$ die Zwei-Schritt Übergangswahrscheinlichkeiten. Die Elemente werden mit $p_{ij}^{(2)}$ bezeichnet (nicht zu verwechseln mit p_{ij}^2). Dies kann man wie folgt einsehen: Nehmen wir an, dass heute (Zeitpunkt 0) die Sonne scheint, also $\mathbb{P}(X_0 = s) = 1$. Dann gilt mit der Definition der bedingten Wahrscheinlichkeit und der Eigenschaft der Markov-Kette:

$$\begin{aligned} \mathbb{P}(X_2 = s | X_0 = s) &= \frac{\mathbb{P}(X_2 = s, X_0 = s)}{\mathbb{P}(X_0 = s)} \\ &= \mathbb{P}(X_2 = s, X_0 = s) \\ &= \sum_{k \in \{s, w, r\}} \mathbb{P}(X_2 = s, X_1 = k, X_0 = s) \\ &= \sum_{k \in \{s, w, r\}} \mathbb{P}(X_1 = k | X_0 = s) \mathbb{P}(X_2 = s | X_1 = k, X_0 = s) \\ &= \sum_{k \in \{s, w, r\}} p_{sk} p_{ks} = p_{ss}^{(2)}. \end{aligned}$$

Für P^n kann man dann die entsprechende Aussage mit vollständiger Induktion zeigen. Insbesondere gilt (dies kann man mit „R“ oder einer anderen Software berechnen):

$$P^5 = \begin{pmatrix} 0.44307 & 0.330475 & 0.226455 \\ 0.340923 & 0.33979 & 0.25098 \\ 0.38961 & 0.34519 & 0.2652 \end{pmatrix},$$

d.h. wenn heute die Sonne scheint, ist die Wahrscheinlichkeit, dass in fünf Tagen ebenfalls die Sonne scheint, gerade 0.44307.

In der Theorie der Markov-Ketten interessiert man sich hauptsächlich dafür, wie die Verteilung von X_n aussieht, wenn n groß ist. Dazu kann man im ersten Anlauf folgende Analyse durchführen: Gilt für je zwei verschiedene Zustände i und j , dass es natürliche Zahlen n und m gibt mit $p_{ij}^{(n)} > 0$ und $p_{ji}^{(m)} > 0$, d.h. man kommt mit positiver Wahrscheinlichkeit

von i irgendwann nach j und wieder zurück, so nennt man die Markov-Kette *irreduzibel*. Stellt man ein Prozessdiagramm auf, so sieht man leicht, dass die Markov-Kette irreduzibel ist, wenn es für je zwei verschiedene Zustände i und j einen Weg von i nach j und wieder zurück gibt. In diesem Fall (siehe z.B. das Wettermodell), ergibt sich langfristig eine *stationäre Verteilung* auf den Zuständen, die nicht mehr davon abhängt, wo die Kette gestartet ist, d.h. es gilt bei drei Zuständen:

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \end{pmatrix}. \quad (1)$$

Dabei gibt π_i die durchschnittliche Zeit an, die die Markov-Kette langfristig im Zustand i verbringt. Diese stationäre Verteilung $\pi = (\pi_i)_{i \in S}$ kann man allgemein aus dem linearen Gleichungssystem

$$\begin{aligned} \pi P &= \pi \\ \sum_{i \in S} \pi_i &= 1 \end{aligned}$$

bestimmen, mit der Nebenbedingung, dass alle $\pi_i \geq 0$ sind. Im Wetterbeispiel also:

$$\begin{aligned} 0.7\pi_1 + 0.3\pi_2 + 0.1\pi_3 &= \pi_1 \\ 0.25\pi_1 + 0.4\pi_2 + 0.4\pi_3 &= \pi_2 \\ 0.05\pi_1 + 0.3\pi_2 + 0.5\pi_3 &= \pi_3 \\ \pi_1 + \pi_2 + \pi_3 &= 1, \end{aligned}$$

mit $\pi_1, \pi_2, \pi_3 \geq 0$. Die Lösung kann man mit dem Gauß-Verfahren bestimmen und ist hier (gerundet):

$$\pi_1 = 0.419, \quad \pi_2 = 0.337, \quad \pi_3 = 0.244.$$

In dem Beispiel scheint also langfristig an knapp 42% der Tage die Sonne. Vergleicht man die stationäre Verteilung π mit P^5 sieht man, dass die Zeilen schon relativ nahe an π sind. Tatsächlich zeigt die Theorie, dass die Konvergenz in (1) exponentiell schnell erfolgt.

Gibt es jedoch absorbierende Zustände, d.h. Zustände in denen man mit Wahrscheinlichkeit 1 bleibt, und sind die Übergangswahrscheinlichkeiten aus den restlichen Zuständen alle positiv, so wird die Markov-Kette irgendwann in einem absorbierenden Zustand landen. Aus einem linearen Gleichungssystem kann man ausrechnen, mit welchen Wahrscheinlichkeiten eine Markov-Kette, die in einem bestimmten (nicht-absorbierenden) Zustand startet, in den einzelnen absorbierenden Zuständen landet (vgl. dazu z.B. Kapitel 5.3 in Tijms (2012)).

2 Modelle zur Ausbreitung von Gerüchten

Im Folgenden soll die Theorie der Markov-Ketten verwendet werden, um die Ausbreitung von Gerüchten in sozialen Netzwerken zu modellieren und zu analysieren. In einem sozialen Netzwerk entsprechen die Mitglieder den Knoten, und es existieren Kanten zwischen zwei Mitgliedern, wenn diese in direktem Kontakt stehen. Interessiert man sich dafür, ob Mitglieder an ein Gerücht glauben oder nicht, so wird jedes Mitglied durch zwei Zustände („glaubt an das Gerücht“, „glaubt nicht an das Gerücht“) repräsentiert. Möchte man ein großes Netzwerk mit z.B. 100 Mitgliedern modellieren, so wird schnell klar, dass hierfür 2^{100} Zustände notwendig sind. Um die Situation zu vereinfachen, muss man also von den einzelnen Mitgliedern abstrahieren und Gruppen bilden. Als Ausgangspunkt kann hier das im Folgenden beschriebene Wright-Fisher-Modell dienen, bei dem man davon ausgeht, dass jedes Mitglied mit jedem anderen in Kontakt steht, und man sich nur dafür interessiert, wieviele Mitglieder an das Gerücht glauben, aber nicht welches diese im einzelnen sind.

2.1 Das Wright-Fisher-Modell

Das Modell von Wright-Fisher geht auf die Arbeiten von Wright (1939) und Moran (1958) zurück. Es wurde ursprünglich in der Populationsgenetik eingeführt, um die Entwicklung der Genhäufigkeit in einer Population haploider Individuen zu studieren. Für eine einführende Darstellung siehe Lanchier (2017), Kapitel 12. Wir verwenden es, um zu untersuchen, ob sich ein Gerücht in einem Netzwerk durchsetzt oder wieder verschwindet. Dazu nehmen wir an, dass ein Netzwerk der Größe $N \in \mathbb{N}$ vorliegt. Jedes Mitglied kann entweder „gesund“ („glaubt nicht an das Gerücht“) oder „infiziert“ („glaubt an das Gerücht“) sein und kennt die Meinung aller anderen Teilnehmer. Der Zustandsraum der zugehörigen Markov-Kette ist $S = \{0, 1, \dots, N\}$ und gibt die Anzahl der „infizierten“ Mitglieder an. Sind heute $i \in S$ Mitglieder infiziert, dann soll mit Wahrscheinlichkeit

$$p = \frac{i}{N}$$

morgen jedes einzelnen Mitglied unabhängig von den anderen infiziert sein. Diese Annahme drückt aus, dass die Wahrscheinlichkeit eines Mitglieds an das Gerücht zu glauben davon abhängt, wieviele andere Mitglieder bereits das Gerücht teilen. Dabei werden die einzelnen Mitglieder nicht unterschieden.

Die Anzahl der infizierten Mitglieder morgen ist also binomialverteilt mit Parameter N und $p = \frac{i}{N}$ und die Wahrscheinlichkeit, dass morgen $j \in S$ Mitglieder infiziert sind, ist demnach gerade

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}.$$

Diese Wahrscheinlichkeit hängt also davon ab, wieviele Mitglieder heute schon infiziert sind. Für ein Netzwerk von $N = 4$ ergibt sich z.B. als Übergangsmatrix

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{81}{256} & \frac{108}{256} & \frac{54}{256} & \frac{12}{256} & \frac{1}{256} \\ \frac{1}{16} & \frac{4}{16} & \frac{6}{16} & \frac{4}{16} & \frac{1}{16} \\ \frac{1}{256} & \frac{12}{256} & \frac{54}{256} & \frac{108}{256} & \frac{81}{256} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}.$$

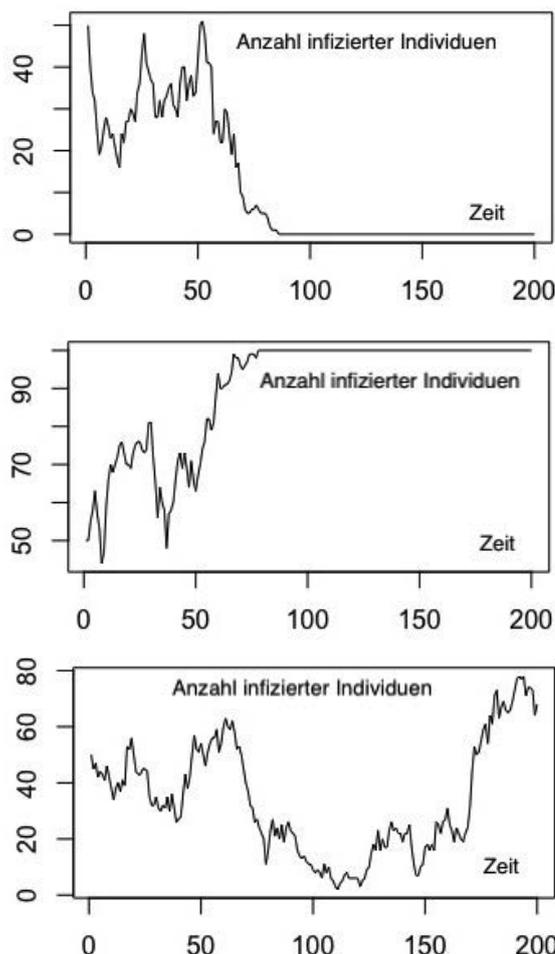


Abb. 2: Simulierte Pfade im Wright-Fisher-Modell mit $N = 100$.

Man sieht schnell, dass es in diesem Modell unabhängig von N immer zwei absorbierende Zustände gibt, nämlich 0 und N , d.h. langfristig ist das Gerücht aus dem Netzwerk verschwunden, oder das gesamte Netzwerk glaubt an das Gerücht. Mit der Software „R“ kann man nun einfach Pfade des Wright-Fisher-Modells simulieren.

Wir setzen dazu $N = 100$ und wählen als Startzustand $i = 50$, also zu Beginn glaubt die Hälfte der Mitglieder an das Gerücht. Abbildung 2 zeigt drei simulierte Pfade. Auf der x -Achse sind die Anzahl der Übergänge abgetragen. Man könnte hier z.B. in Tagen zählen. Auf der y -Achse ist die Anzahl der Mitglieder abgetragen, die zu diesem Zeitpunkt an das Gerücht glauben. Die drei Pfade zeigen mögliche unterschiedliche Verläufe.

Für einen gegebenen Anfangszustand ist es in diesem Modell möglich zu berechnen, mit welcher Wahrscheinlichkeit am Ende das gesamte Netzwerk an das Gerücht glaubt. Diese ist $\frac{i}{N}$, wenn der Anfangszustand i ist. Im vorangegangenen Simulationsbeispiel also gerade $\frac{1}{2}$.

Eine weitere interessante Frage ist die mittlere Zeitdauer bis zur Absorption. Diese ist im Einzelfall nur durch Monte Carlo Simulation zu bestimmen. Theoretisch gibt es dazu Approximationen, wenn das Netzwerk sehr groß ist. Glauben zu Beginn gerade die Hälfte der Mitglieder an das Gerücht, so ist die mittlere Zeit bis zur Absorption etwa $1,386 \times N$. In unserem Beispiel benötigt man also im Mittel etwa 139 Übergänge bis zur Absorption (vgl. Lanchier (2017), Kap 12).

Die „Infektionswahrscheinlichkeit“ p eines Mitglieds kann man auch modifizieren, um der Glaubwürdigkeit des Gerüchts Rechnung zu tragen. So kann man z.B. ein Gewicht $\phi > 0$ einführen und damit

$$p = \frac{i\phi}{i\phi + N - i}$$

definieren. Für $\phi = 1$ entspricht dies gerade dem Ausgangsmodell. Für ϕ gegen Unendlich geht die Wahrscheinlichkeit gegen 1 und für ϕ gegen Null geht sie gegen Null.

2.2 Eigene Überlegungen

Nachteile des Wright-Fisher-Modells sind, wie bereits diskutiert, die Tatsachen, dass einzelne Mitglieder nicht unterschieden werden und dass angenommen wird, dass alle Mitglieder gleichermaßen vernetzt sind.

Deshalb werden im Folgenden Modifikationen des Modells betrachtet, mit dem Ziel, dann die Wahrscheinlichkeit zu bestimmen, dass nach einer beliebigen Anzahl an Übergängen (z.B. Tagen) ein bestimmtes Mitglied an das Gerücht glaubt, wenn zu Beginn das Gerücht von einem festen Mitglied gestreut wird. In den verschiedenen Netzwerken, die betrachtet werden, befinden sich stets $N = 64$ Mitglieder. Die Anzahl wurde gewählt, da die Mitglieder dann leicht gruppiert werden können.

2.2.1 Modell 1:

In Modell 1 sind alle 64 Mitglieder wie im Wright-Fisher-Modell vollständig vernetzt (siehe Abb. 3). Wir nehmen an, dass Mitglied 1 das Gerücht streut, also insbesondere $X_0 = 1$ ist. Wir können dann durch P^n die n -Schritt-Übergangswahrscheinlichkeiten für die Anzahl der Mitglieder, die an das Gerücht glauben, ausrechnen. Da zu Beginn ein Mitglied an das Gerücht glaubt, interessiert uns davon die zweite Zeile: Sie enthält die Wahrscheinlichkeiten $p_{1j}^{(n)}$, dass nach n Übergängen $j = 0, 1, 2, \dots, N$ Mitglieder an das Gerücht glauben. Diese j Mitglieder werden dann zufällig ausgewählt. Die bedingte Wahrscheinlichkeit, dass unser festes Mitglied an das Gerücht glaubt, kann dann mit der *Hypergeometrischen Verteilung* bestimmt werden:

$$\frac{\binom{1}{1} \binom{N-1}{j-1}}{\binom{N}{j}} = \frac{j}{N}.$$

Insgesamt berechnet sich also die Wahrscheinlichkeit, dass das feste Mitglied nach n Übergängen an das Gerücht glaubt, mit dem *Satz über die totale Wahrscheinlichkeit* (dabei ist j/N die bedingte Wahrscheinlichkeit) zu

$$\sum_{j=0}^N \mathbb{P}(X_n = j | X_0 = 1) \frac{j}{N}. \quad (2)$$

Interessanterweise ergibt sich bei dieser Formel unabhängig von n immer $\frac{1}{N}$. Das kann man für konkretes N nachrechnen und durch vollständige Induktion über n zeigen. Eleganter, aber zugegeben auf höherem Niveau, sieht man das wie folgt ein: Zunächst ist im Wright-Fisher-Modell die bedingte Verteilung von X_{n+1} gegeben X_0, \dots, X_n gerade $B(N, \frac{X_n}{N})$. Also gilt

$$\mathbb{E}[X_{n+1} | X_0, \dots, X_n] = N \frac{X_n}{N} = X_n.$$

D.h. die Folge (X_n) ist nicht nur eine Markov-Kette, sondern auch ein *Martingal*. Martingale haben aber einen konstanten Erwartungswert, und da wir hier

mit $X_0 = 1$ starten, ist also immer $\mathbb{E}[X_n] = 1$. Damit bekommen wir mit

$$\begin{aligned} \sum_{j=1}^N \mathbb{P}(X_n = j | X_0 = 1) \frac{j}{N} &= \frac{1}{N} \sum_{j=1}^N \mathbb{P}(X_n = j | X_0 = 1) j \\ &= \frac{1}{N} \mathbb{E}[X_n] = \frac{1}{N} \end{aligned}$$

die Behauptung. Bei 64 Mitglieder beträgt also die Wahrscheinlichkeit, dass nach n Tagen ein bestimmtes Mitglied an das Gerücht glaubt lediglich $\frac{1}{64} = 1,5625\%$. Dies entspricht auch der Wahrscheinlichkeit, dass langfristig alle Mitglieder im Netzwerk an das Gerücht glauben.

2.2.2 Modell 2:

Im Modell 2 sind die 64 Mitglieder in 16 Gruppen aufgeteilt, wobei jeweils vier Gruppen wieder zu einer Übergruppe zusammengefasst werden (siehe Abb. 3). Es gibt also 4 Übergruppen.

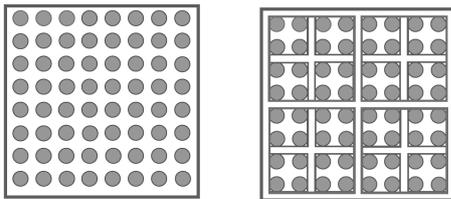


Abb. 3: Modell 1 (links) und Modell 2 (rechts) in graphischer Darstellung.

Das Modell soll eine Situation abbilden, in der nicht alle Mitglieder des Netzwerks in direktem Kontakt stehen, sondern Cliquenbildungen erfolgen. Damit sich ein Gerücht durchsetzen kann, muss es sich zunächst in einer Gruppe langsam durchsetzen. Anschließend wird eine gewisse Dynamik unterstellt, mit der es von einer Gruppe zur nächsten springen kann. Hierzu wird wieder das Wright-Fisher-Modell verwendet. Zu Beginn übernimmt Mitglied 1 in Gruppe 1 das Gerücht und es breitet sich in der ersten Gruppe nach dem Wright-Fisher-Modell aus. Glauben mindestens zwei Mitglieder dieser Gruppe an das Gerücht, so gilt die Gruppe als infiziert und es können weitere Gruppen in dieser Übergruppe nach dem Wright-Fisher-Modell infiziert werden. Die Ausbreitung des Gerüchts innerhalb der einzelnen Gruppen 1, 2, 3 und 4 wird ab diesem Zeitpunkt nicht mehr modelliert. Stattdessen wird (willkürlich) angenommen, dass in einer infizierten Gruppe die

Hälfte der Mitglieder an das Gerücht glauben. Es ergibt sich hier folgende Übergangsmatrix, wenn wir uns auf die erste Übergruppe beschränken:

$$\begin{matrix} & 0_P & 1_P & 2_P & 0_G & 1_G & 2_G & 3_G & 4_G \\ \begin{matrix} 0_P \\ 1_P \\ 2_P \\ 0_G \\ 1_G \\ 2_G \\ 3_G \\ 4_G \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{81}{256} & \frac{108}{256} & \frac{67}{256} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{81}{256} & \frac{108}{256} & \frac{54}{256} & \frac{12}{256} & \frac{1}{256} \\ 0 & 0 & 0 & \frac{1}{16} & \frac{4}{16} & \frac{6}{16} & \frac{1}{16} & \frac{1}{16} \\ 0 & 0 & 0 & \frac{1}{256} & \frac{12}{256} & \frac{54}{256} & \frac{108}{256} & \frac{81}{256} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Dabei ist 0_P (1_P) der Zustand, dass kein (ein) Mitglied in der ersten Gruppe an das Gerücht glaubt. 2_P ist der Zustand, dass mindestens zwei Mitglieder in der ersten Gruppe an das Gerücht glauben. Die Zustände i_G bezeichnen die Situation, dass i Gruppen der ersten Übergruppe an das Gerücht glauben.

Zu beachten ist, dass die Teilmatrix auf den Zuständen $1_G, \dots, 4_G$ stochastisch ist. Befindet sich die Markov-Kette also zum ersten Mal in einem dieser Zustände, so wird nur noch die Übertragung zwischen den einzelnen Gruppen modelliert, nicht aber innerhalb der Gruppe. Die Zeitskala bleibt jedoch ungeachtet davon dieselbe, z.B. findet pro Tag ein Übergang statt. Dadurch wird automatisch die zunehmende Dynamik des Prozesses modelliert (ähnlich wie in der aktuellen COVID-19-Pandemie).

Ist das Zielmitglied nun in der ersten Gruppe, so verwenden wir Modell 1 und die Wahrscheinlichkeit ist gerade $\frac{1}{4}$, dass das Mitglied nach n Schritten an das Gerücht glaubt (unabhängig von n). Ist das Zielmitglied in einer anderen Gruppe dieser Übergruppe, muss das Gerücht zuerst auf diese Gruppe überspringen. Wir nehmen dann an, dass zwei Mitglieder in dieser Gruppe an das Gerücht glauben, also mit Wahrscheinlichkeit $\frac{1}{2}$ dann ein bestimmtes. Ist unser Zielmitglied in der Gruppe 2, 3 oder 4, so ist die Wahrscheinlichkeit, dass es nach n Übergängen an das Gerücht glaubt

$$\left[\sum_{j=1}^4 \mathbb{P}(X_n = j_G | X_0 = 1_P) \frac{j}{4} - \mathbb{P}(X_n = 1_G, X_{n-1} = 2_P | X_0 = 1_P) \frac{1}{4} \right] \frac{1}{2}$$

Die erste Zeile gibt (analog zu Formel (2)) die Wahrscheinlichkeit an, dass die richtige Gruppe in der das Zielmitglied ist, zu den infizierten gehört (diese werden wieder zufällig ausgewählt). Der Faktor $\frac{1}{2}$ hinter der Klammer trägt der Tatsache Rechnung, dass

unabhängig davon die Hälfte der Gruppenmitglieder (per Definition) infiziert sind, unser bestimmtes also mit Wahrscheinlichkeit $\frac{1}{2}$. Allerdings gibt es eine Ausnahme: Wenn zum ersten Mal eine Gruppe infiziert ist, ist es per Definition Gruppe 1. Deshalb dürfen wir hier nicht das Ereignis $\{X_n = 1_G\}$ betrachten, sondern $\{X_n = 1_G, X_{n-1} \neq 2_P\}$. Die Wahrscheinlichkeit in der zweiten Zeile muss daher abgezogen werden. Diese Wahrscheinlichkeiten können in Abhängigkeit von n dann mit einem Programm bestimmt werden. Die Übertragung des Gerüchts auf die Übergruppen kann analog modelliert werden.

Mit dieser Idee ist es nun möglich, verschiedene Cliques zu bilden und die Übertragung von Gerüchten zu modellieren und die entsprechenden Wahrscheinlichkeiten zu berechnen oder simulativ zu bestimmen und zu vergleichen. Wie oben im Wright-Fisher-Modell erwähnt, kann das Gerücht mit einer Glaubwürdigkeit (Gewicht) versehen werden, das die Übertragungswahrscheinlichkeiten erhöht. Generell ist zu beobachten, dass die Wahrscheinlichkeit für die Ausbreitung des Gerüchts abnimmt, je mehr Teilgruppen bestehen.

3 Schlussbemerkungen

Das Thema „Markov-Ketten“ eignet sich als Vertiefung für interessierte und begabte Schüler, als einfache Erweiterung der Bernoulli-Kette. Einführende Bücher sind z.B. Lanchier (2017), Tijms (2012) oder Waldmann und Stocker (2004). Neben der bedingten Wahrscheinlichkeit wird lediglich die Matrixmultiplikation und das Lösen von linearen Gleichungssystemen benötigt. Die Verwendung von Software wie z.B. „R“ erlaubt die Berechnung und Visualisierung von Wahrscheinlichkeiten.

Die vorgestellten Modelle waren Teil eines Projekts, welches die Schüler Leoni und Samuel im Rahmen der Kooperationsphase am Hector Seminar Karlsruhe bearbeitet haben. Das Hector Seminar ist eine Einrichtung zur Förderung hochbegabter junger Menschen mit einem Enrichment-Angebot in den Bereichen Mathematik, Informatik, Naturwissenschaften und Technik (MINT). Das Angebot des Hector Seminars ist eine Ergänzung zum schulischen Angebot der Gymnasien, beginnend von der Klassenstufe

6 bis zur gymnasialen Oberstufe. In der das Hector Seminar abschließenden Kooperationsphase im ersten Oberstufenschuljahr lernen die Jugendlichen an Universitäts- und Forschungsinstituten das wissenschaftliche Arbeiten im MINT-Bereich kennen. Nach der Wahl eines Projektes arbeiten sie eng mit Wissenschaftlern an aktuellen Problemen der jeweiligen Arbeitsgruppen.

Literatur

- Lanchier, N. (2017): Stochastic modeling. Springer-Verlag.
 Moran, P.A.P. (1958): Random processes in genetics. Proc. Cambridge Philos. Soc. 54, 60-71.
 Tijms, H. (2012): Understanding probability. Cambridge University Press 3rd ed.
 Waldmann, K.H. und Stocker U.M. (2004): Stochastische Modelle. Eine anwendungsorientierte Einführung. Springer, Berlin.
 Wright, S. (1939): Statistical genetics in relation to evolution. Nummer 802-803 in Actualités scientifique et industrielles. Hermann.

Anschrift der Verfasser

Prof. Dr. Nicole Bäuerle

Institut für Stochastik

Karlsruher Institut für Technologie

76128 Karlsruhe

baeuerle@kit.edu

Leoni Groll

Albert-Braun Straße 14a

76189 Karlsruhe

Dietmar Gruber

Markgrafen-Gymnasium

Gymnasiumstraße 1-3

76227 Karlsruhe

gruber@hector-seminar.de

Samuel Neukirch

Murgstraße 25

76199 Karlsruhe

Anke Richert

Edith-Stein-Gymnasium

Breitenbachweg 15

75015 Bretten

richert@hector-seminar.de