Das Problem der vollständigen Serie

Eine Bearbeitung unter Verwendung von Markoff-Ketten

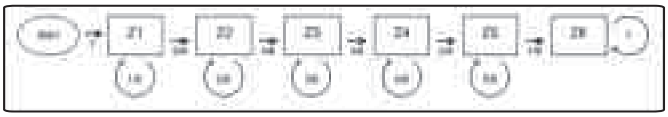
Dr. Ulrich Döring

Wechselt ein System von Beobachtung zu Beobachtung mit konstanten Wahrscheinlichkeiten zwischen endlich vielen Zuständen, so spricht man von einer **Markoff-Kette**. Unter diesem Gesichtspunkt kann man auch das Problem der vollständigen Serie (auch Sammelbilderproblem genannt) betrachten. Ein anschauliches Beispiel dazu ist: Werfen eines Würfels bis alle Augenzahlen von 1 bis 6 mindestens einmal aufgetreten sind. Es handelt sich hierbei um eine klassische Aufgabenstellung aus der Wahrscheinlichkeitsrechnung, mit der sich schon berühmte Mathematiker wie Abraham de Moivre, Leonhard Euler und Pierre Simon Laplace beschäftigt haben.

Es bietet sich an, den Formalismus der Markoff-Ketten als „Bindeglied" zwischen der Linearen Algebra und der Wahrscheinlichkeitsrechnung zu unterrichten. Im Folgenden wird das oben angeführte Würfelbeispiel behandelt. In diesem Zusammenhang ergibt sich eine Reihe von Fragen, von denen drei untersucht werden:

- Wie viele Würfe sind im Mittel notwendig, um eine vollständige Serie zu erhalten?
- Wie sieht die Wahrscheinlichkeitsverteilung dazu aus?
- Wie viele Würfe muss man durchführen, um mit einer Wahrscheinlichkeit von mehr als 50 % eine vollständige Serie zu erhalten?

Der Vorgang lässt sich am besten durch einen sog. **Übergangsgraphen** veranschaulichen:



Der Zustand Z4 gibt dabei z. B. an, dass 4 verschiedene Augenzahlen bis zu diesem Zeitpunkt aufgetreten sind. Mit einer Wahrscheinlichkeit von 4/6 würfelt man dann eine bereits erhaltene Augenzahl und mit einer Wahrscheinlichkeit von 2/6 erzielt man eine neue Augenzahl. Im Zustand Z6 hat man die vollständige Serie erhalten. Einen solchen Endzustand nennt man **absorbierend** (erkennbar am Ringpfeil mit der Wahrscheinlichkeit 1). Alle anderen Zustände nennt man **innere Zustände**.

Computeralgebrasysteme eröffnen die Möglichkeit, Zufallsexperimente zu simulieren. Dies kann im vorliegenden Fall z. B. mithilfe der im Folgenden angegebenen geschachtelten when-Anweisung geschehen:

ziehen(vorhanden,gesamt):=
when(vorhanden=6, gesamt, when(rand int(1,6)>vorhanden,
ziehen(vorhanden+1, gesamt+1), ziehen(vorhanden, gesamt+1)))

Mithilfe der Variablen „vorhanden" wird die Anzahl der unterschiedlichen gewürfelten Augenzahlen gezählt, während mithilfe der Variablen „gesamt" die Anzahl der Würfe ermittelt wird.[1]

Simuliert man jeweils das Werfen von 1000 vollständigen Serien und bestimmt den Mittelwert der benötigten Wurfanzahl, so erkennt man, dass im Mittel etwas unter 15 Würfe notwendig sind, um eine vollständige Serie zu erzielen. Wie sieht nun die theoretische Lösung zu diesem Problem aus?

$$\frac{\mathrm{sum}(\mathrm{seq}(ziehen(0,0),n,1,1000))}{1000} \qquad 14.438$$

$$\frac{\mathrm{sum}(\mathrm{seq}(ziehen(0,0),n,1,1000))}{1000} \qquad 14.532$$

$$\frac{\mathrm{sum}(\mathrm{seq}(ziehen(0,0),n,1,1000))}{1000} \qquad 14.758$$

Die Frage, wie viele Würfe im Mittel notwendig sind, um eine vollständige Serie zu erhalten, ist in der Terminologie der Wahrscheinlichkeitsrechnung die Frage nach dem **Erwartungswert** und in der Terminologie der Markoff-Ketten die Frage nach der **mittleren Schrittzahl** oder **mittleren Wartezeit**. Es werden zwei unterschiedliche Darstellungen des Lösungsweges vorgestellt.

Von zentraler Bedeutung bei der Berechnung einer Markoff-Kette ist die sog. **Übergangsmatrix**. Für das oben angeführte Problem der vollständigen Serie sieht diese Matrix folgendermaßen aus:[2]

$$\begin{array}{r} \textit{nach} \\ S \\ Z1 \\ Z2 \\ U = \textit{von}\ Z3 \\ Z4 \\ Z5 \\ Z6 \end{array} \begin{array}{c} \begin{array}{ccccccc} S & Z1 & Z2 & Z3 & Z4 & Z5 & Z6 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 5/6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2/6 & 4/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3/6 & 3/6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4/6 & 2/6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Dabei steht z. B. in der i-ten Zeile und k-ten Spalte die Wahrscheinlichkeit dafür, vom Zustand Zi zum Zustand Zk zu gelangen. Die Übergangsmatrix ist eine sog. stochastische Matrix, bei der die Summe der Wahrscheinlichkeiten in einer Zeile jeweils 1 ist. Die mittlere Schrittzahl, um von einem der inneren Zustände in den absorbierenden Zustand Z6 zu gelangen, berechnet man jetzt nach der sog. **2. Mittelwertsregel**. Der Grundgedanke sei am Beispiel des Übergangs von Z5 nach Z6 kurz skizziert: Vom Zustand Z5 benötigt man mit der Wahrscheinlichkeit $\frac{1}{6}$ noch 1 Schritt bis Z6 und mit der Wahrscheinlichkeit $\frac{5}{6}$ noch 1 Schritt und zusätzlich die Anzahl Schritte, die von Z5 notwendig sind. Es gilt also:

$m_5 = \frac{1}{6} \cdot 1 + \frac{5}{6} \cdot (1 + m_5) = 1 + \frac{5}{6} \cdot m_5$. Die Argumentation ist auf alle absorbierenden Markoff-Ketten übertragbar und es gilt:

> In einer absorbierenden Markoff-Kette mit n Zuständen und der Übergangsmatrix U gilt für die mittlere Schrittzahl m_i in Schritten von einem inneren Zustand Z_i bis zur Absorption: $m_i = 1 + m_1 \cdot u_{i1} + m_2 \cdot u_{i2} + \ldots + m_i \cdot u_{ii} + \ldots + m_n \cdot u_{in}$. Dabei ist $m_k=0$ für alle absorbierenden Zustände Z_k zu setzen.

Angewendet auf unser Problem resultiert folgendes LGS: Man erhält so die mittleren Schrittzahlen von allen Zuständen aus. Vom Start aus benötigt man im Mittel 14,7 Würfe, um eine vollständige Serie zu erzielen.

$$\text{solve}\left(\begin{cases} ms=1+m1 \\ m1=1+\frac{1}{6}\cdot m1+\frac{5}{6}\cdot m2 \\ m2=1+\frac{2}{6}\cdot m2+\frac{4}{6}\cdot m3 \\ m3=1+\frac{3}{6}\cdot m3+\frac{3}{6}\cdot m4 \\ m4=1+\frac{4}{6}\cdot m4+\frac{2}{6}\cdot m5 \\ m5=1+\frac{5}{6}\cdot m5 \end{cases},\{ms,m1,m2,m3,m4,m5\}\right)$$

$$m1=\frac{137}{10} \text{ and } m2=\frac{25}{2} \text{ and } m3=11 \text{ and } m4=9 \text{ and } m5=6 \text{ and } ms=\frac{147}{10}$$

2. Bei komplexeren Markoff-Ketten ist ein Lösungsweg zu bevorzugen, der nur auf der Matrizenrechnung basiert. Es besteht ein einfacher Zusammenhang zwischen dem LGS und dem Teil der Übergangsmatrix U, der zu den inneren Zuständen gehört:

$$\begin{pmatrix} m_S & m_1 & m_2 & m_3 & m_4 & m_5 \end{pmatrix} = \begin{pmatrix} m_S & m_1 & m_2 & m_3 & m_4 & m_5 \end{pmatrix} \cdot$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 5/6 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 2/3 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 5/6 \end{pmatrix}$$

$$+\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Der Zusammenhang mit dem LGS wird schnell deutlich, wenn man die rechte Seite ausführlich schreibt.

Die mittleren Schrittzahlen m_s, m_1, ..., m_5 werden zu einem Vektor $\vec{m}$ zusammengefasst. Die Matrix, die zu den inneren Zuständen gehört, taucht in transponierter Form auf (weil hier $\vec{m}$ als Zeilenvektor verwendet wird) und wird mit Q bezeichnet. $\vec{1}$ ist der Vektor mit so vielen Einsen, wie es innere Zustände gibt. Die Matrizengleichung lässt sich folgendermaßen formalisieren und lösen:

$\vec{m} = \vec{m} \cdot Q + \vec{1} \Rightarrow \vec{m} = \vec{1} \cdot (E-Q)^{-1} = \vec{1} \cdot F$. Dabei ist $F=(E-Q)^{-1}$ die sog. **Fundamentalmatrix**. Die Berechnung des Zeilenvektors $\vec{m}$ mit dem TI-NspireCAS zeigt der folgende Ausschnitt.

$$\begin{bmatrix} \frac{147}{10} & \frac{137}{10} & \frac{25}{2} & 11 & 9 & 6 \end{bmatrix}$$

Wie sieht nun die Wahrscheinlichkeitsverteilung zum Problem der vollständigen Serie aus?

Geht man von der Anfangsverteilung v0=[1,0,0,0,0,0,0] aus, so kann man unter Verwendung der Übergangsmatrix „matu" mit dem CAS-Befehlen v0·matu^6, v0·matu^7 die Wahrscheinlichkeiten berechnen, mit denen sich das System nach 6, 7, … Würfen in den Zuständen S, Z1, Z2, ..., Z6 befindet. Im Folgenden sind die ersten beiden Zeilen der entstandenen Matrizen angegeben:

$v0\cdot matu^6$	[0. 0.000129 0.019933 0.200617 0.501543 0.231481 0.015432]
$v0\cdot matu^7$	[0. 0.000021 0.006752 0.120029 0.460105 0.358882 0.054012]

Man erkennt, dass die Wahrscheinlichkeit dafür, dass man schon mit 6 Würfen eine vollständige Serie erzielt hat, ungefähr 1,5 % beträgt. Nach 7 Würfen hat man bereits mit einer Wahrscheinlichkeit von ca. 5,4 % alle Zahlen gewürfelt. Wünschenswert wäre, dass TI den seq-Befehl auch für das Arbeiten mit Matrizen verfügbar macht, damit alle Absorptionswahrscheinlichkeiten durch eine Eingabe ermittelt werden können.

Die Zahlenwerte für die Wahrscheinlichkeiten werden jetzt in die Tabellenkalkulation übertragen. Der abgebildete Ausschnitt von Tab. 1 zeigt in der 1. Spalte die Wurfanzahlen 6, 7, …, 40 (erzeugt mithilfe der Eingabe =seq(n,n,6,40)) und in der 2. Spalte die Absorptionswahrscheinlichkeiten nach 6, 7, …, 40 Würfen. Die Datenlisten werden hier mit „lin" bzw. „liw" bezeichnet. Anschließend geht man auf den Menuepunkt 3: Daten, 5: Häufigkeitsplot und ordnet zu: Datenliste: lin, Häufigkeitsliste: liw. Man erhält die sog. **kumulierte Wahrscheinlichkeitsverteilung** (s. Abb. 1). Sollte das Histogramm nicht so aussehen, wählt man unter Histogrammeigenschaften die Breite 1 und die Ausrichtung -0.5.

	lin	liw	lid
◆	=seq(n,n,6,40)		
1	6	0.015432	0.015432
2	7	0.054012	0.03858
3	8	0.114026	0.060014
4	9	0.188043	0.074017
5	10	0.271812	0.083769
6	11	0.356206	0.084394
7	12	0.437816	0.08161
8	13	0.513858	0.076043
9	14	0.582845	0.068987
10	15	0.644213	0.061368
11	16	0.698004	0.053791
12	17	0.744832	0.046828
13	18	0.784707	0.039875
14	19	0.818925	0.034218
15	20	0.848008	0.029083

Tab 1: Wurfzahlen n, Absorptionswahrscheinlichkeiten und Wahrscheinlichkeiten für eine vollständige Serie mit genau n Würfen.

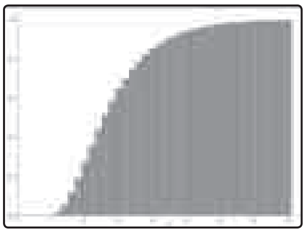


Abb. 1: Kumulierte Wahrscheinlichkeitsverteilung zum Problem der vollständigen Serie.

Man erkennt an der kumulierten Wahrscheinlichkeitsverteilung, dass bereits nach dem 13. Wurf die Wahrscheinlichkeit größer als 50 % ist, eine vollständige Serie zu würfeln. Nach 40 Würfen beträgt die Wahrscheinlichkeit für das Erzielen einer vollständigen Serie bereits gerundet 99,6 %.

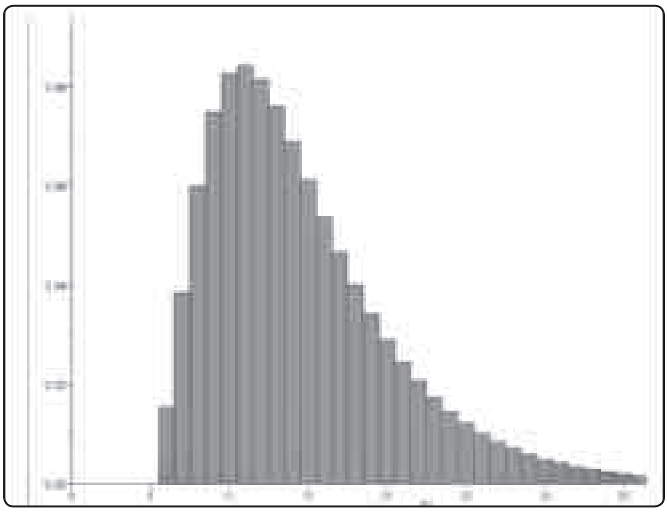


Abb. 2: Wahrscheinlichkeitsverteilung zum Problem der vollständigen Serie.

Um zu der oben abgebildeten **Wahrscheinlichkeitsverteilung** zu gelangen, müssen mithilfe der Tabellenkalkulation die Differenzen zwischen den in der 2. Spalte angegebenen Absorptionswahrscheinlichkeiten ermittelt werden. Wenn man die Absorptionswahrscheinlichkeit für den (n-1)-ten Wurf von derjenigen für den n-ten Wurf subtrahiert, erhält man die Wahrscheinlichkeit dafür, dass man genau mit dem n-ten Wurf eine vollständige Serie erzielt. Die Werte für diese Wahrscheinlichkeiten sind in der 3. Spalte der Tab. 1 unter dem Listennamen „lid" angegeben.

Mithilfe der Datenlisten „lin" und „lid" kann man in analoger Weise wie vorstehend beschrieben das in Abb. 2 angeführte Histogramm erstellen. Man erkennt, dass bei 11 Würfen die Wahrscheinlichkeit am größten ist, eine vollständige Serie zu erzielen. Sie beträgt gerundet 8,4 %.

Die hier vorgestellte Vorgehensweise lässt sich auf alle Markoff-Ketten mit **einem** absorbierenden Zustand anwenden. Beispiele dafür sind radioaktive Zerfallsreihen und Spiele mit einem Endzustand.

Autor:
Dr. Ulrich Döring
Willi-Graf-Gymnasium, Berlin (D)
doc.doe@gmx.de

Literaturverweis:

[1] Eine Verallgemeinerung auf eine vollständige Serie von n Elementen ist dadurch möglich, indem man die Zahl „6" im Term durch „n" ersetzt und die Variable n als Argument integriert:
ziehen(vorhanden,gesamt,n):=
when(vorhanden=n, gesamt, when(rand int(1,n)>vorhanden,
ziehen(vorhanden+1, gesamt+1,n), ziehen(vorhanden, gesamt+1,n)))

[2] Die Definition der Übergangsmatrix ist in der Literatur nicht einheitlich. Teilweise wird auch die transponierte der oben angeführten Matrix als Übergangsmatrix bezeichnet.

Krake Paul – ein neuer Nostradamus?

Dr. Andreas Pallack

Abbildung 1: Prophezeiung aus der Bildzeitung

Bekomme ich nächstes Jahr eine 5 in Mathematik? Betrügt mich meine Frau? Soll ich mehr Geld ins Lotto-Spiel investieren? – Die Zukunft – oder zumindest Antworten auf einige drängende Fragen zu kennen reizt den Menschen. Die hier vorgestellten Fragen wurden jedoch nicht irgendeiner Wahrsagerin geschickt, sondern dem König der Wahrsager – der Krake Paul. Er tippte sämtliche Fußballtore und das Endspiel bei der WM 2010 richtig. Und so hat er das gemacht: Ihm wurden zwei Kästen mit Muschelfleich gereicht, an denen die Fahnen der spielenden Nationen befestigt wurden. Paul entschied sich für eine der Seiten – diese Seite sollte das Spiel gewinnen. Insgesamt tippte er acht Spiele nacheinander richtig.

Mathematikprofessor: ... ich geh in Rente

Der Mathematiker David Spiegelhalter sagte in einem Interview: „Es ist wirklich Zeit, in Rente zu gehen, zusammen mit meiner ganzen Wissenschaft." Eine seiner Hauptaufgaben sei es bisher gewesen, Schülerinnen und Schüler über die Risiken des Glücksspiels aufzuklären. Nun aber erwiesen sich Pauls Vorhersagen „ohne irgendeine Form nostradamischer Zweideutigkeit" jedes Mal als richtig. So viel Glück könne man einfach nicht haben – zitiert ihn ein Tagesblatt.