

# Von Pferden, Ziegen und unmöglichen Würfeln

STEFAN GÖTZ

## Zusammenfassung

In dieser Arbeit werden stochastische Problemstellungen mit Methoden der Analysis und / oder der Linearen Algebra behandelt. Dazu besuchen wir das traditionelle Pferderennen in Siena, eine Talkshow in den USA, suchen zwei Würfel, die miteinander den strengen Vorstellungen von LAPLACE genügen sollen u. a. m. Die Vernetzung verschiedener Gebiete (in) der Mathematik trägt immer wieder zu entscheidenden Fortschritten in dieser Wissenschaft bei, insofern ist sie typisch für die mathematische Arbeitsweise. Eine kleine Kostprobe davon kann auch im Mathematikunterricht angeboten werden, wie exemplarisch eben an diesen drei Gebieten der Schulmathematik gezeigt werden soll. Diesbezügliche Querverweise sind **fett** gedruckt.

## 1 Die Pferde (nach [GG])

### 1.1 Die Ausgangssituation

Der Palio (manchmal auch: das Palio, ital. „Il Palio“) ist eines der härtesten Pferderennen der Welt. Es wird auf dem zentralen Platz in Siena ausgetragen, der Piazza del Campo. Im Rennen zu Ehren der Jungfrau Maria treten die (Vertreter der) heutigen 17 Contraden, das sind die Stadtteile Sienas, gegeneinander an. Es wird seit dem Mittelalter zweimal jährlich ausgetragen (2. Juli und 16. August) und ist „das“ Ereignis in Siena.

Der Name „Palio“ kommt aus dem lateinischen „pallium“ und bedeutet „Tuch“ oder „Umhang“. Später bekam das Wort auch die Bedeutung „Fahne“ bzw. „Standarte“. Der Sieger des Palio bekommt als Preis eine bunte Standarte, die jährlich neu gestaltet wird.

Aus dem Roman „Der Palio der toten Reiter“ von CARLO FRUTTERO und FRANCO LUCENTINI, der im Verlag Piper, München 1986 erschienen ist ([FRU]), sei dazu Folgendes zitiert (S. 18f.):

*„Für ein Konversationshandbuch über den Palio*

*Guidobaldo*: ‚Siena hat siebzehn Contraden, aber nur zehn davon beteiligen sich am Palio.‘

*Valeria*: ‚Ach wirklich? Und wieso?‘

*Guidobaldo*: ‚Die Rennbahn um den Platz ist eng, unregelmäßig, voller Auf und Ab und gefährlicher Kurven (lange Fußnote über die mörderische San-Martino-Kurve). Was meinen Sie, was passieren würde, wenn da siebzehn Pferde gleichzeitig losgaloppierten? Sie würden nicht mal zwanzig Meter weit kommen, ohne zu stürzen.‘

*Valeria* (sich die Hände vor die Augen schlagend): ‚Die armen Tiere!‘

*Guidobaldo*: ‚Deswegen machen bei jedem Palio nur zehn Contraden mit, und die sieben anderen haben das Recht, im nächsten Jahr mitzumachen, und so weiter.‘

*Anwalt*: ‚Und die übrigen drei, die bis zehn noch fehlen?‘

*Guidobaldo*: ‚Die werden aus den zehn Teilnehmern des letzten Jahres ausgelost.‘

Ergänzt sei noch, dass die beiden jährlichen Rennen *unabhängig voneinander* betrachtet werden. Das eben vorgestellte Verfahren zur Bestimmung der teilnehmenden Contraden (an einem bestimmten Rennen) wird auf die Rennen von Juli auf Juli und auf die von August auf August *getrennt* angewendet.

## 1.2 Das Modell

Wir wollen nun das im vorigen Abschnitt beschriebene Auslosungsverfahren zur Bestimmung der teilnehmenden Contraden an einem bestimmten Rennen modellieren und fassen dazu eine *bestimmte Contrade* ins Auge.

Es sei die Zufallsvariable

$$X_i := \begin{cases} 1, & \text{falls die Contrade am } i\text{-ten Rennen teilnimmt,} \\ 0, & \text{falls die Contrade am } i\text{-ten Rennen nicht teilnimmt,} \end{cases}$$

mit gewissen *Anfangswahrscheinlichkeiten*  $p_1 := P(X_1 = 1)$  und  $q_1 := P(X_1 = 0)$  definiert. Für diese gilt natürlich  $0 \leq p_1, q_1 \leq 1$  und  $p_1 + q_1 = 1$ .

Die Auswahlregeln für die Teilnahme am Palio (siehe [FRU]) finden in den sogenannten (bedingten) *Übergangswahrscheinlichkeiten* ihren Ausdruck:

$$\begin{aligned} p_{00} &:= P(X_{n+1} = 0 | X_n = 0) = 0 \\ p_{01} &:= P(X_{n+1} = 1 | X_n = 0) = 1 - p_{00} \\ p_{10} &:= P(X_{n+1} = 0 | X_n = 1) = \frac{7}{10} \\ p_{11} &:= P(X_{n+1} = 1 | X_n = 1) = \frac{3}{10} (= 1 - p_{10}). \end{aligned}$$

Übersetzt heißt das, dass die Nichtteilnahme an zwei Rennen hintereinander unmöglich ist, dagegen die Teilnahme unmittelbar nach einer Nichtteilnahme sicher, zwei Teilnahmen hintereinander passieren mit Wahrscheinlichkeit  $\frac{3}{10} = \frac{\binom{9}{2}}{\binom{10}{3}}$ , schließlich ist eine Nichtteilnahme unmittelbar nach einer Teilnahme mit Wahrscheinlichkeit  $\frac{7}{10}$  möglich.

Weiters entnehmen wir [FRU]:

$$\begin{aligned} & P(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_1 = i_1) = \\ & = P(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n i_{n+1}} \quad (i_k \in \{0, 1\} \forall k = 1, \dots, n+1), \end{aligned}$$

das heißt also, dass nur das unmittelbar vorhergehende Rennen auf das nachfolgende einen Einfluss hat, nicht aber die Rennen davor.

ACHTUNG: In [FRU] finden wir *keine konkreten Werte* für  $p_1$  bzw.  $q_1$ !

### 1.3 Die Fairness des Auslosungsverfahrens

Die nun folgenden Überlegungen drehen sich alle um die *Frage*: Ist das Auslosungsverfahren auf lange Sicht „fair“?

Mit anderen Worten und salopp: Gilt  $\lim_{n \rightarrow \infty} r_n = \frac{\binom{16}{9}}{\binom{17}{10}} = \frac{10}{17}$ ? (Dabei ist  $r_n$  die relative Häufigkeit der Teilnahmen „unserer“ Contrade nach  $n$  Rennen.)

Genauer werden wir die *Erwartungswerte* von  $X_n$ ,  $Y_n := \sum_{i=1}^n X_i$  (beschreibt die Gesamtanzahl der Teilnahmen der Contrade nach  $n$  Rennen) und  $Z_n := \frac{1}{n} \cdot Y_n$  berechnen, das Gesetz der großen Zahlen stellt dann eine Beziehung zwischen diesen (theoretischen) Parametern und der empirischen relativen Häufigkeit  $r_n$  her.

Fairness bestehe genau dann, wenn  $\lim_{n \rightarrow \infty} EZ_n = \frac{10}{17}$  gilt.

Im Folgenden werden wir auf verschiedene Arten dieses Ergebnis der Fairness des Auslosungsverfahrens herleiten.

#### 1.3.1 Eine Rekursionsformel

Wir schreiben nun  $p_n := P(X_n = 1)$ . Damit ergibt sich

$$EX_n = 1 \cdot P(X_n = 1) + 0 \cdot P(X_n = 0) = p_n,$$

$$EY_n = \sum_{k=1}^n p_k \quad \text{und}$$

$$EZ_n = \frac{1}{n} \cdot \sum_{k=1}^n p_k.$$

Wir beweisen

$$\lim_{n \rightarrow \infty} p_n = \frac{10}{17},$$

damit auch:

$$\lim_{n \rightarrow \infty} EZ_n = \lim_{n \rightarrow \infty} \frac{1}{n} \cdot \sum_{k=1}^n p_k = \frac{10}{17}$$

nach einem bekannten Satz aus der Analysis (**Grenzwertsatz von CAUCHY**, siehe z. B. [HEU], S. 177): Wenn eine reelle Folge  $a_n$  für  $n$  gegen Unendlich gegen  $a$  konvergiert, dann strebt das arithmetische Mittel der Folgenglieder  $\frac{\sum_{k=1}^n a_k}{n}$  ebenfalls gegen  $a$  (wieder für  $n$  gegen Unendlich).

Mit  $p := p_{11} = \frac{3}{10}$  ergibt sich aus den Regeln für die Teilnahme am Palio die *Rekursionsformel*

$$\begin{aligned} p_{n+1} &= P(X_{n+1} = 1) = \\ &= P(X_{n+1} = 1 | X_n = 1) \cdot p_n + P(X_{n+1} = 1 | X_n = 0) \cdot (1 - p_n) = \\ &= p_{11} \cdot p_n + p_{01} \cdot (1 - p_n) = \\ &= p \cdot p_n + 1 \cdot (1 - p_n) = \\ &= 1 - (1 - p) \cdot p_n \end{aligned} \tag{1}$$

mit Hilfe des *Satzes von der totalen Wahrscheinlichkeit*.

Sie liefert *einerseits*

$$\begin{aligned} p_{n+1} &= 1 - (1 - p) \cdot p_n = 1 - p_n + p \cdot p_n, \\ p_{n+1} - p_n &= 1 - 2 \cdot p_n + p \cdot p_n = 1 - (2 - p) \cdot p_n \end{aligned}$$

[für  $n = 1$  ergibt sich speziell

$$p_2 - p_1 = 1 - (2 - p) \cdot p_1];$$

*andererseits*

$$p_{n+1} - p_n = (p - 1) \cdot (p_n - p_{n-1}) = \dots = (p - 1)^{n-1} \cdot (p_2 - p_1)$$

wegen

$$\begin{aligned} p_{n+1} &= 1 - (1 - p) \cdot p_n \\ p_n &= 1 - (1 - p) \cdot p_{n-1} \end{aligned}$$

---

$$p_{n+1} - p_n = (1 - p) \cdot (p_{n-1} - p_n) = (p - 1) \cdot (p_n - p_{n-1}).$$

Gleich- und Einsetzen liefert

$$1 - (2 - p) \cdot p_n = (p - 1)^{n-1} \cdot [1 - (2 - p) \cdot p_1].$$

Mit

$$1 - (p-1)^{n-1} \cdot [1 - (2-p) \cdot p_1] = (2-p) \cdot p_n$$

folgt die gewünschte *explizite Formel* für  $p_n$ :

$$\begin{aligned} EX_n = p_n &= \frac{1}{2-p} - \frac{(p-1)^{n-1}}{2-p} \cdot [1 - (2-p) \cdot p_1] = \\ &= \frac{1}{2-p} - (p-1)^{n-1} \cdot \left( \frac{1}{2-p} - p_1 \right) = \\ &= \frac{1}{2-p} + (p-1)^{n-1} \cdot \left( p_1 - \frac{1}{2-p} \right). \end{aligned} \tag{2}$$

Daraus erkennen wir sofort

$$\lim_{n \rightarrow \infty} p_n = \frac{1}{2-p} = \frac{1}{2 - \frac{3}{10}} = \frac{10}{17}$$

wegen

$$\lim_{n \rightarrow \infty} (p-1)^{n-1} = 0, \quad \text{weil } |p-1| = \left| -\frac{7}{10} \right| < 1$$

(Konvergenzkriterium für geometrische Folgen).

Damit haben wir auch — wie erwähnt —

$$\lim_{n \rightarrow \infty} EZ_n = \lim_{n \rightarrow \infty} \frac{1}{n} \cdot \sum_{k=1}^n p_k = \frac{10}{17}$$

gezeigt.

*Explizit* berechnen wir mit (2)

$$\begin{aligned} EY_n = \sum_{k=1}^n EX_k &= \frac{n}{2-p} + \left[ \sum_{k=0}^{n-1} (p-1)^k \right] \cdot \left( p_1 - \frac{1}{2-p} \right) = \\ &= \frac{n}{2-p} + \frac{1 - (p-1)^n}{1 - (p-1)} \cdot \left( p_1 - \frac{1}{2-p} \right) = \\ &= \frac{n}{2-p} + \frac{1 - (p-1)^n}{2-p} \cdot \left( p_1 - \frac{1}{2-p} \right) \end{aligned}$$

mit Hilfe der Summenformel für die endliche geometrische Reihe und schließlich

$$EZ_n = \frac{1}{n} \cdot EY_n = \frac{1}{2-p} + \frac{1}{n} \cdot \frac{1 - (p-1)^n}{2-p} \cdot \left( p_1 - \frac{1}{2-p} \right)$$

bzw. (unabhängig von  $p_1$ !)

$$\lim_{n \rightarrow \infty} EZ_n = \frac{1}{2-p} = \frac{10}{17}.$$

### 1.3.2 Eine inhomogene lineare Differenzgleichung erster Ordnung

Wir können (1) auch als inhomogene lineare Differenzgleichung erster Ordnung sehen. Die Lösungen einer solchen Gleichung findet man bekanntlich durch *Addieren* einer *speziellen Lösung* der *inhomogenen* Differenzgleichung mit der *allgemeinen Lösung* der zugehörigen *homogenen* Differenzgleichung. — Vgl. dazu das Kapitel „**Mathematische Beschreibung dynamischer Systeme und Prozesse**“ — insbesondere Beispiel 805 auf S. 220 — aus [GR7]!

Die allgemeine Form einer solchen Gleichung lautet  $y_{n+1} = a + b \cdot y_n$  mit  $b \neq 1$ , sie hat die *konstante Lösung*  $y_n = c$  mit  $c = a + b \cdot c$ , also  $c = \frac{a}{1-b}$ . Bei uns ist  $a = 1$  und  $b = p - 1 \leq 0$ , daher  $c = \frac{1}{2-p}$ .

Die *homogene* Gleichung  $z_{n+1} = -(1-p) \cdot z_n = (p-1) \cdot z_n$  hat die *allgemeine Lösung*  $z_n = \mu \cdot (p-1)^n$  ( $\mu \in \mathbb{R}$ ), sodass wir nur mehr  $\mu$  aus der Anfangsbedingung bestimmen müssen:  $p_1 = y_1 + z_1 = \frac{1}{2-p} + \mu \cdot (p-1)$ , also  $\mu = \left(p_1 - \frac{1}{2-p}\right) \cdot \frac{1}{p-1}$ .

Insgesamt erhalten wir so

$$\begin{aligned} p_n = y_n + z_n &= \frac{1}{2-p} + \left(p_1 - \frac{1}{2-p}\right) \cdot \frac{1}{p-1} \cdot (p-1)^n = \\ &= \frac{1}{2-p} + (p-1)^{n-1} \cdot \left(p_1 - \frac{1}{2-p}\right), \end{aligned}$$

also wiederum (2).

### 1.3.3 Eine homogene lineare Differenzgleichung zweiter Ordnung

- Eine Teilnahme beim  $(n+1)$ -ten Rennen folgt (in einander ausschließender Weise)
  - entweder aus einer Teilnahme beim  $n$ -ten Rennen und darauffolgendem positiven Losentscheid
  - oder aus einer Nichtteilnahme beim  $n$ -ten Rennen, der notwendigerweise eine Teilnahme am  $(n-1)$ -ten Rennen und ein negativer Losentscheid vorangegangen sein müssen:

$$p_{n+1} = p \cdot p_n + (1-p) \cdot p_{n-1}.$$

- Eine formale Herleitung aus der Rekursionsformel (1) sieht so aus:

$$\begin{aligned} p_{n+1} &= 1 - p_n + p \cdot p_n = 1 - [1 - (1-p) \cdot p_{n-1}] + p \cdot p_n = \\ &= p \cdot p_n + (1-p) \cdot p_{n-1}. \end{aligned}$$

In beiden Fällen erhalten wir also eine homogene lineare Differenzgleichung zweiter Ordnung.

Mit dem Ansatz  $p_n = \lambda^n$  erhalten wir, wenn wir ihn in die Differenzgleichung einsetzen, die sog. *charakteristische Gleichung*

$$\lambda^2 - p \cdot \lambda - (1 - p) = 0$$

mit den Lösungen

$$\lambda_{1,2} = \frac{p}{2} \pm \sqrt{\frac{p^2}{4} + 1 - p} = \left\{ \begin{array}{l} 1 \\ -(1 - p) \end{array} \right.$$

Für  $\lambda_1 \neq \lambda_2$  hat eine derartige Differenzgleichung die *allgemeine Lösung*

$$c_1 \cdot \lambda_1^n + c_2 \cdot \lambda_2^n \quad \text{mit} \quad c_i \in \mathbb{R} \quad (i = 1, 2).$$

(Vgl. dazu in [GR7] S. 232 „Differenzgleichungen 2. Ordnung mit einer Variablen“!)

$c_1$  und  $c_2$  bestimmen wir aus den beiden Anfangsbedingungen für  $n = 1$  und  $n = 2$ : Es ergibt sich  $p_1$  bzw.  $p_2 = 1 - (1 - p) \cdot p_1$ .

Also:

$$\begin{aligned} p_1 &= c_1 \cdot 1 + c_2 \cdot (p - 1) \\ 1 - (1 - p) \cdot p_1 &= c_1 \cdot 1^2 + c_2 \cdot (p - 1)^2 \end{aligned}$$

Dieses lineare Gleichungssystem hat die eindeutig bestimmten Lösungen

$$c_1 = \frac{1}{2 - p} \quad \text{und} \quad c_2 = \frac{1}{(1 - p) \cdot (2 - p)} - \frac{p_1}{1 - p}.$$

Insgesamt erhalten wir aus

$$\begin{aligned} p_n &= c_1 \cdot 1^n + c_2 \cdot (p - 1)^n = \\ &= \frac{1}{2 - p} + \left( p_1 - \frac{1}{2 - p} \right) \cdot (p - 1)^{n-1} \end{aligned}$$

wiederum die explizite Formel (2) für  $p_n$ .

### 1.3.4 Die Teilnahme am Palio als MARKOFF-Kette

Die *Idee* ist nun, mögliche Abfolgen von Teilnahme bzw. Nichtteilnahme einer bestimmten Contrade am Palio als sogenannte MARKOFF-Kette darzustellen.

In Anspielung auf  $p_n := P(X_n = 1)$  definieren wir nun  $q_n := 1 - p_n = P(X_n = 0)$ .

Die Regeln für die Teilnahme am Palio sehen damit so aus:

$$\begin{aligned} p_{n+1} &= p \cdot p_n + q_n \\ q_{n+1} &= (1 - p) \cdot p_n \end{aligned}$$

Mit  $p_n = 1 - q_n$  erhalten wir jeweils Gleichung (1):

$$\begin{aligned} p_{n+1} &= (1 - p_n) + p \cdot p_n = 1 - (1 - p) \cdot p_n \\ 1 - p_{n+1} &= (1 - p) \cdot p_n . \end{aligned}$$

In Matrixschreibweise lautet das obige System:

$$\begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix} = \begin{pmatrix} p & 1 \\ 1 - p & 0 \end{pmatrix} \cdot \begin{pmatrix} p_n \\ q_n \end{pmatrix} = \begin{pmatrix} \frac{3}{10} & 1 \\ \frac{7}{10} & 0 \end{pmatrix} \cdot \begin{pmatrix} p_n \\ q_n \end{pmatrix} .$$

Wir interpretieren  $\pi_n := \begin{pmatrix} p_n \\ q_n \end{pmatrix}$  als Vektor, der die Teilnahme der betrachteten Contrade am  $n$ -ten Rennen beschreibt. Die Multiplikation von  $\pi_n$  mit  $T := \begin{pmatrix} p & 1 \\ 1 - p & 0 \end{pmatrix} = \begin{pmatrix} \frac{3}{10} & 1 \\ \frac{7}{10} & 0 \end{pmatrix}$  liefert den Vektor  $\pi_{n+1} = \begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix}$ , der die Teilnahme am  $(n + 1)$ -ten Rennen beschreibt.

*Bemerkung:* Die Eintragungen von  $T$  sind gerade die Übergangswahrscheinlichkeiten:

$$\begin{pmatrix} p_{11} & p_{01} \\ p_{10} & p_{00} \end{pmatrix} = T .$$

*Definition:* Ein Vektor

$$\begin{pmatrix} p \\ q \end{pmatrix} \quad \text{mit} \quad 0 \leq p, q \leq 1 \quad \text{und} \quad p + q = 1$$

heißt *Wahrscheinlichkeitsvektor*.

Mit  $\pi$  ist stets auch  $T \cdot \pi$  ein Wahrscheinlichkeitsvektor:

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

Zur *Terminologie*:

- Eine Folge von Zufallsvariablen  $\langle X_n \rangle$  heißt *stochastischer Prozess*. Der Index  $n$  kann dabei oft als Zeit(punkt) gedeutet werden.
- Wenn der „Zustand“ zum Zeitpunkt  $n + 1$  nur von dem zum Zeitpunkt  $n$  abhängt, wie das bei uns der Fall ist:  $\pi_{n+1} = T \cdot \pi_n$ , dann ist der stochastische Prozess der  $X_n$  eine *MARKOFF-Kette*.
- $T$  hängt *nicht* von  $n$  ab (die Teilnahmeregeln für den Palio ändern sich nicht): Die MARKOFF-Kette ist *homogen*.
- $T$  wird die *Übergangsmatrix* der MARKOFF-Kette genannt.



Also: Aus  $\pi_1 = \begin{pmatrix} p_1 \\ q_1 \end{pmatrix}$  (Teilnahme am ersten Rennen) ergibt sich mit

$$\pi_n = \begin{pmatrix} p_n \\ q_n \end{pmatrix} = T \cdot \pi_{n-1} = T \cdot T \cdot \pi_{n-2} = T^2 \cdot \pi_{n-2} = \dots = T^{n-1} \cdot \pi_1$$

die Teilnahmewahrscheinlichkeit beim  $n$ -ten Rennen.

Wir wollen nun die Matrix  $T$  untersuchen:

- Die Eigenwerte von  $T$  ergeben sich aus der Gleichung

$$\det(\lambda \cdot I - T) = \begin{vmatrix} \lambda - p & -1 \\ -(1-p) & \lambda \end{vmatrix} = \lambda^2 - p \cdot \lambda - (1-p) = 0,$$

wobei  $I$  die Einheitsmatrix meint.

*Bemerkung:*

Das ist die charakteristische Gleichung der Differenzengleichung von eben!

Dementsprechend bekommen wir auch hier als Lösungen

$$\lambda_{1,2} = \frac{p}{2} \pm \sqrt{\frac{p^2}{4} + 1 - p} = \begin{cases} 1 \\ -(1-p) \end{cases},$$

das sind die Eigenwerte  $\lambda_1$  und  $\lambda_2$  von  $T$ .

- Eigenvektoren  $\xi_i$  von  $T$  erhalten wir wegen

$$\lambda_i \cdot \xi_i = T \cdot \xi_i \quad (i = 1, 2), \quad \xi_i \neq \vec{0},$$

wie üblich durch Lösen der Systeme

$$(T - \lambda_i \cdot I) \cdot \xi_i = \vec{0} \quad (i = 1, 2).$$

\* Für  $\lambda_1 = 1$  ist

$$\begin{pmatrix} \frac{3}{10} - 1 & 1 \\ \frac{7}{10} & 0 - 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -\frac{7}{10} & 1 \\ \frac{7}{10} & -1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

woraus wir

$$-\frac{7}{10} \cdot x_1 + x_2 = 0$$

schließen.

Für  $x_2 = 1$  ist  $x_1 = \frac{10}{7}$ . Wir wählen den Eigenvektor

$$\xi_1 = \frac{7}{17} \cdot \begin{pmatrix} \frac{10}{7} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{10}{17} \\ \frac{1}{17} \end{pmatrix} = \begin{pmatrix} \frac{1}{2-p} \\ \frac{1-p}{2-p} \end{pmatrix}.$$

\* Für  $\lambda_2 = -\frac{7}{10}$  sehen wir:

$$\begin{pmatrix} \frac{3}{10} - \left(-\frac{7}{10}\right) & 1 \\ \frac{7}{10} & 0 - \left(-\frac{7}{10}\right) \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \\ = \begin{pmatrix} 1 & 1 \\ \frac{7}{10} & \frac{7}{10} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

also

$$y_1 + y_2 = 0.$$

Für  $y_1 = 1$  ist  $y_2 = -1$ . Wir wählen den Eigenvektor

$$\xi_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Untersuchen wir nun das *Langzeitverhalten* von  $\pi_1 = \begin{pmatrix} p_1 \\ q_1 \end{pmatrix}$ : Dazu sei  $\xi \in \mathbb{R}^2$  beliebig und betrachte  $\langle T^n \cdot \xi \rangle$ .

Die Basis  $\{\xi_1, \xi_2\}$  des  $\mathbb{R}^2$  aus Eigenvektoren liefert  $\xi = c_1 \cdot \xi_1 + c_2 \cdot \xi_2$ , daraus folgt wieder mit dem **Konvergenzkriterium für geometrische Folgen**

$$T^n \cdot \xi = c_1 \cdot T^n \cdot \xi_1 + c_2 \cdot T^n \cdot \xi_2 = c_1 \cdot \xi_1 + (p-1)^n \cdot c_2 \cdot \xi_2 \longrightarrow c_1 \cdot \xi_1 \quad (n \rightarrow \infty).$$

Für *jeden* Wahrscheinlichkeitsvektor  $\xi = \pi_1 = \begin{pmatrix} p_1 \\ q_1 \end{pmatrix}$  ist auch der Grenzwert  $c_1 \cdot \xi_1$  ein solcher; daraus folgt  $c_1 = 1$  und

$$T^n \cdot \pi_1 \rightarrow \xi_1 = \pi_\infty = \begin{pmatrix} \frac{10}{17} \\ \frac{7}{17} \end{pmatrix} = \begin{pmatrix} \frac{1}{\frac{2-p}{1-p}} \\ \frac{1-p}{2-p} \end{pmatrix} \quad (n \rightarrow \infty).$$

Aus

$$T \cdot \xi_1 = \lambda_1 \cdot \xi_1 = 1 \cdot \xi_1 = \xi_1$$

sehen wir, dass die *Teilnahmewahrscheinlichkeit*  $\frac{1}{2-p} = \frac{10}{17}$  einer bestimmten *Contrade zeitlich invariant* ist.

Daher heißt  $\pi_\infty = \xi_1 = \begin{pmatrix} \frac{10}{17} \\ \frac{7}{17} \end{pmatrix} = \begin{pmatrix} \frac{1}{\frac{2-p}{1-p}} \\ \frac{1-p}{2-p} \end{pmatrix}$  auch *stationärer Wahrscheinlichkeitsvektor* der vorliegenden MARKOFF-Kette.

*Beachte:* Nur  $\xi_1$  ist Eigenvektor und zugleich Wahrscheinlichkeitsvektor!

Also ist für diese Auslosungsregeln  $\frac{10}{17}$  die *einzig* zeitlich invariante Teilnahmewahrscheinlichkeit. *Jede andere anfängliche Teilnahmewahrscheinlichkeit konvergiert „im Laufe der Zeit“ gegen diesen Wert.*

## 1.4 Moral

Alternativ könnten die gegebenen Auslosungsregeln durch folgendes Procedere ersetzt werden: Jedes Jahr werden von neuem durch Losentscheid zehn der siebzehn Contraden bestimmt am Palio teilzunehmen. Die „Fairness“ im Sinne der langfristigen Tendenz der Erwartungswerte wäre natürlich gewahrt, aber zwei oder mehr aufeinanderfolgende Nichtteilnahmen könnten dann eine bestimmte Contrade treffen. Das heißt das tatsächlich verwendete Verfahren sichert die *Minimierung der Frustration einer bestimmten Contrade* im Vergleich zur eben diskutierten Alternative.

## 1.5 Didaktischer Kommentar

Es handelt sich hierbei um einen *(hoch-)schuldidaktischen Beitrag aus folgenden Gründen:*

Die verschiedenen Übersetzungsvarianten der Auslosungsregeln für die Teilnahme am Palio können als vielfältiger *Modellbildungsprozess* angesehen werden. Dabei ist es hier so, dass das Ergebnis *nicht* von der Modellwahl abhängt, sehr wohl aber die zum Einsatz kommenden mathematischen Werkzeuge.

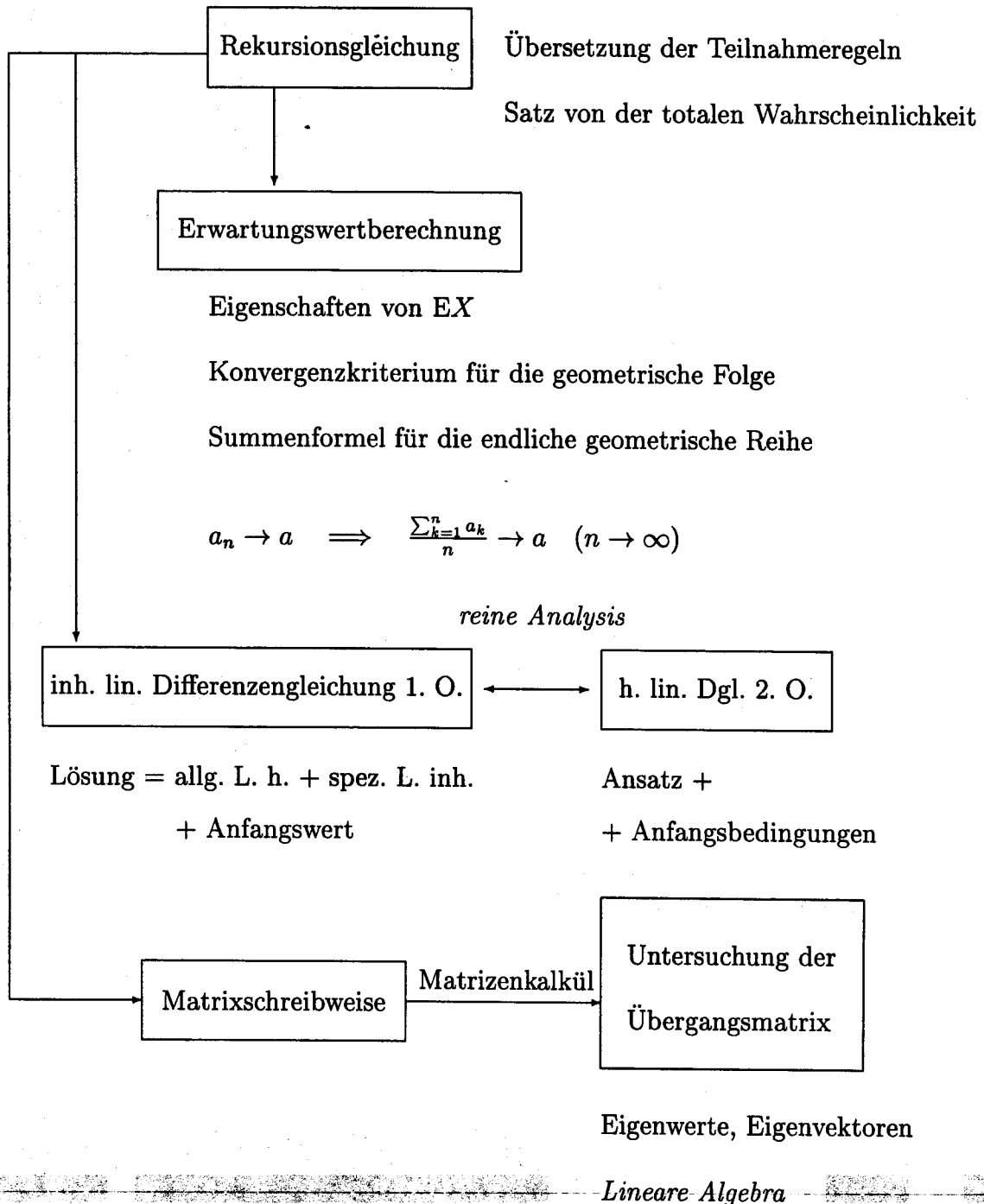
Die behandelte Thematik ist eine *anwendungsorientierte*, wenngleich schon gesagt werden muss, dass die Fairness des Auswahlverfahrens auch ohne mathematischen „Beweis“ angenommen werden konnte. Die Empirie alleine spricht dafür (das Verfahren wird schon sehr lange eingesetzt), und außerdem: Warum sollte es nicht fair sein? — Es gibt keinen Hinweis darauf. Dennoch: Hier steht nicht das Ergebnis, sondern der Weg dorthin im Vordergrund. Das *Wechselspiel Analysis — Lineare Algebra — Stochastik* kann so an einer (stochastischen) Situation demonstriert werden.

Auch als Beitrag zum *problem solving* kann das eben Gesagte gesehen werden. Eine klare Fragestellung zieht realistische Möglichkeiten der Beantwortung mit sich.

Der Einsatz von *Computeralgebrasystemen* kann zu einer Vermutung über das Langzeitverhalten des stochastischen Prozesses führen. Freilich muss ein Beweis mit anderen Mitteln geführt werden. Insgesamt ist das eine durchaus gängige Ausformung des Mathematik Treibens.

Außerdem ist eine *Einführung in verschiedene mathematische Themen* (Differenzgleichungen, Matrizenrechnung) gegeben. Liegen diese oder einzelne davon schon vor, können umgekehrt *fertige Theorien* erfolgreich angewendet werden: das Wissen über den Lösungsraum von bestimmten Typen von Differenzgleichungen, die Eigenwerte und -vektoren von Matrizen, das Konvergenzverhalten von (bestimmten) Folgen und Reihen etc.

Das folgende Diagramm soll eine Übersicht über die verwendeten mathematischen Werkzeuge geben.



Insgesamt: Vielfalt der Modelle, der Beschreibungsmöglichkeiten!

## Literatur

- [FRU] Fruttero, Carlo und Lucentini, Franco: *Der Palio der toten Reiter*. Verlag Piper, München 1986.
- [GG] Götz, Stefan und Grosser, Michael: *Über das Pferderennen in Siena*. Erschienen in: *Mathematische Semesterberichte* **46** (1999), S. 77–92.
- [GR7] Götz, Stefan und Reichel, Hans-Christian (Hrsg.): *Lehrbuch der Mathematik 7* von Robert Müller und Günter Hanisch. öbv&hpt, Wien 2003.
- [HEU] Heuser, Harro: *Lehrbuch der Analysis. Teil 1*. B. G. Teubner, Stuttgart 1986 (4., durchgesehene Auflage).

## 2 Die Ziegen

*Problem:* Folgendes Angebot ist bei einer Show in den USA tatsächlich gemacht worden: Hinter drei (geschlossenen) Türen befinden sich zwei Ziegen und ein Auto. Der Kandidat, der nicht weiß, hinter welcher Tür sich das Auto befindet, wählt eine Tür, z. B. die linke. Befindet sich dahinter das Auto, so hat es der Kandidat gewonnen, andernfalls schweigt die Kunde, ob auch eine Ziege mitgenommen werden kann. Dieser andere Fall wird jedenfalls als Verlust für den Kandidaten angesehen. Der Spielleiter (der schon weiß, wo sich das Auto befindet) öffnet eine andere Tür, hinter der sich eine Ziege befindet. Der Kandidat hat jetzt die Möglichkeit, bei seiner gewählten Tür (der linken) zu bleiben oder auf die noch geschlossene zu wechseln. Wie sind die beiden Strategien (bleiben oder wechseln) zu bewerten?

### 2.1 Erste Lösung

Siehe Abbildung 1: in zwei von drei Fällen führt *Wechseln* zum Auto!

### 2.2 Zweite Lösung

Bleiben			Wechseln		
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
Z	Z	A	Z	Z	A
1	1	1	1	1	1
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$\underbrace{\hspace{2em}}_{\frac{2}{3}}$			$\underbrace{\hspace{2em}}_{\frac{2}{3}}$		

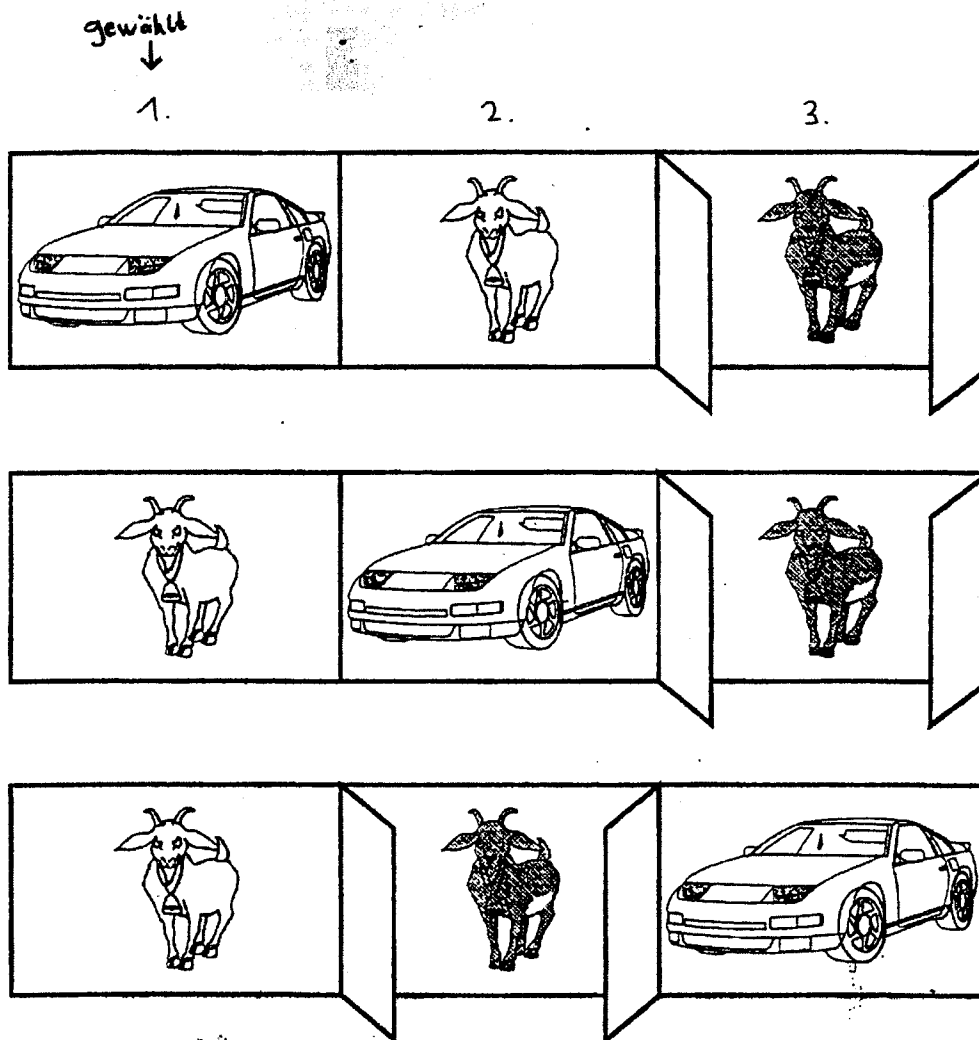


Abbildung 1: Das Drei-Türen-Problem

### 2.3 Dritte Lösung nach BAYES (nach [VW])

Wir bezeichnen mit  $\theta_j$  den Zustand „Das Auto steht hinter der Tür  $j$ .“.

*A priori* setzen wir die drei möglichen Zustände gleichwahrscheinlich an:

$$P(\theta_j) = \frac{1}{3} \quad (j = 1, 2, 3) .$$

Die in Rede stehende *Situation* sei die folgende: Kandidat tippt auf Tür 1, Tür 3 wird geöffnet (Ereignis  $x = 3$ ).

Damit sind die sogenannten (bedingten) Vorwärtswahrscheinlichkeiten bis auf eine festgelegt:

$$\begin{aligned} P(x = 3|\theta_3) &= 0 , \\ P(x = 3|\theta_2) &= 1 \quad \text{und} \\ P(x = 3|\theta_1) &= \frac{1}{2} . \end{aligned}$$

Die letzte Wahrscheinlichkeit enthält eine Annahme über das Verhalten des Spielers: Nur in diesem Fall hat er die Freiheit, zwischen zwei (Ziegen-)Türen zu wählen. Wenn wir die korrespondierende Wahrscheinlichkeit also mit  $\frac{1}{2}$  bewerten, so heißt das, dass der Spielleiter in diesem Fall immer durch Losentscheid (etwa mittels einer fairen Münze) die eine oder andere Tür auswählt.

*A posteriori* erhalten wir mit Hilfe des BAYES'schen Theorems

$$P(\theta_j|x = 3) = \frac{P(x = 3|\theta_j) \cdot P(\theta_j)}{\sum_{i=1}^3 P(x = 3|\theta_i) \cdot P(\theta_i)} \quad j = 1, 2, 3 ;$$

das sind die sogenannten Rückwärtswahrscheinlichkeiten, damit folgt konkret

$$\begin{aligned} P(\theta_1|x = 3) &= \frac{P(x = 3|\theta_1) \cdot P(\theta_1)}{P(x = 3|\theta_1) \cdot P(\theta_1) + P(x = 3|\theta_2) \cdot P(\theta_2) + P(x = 3|\theta_3) \cdot P(\theta_3)} = \\ &= \frac{\frac{1}{2}}{\frac{1}{2} + 1 + 0} = \frac{\frac{1}{2}}{\frac{3}{2}} = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3} , \\ P(\theta_2|x = 3) &= \frac{P(x = 3|\theta_2)}{\frac{3}{2}} = \frac{1}{\frac{3}{2}} = \frac{2}{3} \quad \text{und} \\ P(\theta_3|x = 3) &= 0 . \end{aligned}$$

---

D. H. EIN WECHSEL VON TÜR 1 AUF TÜR 2 VERDOPPELT DIE AUS-SICHT, DAS AUTO ZU GEWINNEN!

### 2.3.1 Variationen über das Verhalten des Spielleiters

- Angenommen, der Spielleiter öffnet *immer* die dritte Tür, wenn das Auto hinter der ersten Tür ist und der Kandidat ebendiese gewählt hat, das heißt also:  $P(x = 3|\theta_1) = 1$ . Wir bekommen damit

$$P(\theta_1|x = 3) = \frac{1}{1 + 1 + 0} = \frac{1}{2},$$

$$P(\theta_2|x = 3) = \frac{1}{2} \quad \text{und}$$

$$P(\theta_3|x = 3) = 0.$$

In diesem Fall (und nur in diesem, wie wir sehen werden) ist es *egal*, ob man wechselt oder bleibt. Aus der Sicht des Kandidaten stellt sich die Situation folgendermaßen dar: Der Spielleiter kann aus zwei Gründen die dritte Tür geöffnet haben: Entweder weil er muss (das wäre dann der Fall, wenn das Auto hinter der zweiten Tür verborgen ist) oder weil sein Verhalten ihn dazu zwingt (siehe Annahme!).

- Angenommen, der Spielleiter öffnet *niemals* die dritte Tür, wenn das Auto hinter der ersten Tür ist und der Kandidat ebendiese gewählt hat, das heißt also  $P(x = 3|\theta_1) = 0$ . Wir erkennen so

$$P(\theta_1|x = 3) = \frac{0}{0 + 1 + 0} = 0,$$

$$P(\theta_2|x = 3) = \frac{1}{1} = 1 \quad \text{und}$$

$$P(\theta_3|x = 3) = 0.$$

In diesem Fall (und nur in diesem, wie wir sehen werden) ist es *sicher*, dass das Auto hinter der zweiten Tür steht, denn der Spielleiter öffnet die dritte Tür nur, wenn ihn die Situation dazu zwingt. Und dies ist nur für  $\theta_2$  der Fall.

- *Allgemein* können wir  $P(x = 3|\theta_1) =: p \in [0, 1]$  modellieren und erhalten

$$P(\theta_1|x = 3) = \frac{p}{p + 1 + 0} = \frac{p}{1 + p} \quad (\text{„bleiben“}),$$

$$P(\theta_2|x = 3) = 1 - \frac{p}{1 + p} = \frac{1 + p - p}{1 + p} = \frac{1}{1 + p} \quad (\text{„wechseln“}) \quad \text{und}$$

$$P(\theta_3|x = 3) = 0.$$

Wir vergleichen nun die Graphen der Funktionen (siehe Abbildung 2)

$$b(p) := \frac{p}{1 + p} \quad \text{und} \quad w(p) := \frac{1}{1 + p}$$

und erkennen wegen  $0 \leq p \leq 1$ :

$$w(p) \geq b(p) \quad \forall p \in [0, 1].$$



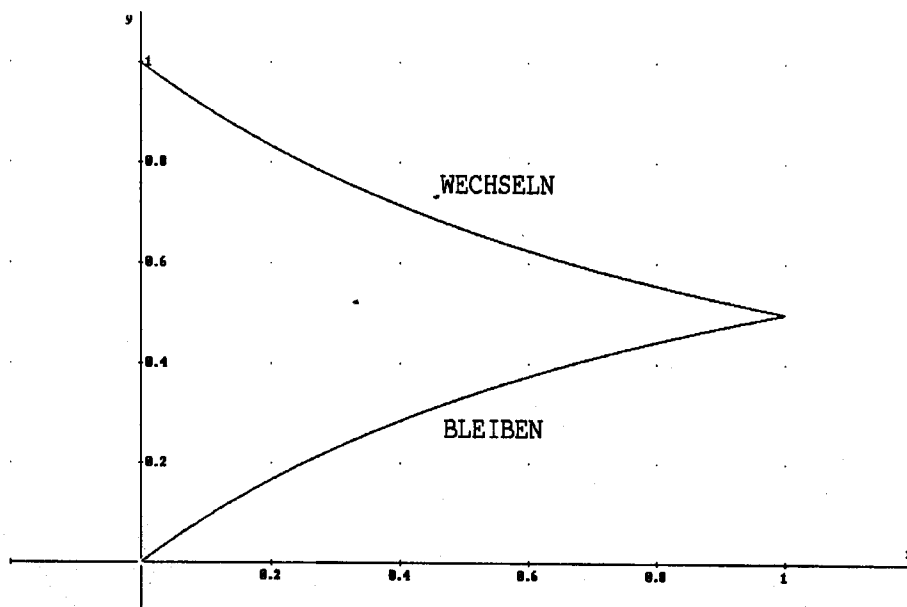


Abbildung 2: Zum Vergleich „Wechseln“ – „Bleiben“

## 2.4 Moral

WECHSELN IST NIE SCHLECHTER ALS BLEIBEN!

*Bemerkung:* Durch Beobachten des Verhaltens des Spielleiters über eine längere Zeitperiode (d. h. viele Shows ansehen!) hinweg könnte man zu einem Schätzwert für  $p$  kommen.

## 2.5 Didaktischer Kommentar

Die Variation des Parameters  $p$  führt zu einer *Modellbildung* des „Verhaltens“ des Spielleiters, was eine exotische Stellung dieser Darstellung einer (stochastischen) Situation in der (Schul-)Mathematik zur Folge hat. Unser Unwissen bzw. das des Kandidatens wird mittels wahrscheinlichkeitstheoretischer Werkzeuge beschrieben und bewertet.

Das ist typisch für die BAYES'sche Sichtweise: *Vorwissen* [hier: jede Tür kommt zu Beginn gleichermaßen in Frage, auch diese (*A-priori*-)Einschätzung kann variiert werden] und *Daten* (hier: der Spielleiter öffnet die dritte Tür) beeinflussen die *A-posteriori*-Einschätzung der „Welt“ (das sind hier die drei Türen).

Die Einführung der Variablen  $p$  suggeriert die Betrachtung der **Graphen der Funktionen**  $w$  und  $b$  bringt damit einen Überblick über die Situation. Also hilft die analytische „Brille“ das in Rede stehende Problem in seiner Gesamtheit zu erfassen.

## Literatur

- [VW] Vancsó, Ödön und Wickmann, Dieter: *Das Drei-Türen-Problem in bayescher Sicht*. In: Beiträge zum Mathematikunterricht 1999. Vorträge auf der 33. Tagung für Didaktik der Mathematik vom 1. bis 5.3.1999 in Bern. Verlag Franzbecker, Hildesheim, Berlin 1999, S. 551–554.

*Wer mehr über die BAYES'sche Sichtweise erfahren möchte:*

- [BOR] Borovcnik, Manfred: *Stochastik im Wechselspiel von Intuitionen und Mathematik*. Lehrbücher und Monographien zur Didaktik der Mathematik, Band 10. Herausgegeben von N. Knoche und H. Scheid. BI Wissenschaftsverlag, Mannheim/Leipzig/Wien/Zürich 1992.
- [GÖ1] Götz, Stefan: *Bayes-Statistik mit DERIVE*. In: Stochastik in der Schule Band 20 (2000), Heft 3, S. 2–14.
- [GÖ2] Götz, Stefan: *Klassische und Bayesianische Behandlung von Stochastikaufgaben aus österreichischen Schulbüchern*. In: M. Borovcnik, J. Engel und D. Wickmann (Hrsg.): *Anregungen zum Stochastikunterricht*:
- Die NCTM-Standards 2000
  - Klassische und Bayessche Sichtweise im Vergleich

Bericht von zwei Arbeitskreistagungen des Arbeitskreises „Stochastik in der Schule“ in der Gesellschaft für Didaktik der Mathematik e.V. vom 29./30. Oktober 1999 und 10.-12. November 2000 in Berlin. Verlag Franzbecker, Hildesheim, Berlin 2001, S. 147–162.

- [GÖ3] Götz, Stefan: *Bayes-Statistik mit DERIVE, Teil 2*. In: Stochastik in der Schule Band 22 (2002), Heft 2, S. 14–22.
- [VI] Viertl, Reinhard: *Einführung in die Stochastik. Mit Elementen der Bayes-Statistik und der Analyse unscharfer Information*. Reihe: Springers Lehrbücher der Informatik. Springer, Wien u. a. 2003 (3., überarb. u. erw. Aufl.).
- [WI1] Wickmann, Dieter: *Bayes-Statistik. Einsicht gewinnen und entscheiden bei Unsicherheit*. Mathematische Texte, Band 4. Herausgegeben von N. Knoche und H. Scheid. BI Wissenschaftsverlag, Mannheim/Wien/Zürich 1990.
- 
- [WI2] Wickmann, Dieter: *Zur Begriffsbildung im Stochastikunterricht*. In: *Journal für Mathematik-Didaktik* 19 (1998), Heft 1, S. 46–80.

### 3 Die Würfel (nach [PL], S. 91f.)

#### 3.1 Das Problem

Würfelt man mit zwei fairen Würfeln und betrachtet die Augensumme, so sind die einzelnen Ausprägungen  $2, 3, \dots, 12$  bekanntlich nicht gleichwahrscheinlich. Die Frage ist nun, ob die beiden Würfel (oder nur einer davon) so manipuliert werden könnten, dass sich die Gleichwahrscheinlichkeit für die Augensumme einstellt. Die Antwort ist negativ.

Um dies einzusehen, müssen wir sogenannte (**wahrscheinlichkeits**)**erzeugende Funktionen** kennenlernen (nach [BOS], S. 80f.):

*Definition:* Eine diskrete Zufallsvariable  $X$  habe den Wertevorrat  $W_X = \mathbb{N}$  und besitze die Verteilung  $[i, P(X = i)]$ ,  $i = 0, 1, 2, \dots$ . Dann heißt

$$G_X(t) = \sum_{i=0}^{\infty} t^i \cdot P(X = i)$$

erzeugende Funktion  $G_X$  der Zufallsvariablen  $X$ .

Wozu dienen diese Funktionen? — Sie können  $X$  vollständig beschreiben.

Es ist nämlich  $|G_X(t)| \leq \sum_{i=0}^{\infty} P(X = i) = 1$  für  $|t| \leq 1$ .

Daraus schließen wir  $G_X(0) = P(X = 0)$  ( $t^0 = 1!$ ). (Formales) Differenzieren ergibt  $G'_X(t) = \sum_{i=1}^{\infty} i \cdot t^{i-1} \cdot P(X = i)$ , für  $t = 0$  ist  $G'_X(0) = P(X = 1)$ . Nochmaliges Differenzieren liefert  $G''_X(t) = \sum_{i=2}^{\infty} i \cdot (i-1) \cdot t^{i-2} \cdot P(X = i)$ , was die nächste Wahrscheinlichkeit mit sich bringt:  $G''_X(0) = 2 \cdot P(X = 2)$ .

Allgemein halten wir

$$G_X^{(n)}(0) = n! \cdot P(X = n) \quad \text{für } n = 0, 1, 2, \dots \text{ fest.}$$

Weiters ist (wie man durch Nachrechnen leicht selbst erkennt)

- $\mu = E(X) = G'_X(1)$  und
- $\sigma^2 = D^2(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2$ .

Für die *Binomialverteilung* erhalten wir z. B.

$$\begin{aligned} G_X(t) &= \sum_{i=0}^n t^i \cdot \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} = \sum_{i=0}^n \binom{n}{i} \cdot (p \cdot t)^i \cdot (1-p)^{n-i} = \\ &= [p \cdot t + (1-p)]^n = [p \cdot (t-1) + 1]^n \end{aligned}$$

mittels des **Binomischen Lehrsatzes**.

So bekommen wir z. B. die bekannte Formel

$$\mu = G'_X(1) = n \cdot [p \cdot (t-1) + 1]^{n-1} \cdot p \Big|_{t=1} = n \cdot p.$$

### 3.2 Die Antwort

Die Zufallsvariablen  $X_1, X_2$

- beschreiben die (geworfenen) *Augenzahlen* der beiden Würfel und
- sind (stochastisch) *unabhängig* voneinander:

$$P(X_1 = i, X_2 = j) = P(X_1 = i) \cdot P(X_2 = j) \quad \forall i, j \in \{1, \dots, 6\}.$$

Wir setzen

$$P(X_1 = i) =: p_i \quad \text{und} \quad P(X_2 = j) =: q_j, \quad i, j = 1, \dots, 6; \quad p_i = q_j = 0 \quad \text{sonst.}$$

Damit schreiben wir

$$G_{X_1}(t) = \sum_{k=1}^6 p_k \cdot t^k \quad \text{und} \quad G_{X_2}(t) = \sum_{l=1}^6 q_l \cdot t^l.$$

Jetzt gehen wir auf die *Augensumme*  $X_1 + X_2$  los: Dazu definieren und berechnen wir für  $m = 2, \dots, 12$

$$\begin{aligned} r_m &:= P(X_1 + X_2 = m) = \sum_{n=1}^{m-1} P(X_1 = n, X_2 = m - n) = \\ &= \sum_{n=1}^{m-1} P(X_1 = n) \cdot P(X_2 = m - n) = \sum_{n=1}^{m-1} p_n \cdot q_{m-n}. \end{aligned}$$

Ihre *erzeugende Funktion* ist

$$\begin{aligned} G_{X_1+X_2}(t) &= \sum_{m=2}^{12} r_m \cdot t^m = \sum_{m=2}^{12} \left( \sum_{n=1}^{m-1} p_n \cdot q_{m-n} \right) \cdot t^m = \\ &= \sum_{k=1}^6 p_k \cdot t^k \cdot \sum_{l=1}^6 q_l \cdot t^l = G_{X_1}(t) \cdot G_{X_2}(t). \end{aligned}$$

Angenommen, eine *Gleichverteilung* für  $X_1 + X_2$  ist *möglich*, dann verzeichnen wir  $r_m = \frac{1}{11} \quad \forall m = 2, \dots, 12$  und damit

$$G_{X_1+X_2}(t) = \frac{1}{11} \cdot \sum_{m=2}^{12} t^m = G_{X_1}(t) \cdot G_{X_2}(t) = \sum_{k=1}^6 p_k \cdot t^k \cdot \sum_{l=1}^6 q_l \cdot t^l,$$

also muss

$$\frac{1}{11} \cdot \sum_{m=2}^{12} t^m = \sum_{k=1}^6 p_k \cdot t^k \cdot \sum_{l=1}^6 q_l \cdot t^l \quad \text{gelten.}$$

Division durch  $t^2 \neq 0$  liefert

$$\frac{1}{11} \cdot \sum_{m=2}^{12} t^{m-2} = \sum_{k=1}^6 p_k \cdot t^{k-1} \cdot \sum_{l=1}^6 q_l \cdot t^{l-1} . \quad (3)$$

Insbesondere ist also wegen  $0 \leq p_i, q_j \leq 1 \forall i, j = 1, \dots, 6$

$$p_6 > 0 \quad \text{und} \quad q_6 > 0 \quad (\text{„Koeffizientenvergleich“ f\u00fcr } t^{10}) ,$$

woraus wir

$$\lim_{t \rightarrow \pm\infty} \sum_{k=1}^6 p_k \cdot t^{k-1} = \pm\infty \quad (\text{Polynom ungeraden Grades})$$

erkennen.

Daher

$$\exists t_0 \in \mathbb{R} \quad \text{mit} \quad \sum_{k=1}^6 p_k \cdot t_0^{k-1} = 0 \quad (\text{Nullstellensatz}) .$$

Eingesetzt in (3) ergibt sich so

$$\frac{1}{11} \cdot \sum_{m=2}^{12} t_0^{m-2} = \sum_{k=1}^6 p_k \cdot t_0^{k-1} \cdot \dots = 0 ,$$

also

$$\frac{1}{11} \cdot \frac{t_0^{11} - 1}{t_0 - 1} = 0 \quad (\text{Summenformel f\u00fcr die endliche geometrische Reihe}) .$$

Dies bedeutet einen *Widerspruch*, wie wir in vielerlei Hinsicht einsehen k\u00f6nnen:

- $t_0^{11} = 1$  bedeutet  $t_0 = 1$  in  $\mathbb{R}$ , aber  $\sum_{k=1}^6 p_k \cdot t_0^{k-1} = \sum_{k=1}^6 p_k = 1 \neq 0!$
- Der **Graph** der reellen Funktion  $f$  mit  $f(t) := \frac{t^{11}-1}{t-1}$  enth\u00e4lt *keine Nullstelle*, wie Abbildung 3 zeigt.
- Es ist  $f(1) = \lim_{t \rightarrow 1} \frac{t^{11}-1}{t-1} = \lim_{t \rightarrow 1} (1+t+t^2+\dots+t^{10}) = 11$ , es handelt sich hierbei also um eine **hebbare Unstetigkeitsstelle**, siehe z. B. [RMH6], S. 254.

### 3.3 Moral

Auch F\u00e4lschen mit allen nur m\u00f6glichen Freiheiten f\u00fchrt nicht immer zum Ziel!

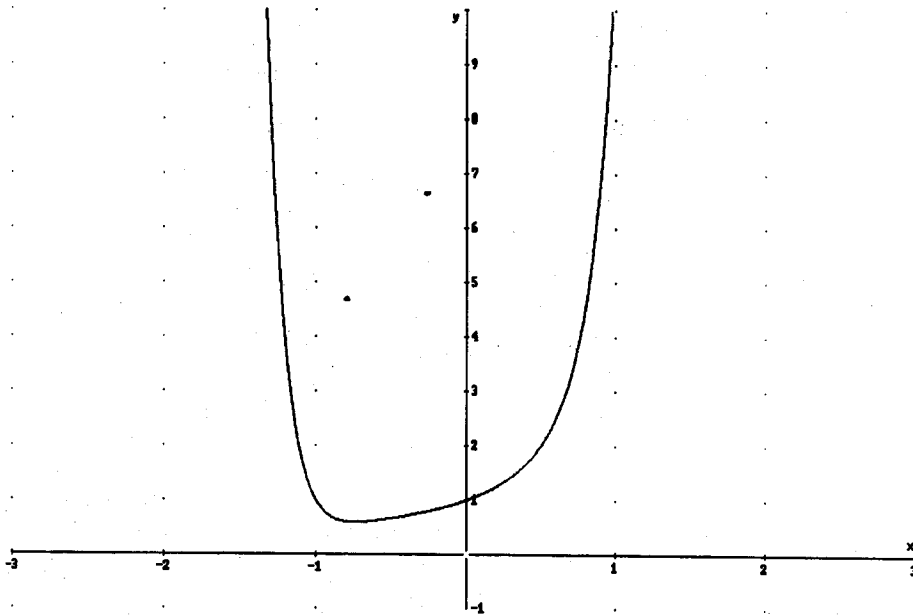


Abbildung 3: Die Funktion  $f$  hat keine Nullstelle.

### 3.4 Didaktischer Kommentar

*Unmöglichkeitsbeweise* in der Schulmathematik sind aus vielerlei Gründen *schwierig* schon vom *Ansatz* her: „Zeigen, dass es etwas nicht gibt“, dies muss man erst einmal einsehen.

Der Beginn des damit verbundenen *indirekten* Beweises: „Ang. das gibt es doch, ...“ trägt nicht gerade zur leichteren Einsicht bei.

Und schließlich ist der erzeugte *Widerspruch* hier nicht eindeutig. Welches Argument soll im Unterricht verwendet werden? Wird die Widerspruchsfindung durch die Fülle der Einwände überzeugender?

Die eben aufgezählten Aspekte zeigen die Notwendigkeit, erst ein prinzipielles Verständnis für die Vorgehensweise bei den SchülerInnen zu erwecken. Wie kann das geschehen? — Z. B. durch Vorübungen an konkreten einfacheren Beispielen wie z. B. in [GR5] im ersten Kapitel: „Die Sprache der Mathematik“.

*Inhaltlich* scheint die Problemstellung relativ einfach und klar zu sein.

Die gewählte *Methode* bringt erzeugende Funktionen mit sich, die aus der Wahrscheinlichkeitstheorie direkt in die Analysis führen. Und darin liegt auch ihr großer Wert: Damit stehen (mächtige) Werkzeuge (der Analysis) zur Verfügung, vgl. etwa  $[x_i | P(X = x_i)]$  mit  $G_X(t)$ . Allerdings ist es oft nicht einfach, eine geschlossene Darstellung von  $G_X$  zu finden. Dies ist ein Ausdruck des *Satzes von der Erhaltung der Schwierigkeit*.

Hier liegt nun ein höchst *endliches* Problem vor, die vorkommenden Summen

sind überschaubar, die geschlossene Form von  $G_{X_1+X_2}$  liefert eine sehr prominente Formel: die *Summenformel für die endliche geometrische Reihe*.

Die *Widerspruchsfindung* schließlich kann auf *unterschiedlicher Argumentationsbasis* stattfinden: wahrscheinlichkeitstheoretisch, naiv-analytisch oder (profunder-)analytisch haben wir gesehen. Die oben aufgezeigte damit verbundene Schwierigkeit kann auch eine Chance bedeuten: Verschiedene Ebenen finden i. Allg. auch verschiedene AbnehmerInnen, die Verbreitung der Akzeptanz unter den SchülerInnen für diese Argumentationsweise kann so also auch zunehmen.

## Literatur

- [BOS] Bosch, Karl: *Elementare Einführung in die Wahrscheinlichkeitsrechnung*. vieweg studium Basiswissen. Friedr. Vieweg & Sohn, Braunschweig / Wiesbaden 1986 (5., durchgesehene Auflage).
- [GR5] Götz, Stefan und Reichel, Hans-Christian (Hrsg.): *Mathematik Lehrbuch 5* von Robert Müller und Günter Hanisch. öbv&hpt, Wien 2004.
- [PL] Plachky, Detlef: *Mathematische Grundbegriffe und Grundsätze der Stochastik*. Springer, Berlin u. a. 2001.
- [RMH6] Reichel, Hans-Christian, Müller, Robert und Hanisch, Günter: *Lehrbuch der Mathematik 6*. öbv&hpt, Wien 2002 (5. Auflage).

## 4 Lineare Regression — einmal ein wenig anders (nach [MEY], vgl. auch [SCH], S. 30f.)

### 4.1 Das (wohlbekannte) Problem

Es werden zweidimensionale kardinalskalierte Daten — Datenpaare —  $\begin{pmatrix} x_i \\ y_i \end{pmatrix}$  ( $i = 1, \dots, n$ ) erhoben.

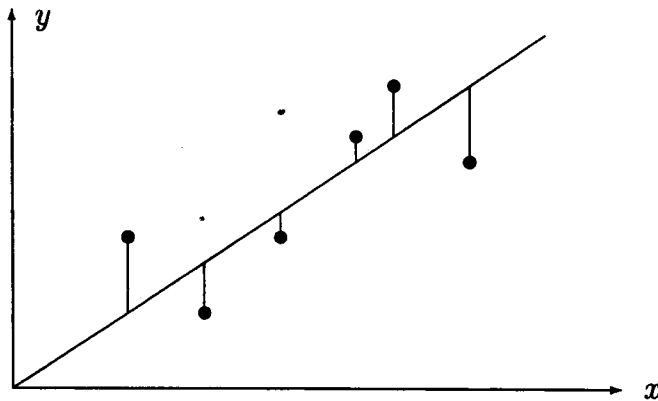
Gesucht ist diejenige Gerade („Ausgleichs-“ oder „Regressionsgerade“)  $y = ax + b$ , die die Daten „möglichst gut“ annähert.

*Die x-Werte stehen dabei fest.*

### 4.2 Der erste wohlbekannte Ansatz

Zuerst wollen wir fordern, dass *die Summe der vertikalen Abstände verschwinden möge*, also:  $\sum_{i=1}^n (a \cdot x_i + b - y_i) = 0$ , gesucht sind dann die Parameter  $a$  und  $b$ ,

die empirischen Regressionskoeffizienten. Die folgende Abbildung illustriert die Situation:



Gesucht ist also eine solche Ausgleichsgerade, bei der die Abweichungen nach oben gerade durch diejenigen nach unten kompensiert werden.

Als zentrale Idee für das Folgende stellt sich nun eine günstige Wahl des Koordinatensystems heraus: Wir setzen den „Schwerpunkt“  $\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$  in den Ursprung des neuen Koordinatensystems.

Die neuen Koordinaten der Daten lauten somit  $\begin{pmatrix} u_i \\ v_i \end{pmatrix} = \begin{pmatrix} x_i - \bar{x} \\ y_i - \bar{y} \end{pmatrix}$ . Dann ist  $\bar{u} = \bar{v} = 0$ , wie man leicht nachrechnet. Mit

$$U := \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \quad V := \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \quad \text{und} \quad E := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

ist daher

$$U \cdot E = V \cdot E = 0,$$

mit „ $\cdot$ “ ist (hier) das (Standard-)Skalarprodukt zwischen zwei Vektoren gemeint. In den neuen Koordinaten und unter Berücksichtigung der eben festgestellten Identität sieht unsere Forderung nun so aus:

$$\sum_{i=1}^n (a \cdot u_i + b - v_i) = a \cdot \underbrace{\sum_{i=1}^n u_i}_{=0} + b \cdot n - \underbrace{\sum_{i=1}^n v_i}_{=0} = b \cdot n \stackrel{!}{=} 0,$$

sie wird also stark verkürzt.

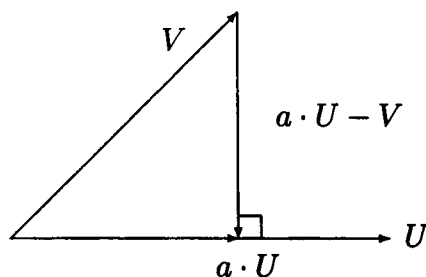
Die (vielleicht überraschende) *Conclusio* daraus ist  $b = 0$ , übersetzt heißt das: Jede Gerade durch den Schwerpunkt bringt die Summe der vertikalen Abweichungen zum Verschwinden!



### 4.3 Eine ebenfalls wohlbekannte Verfeinerung

Um eine eindeutige Ausgleichsgerade zu definieren fordern wir nun zusätzlich, dass die *Summe der vertikalen Abstandsquadrate minimal* werden soll („Methode der kleinsten Quadrate“, nach C. F. GAUSS 1809 und A. M. LEGENDRE 1806). Wegen  $b = 0$  ist  $\sum_{i=1}^n (a \cdot u_i - v_i)^2$  zu minimieren, suche  $a$  so, dass  $a \cdot U - V$  möglichst kurz wird, also

$V$  wird auf  $U$  projiziert:



Damit ist

$$\begin{aligned} (a \cdot U - V) \cdot a \cdot U &= 0 \\ a^2 \cdot U^2 - a \cdot U \cdot V &= 0 \quad | : a \neq 0 \\ a \cdot U^2 - U \cdot V &= 0, \end{aligned}$$

also

$$a = \frac{U \cdot V}{U^2} \quad \text{— was heißt das?}$$

Die *Rückübersetzung* in die ursprünglichen Koordinaten bringt ein altbekanntes Ergebnis (siehe auch [RMH8], S. 211):

$$\begin{aligned} a &= \frac{\begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} \cdot \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}}{\begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \\ &= \dots = \frac{\left( \sum_{i=1}^n x_i \cdot y_i \right) - n \cdot \bar{x} \cdot \bar{y}}{\left( \sum_{i=1}^n x_i^2 \right) - n \cdot \bar{x}^2} = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} \end{aligned}$$

und wegen

$$\sum_{i=1}^n (a \cdot x_i + b - y_i) = 0$$

ist

$$b = \frac{\sum_{i=1}^n y_i}{n} - a \cdot \frac{\sum_{i=1}^n x_i}{n} = \bar{y} - a \cdot \bar{x}$$

(beachte: der Schwerpunkt liegt *jedenfalls* auf der Ausgleichsgeraden).

#### 4.4 Wie gut waren wir? — Die Güte der Anpassung

- Wenn alle Daten auf einer Geraden liegen, dann gilt (und vice versa):

$$V = a \cdot U \iff \cos(U, V) = \pm 1.$$

In diesem Fall liegt *minimale* Abweichung vor. „Geometrisch“ interpretieren wir, dass  $V$  und  $U$  „**dieselbe Richtung**“ (aber nicht unbedingt dieselbe Orientierung!) haben.

- Daraus schließen wir, dass *maximale* Abweichung dann gegeben ist, wenn  $V$  und  $U$  „**aufeinander senkrecht**“ stehen:

$$V \perp U \iff \cos(U, V) = 0.$$

Daher sehen wir allgemein, dass

$$\cos(U, V) = \frac{U \cdot V}{|U| \cdot |V|}$$

ein gutes Maß für die Abweichung der Daten von der Ausgleichsgeraden ist.

Wieder passiert die *Rückübersetzung* in die alten Koordinaten, um zu einem bekannten Ergebnis zu kommen:

$$\begin{aligned} \frac{U \cdot V}{|U| \cdot |V|} &= \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \\ &= \dots = \frac{\left( \sum_{i=1}^n x_i \cdot y_i \right) - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2} \cdot \sqrt{\sum_{i=1}^n y_i^2 - n \cdot \bar{y}^2}} = \\ &= \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{\sqrt{n \cdot \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} \cdot \sqrt{n \cdot \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2}} =: r \end{aligned}$$

(PEARSON'scher Korrelationskoeffizient, siehe auch [RMH8], S. 213).

## 4.5 Zu guter Letzt

Aus der *einen* Bedingung  $\sum_{i=1}^n [(a \cdot u_i + b) - v_i]^2$  zu minimieren bekommt man  $a$  und  $b$  gleich! Die schrittweise Annäherung in zwei Etappen ist aber didaktisch nicht unklug.

Die *Regression von X bezüglich Y*  $X = a^*Y + b^*$  führt auf die analoge Bedingung  $\sum_{i=1}^n [(a^* \cdot y_i + b^*) - x_i]^2$  zu minimieren. „Geometrisch“ heißt das eine **Spiegelung an der ersten Medianen**:  $U \leftrightarrow V$ . Damit bekommen wir sofort den gesuchten Parameter  $a^*$  (und daraus auch  $b^*$ ):

$$a^* = \frac{U \cdot V}{V^2}$$

und wir erkennen mit

$$a \cdot a^* = \frac{U \cdot V}{U^2} \cdot \frac{U \cdot V}{V^2} = \frac{(U \cdot V)^2}{U^2 \cdot V^2} = \left( \frac{U \cdot V}{|U| \cdot |V|} \right)^2 = r^2$$

das *Bestimmtheitsmaß*.

Wenn  $r = 1$ , dann  $r^2 = 1$  und damit  $a \cdot a^* = 1$ , also  $a^* = \frac{1}{a}$ .

## 4.6 Didaktischer Kommentar

Die grundlegende Idee ist hier natürlich, zu optimieren *ohne* (partiell) zu differenzieren, wie das sonst in vielen gängigen Lehrbüchern der Fall ist. Dazu *abstrahieren* wir von geometrischen, anschaulichen Begriffen bzw. Größen wie „Vektor“, „Länge von Vektoren“, „Winkel zwischen zwei Vektoren“, „Projektion eines Vektors auf einen anderen“, „Schwerpunkt“ oder „Orthogonalität“.

*Beachte dabei:* Einzig die „Abstände“ behalten ihre ursprüngliche, anschauliche Bedeutung bei!

Die Lineare Regression im Mathematikunterricht trägt einerseits in der *Stochastik* zur Veranschaulichung und Interpretation von (scheinbaren) Zusammenhängen von Daten bei. Andererseits — und darauf liegt hier der Schwerpunkt — werden bei dieser Herleitung der entsprechenden Formeln grundlegende *geometrische Vorstellungen verallgemeinert*. Und das ist eine ganz typische Methode in der Mathematik! *Keinesfalls* natürlich soll die Lineare Regression als Vehikel zum *händischen* Auswerten von komplizierten Ausdrücken dienen, hier ist eine echte Indikation des Computereinsatzes im Mathematikunterricht zu konstatieren.

Die *Motivation* der Methode der kleinsten Quadrate ist heutzutage nicht mehr ganz einfach: Warum minimiert man nicht die Summe der Abstandsquadrate an sich? Vieles deutet darauf hin, dass ursprünglich rechentechnische Gründe für diese Vorgangsweise (die der kleinsten Quadrate) gesprochen haben, weil sie

eben eine gewisse Vereinfachung darstellt, die aber im Zeitalter des Computers nicht mehr aufrecht zu erhalten sind. Also bleiben *historische* Gründe, und die Methode ist nach wie vor (in den Anwendungen) *weit verbreitet*.

*Innermathematisch* ist Folgendes in diesem Zusammenhang von Bedeutung: Unterstellt man den Residuen (Abweichungen) eine  $N(0, \sigma^2)$ -Verteilung, dann führt die *Maximum-Likelihood-Methode* zur Schätzung von  $a$  und  $b$  auf genau diese Minimalbedingung ([HOF] und [STA], S. 262). Natürlich ist diese Motivation für den Mathematikunterricht selbst (i. Allg.) nicht brauchbar, für das Hintergrundwissen der Lehrenden aber allemal!

Insgesamt liegt hier insofern eine ertragreiche Verquickung der Linearen Algebra mit der Stochastik vor, weil die damit verbundene Abstraktion von anschaulichen Begriffen und Größen in der Mathematik eine weit über das konkrete Anliegen hinaus führende grundlegende Bedeutung hat.

## Literatur

- [HOF] Hofbauer, Franz: *Mündliche Mitteilung* am 9. April 2004 im Hause.
- [MEY] Meyer, Jörg: *Begriffe und Methoden der Vektorgeometrie in der Stochastik*. Skriptum zur Herbsttagung des Arbeitskreises „Stochastik in der Schule“ in der Gesellschaft für Didaktik der Mathematik vom 7.11.–9.11. 2003 in der Reinhardswaldschule bei Kassel. Hameln, 2003.
- [RMH8] Reichel, Hans-Christian, Müller, Robert und Hanisch, Günter: *Lehrbuch der Mathematik 8*. öbv&hpt, Wien 1999 (3. Auflage).
- [SCH] Scheid, Harald: *Wahrscheinlichkeitsrechnung*. Mathematische Texte, Band 6. Herausgegeben von N. Knoche und H. Scheid. BI-Wissenschaftsverlag, Mannheim/Leipzig/Wien/Zürich 1992.
- [STA] Stahel, Werner A.: *Statistische Datenanalyse. Eine Einführung für Naturwissenschaftler*. Friedr. Vieweg & Sohn, Braunschweig / Wiesbaden 2002 (4., verbesserte Auflage).

### *Anschriften des Verfassers:*

Ao. Univ.-Prof. Mag. Dr. Stefan Götz  
Institut für Mathematik  
Universität Wien  
Nordbergstraße 15 (UZA 4)  
A-1090 Wien  
Stefan.Goetz@univie.ac.at

Akademisches Gymnasium  
Wien I.  
Beethovenplatz 1  
A-1010 Wien

Den Kollegen Raith und Piskernig (beide Universität Wien) danke ich für die Unterstützung bei der Einbindung der Abbildungen ins  $\LaTeX$ -file.