Der PALIO, das Pferderennen von Siena — Ausgangspunkt für Modelle von Auswahlprozessen Einstieg zum Thema „Markoff-Ketten"

Hans HUMENBERGER, Universität Dortmund

Zusammenfassung

Bei jedem „Palio" müssen die jeweiligen 10 Teilnehmer aus 17 möglichen Kandidaten gewählt werden. Es wird gezeigt, wie mögliche Auswahlvorgänge modelliert werden können und wie diese Fragestellung in natürlicher Weise das Tor zu Markoff-Ketten öffnet. CAS-Einsatz ermöglicht den Schülerinnen und Schülern selbständiges Entdecken einiger Phänomene, wodurch in ihnen ein Begründungsbedürfnis geweckt werden kann.

1 Einleitung

In Siena (Toskana, Norditalien) findet jährlich ein sehr berühmtes Pferderennen mit einer schon jahrhundertelangen Tradition statt, der sogenannte „Palio di Siena" (kurz: Palio): Ein waghalsiges Rennen auf ungesattelten Pferden rund um die *Piazza del Campo*; es dauert jeweils nur ca. $1\frac{1}{2}$ Minuten, wobei allerdings dieses Rennen von zahlreichen traditionellen (tagelangen!) Zeremonien umrahmt wird (vgl. Abb. 1).

Abb. 1: Der Palio

Vorformen reichen bis ins 11. Jhdt. zurück; seit ungefähr der Mitte des 17. Jhdt. findet der Palio in einer zur heutigen ähnlichen Form statt (jeweils am 2. Juli), seit 1802 sogar zweimal pro Jahr[1] (2. Juli und 16. August), wobei aber die Juli-Rennen und die August-Rennen zwei eigenständige „Schienen" darstellen. Das Juli- und das August-Rennen des selben Jahres hängen nicht zusammen, wohl aber die einzelnen Juli-Rennen verschiedener Jahre bzw. die August-Rennen verschiedener Jahre in einem noch genauer zu beschreibenden Sinn.

[1]Genaueres zur Geschichte findet man z. B. unter `http://www.comune.siena.it`

In Siena gibt es (seit 1729) 17 Stadtbezirke („Contraden"), und jeder Stadtbezirk wird durch ein Gespann Reiter-Pferd vertreten: Doch dabei ergibt sich eine gewisse „Misere": Die Platzverhältnisse auf der genau festgelegten „Renn-Strecke" im alten Siena sind so eng, dass es zu gefährlich wäre, das Rennen mit 17 Pferden zu bestreiten. Man hat sich schon seit sehr langer Zeit darauf geeinigt, bei jedem Palio nur 10 Reitergespanne (und damit nur 10 Contraden) gegeneinander antreten zu lassen.

Damit erhebt sich natürlich die Frage, *welche* 10 der 17 Contraden je ein Reiter-Pferd-Gespann zum Palio entsenden dürfen (es ist also ein gerechtes Auswahlverfahren nötig!), um die Chance zu haben, die beliebte Trophäe für sich und den Heimatbezirk zu erobern: sie ist ein Banner, das früher aus wertvollem roten Samt gemacht wurde und „Pallium" hieß, woraus der Name „Palio" für das Rennen resultiert.

2 Modelle für die Auswahl – mögliche Schülervorschläge

Schülerinnen und Schüler werden hier schnell an ihre Kinderzeit erinnert, wo es sicher oft Situationen gab, in denen eine bestimmte Anzahl von Kindern vorhanden war, aber ein geplantes Spiel nur weniger Teilnehmer zuließ: man brauchte ein geeignetes Auswahl- bzw. Auslosungsverfahren, *wer* denn nun zu den (glücklichen) Teilnehmern und wer zunächst zum Zuschauer verurteilt war.

Aufgabe: *Mache konkrete Vorschläge für ein Auswahlverfahren und berechne die Wahrscheinlichkeit, mit der eine bestimmte Contrade am Palio teilnimmt.*

Die Aufgabe fordert also dazu auf, den Wahlvorgang zu *modellieren*, und eignet sich sehr gut für Gruppenarbeit. Im Folgenden seien einige mögliche Schülerantworten (samt stochastischer Analyse) relativ ausführlich dargestellt: die Schüleraktivitäten (selbständiges Bearbeiten - dabei gute Wiederholungsmöglichkeit elementarer Stochastik) sollen dabei im Vordergrund stehen[2]!

2.1 Die 17 Contraden „liegen" in einer Urne

2.1.1 10 einzelne Ziehungen ohne Zurücklegen

17 Kugeln liegen in einer Urne; die Kugeln enthalten zusammengerollte Zettel als „Innenleben", auf denen die Namen der 17 Contraden stehen (darunter unsere Heimat-Contrade C). Eine „Glücksfee" (Kinder, Politiker, Prominente) zieht aus der Urne ohne Zurücklegen (vgl. z. B. die Auslosung zur Fußball WM). Dabei interessiert uns wohl am meisten, mit welcher Wahrscheinlichkeit *wir* am Palio bei diesem Auswahlverfahren teilnehmen dürfen.

$Z_i = C$ bedeutet: „bei der i-ten Ziehung wird C gezogen."

C kann bei der 1. Ziehung, 2. Ziehung, ..., 10. Ziehung drankommen, wobei dies einander ausschließende Ereignisse sind.

Wie groß sind die Wahrscheinlichkeiten $\mathrm{P}(Z_i = C)$ für $i = 1, \ldots, 10$?

Man „sieht" in Tabelle 1: alle Werte $\mathrm{P}(Z_i = C)$ können zu $\frac{1}{17}$ gekürzt werden! Die Bedeutung der involvierten Brüche (Wahrscheinlichkeiten) sollte unmittelbar klar sein: z. B. bedeutet „$Z_2 = C$", dass beim ersten Mal nicht C (Wahrscheinlichkeit $\frac{16}{17}$) und beim zweiten Mal C gezogen wird ($\frac{1}{16}$); analog bei den anderen Zeilen.

$\mathrm{P}(Z_1 = C)$	$\frac{1}{17}$
$\mathrm{P}(Z_2 = C)$	$\frac{16}{17} \cdot \frac{1}{16} = \frac{1}{17}$
$\mathrm{P}(Z_3 = C)$	$\frac{16}{17} \cdot \frac{15}{16} \cdot \frac{1}{15} = \frac{1}{17}$
$\vdots$	$\vdots$
$\mathrm{P}(Z_{10} = C)$	$\frac{16}{17} \cdot \frac{15}{16} \cdot \ldots \cdot \frac{8}{9} \cdot \frac{1}{8} = \frac{1}{17}$

Tab. 1: Wahrscheinlichkeiten für die Ziehung von C bei den einzelnen Ziehungen

Für die Wahrscheinlichkeit, dass C *„überhaupt gezogen wird"*, ergibt sich daher:

$$\boxed{\mathrm{P}(C \text{ wird gezogen}) = \frac{1}{17} + \ldots + \frac{1}{17} = \frac{10}{17}}\,.$$

Ebenso wäre denkbar, zuerst jene 7 Contraden zu ziehen, die NICHT teilnehmen; die restlichen 10 in der Urne verbleibenden nehmen am Palio teil:

$$\boxed{\mathrm{P}(C \text{ nimmt teil}) = 1 - \underbrace{\left(\frac{1}{17} + \ldots + \frac{1}{17}\right)}_{\frac{7}{17}} = \frac{10}{17}}$$

Ein weiterer möglicher Auswahlvorschlag[3]: Die Contraden der 1., 3., 5., 7., 9., 10., 11., 13., 15., 17. Ziehung sollen teilnehmen (oder einer **beliebigen** anderen vorgegebenen Auswahl von 10 aus den 17 Ziehungen).

2.1.2 Eine Ziehung — „10 Kugeln mit einem Griff"

Dies ist mathematisch natürlich äquivalent zu den einzelnen Ziehungen; trotzdem sei diese Möglichkeit extra erwähnt, denn:

- der *Vorgang* ist doch ein anderer,
- Schüler kommen zum Wert $\frac{10}{17}$ eventuell auf einem *anderen Weg*.

Es gibt $\binom{17}{10}$ Möglichkeiten, 10 Kugeln (Contraden) aus den 17 mit einem Griff auszuwählen.

[2] Auch im Abschnitt über Markoff-Ketten werden mögliche selbständige Schüleraktivitäten immer wieder hervorgehoben.

[3] Auch dieser wäre gerecht; jede Contrade hätte eine Teilnahmewahrscheinlichkeit von $\frac{10}{17}$.

Wie viele davon enthalten die Contrade C? ($\rightarrow$ „hypergeometrische Verteilung"; aber auch ohne diesen Begriff, d. h. ohne $\binom{1}{1}$ im Zähler, kann leicht begründet werden):

$$\boxed{\mathrm{P}(C \text{ wird gewählt}) = \frac{\binom{1}{1}\cdot\binom{16}{9}}{\binom{17}{10}} = \frac{10}{17}}\,.$$

Ein analoger Schülervorschlag: mit einem Griff 7 aus den 17 Kugeln (Contraden) herausziehen, die *nicht* teilnehmen:

$$\boxed{\mathrm{P}(C \text{ nimmt teil}) = 1 - \frac{\binom{1}{1}\cdot\binom{16}{6}}{\binom{17}{7}} = \frac{10}{17}}\,.$$

Bei den bis jetzt angesprochenen Auswahlverfahren gibt es wahrscheinlich keine Diskussionen über mögliche Ungerechtigkeiten, bei der folgenden vielleicht aber doch.

2.2 Die Urne enthält „Lose" — Ziehen ohne Zurücklegen

Eine andere Möglichkeit der Auslosung wäre: In einer Urne sind wieder 17 Kugeln. 10 davon enthalten einen Zettel mit der Aufschrift „Teilnahme" (T) und 7 einen Zettel mit „Nicht-Teilnahme" ($\neg T$), also 10 *Gewinnlose* und 7 *Nieten.* Jede Contrade zieht eine Kugel (T oder $\neg T$; ohne Zurücklegen!) aus der Urne.

Die Contraden „liegen" hier also nicht (passiv) in der Urne, sondern sie (d. h. ein jeweiliger Vertreter) ziehen selbst ein Los.

Hierbei sind nun aber Streitigkeiten durchaus denkbar: Welche Contrade darf als erste ziehen? Welche muss als letzte ziehen?

Da *ohne Zurücklegen* gezogen wird, finden die einzelnen Contraden bei ihren Ziehungen *verschieden bestückte Urnen* vor, was die Frage nach der Gerechtigkeit aufwirft.

Wir setzen „$C_i = T$" … „die an i-ter Stelle ziehende Contrade zieht ein T, d. h. nimmt teil."

und wollen die Wahrscheinlichkeiten $\boxed{\mathrm{P}(C_i = T)}$ für $1 \leq i \leq 17$ berechnen.

Klarerweise ist $\boxed{\mathrm{P}(C_1 = T) = \frac{10}{17}}$.

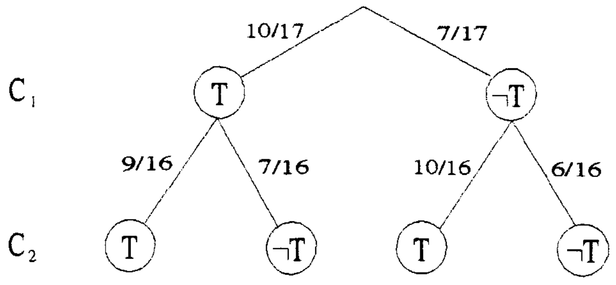

Abb. 2: Baumdiagramm für die 2. Ziehung

Die Contrade C_2 kann nun 2 verschieden bestückte Urnen vorfinden, je nach $C_1 = T$ oder $C_1 = \neg T$ - siehe Abb. 2.

$$\boxed{\mathrm{P}(C_2 = T)} = \frac{9}{16}\cdot\frac{10}{17} + \frac{10}{16}\cdot\frac{7}{17} = \boxed{\frac{10}{17}}\,.$$

Es liegt nun die Vermutung nahe, dass nicht nur $\mathrm{P}(C_1 = T) = \mathrm{P}(C_2 = T) = \frac{10}{17}$, sondern sogar $\mathrm{P}(C_i = T) = \frac{10}{17}$ für alle $1 \leq i \leq 17$ ist. Dann bräuchten sich die Contraden keinerlei Gedanken um die Ziehungsreihenfolge zu machen, jeder Streit wäre sinnlos und unbegründet!

Aus einem 3-stufigen Baumdiagramm wie in Abb. 3 erhält man unmittelbar:

$$\mathrm{P}(C_3 = T) = \frac{10}{17}\cdot\frac{9}{16}\cdot\frac{8}{15} + \frac{10}{17}\cdot\frac{7}{16}\cdot\frac{9}{15} + \frac{7}{17}\cdot\frac{10}{16}\cdot\frac{9}{15} + \frac{7}{17}\cdot\frac{6}{16}\cdot\frac{10}{15} = \frac{10}{17}$$

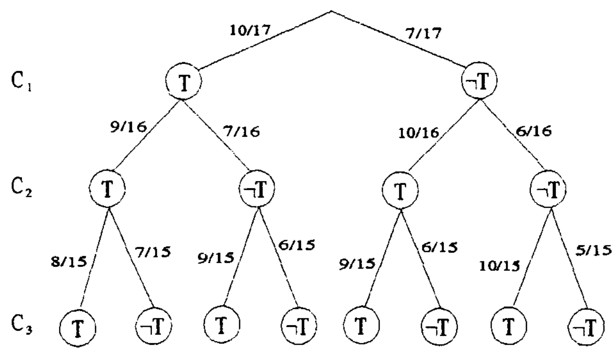

Abb. 3: Baumdiagramm für die 3. Ziehung

Die Vermutung hat sich also bestätigt, aber in dieser Art weiter zu machen scheint ein mühsames Unterfangen zu werden …??

Ein einfaches „Symmetrieargument" kann schnell Klarheit verschaffen: Da bei den Ziehungen kein T- bzw. $\neg T$-Los „bevorzugt" ist, wird an jeder festen Stelle i *jedes Los mit der selben*[4] *Wahrscheinlichkeit* gezogen (für alle $1 \leq i \leq 17$). Daher ist $\mathrm{P}(C_i = T) = \mathrm{P}(C_i = T_1) + \ldots + \mathrm{P}(C_i = T_{10}) = \frac{10}{17}$ für alle $1 \leq i \leq 17$.

Durch *Umdeutung des Ziehungsvorganges* der 17 Kugeln kann auch so argumentiert werden:

Das Ziehen der 17 Kugeln (Lose: 10 T, 7 $\neg T$) durch die 17 Contraden $C_1, \ldots, C_{17}$ bedeutet:

die 10 T-Kugeln und die 7 $\neg T$-Kugeln in eine bestimmte Reihenfolge zu bringen, eine spezielle Anordnung der 17 Kugeln zu erzeugen (siehe Abb. 4).

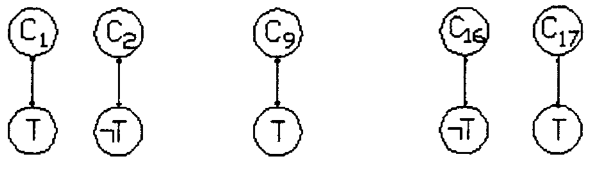

Abb. 4: Umdeutung des Ziehungsvorganges als Anordnungsproblem

Es ergibt sich die Äquivalenz:

C_i nimmt teil $\Leftrightarrow$ an der Stelle i ist ein T.

Für die weitere Argumentation gibt es zwei Möglichkeiten:

- Da bei Anordnungen keine Stelle bevorzugt ist, gibt es genau so viele Anordnungen mit einem T an 1. Stelle wie mit einem T an jeder anderen Stelle $1 \leq i \leq 17$. Daher ist wegen $\mathrm{P}(C_1 = T) = \frac{10}{17}$ (siehe oben) für alle $1 \leq i \leq 17$ klar: $\mathrm{P}(C_i = T) = \mathrm{P}(C_1 = T) = \frac{10}{17}$.
- Konkrete Berechnung: Für $\mathrm{P}(C_i = T)$ zählen wir die Möglichkeiten, wie viele Anordnungen es überhaupt gibt („möglich“) und wie viele davon ein T an der Stelle i haben („günstig“).

#{mögliche Anordnungen}: $\binom{17}{10}$
(10 von 17 Stellen mit T belegen)

#{günstige Anordnungen}:
(T an der Stelle i, 9 von den 16 weiteren Stellen mit T belegen): $\binom{16}{9}$, d. h.

$$\boxed{\mathrm{P}(C_i = T) = \frac{\binom{16}{9}}{\binom{17}{10}} = \frac{10}{17}} \quad \forall\, i = 1, \ldots, 17 \,.$$

[4] Nämlich, da es 17 Lose gibt, klarerweise $\frac{1}{17}$.

Wir haben damit mehrfach begründet: Die **Ziehungsreihenfolge** ist wirklich **irrelevant**, auch für die als letzte ziehende Contrade C_{17} gilt $\mathrm{P}(C_{17} = T) = \frac{10}{17}$, obwohl C_{17} gar keine Wahlmöglichkeit mehr hat und die in der Urne verbleibende Kugel nehmen muss!

Alle bis jetzt beschriebenen Modelle des Auswählens haben gemeinsam:

- Für jede Contrade beträgt die Teilnahmewahrscheinlichkeit (bei jedem Wahlvorgang, jedes Jahr) $\frac{10}{17}$ – das Auswahlverfahren ist in diesem Sinn also *gerecht*, keine Contrade wird bevorzugt.
- Die Auslosung geschieht jedes Mal „neu“ unter den selben Bedingungen: Die *Konstellationen* des Vorjahres (z. B. ob eine Contrade am Palio teilgenommen hat oder nicht, ob sie vielleicht sogar gewonnen hat etc.) beeinflusst die Teilnahmewahrscheinlichkeit einer Contrade am diesjährigen Palio nicht. Jede Contrade stellt sich jedes Jahr der Wahl „neu“ (*Unabhängigkeit* der einzelnen Wahlvorgänge; „BERNOULLI-Kette“).

3 Oftmalige Nicht-Teilnahme hintereinander erzeugt Frustration – ein möglicher Einstieg zu „Markoff-Ketten“

Die beschriebene *Unabhängigkeit* der jährlichen Auswahlvorgänge („Bernoulli-Kette“) ist ein entscheidender Nachteil in den obigen Modellen: Wenn jedes Jahr nach dem Prinzip „neues Jahr, neues Rennen, neues Glück“ verfahren würde, bestünde durchaus die Möglichkeit, dass eine bestimmte Contrade über mehrere (u. U. sogar sehr viele) Jahre am Palio nicht teilnehmen könnte, was natürlich zu einer sehr hohen Frustration führte.

Wie könnte diese für alle beteiligten Contraden unerwünschte Frustrationsmöglichkeit beseitigt werden?

Hier werden die Schülerinnen und Schüler schnell naheliegende Modelle vorschlagen: Man könnte die 7 Nichtteilnehmer eines Rennens zu *Fixstartern* für das nächste Rennen erklären! Wie könnten dann die restlichen 3 Startplätze für den nächsten Palio gerecht vergeben werden? Auf diese Frage könnten verschiedene Vorschläge kommen:

- Die Contraden auf den *ersten* 3 Plätzen eines Palio sollen auch im nächsten starten dürfen: *Belohnung* für gute Leistung.

  Oder: Die Contraden auf den *letzten* 3 Plätzen eines Palio sollen auch im nächsten starten dürfen: *Trost* für verpatzte Rennen.

- Aus den 10 Teilnehmern eines Palio sollen (unabhängig von ihren Erfolgen) 3 gewählt werden, die auch am nächsten teilnehmen dürfen. Dies kann z. B. nach einem Verfahren des Abschnitts 2 geschehen: jede an einem Palio teilnehmende Contrade hätte dann eine $\frac{3}{10}$-Chance, auch beim nächsten Mal dabei zu sein.

Dieses letzte Modell wird in Siena gewählt! In dieser Weise hängen die Juli-Rennen untereinander und die August-Rennen untereinander zusammen. Die Juli- und die August-Rennen sind aber zwei voneinander unabhängige „Stränge"[5].

Dadurch wird erreicht, dass keine Contrade zweimal hintereinander pausieren muss: nach einer Nicht-Teilnahme folgt mit Sicherheit eine Teilnahme[6]; nach einer Teilnahme folgt mit Wahrscheinlichkeit $\frac{3}{10}$ eine erneute Teilnahme (mit Wahrscheinlichkeit $\frac{7}{10}$ eine Nicht-Teilnahme).

Dieses Auswahlverfahren ist keine Bernoulli-Kette mehr, sondern eine so genannte „MARKOFF-Kette": Die Teilnahmewahrscheinlichkeiten einer Contrade C am Rennen $n+1$ hängen (nur) davon ab, ob C beim Rennen n dabei war oder nicht.

Bemerkung: In das Thema „Markoff-Ketten" fließen die drei klassischen Gebiete der Oberstufenmathematik (Stochastik, Lineare Algebra, Analysis) ein, was insbesondere im Hinblick auf das Prinzip der „Vernetzung" Bedeutung hat: neben Wahrscheinlichkeiten (Stochastik) werden auch Matrizen und Vektoren (Lineare Algebra) und der Grenzwertbegriff (Analysis) eine Rolle spielen.

Ein Großteil des Folgenden dient der Beantwortung der Frage: **Ist die Palio-Auswahl „gerecht"? Mit welcher Wahrscheinlichkeit nehmen wir (die Contrade C) am Palio nach 1 (3, 10, 100, ...) Jahren teil?**

Dabei liegt es nahe, unter „gerecht" zu verstehen, dass sich **auf lange Sicht, d. h. für große n,** *für jede Contrade die selbe Teilnahmewahrscheinlichkeit* ergibt, und zwar unabhängig davon, ob die Contrade beim ersten Rennen (das nach diesem System abläuft) teilgenommen hat oder nicht. Andernfalls bedeutete eine (Nicht-)Teilnahme von C beim ersten Rennen eine Verbesserung (bzw. Verschlechterung) ihrer Teilnahmechancen „auf ewig", was klarerweise nicht wünschenswert wäre. Von so einer *gerechten*, für je zwei Contraden gleichen *Teilnahmewahrscheinlichkeit auf lange Sicht* erwarten wir natürlich, dass sie den Wert $\frac{10}{17}$ hat, auch aus folgender Überlegung über die **relative Teilnahmehäufigkeit** bei den ersten n Rennen: Bei n Rennen (n groß) gibt es insgesamt $10 \cdot n$ Rennteilnahmen, so dass im „gerechten" Fall bei 17 Konkurrenten jedem $\frac{10}{17} \cdot n$ Rennteilnahmen zustehen.

Zunächst definieren wir Zufallsgrößen X_n mit den beiden möglichen Ausprägungen T und $\neg T$:

$$X_n = \begin{cases} T, & \text{falls } C \text{ am Rennen } n \text{ teilnimmt} \\ \neg T, & \text{falls } C \text{ am Rennen } n \text{ \textit{nicht} teilnimmt} \end{cases}$$

und interessieren uns für die Wahrscheinlichkeitsverteilungen von X_n.

Bemerkung: Die *Übergangswahrscheinlichkeiten* zwischen den beiden Zuständen $T \leftrightarrow \neg T$ (von einem Rennen zum nächsten) lassen sich z. B. auch in einem *gerichteten Graphen* sehr übersichtlich darstellen – siehe Abb. 5.

Mit der Abkürzung

[5] Für eine prinzipiell andere (und eher hochschulmathematisch gehaltene) Analyse dieses Auswahlverfahrens siehe GÖTZ/GROSSER 1999.

[6] Auch folgendes System würde dies leisten (möglicher Schülervorschlag): Man nummeriert die Contraden von 1-17, schreibt die Zahlen 1–17 (also den „Block aus 17 Contraden") oft hintereinander und nimmt jedes Jahr „der Reihe nach" 10 Contraden (1-10; 11-3; 4-13; 14-6; 7-16; 17-9; ...). Dadurch wäre nichts Zufälliges mehr im System, sondern jede Contrade hätte ihre vollständig determinierte 17-Jahre-periodische T-$\neg T$-Folge auf „ewig" (jeweils bestehend aus 10 T's und 7 $\neg T$'s in einer bestimmten Abfolge); jene der Contrade 1 wäre z. B. $(T, T, \neg T, T, \neg T, T, T, \neg T, T, \neg T, T, T, \neg T, T, \neg T, T, \neg T, \mid T, T, \ldots)$; die der anderen Contraden wären analog, nur entsprechend „versetzt".

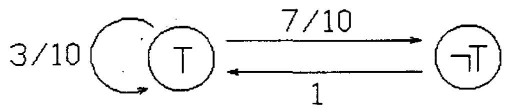

Abb. 5: Gerichteter Graph

$p_n := P(X_n = T),\ q_n := 1 - p_n = P(X_n = \neg T)$ interessiert uns also die Folge $(p_n)_{n \in \mathbb{N}}$ (damit ist auch die Folge $(q_n)_{n \in \mathbb{N}}$ festgelegt).

Für große n sollte $p_n \approx \frac{10}{17}$ sein, und zwar *unabhängig* davon, ob C beim Anfangsrennen dabei war oder nicht – siehe oben.

So ohne weiteres können wir über die Werte p_n bzw. q_n keine direkten Aussagen machen, hängen sie doch davon ab, ob C beim Rennen davor teilgenommen hat oder nicht. Der „Mechanismus" dieser Abhängigkeit von einem Rennen n zum nächsten Rennen $n+1$ ist uns bekannt und für alle n der selbe; er ist im Baumdiagramm in Abb. 6 dargestellt.

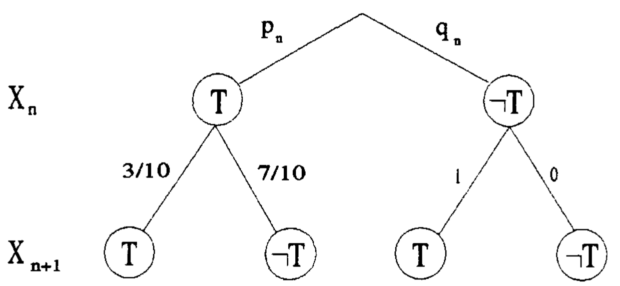

Abb. 6: 2-stufiges Baumdiagramm für die Zufallsgrößen X_n bzw. X_{n+1}

Die Pfadregeln ergeben:

$$\begin{aligned} p_{n+1} &= \frac{3}{10} \cdot p_n + 1 \cdot q_n && (1) \\ q_{n+1} &= \frac{7}{10} \cdot p_n + 0 \cdot q_n && (2) \end{aligned}$$

Mit Hilfe dieser beiden Iterationsformeln können wir nun ausgehend von einem bestimmten Anfangsrennen, an dem wir teilgenommen haben (oder auch nicht: je nachdem ist dann $p_0 = 1$ oder $p_0 = 0$) schrittweise die Teilnahmewahrscheinlichkeiten p_i bzw. q_i konkret berechnen: $(p_0, q_0) \to (p_1, q_1) \to (p_2, q_2) \to (p_3, q_3) \to \ldots$. Dies sollten Schülerinnen und Schüler (noch ohne Computer) bis zum Index 3 oder 4 wirklich ausführen. Sie merken schnell, dass hierbei wiederholt das selbe zu tun ist, was immer ein Anzeichen dafür ist, dass Computer gut einsetzbar sein müssten.

Obiges Gleichungssystem kann man auch mittels einer Matrix schreiben:

$$\underbrace{\begin{pmatrix} \frac{3}{10} & 1 \\ \frac{7}{10} & 0 \end{pmatrix}}_{:=A} \cdot \underbrace{\begin{pmatrix} p_n \\ q_n \end{pmatrix}}_{:=\vec{\pi}_n} = \underbrace{\begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix}}_{:=\vec{\pi}_{n+1}} \qquad (3)$$

Nun werden die Verbindungen zur Linearen Algebra schon deutlich.

Bemerkungen:

- Matrizenrechnung wird ab nun vorausgesetzt, obwohl sie nicht an jeder Oberstufe unterrichtet wird (es wäre u. U. auch möglich, sie an dieser Stelle einzuführen).

- Es ist nicht denknotwendig, hier Matrizenrechnung und Markoff-Ketten ins Spiel zu bringen, man fände bei Prozessen mit nur 2 möglichen Zuständen auch das Auslangen mit einfachen Rekursionsgleichungen (hier: lineare Differenzengleichungen 1. Ordnung), um die Konvergenz $p_n \to \frac{10}{17}$ bzw. eine explizite Darstellung für p_n herauszufinden; dies wird hier aber absichtlich nicht im Detail verfolgt[7]. Obwohl es im 2×2-Fall also „einfacher" ginge (siehe Fußnote), bleiben wir hier bei *„Markoff-Ketten in Matrixdarstellung"*, weil wir einen möglichen *Einstieg* in das in Rede stehende *allgemeine* Thema bieten

[7]Hier nur eine Kurzfassung: Aus der Gleichung (1) folgt mit $q_n = 1 - p_n$ unmittelbar $p_{n+1} = -\frac{7}{10}p_n + 1$ (*), eine lineare Differenzengleichung 1. Ordnung, deren klassische Lösung, d. h. *explizite* Darstellung für p_n, mit der Summenformel für geometrische Reihen zu tun hat; alternativ dazu könnte kürzer auch folgendermaßen argumentiert werden: der einzig mögliche (dadurch noch nicht nachgewiesene!) Grenzwert p ergibt sich aus der Gleichung $p = -\frac{7}{10}p + 1$, also $p = \frac{10}{17}$. Daher liegt es besonders nahe, die Abweichungen $p_n - \frac{10}{17}$ zu untersuchen (und als Nullfolge zu entlarven): Aus (*) erhält man sofort $(p_{n+1} - \frac{10}{17}) = (-\frac{7}{10}) \cdot (p_n - \frac{10}{17})$, wodurch klar ist, dass die Differenzen $p_n - \frac{10}{17}$ eine geometrische Folge mit dem Faktor $-\frac{7}{10}$ bilden. Dadurch ist nicht nur die Konvergenz $p_n \to \frac{10}{17}$, sondern auch die Konvergenzgeschwindigkeit klar, und dass die Werte p_n abwechselnd über bzw. unter dem Grenzwert $p = \frac{10}{17}$ liegen. Wem nur an der Klärung $p_n \to p$ bei 2 möglichen Zuständen gelegen ist, kann nach der Fußnote die Lektüre abbrechen, denn der Rest widmet sich dem Matrizenkalkül, der auch bei mehr als 2 Zuständen anwendbar ist. Analoges könnte man nämlich bei allen solchen „Prozessen" mit nur 2 möglichen Zuständen machen: aus $p_{n+1} = a \cdot p_n + b$ mit $a \neq 1$ folgt allgemein: $\left(p_{n+1} - \frac{b}{1-a}\right) = a \cdot \left(p_n - \frac{b}{1-a}\right)$, woraus die Konvergenz von (p_n) gegen $\frac{b}{1-a}$ für $|a| < 1$ ablesbar ist. Die explizite Darstellung für p_n ergibt sich – falls gewünscht – unmittelbar als $p_n = a^n \cdot \left(p_0 - \frac{b}{1-a}\right) + \frac{b}{1-a}$.

wollen und dort eben Matrizen nötig sind (die „einfache“ matrixfreie Methode funktioniert bei $n \geq 3$ Zuständen eben nicht mehr so einfach). Die im 2×2-Fall erworbenen und verstandenen Erkenntnisse können dann unmittelbar auch auf $n \geq 3$ „Zustände“ übertragen werden (nur $n \times n$- statt 2×2-Matrizen). Außerdem wird dadurch eine Vernetzung Stochastik – Lineare Algebra (auch Analysis) erreicht (weitere Argumente für diesen Zugang – siehe unten: selbständiges Entdecken von Phänomenen, CAS-Einsatz etc.).

- Der Palio bietet zudem *inhaltlich* einige Vorteile: Realitätsbezug; die Markoff-Bedingungen sind wirklich erfüllt und ändern sich über Jahrhunderte nicht („homogene Markoff-Kette“), so dass die Untersuchung des „Langzeitverhaltens“ wirklich Sinn macht (im Gegensatz zu manchen anderen Anwendungsbeispielen).

Zurück zur Matrixdarstellung (3):

Der (Spalten-)Vektor $\vec{\pi}_n := \begin{pmatrix} p_n \\ q_n \end{pmatrix}$ beschreibt offenbar die Teilnahmewahrscheinlichkeit der Contrade beim n-ten Rennen: m. a. W. er beschreibt die Wahrscheinlichkeits*verteilung* der Zufallsgröße X_n (die ja nur zwei mögliche Ausprägungen hat). Der Vektor $\vec{\pi}_n$ heißt deshalb auch „Wahrscheinlichkeitsvektor zum Zeitpunkt n“ oder „Verteilung zum Zeitpunkt n“.

Multipliziert man ihn (von links) mit der Matrix A, so erhält man den Vektor $\vec{\pi}_{n+1}$, der die Teilnahmewahrscheinlichkeiten am $(n + 1)$-ten Rennen beschreibt.

Die Matrix A sorgt für den Vektor-Übergang $\vec{\pi}_n \to \vec{\pi}_{n+1}$ und heißt deshalb *Übergangsmatrix.*

Was hat man von dieser Erkenntnis, dass das Gleichungssystem in Matrixform geschrieben werden kann? Es eröffnet eine Möglichkeit, ausgehend von $\vec{\pi}_0 = \begin{pmatrix} p_0 \\ q_0 \end{pmatrix}$ **direkt** zu $\vec{\pi}_n = \begin{pmatrix} p_n \\ q_n \end{pmatrix}$ zu kommen, ohne n-mal die mühselige Iteration durchführen zu müssen:

$$\begin{aligned}
\vec{\pi}_1 &= A \cdot \vec{\pi}_0 \\
\vec{\pi}_2 &= A \cdot \vec{\pi}_1 = A \cdot (A \cdot \vec{\pi}_0) = A^2 \cdot \vec{\pi}_0 \\
\vec{\pi}_3 &= A \cdot \vec{\pi}_2 = A \cdot \left(A^2 \cdot \vec{\pi}_0\right) = A^3 \cdot \vec{\pi}_0 \\
&\vdots \\
\vec{\pi}_n &= A \cdot \vec{\pi}_{n-1} = \ldots = A^n \cdot \vec{\pi}_0
\end{aligned}$$

Um $\vec{\pi}_n$ zu erhalten, muss man also „nur“ A^n bilden und mit $\vec{\pi}_0$ multiplizieren.

Bemerkung: Diese Arbeit wird vom Computer – natürlich auch iterativ (schrittweise) – erledigt. Durch die Matrix-*Darstellung* wird also nicht Rechenarbeit erspart, sondern nur eine „CAS-taugliche“ Form gewählt, so dass das Mühselige an den Computer übertragen werden kann, ohne selbst ein Programm dafür schreiben zu müssen.

Man „erlebt“ die Bedeutung der Matrizenmultiplikation, des Potenzierens von Matrizen, der Matrix-Vektor-Multiplikation, der Assoziativität. Ohne Computer kann man natürlich keine hohen Matrixpotenzen berechnen, sondern nur solche von sehr niederer Ordnung; mit einem CAS (diese haben alle die Möglichkeit, Matrizen in Sekundenbruchteilen zu potenzieren) steht hingegen einer weitgehend selbständigen Entdeckung einiger Phänomene durch die Schülerinnen und Schüler nichts im Wege (im Folgenden genauer beschrieben, z. B. spezielle Eigenschaften von A^n bzw. $\vec{\pi}_n$ für große n).

Zurück zu unserer Contrade: wie groß ist die Wahrscheinlichkeit, dass sie in 3 Jahren teilnehmen / nicht teilnehmen wird?

Wir brauchen dafür $A^3 \approx \begin{pmatrix} 0.45 & 0.79 \\ 0.55 & 0.21 \end{pmatrix}$ (grobe 2-stellige Rechnung, d. h. 2 signifikante Ziffern, genügt einstweilen).

Sie habe im Palio dieses Jahres teilgenommen, d. h. $\vec{\pi}_0 = \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; dann ergibt sich:

$$\vec{\pi}_3 = \begin{pmatrix} p_3 \\ q_3 \end{pmatrix} = A^3 \cdot \vec{\pi}_0 \approx \begin{pmatrix} 0.45 & 0.79 \\ 0.55 & 0.21 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.45 \\ 0.55 \end{pmatrix} .$$

Die gesuchte Teilnahmewahrscheinlichkeit beträgt also $p_3 \approx 0.45$.

Welche Werte haben p_3 bzw. q_3, wenn die Contrade in diesem Jahr nicht teilgenommen hat, d. h. bei $\vec{\pi}_0 = \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$? Analog:

$$\vec{\pi}_3 = \begin{pmatrix} p_3 \\ q_3 \end{pmatrix} \approx \begin{pmatrix} 0.45 & 0.79 \\ 0.55 & 0.21 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.79 \\ 0.21 \end{pmatrix} .$$

Die Teilnahmewahrscheinlichkeit beträgt hier $p_3 \approx 0.79$ und unterscheidet sich von obiger doch beträchtlich. Die diesjährige Teilnahme bzw. Nicht-Teilnahme wirkt sich also noch stark auf die Teilnahmechancen in drei Jahren, d. h. auf den Wahrscheinlichkeitsvektor $\vec{\pi}_3$ aus.

Die selben Fragen für 10 Jahre (nur noch mit CAS sinnvoll): $A^{10} \approx \begin{pmatrix} 0.60 & 0.57 \\ 0.40 & 0.43 \end{pmatrix}$.

Teilnahme in diesem Jahr, d. h. $\vec{\pi}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$:

$$\vec{\pi}_{10} = \begin{pmatrix} p_{10} \\ q_{10} \end{pmatrix} \approx \begin{pmatrix} 0.60 & 0.57 \\ 0.40 & 0.43 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.60 \\ 0.40 \end{pmatrix}$$

Keine Teilnahme in diesem Jahr, $\vec{\pi}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$:

$$\vec{\pi}_{10} = \begin{pmatrix} p_{10} \\ q_{10} \end{pmatrix} \approx \begin{pmatrix} 0.60 & 0.57 \\ 0.40 & 0.43 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.57 \\ 0.43 \end{pmatrix}$$

Nach 10 Jahren unterscheiden sich die beiden Teilnahmewahrscheinlichkeiten p_{10} bzw. die beiden Wahrscheinlichkeitsvektoren $\vec{\pi}_{10}$ (in Bezug auf diesjährige Teilnahme/Nicht-Teilnahme) nur mehr wenig: $p_{10} \approx 0.60$ und $p_{10} \approx 0.57$. Wird dieser Unterschied mit wachsendem n tatsächlich immer geringer? Um diese Vermutung experimentell zu bestätigen, wählen wir bei größeren n vorsichtshalber eine höhere Genauigkeit, z. B. die CAS-Standardeinstellung mit 6 signifikanten Ziffern:

Obige Fragen für z. B. $n = 100$ Jahre:
$A^{100} \approx \begin{pmatrix} 0.588235 & 0.588235 \\ 0.411765 & 0.411765 \end{pmatrix}$

$$\vec{\pi}_{100} = \begin{pmatrix} p_{100} \\ q_{100} \end{pmatrix} = A^{100} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \approx \begin{pmatrix} 0.588235 \\ 0.411765 \end{pmatrix}$$

$$\vec{\pi}_{100} = \begin{pmatrix} p_{100} \\ q_{100} \end{pmatrix} = A^{100} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.588235 \\ 0.411765 \end{pmatrix}$$

Nach 100 Jahren unterscheiden sich die beiden Wahrscheinlichkeitsvektoren $\vec{\pi}_{100}$ (in Bezug auf dieses Jahr T bzw. $\neg T$) praktisch nicht mehr!

Auch bei jedem anderen „Wahrscheinlichkeits-Startvektor" $\vec{\pi}_0 = \begin{pmatrix} p_0 \\ q_0 = 1 - p_0 \end{pmatrix}$, der die Anfangs-Teilnahmewahrscheinlichkeit beschreibt, hätte sich – weil jede Zeile in A^{100} aus (fast) identischen Werten besteht und $q_0 = 1 - p_0$ ist – der selbe Vektor $\vec{\pi}_{100}$ ergeben:

$$\vec{\pi}_{100} = \begin{pmatrix} p_{100} \\ q_{100} \end{pmatrix} = A^{100} \cdot \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} \approx \begin{pmatrix} 0.588235 \\ 0.411765 \end{pmatrix}$$

Berechnet man auch A^{101} und $\vec{\pi}_{101}$, so sieht man, dass sich A^{100} von A^{101} bzw. $\vec{\pi}_{100}$ von $\vec{\pi}_{101}$ (bei 6 signifikanten Ziffern) nicht mehr unterscheiden.

Der experimentelle Befund lässt vermuten, dass die Folge der Matrixpotenzen (A^n) gegen eine Grenzmatrix G konvergiert, deren Eigenschaft man in $A^{100} \approx G$ schon ablesen kann:

- G hat Zeilen aus jeweils gleichen Werten;
- oder: G hat identische Spalten.

Wenn dies richtig ist, dann konvergieren auch die Wahrscheinlichkeitsvektoren $\vec{\pi}_n$ gegen einen eindeutig bestimmten Grenzvektor $\vec{\pi} \approx \begin{pmatrix} 0.588235 \\ 0.411765 \end{pmatrix}$, der nichts anderes ist als eine Spalte von G; diese **Konvergenz** $\vec{\pi}_n \to \vec{\pi}$ ist dann sogar **unabhängig vom Startvektor** $\vec{\pi}_0$ (siehe oben).

Diese spezielle Konvergenz der $\vec{\pi}_n$ (Unabhängigkeit von $\vec{\pi}_0$ [8]; genaue *B*edingungen dafür: siehe unten) können Schülerinnen und Schüler einfach durch Probieren am Computer selbständig entdecken und diesbezügliche Vermutungen äußern.

Es bleibt die Frage, was es mit den Werten 0.588235 und 0.411765 auf sich hat?

Auf lange Sicht hatten wir doch (bei Gerechtigkeit des Auswahlverfahrens) eine Teilnahmewahrscheinlichkeit von $\frac{10}{17}$ erwartet; tatsächlich gilt: $\frac{10}{17} \approx 0.588235$ und $\frac{7}{17} \approx 0.411765$.

Dadurch drängt sich natürlich sofort die Vermutung auf, dass diese Werte des Grenzvektors $\vec{\pi} \approx \vec{\pi}_{100}$ vielleicht auch einfacher zu bekommen sein müssten, ohne die Grenzmatrix $G \approx A^{100}$ zu berechnen.

Tatsächlich, allein die gesicherte *Existenz einer Grenzmatrix* $G = \lim_{n\to\infty} A^n$ *mit Zeilen aus jeweils gleichen Werten*, ohne sie konkret zu kennen, hat entscheidende Konsequenzen (deswegen drängt sich die Frage nach Bedingungen dafür ja geradezu auf – Exaktifizierung später):

Wenn es nämlich eine *solche* Grenzmatrix G gibt, so gibt es „klarerweise"[9] auch **den** Grenzvektor $\vec{\pi}$, und zwar eindeutig (unabhängig von $\vec{\pi}_0$!) – siehe oben.

[8] Das Entdecken dieser Unabhängigkeit erfordert mehr als das bloße Sehen, dass die Matrixpotenzen gegen eine spezielle Grenzmatrix zu konvergieren scheinen.

[9] Formal gezeigt im nächsten Abschnitt.

Weiters ist „klar“ (kann auch formal exaktifiziert werden – nächster Abschnitt): dieser Grenzvektor $\vec{\pi}$ darf sich durch Multiplikation mit A natürlich nicht mehr verändern, d. h. es muss gelten: $A \cdot \vec{\pi} = \vec{\pi}$; $\vec{\pi}$ muss also *„Fixvektor“* von A sein. Dies führt zur Fixvektorgleichung

$$\begin{pmatrix} \frac{3}{10} & 1 \\ \frac{7}{10} & 0 \end{pmatrix} \cdot \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix},$$

wobei für die Wahrscheinlichkeiten $p, q \geq 0$ noch $p + q = 1$ gelten muss. Die leicht verifizierbare Lösung ist

$$\vec{\pi} = \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \frac{10}{17} \\ \frac{7}{17} \end{pmatrix}.$$

In diesem Zusammenhang geben wir folgende

Definition: Ein Vektor mit nichtnegativen Einträgen und Spaltensumme 1 heißt *„stochastischer Vektor“* (wie z. B. die Vektoren $\vec{\pi}_0$, $\vec{\pi}_n$ bzw. $\vec{\pi}$). Eine $n \times n$-Matrix mit nichtnegativen Einträgen und alle Spaltensummen $= 1$ heißt *„stochastische Matrix“* (wie z. B. die Matrizen A, A^n, G) [10].

Eine entscheidende Frage ist nun: kann es *andere* stochastische Fixvektoren als den gesuchten Grenzvektor $\vec{\pi}$ geben? Wenn nicht, so könnten wir $\vec{\pi}$ einfach als **den** stochastischen Fixvektor berechnen.

Diese Frage kann mit einem „erlösenden NEIN“ beantwortet werden: denn jeder stochastische Fixvektor $\vec{s}$ ist klarerweise auch Grenzvektor – nämlich bei Startvektor $\vec{\pi}_0 = \vec{s}$ – und der Grenzvektor $\vec{\pi}$ ist ja *eindeutig* (unabhängig vom stochastischen Startvektor $\vec{\pi}_0$ – siehe oben; formal wieder im nächsten Abschnitt).

Wir haben dadurch nun die Absicherung: der unbekannte (stochastische) Grenzvektor $\vec{\pi}$ ist einfach zu bestimmen, nämlich als **der** (eindeutig existierende, stochastische) Fixvektor von A.

Für den Palio bedeutet dies: wie auch immer der Startvektor bzw. die Startverteilung $\vec{\pi}_0 = \begin{pmatrix} p_0 \\ 1 - p_0 = q_0 \end{pmatrix}$ „aussieht“, beim Palio-Auswahlmodus ist die *Teilnahmewahrscheinlichkeit nach „vielen“ Jahren* davon (nahezu) unabhängig und nähert sich dem Wert $\frac{10}{17}$.

[10] Es ist eine einfache Übungsaufgabe zu zeigen: Wenn A, B, $\vec{x}$ stochastisch sind, so auch $\vec{y} = A \cdot \vec{x}$ und $C = A \cdot B$!

Die **Frage** nach der **Gerechtigkeit** des Palio-Auswahlverfahrens kann also **positiv beantwortet** werden.

4 Übergangsmatrizen und Formales zu den obigen Plausibilitätsbetrachtungen

4.1 Aufstellen von Übergangsmatrizen

Die Elemente p_{ij} der Übergangsmatrix $A = (p_{ij})$ („Übergangswahrscheinlichkeiten“) sind nicht nur im 2×2-Fall durch folgendes Schema leicht zu erhalten (nur zwei mögliche Zustände $Z_1 = T$ und $Z_2 = \neg T$):

		von	
		T	$\neg T$
nach	T	$\frac{3}{10}$	1
	$\neg T$	$\frac{7}{10}$	0

$$A = \begin{pmatrix} \frac{3}{10} & 1 \\ \frac{7}{10} & 0 \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

Genau so bekäme man auch die entsprechende $k \times k$-Übergangsmatrix A bei $k \geq 2$ Zuständen $\{Z_1, \ldots, Z_k\}$:

			von	
		Z_1	$\ldots$	Z_k
nach	Z_1	p_{11}	$\cdots$	p_{1k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	Z_k	p_{k1}	$\cdots$	p_{kk}

p_{ij} ist die Wahrscheinlichkeit, dass „sich der Prozess im Zustand Z_i befindet, wenn er sich davor im Zustand Z_j befunden hat“; man beachte die Rückwärtsrichtung des Pfeiles in: $p_{ij} := \mathrm{P}(Z_i \leftarrow Z_j)$ [kurz: $\mathrm{P}(i \leftarrow j)$] .

4.2 Obige Plausibilitätsbetrachtungen in formaler Sichtweise

Die Inhalte der in den letzten Absätzen salopp formulierten Plausibilitätsbetrachtungen können auch formalisiert (exaktifiziert) werden; Vier einfache Lemmata werden schließlich zu dem Satz „zusammengefügt", der die formalisierte Rechtfertigung dafür ist, im Fall der Existenz von G mit Zeilen aus jeweils gleichen Werten, den unbekannten (stochastischen) **Grenzvektor** $\vec{\pi}$ einfach zu bestimmen, nämlich als **den** (eindeutig existierenden, stochastischen) **Fixvektor** von A.

Lemma 1: $\exists \lim_{n\to\infty} A^n =: G$ (nicht notwendig mit identischen Spalten) $\Rightarrow$ Konvergenz der Folge $(\vec{\pi}_n)$ (für verschiedene Startvektoren $\vec{\pi}_0$ möglicherweise verschiedene Grenzvektoren $\vec{\pi}$) !

Beweis: ($\vec{\pi}_0$ ist ein *konstanter* Vektor !)

$$\lim_{n\to\infty} \vec{\pi}_n = \lim_{n\to\infty} (A^n \cdot \vec{\pi}_0) = \underbrace{\left(\lim_{n\to\infty} A^n\right)}_{G} \cdot \vec{\pi}_0 =$$

$$= G \cdot \vec{\pi}_0 =: \vec{\pi}\ . \quad \blacksquare$$

Definition: Ein Vektor $\vec{s}$ heißt *Fixvektor* von A, wenn $A \cdot \vec{s} = \vec{s}$ gilt.

Lemma 2: Jeder *Grenzvektor* $\vec{\pi}$ ist auch *Fixvektor.*

Beweis: (A ist *konstant*!) $\boxed{A \cdot \vec{\pi}} =$

$$A \cdot (G \cdot \vec{\pi}_0) = (A \cdot G) \cdot \vec{\pi}_0 = \left(A \cdot \lim_{n\to\infty} A^n\right) \cdot \vec{\pi}_0 =$$

$$\underbrace{\left(\lim_{n\to\infty} A^{n+1}\right)}_{G} \cdot \vec{\pi}_0 = G \cdot \vec{\pi}_0 = \boxed{\vec{\pi}}\ . \quad \blacksquare$$

Lemma 3: Jeder *Fixvektor* $\vec{s}$ ist auch *Grenzvektor* (bei Startvektor $\vec{\pi}_0 = \vec{s}$).

Beweis: $A \cdot \vec{s} = \vec{s} \Rightarrow A^2 \cdot \vec{s} = \vec{s} \Rightarrow \dots \Rightarrow A^n \cdot \vec{s} = \vec{s}$ ($\forall n \in \mathbb{N}$). Daher ist $\boxed{\lim_{n\to\infty} (A^n \cdot \vec{s}) = \vec{s}}$ und $\vec{s}$ also Grenzvektor. $\blacksquare$

Lemma 4: $\exists \lim_{n\to\infty} A^n =: G$ mit Zeilen aus jeweils gleichen Werten $\Rightarrow$ der **Grenzvektor** $\vec{\pi}$ **ist unabhängig von** $\vec{\pi}_0$, d. h. **eindeutig!** ($\vec{\pi}_0$, $\vec{\pi}$ beide stochastisch)

Beweis: $\exists\, \vec{\pi} = G \cdot \vec{\pi}_0$ nach Lemma 1.

$$\vec{\pi} = G \cdot \vec{\pi}_0 = \begin{pmatrix} a & a \\ b & b \end{pmatrix} \cdot \begin{pmatrix} p_0 \\ 1 - p_0 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$\Rightarrow$ $\vec{\pi}$ ist immer eine G-Spalte — unabhängig vom konkreten Wert $0 \le p_0 \le 1$. $\blacksquare$

Aus diesen vier Lemmata setzt sich der Beweis des folgenden Satzes zusammen:

Satz 1: $\exists \lim_{n\to\infty} A^n =: G$ mit Zeilen aus jeweils gleichen Werten $\Rightarrow$

eindeutiger Grenzvektor $\vec{\pi}$ **= eindeutiger Fixvektor** $\vec{s}$ ($\vec{\pi}$, $\vec{s}$, $\vec{\pi}_0$ stochastisch)

Beweis: Nach Lemma 4 existiert der Grenzvektor $\vec{\pi}$ für alle $\vec{\pi}_0$, sogar unabhängig von $\vec{\pi}_0$, d. h. er ist eindeutig. Nach Lemma 2 ist $\vec{\pi}$ auch Fixvektor. Umgekehrt ist nach Lemma 3 jeder Fixvektor $\vec{s}$ auch Grenzvektor, der nach Lemma 4 eindeutig $\vec{\pi}$ ist: $\vec{s} = \vec{\pi}$! $\blacksquare$

Bemerkungen:

- Andere mögliche Zugänge und Vertiefungen innerhalb des Matrix-Kalküls (Eigenwerte, Eigenvektoren, Diagonalisierung etc.) seien hier absichtlich ausgeblendet, um das Elementare im Vordergrund zu lassen.
- Man kann Markoff-Ketten auch in Verbindung mit *gerichteten Graphen* betrachten (vgl. Abb. 5), wobei diese eine ebenfalls sehr übersichtliche Form der *Darstellung* jenes Informationsgehaltes sind, der auch in der Übergangsmatrix (durch die Übergangswahrscheinlichkeiten) steckt. Man könnte Markoff-Ketten sogar ausschließlich durch *gerichtete Graphen* darstellen und bearbeiten (z. B. ENGEL 1976), wobei dann allerdings die Verbindung zur Linearen Algebra nicht gegeben wäre.

5 Der Satz von Markoff

Dies alles erhielten wir unter der Voraussetzung, dass eine wie oben beschriebene Grenzmatrix G existiert, so dass die Frage auf der Hand liegt: Wie kann man es einer stochastischen Matrix A ansehen, ob die Matrixpotenzen A^n gegen eine Grenzmatrix G mit Zeilen aus jeweils gleichen Werten konvergieren? (Hinreichendes Kriterium?)

Bemerkung: Hier eröffnet sich mit CAS ein weites *Probierfeld* mit stochastischen Matrizen (insbesondere 2×2, evtl. auch 3×3, 4×4, ...): Finde stochastische Matrizen, so dass A^n gegen eine solche Grenzmatrix G konvergiert (bzw. nicht konvergiert). Besonders für den Fall von 2×2-Matrizen ist eine Vermutung leicht zu erhalten: bis auf zwei (leicht zu findende) Ausnahmen konvergiert (A^n) für *alle* stochastischen

2×2-Matrizen A gegen eine Grenzmatrix G mit den genannten Eigenschaften. Dazu nun der „entscheidende“

Satz von Markoff:
Sei A eine stochastische Matrix. Dann gilt:

$$\boxed{\begin{array}{l}\exists\, n \geq 1:\ A^n \text{ hat (mindestens) } \textbf{eine Zeile} \text{ mit} \\ \textbf{nur positiven} \text{ Elementen (d. h. keine Nullen)}\end{array}}$$

$$\Downarrow$$

$$\boxed{\begin{array}{c}\exists \lim\limits_{n\to\infty} A^n =: G \\ \text{mit Zeilen aus jeweils gleichen Werten}\end{array}}$$

Der *allgemeine* Beweis dieses Satzes ist im Schulunterricht nicht sinnvoll; die Formulierung, die Diskussion seiner Bedeutung und seine Verwendung jedoch sehr wohl! Für den Spezialfall von 2×2-Matrizen gibt es verschiedene einfache Beweise; ein besonders einfacher ist hier dargestellt[11].

Zur Vorbereitung des Beweises ist es günstig, die ersten Potenzen einer stochastischen 2×2-Matrix $A = \begin{pmatrix} 0,30 & 0,60 \\ 0,70 & 0,40 \end{pmatrix}$ etwas genauer zu betrachten A^2, A^3, A^4 und die Entwicklung der „Zeilenintervalle“ etwas genauer zu beobachten:

$$\begin{pmatrix} 0,51 & 0,42 \\ 0,49 & 0,58 \end{pmatrix}, \begin{pmatrix} 0,447 & 0,474 \\ 0,553 & 0,526 \end{pmatrix}, \begin{pmatrix} 0,466 & 0,458 \\ 0,534 & 0,542 \end{pmatrix}$$

Man kann dabei schon einiges über die einzelnen Zeilen in den Matrixpotenzen ablesen und dadurch auf genauer zu untersuchende Vermutungen kommen, die den Beweis schon „vorwegnehmen“:

Die „Zeilenintervalle“ scheinen sich *zusammenzuziehen*:

- das Intervall der 1. Zeile von A^{n+1} ist ganz im 1. Zeilenintervall von A^n enthalten: $[0,30; 0,60] \supseteq [0,42; 0,51] \supseteq [0,447; 0,474] \supseteq [0,458; 0,466]$!

- sucht man nach einem bestimmten Muster, nach dem sich diese Intervalle entwickeln, so kann rasch ins Auge springen, dass alle Zeilenintervall**längen** ca. das 0,3-fache der vorhergehenden sind; die Längen der 1. Zeilenintervalle betragen der Reihe nach: $0,3 \to 0,09 \to 0,027 \to 0,008$. (Beide Punkte analog für die 2. Zeilen)!

[11] Ich danke Herrn W. Hansen (Universität Bielefeld) für Anregungen zu diesem Beweis. Es ist damit auch schon der Weg zum Beweis des allgemeinen Falls angedeutet: „Intervallschachtelung“.

Diese Beobachtungen regen zu Vermutungen an, die im Folgenden formuliert und bewiesen werden.

Lemma 5: Sei $A = \begin{pmatrix} p & q \\ 1-p & 1-q \end{pmatrix}$ eine stochastische 2×2-Matrix (d. h. $p, q \in [0;1]$) und $A^n =: \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}$; dann gilt in der „Folge der Zeilenintervalle“ für alle $n \geq 1$:

1. jedes „neue“ **Zeilenintervall** $[a_{n+1}, b_{n+1}]$ liegt ganz im „alten“ $[a_n, b_n]$ und analog: $[c_{n+1}, d_{n+1}] \subseteq [c_n, d_n]$!
2. die „neuen“ **Zeilenintervall-Längen** in A^{n+1} betragen das $(p-q)$-Fache der „alten“ in A^n !

Beweis: Wir berechnen zunächst

$$\boxed{\begin{pmatrix} a_{n+1} & b_{n+1} \\ c_{n+1} & d_{n+1} \end{pmatrix}} = A^{n+1} = A^n \cdot A =$$

$$= \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \cdot \begin{pmatrix} p & q \\ 1-p & 1-q \end{pmatrix} =$$

$$\boxed{\begin{pmatrix} a_n p + b_n(1-p) & a_n q + b_n(1-q) \\ c_n p + d_n(1-p) & c_n q + d_n(1-q) \end{pmatrix}}$$

und erhalten dadurch unmittelbar:

1. **Zeilenintervalle:** a_{n+1}, b_{n+1} sind „konvexe Linearkombinationen“ von a_n, b_n:

   $$a_{n+1} = a_n p + b_n(1-p)\ , \quad b_{n+1} = a_n q + b_n(1-q)\ ,$$

   womit sich $a_{n+1}, b_{n+1} \in [a_n, b_n]$ ergibt (analog für die 2. Zeile).

2. **Zeilenintervall-Längen:** Man errechnet unmittelbar:

   $$\begin{aligned} b_{n+1} - a_{n+1} &= (p-q)(b_n - a_n)\ , \\ d_{n+1} - c_{n+1} &= (p-q)(d_n - c_n)\ , \end{aligned}$$

   womit obige Behauptung gezeigt ist. Wendet man diese Erkenntnis beginnend bei n

schrittweise $n-1$-mal an, so erhält man daraus sogar weiter gehend

$$b_n - a_n = (p-q)^{n-1}\underbrace{(b_1 - a_1)}_{q-p} = -(p-q)^n ,$$
$$d_n - c_n = (p-q)^{n-1}\underbrace{(d_1 - c_1)}_{p-q} = (p-q)^n . \quad \blacksquare$$

Der Satz von Markoff für 2×2-Matrizen:

Sei $A = \begin{pmatrix} p & q \\ 1-p & 1-q \end{pmatrix}$ eine stochastische 2×2-Matrix, d. h. $p, q \in [0;1]$; dann gilt:

Genau für $(p,q) \neq (1,0)$ und $(p,q) \neq (0,1)$ hat die Grenzmatrix $\lim_{n\to\infty} A^n =: G = \begin{pmatrix} \alpha & \alpha \\ \beta & \beta \end{pmatrix}$ Zeilen aus jeweils gleichen Werten.

Beweis:

1. • Für $(p,q) = (1,0)$ ist $A = E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ und daher $A^n = E = G$, d. h. G hat nicht die geforderte Gestalt.
   • Für $(p,q) = (0,1)$ ist $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ und $A^{2n} = E$ bzw. $A^{2n+1} = A$ für alle n; daher existiert G nicht.

2. Sei nun $(p,q) \neq (1,0) \quad \wedge \quad (p,q) \neq (0,1)$ $\Rightarrow \quad |p-q| < 1$.

   In $A^n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}$ liegt nach Lemma 5.1 jedes „neue Zeilenintervall" ganz im vorherigen; nach Lemma 5.2 gehen die Intervall-Längen ($|p-q| < 1!$) gegen 0:

$$\lim_{n\to\infty} (b_n - a_n) = \lim_{n\to\infty} (-(p-q)^n) = 0$$
$$\lim_{n\to\infty} (d_n - c_n) = \lim_{n\to\infty} ((p-q)^n) = 0$$

   Daraus folgt (*Intervallschachtelung!*) die Existenz und Gleichheit der Limiten:

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n =: \alpha ;$$
$$\lim_{n\to\infty} c_n = \lim_{n\to\infty} d_n =: \beta . \quad \blacksquare$$

Bemerkung: Beim allgemeinen Satz ist der Beweis, dass jedes Zeilenintervall ganz im nächsten liegt, genau so einfach („konvexe Linearkombination"); dass die Zeilenintervall-Längen gegen 0 konvergieren, ist doch aufwändiger zu zeigen.

Erfahrungen aus dem Unterricht:

Herr Florian Rösner hat in einem Leistungskurs der Klasse 12 (wirtschaftlich orientierter Zweig des Friedrich-List-Berufskollegs in Hamm) diesen Zugang zu Markoff-Ketten ausprobiert und dem Autor nur Positives darüber berichtet (die Schülerinnen und Schüler waren im Rahmen einer Kultur-Woche in der Toskana sogar vor Ort in Siena und haben einen Palio gesehen). Ein anschließender Unterrichtsschwerpunkt lag vor allem bei *selbständigem Experimentieren mit CAS*. Da die Schülerinnen und Schüler mit CAS umgehen konnten, haben auch Leistungsschwächere herausgefunden (etwas langsamer als die leistungsstärkeren, aber doch selbständig!), dass die Matrixpotenzen A^n gegen eine spezielle Grenzmatrix zu konvergieren scheinen, wodurch ein Begründungsbedürfnis entstanden ist. Dies ist zunächst durch heuristische Plausibilitätsbetrachtungen und anschließend in Form eines (anderen) Beweises für den Fall von 2×2-Matrizen geschehen. Weitere Anwendungsaufgaben aus dem Bereich Wirtschaft-Marketing ergänzten den Lehrgang.

Literatur:

1. ENGEL, A. (1976): Wahrscheinlichkeitsrechnung und Statistik (Band 2). Klett, Stuttgart.
2. GÖTZ, S. u. M. GROSSER (1999): Über das Pferderennen in Siena. In: Mathematische Semesterberichte **46**, 1, 77–92.
3. JACOBS, M. u. H. SORGE (1994): In welchem Laden kaufst Du morgen ein? – Markow-Ketten und Entscheidungen. In: mathematik lehren **63**, 47 - 50.
4. LAMBACHER-SCHWEIZER (2001): Leistungskurs: Lineare Algebra mit analytischer Geometrie. Klett, Stuttgart.
5. LEHMANN, E. (1973): Endliche homogene Markoffsche Ketten. Beiträge für den mathematischen Unterricht (10). Bayerischer Schulbuchverlag, München.
6. LEHMANN, E. (1986): Markow-Ketten. In: Der Mathematikunterricht **32**, 5, 60–92.
7. MEYER, D. (1998): Markoff-Ketten. In: Mathematik in der Schule **36**, 12, 661–670, 675–680.
8. SCHEID, H. (1986): Stochastik in der Kollegstufe. BI, Mannheim (jetzt: Spektrum).
9. WIRTHS, H. (1997): Markow-Ketten – Brücke zwischen Analysis, linearer Algebra und Stochastik. In: Math. Schule **35**, 11, 601–613.

Anschrift des Verfassers: Hans HUMENBERGER, IEEM, FB Mathematik, Universität Dortmund, D - 44221 Dortmund.

Mail: `hans.humenberger@math.uni-dortmund.de`